Using Working Memory to Address the Validity-Diversity Dilemma: Incremental Validity and Subgroup Differences Compared to GMA

Devin Matthew Carter

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Psychology
Area: Industrial/Organizational Psychology

E. Scott Geller, Co-Chair
Neil M. A. Hauenstein, Co-Chair
Roseanne J. Foti
Brett D. Jones

March 26th, 2019
Blacksburg, Virginia

Keywords: Working Memory, Intelligence, Performance, Learning, Race Differences

Using Working Memory to Address the Validity-Diversity Dilemma: Incremental Validity and Subgroup Differences Compared to GMA

Devin Matthew Carter

ABSTRACT

General mental ability (GMA) has been found to be the best predictor of job knowledge and job performance, and it is widely-used for personnel selection decisions. However, the use of GMA in selection is a concern for practitioners because of the large Black-White race differences associated with GMA tests. The use of GMA tests, therefore, results in adverse impact when basing decisions on predicted performance. In order to address this validity-diversity tradeoff, a more specific cognitive ability is examined – working memory (WM). Two-hundred participants (50% Black, 50% White) were given measures of GMA and WM before being presented with learning opportunities meant to teach them novel information. The participants were then instructed to complete tasks which apply this newly learned knowledge. WM was examined in terms of how much additional variance was accounted for in task knowledge and task performance after controlling for GMA. In addition, race group differences of WM were compared to those of GMA. Results indicated that WM was able to account for significant additional variance in knowledge and performance, and that this relationship have been moderated by task complexity. WM exhibited slightly smaller absolute race differences as well, but these reductions were nonsignificant. Results are discussed in terms of the possible use of WM in a selection context.

Using Working Memory to Address the Validity-Diversity Dilemma: Incremental Validity and Subgroup Differences Compared to GMA

Devin Matthew Carter

GENERAL AUDIENCE ABSTRACT

General mental ability (GMA), or general intelligence, has some of the largest correlations with both job knowledge and job performance. However, Black applicants tend to score lower on GMA tests compared to White applicants. Therefore, when using GMA tests to select applicants, a higher proportion of Whites than Blacks are selected for a given job. This study aimed to examine whether someone's ability to hold and manipulate information in the midst of distracting tasks (i.e., working memory, WM) would also have large correlations with knowledge and performance. In addition, this study aimed to determine whether Black and White test takers had differences in mean WM scores that were smaller than those exhibited by GMA tests. Two-hundred participants (50% Black, 50% White) were given measures of GMA and WM before being presented with learning opportunities meant to teach them novel information. The participants were then instructed to complete tasks which apply this newly learned knowledge. Results indicated that the correlations between WM and task knowledge/performance were almost as large as those of GMA. In addition, there were smaller differences in means between Black and White test takers, although this difference was nonsignificant. Results are discussed in terms of the possible use of WM in a job selection context.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

General mental ability (GMA), sometimes referred to as cognitive ability, aptitude, IQ, or general intelligence, has been used by I-O researchers and practitioners for decades to explain a wide range of organizational outcomes. Some of these outcomes include counterproductive work behaviors (CWB; Dilchert, Ones, Davis, & Rostow, 2007), job knowledge (Colquitt, LePine, Noe , 2000; Hunter, 1986; McCloy, Campbell, & Cudeck, 1994), training performance (Ree & Carretta, 1998; Ree & Earles, 1991; Schmidt, 2002), and both subjective and objective job performance (Hunter, 1986; Roth, Huffcutt, & Bobko, 2003).

Perhaps most notable is Schmidt and Hunter’s (1998) estimate that GMA accounted for 26% of the variance in overall job performance when validity coefficients are corrected for measurement error and range restriction. GMA is considered to be among the best antecedents of both training and job performance (Ree & Earles, 1991; Ree, Earles, & Teachout, 1994; Schmidt & Hunter, 2004), and much of the past field research has focused on identifying other constructs that add incremental validity to GMA (Avis, Kudisch, & Fortunato, 2002; Huffcutt, Roth, & McDaniel, 1996; Stanhope & Surface, 2014).

Specific cognitive abilities have been explored in this attempt to find constructs that explain incremental variance in outcomes beyond GMA. Although early attempts to find specific abilities were mostly unsuccessful (e.g., Carey, 1994; McHenry, Hough, Toquam, Hanson, & Ashworth, 1990; Morales & Ree, 1992; Ree, Earles, & Teachout, 1994), recent research has indicated that working memory (WM) assessments may reliably account for variance in performance beyond GMA.

Carter, Hauenstein, and Geller (2018) examined the incremental validity of WM scores in accounting for variance above GMA for both task knowledge and task performance. Specifically, they gave undergraduate participants two measures of WM and a measure of fluid reasoning (Gf). Participants were then given a training video that taught them how to properly format an APA-style reference page, followed by measures of APA reference-page knowledge and a measure of task performance which applied this knowledge. The researchers found support for the hypothesis that WM has both direct and indirect effects on task performance through task knowledge after controlling for GMA

The purpose of this research is fourfold. First, this study aimed to replicate the Carter et al. (2018) findings and provide supporting evidence for the incremental validity in task knowledge and performance explained by WM assessments. Second, this study attempted to provide further support for Carter et al.'s mediating model whereby WM has both direct and indirect effects on task performance. Third, this research extended the prior Carter et al. study by investigating whether the predictive accuracy of WM assessments are affected by the alignment between the WM assessment and the nature of the performance task. Finally, the current study further extended the Carter et al. study by examining potential Black-White race effects in two ways; first, by testing whether race moderates the proposed mediating model, and second, by studying racial subgroup differences on WM assessments.

Working Memory

Baddeley and Hitch's (1974) multicomponent model of WM is one of the earliest models of WM. They proposed WM is short-term memory (STM) at work – the ability of a person to maintain information in STM while processing other tasks simultaneously. Individuals with better WM are not only more able to recall stored information in the presence of distracting stimuli, but they also have greater processing efficiency on simultaneous tasks (Daneman & Carpenter, 1980).

Although Baddeley and Hitch's model of WM remains influential across multiple disciplines, a new model of WM is gaining popularity in the cognitive and neuroscience literature. Oberauer (2002) proposed a model of WM which involves three embedded components of WM. He proposed that individuals focus their attention on a specific piece of information when it is presented. This is called the Focus of Attention (FoA). The FoA is embedded in the region of direct access (RDA) that includes units of information not being directly attended to, but these units are accessible due to their importance, their relatedness, or the state of activation. Lastly, the activated long-term memory (ALTM) is further distant from the FoA and contains information that may be linked to the items in the FoA or RDA, but at a more distant level of activation than the RDA. Because it is more current and better conveys the activation of information in WM, Oberaurer's three-embedded-components model will be referenced throughout this proposed study.

Measurement of WM. The most commonly-used WM tasks are complex span tasks such as the Operation Span Task (OSPAN) and the Reading Span Task (RSPAN). In the RSPAN, for example, a sentence is presented to the participant and the participant must determine whether the sentence semantically makes sense (e.g., "My friend and I jumped a pickle last weekend"). After they determine whether or not the sentence makes sense, participants are quickly presented with a target letter they are told to remember throughout the trial. They are then presented with another sentence, followed by another target letter, and so on. After about three to seven letters are presented, participants are asked to recall and report the target letters in the order they were presented.

The three-embedded-components model (Oberauer, 2009) explains the processes associated with completing a task such as the RSPAN. When a sentence is presented, the participant focuses his/her attention on the words s/he is reading (i.e., FoA), linking each word to the other words concurrently maintained in the RDA. When the first target letter is presented, it becomes the focus of attention and the preceding sentence transitions into ALTM (where it is likely inaccessible due to its insignificance and lack of further activation). When the second sentence is presented, the first target letter is stored in the ALTM but is still activated due to the task instructions to remember the target letter presented after each sentence. Once a second target letter is presented, the first target letter is retrieved from ALTM and is maintained in the RDA while the second target letter is in the FoA. When the third sentence is presented both the first and second target letters are stored in the ALTM until the third target letter is presented. This process continues until recall of the target letters is requested, at which time the letters are reported and the participant no longer invests effort to keep the target letters for that block activated. The OSPAN is a similar task, but instead of sentences, participants are presented with a simple arithmetic problem and asked to solve it before the target letter is presented.

WM and g. When Carroll (1993) conducted his initial factor analysis of cognitive abilities, few studies included WM assessments. Therefore, WM was not included in the Cattell-Horn-Carroll model (CHC; Kaufman, 2009) of intelligence, which has been the most widely-accepted and applied model of intelligence to date. In this CHC model, psychometric GMA is the overarching third-order "general intelligence," as proposed by Spearman (1904). Psychometric GMA reflects the positive manifold of success across cognitive tasks.

Hierarchically beneath GMA in the CHC model are second-order and first-order specific abilities. Some researchers consider WM to be a first-order and more specific ability in the GMA hierarchy (e.g., Schneider & McGrew, 2012). Although, Bosco, Allen, and Singh (2015) do not consider WM to be part of the GMA hierarchy at all. Further, Kane, Conway, Hambrick and Engle (2007) suggest that WM is a higher-order function that has more similarities to GMA than differences, suggesting WM is not a specific ability on lower levels of the GMA hierarchy.

Whether or not WM is considered to be part of the hierarchy of GMA is not relevant when considering WM as a construct that can account for additional variance in task knowledge or performance. More relevant is how highly WM is correlated with estimates of GMA. Traditional tests of GMA, especially those used in the selection context, do not include WM assessments. Instead, the constructs being assessed are primarily measures of comprehension knowledge (e.g., seven of the ASVAB subtests) or fluid reasoning (e.g., Raven's Progressive Matrices) that are used to represent GMA because they are highly saturated with GMA. However, WM scores are not used to measure GMA because WM is *not* highly saturated with

GMA. For this reason WM should be considered a potential incremental predictor of task knowledge and performance, regardless of the theoretical status of WM.

Incremental Validity of WM Scores

Several studies have shown that WM affects the learning of computational skills (Wilson & Swanson, 2001) and math skills (Alloway & Passolunghi, 2011; Gersten, Jordan, & Flojo, 2005; Swanson, 2004), the ability to remember lengthy instructions (Gathercole, Lamont, & Alloway, 2006), learn visually-presented material (Logie, Gilhooly, & Wynn, 1994), and learn a second language (DeKeyser & Koeth, 2011; Linck & Weiss, 2015; Wen & Skehan, 2011). However, the fundamental question is the extent to which WM scores explain incremental variance in outcomes beyond that accounted for by GMA. Research on the incremental validity of WM scores are less common than studies demonstrating criterion-related validity for only WM scores. Nonetheless, there is evidence that WM scores provide incremental variance for learning outcomes (Alloway, 2009; Alloway & Alloway, 2010; Carter et al., 2018; Krumm, Ziegler, & Buehner, 2008; Luo, Thomspon, & Detterman, 2006), incremental variance in task performance (Bosco et. al., 2015; Carter et al., 2018), and incremental variance in job performance (Bosco et. al., 2015).

A Partial Mediation Model

Hunter (1986) proposed a partially-mediated model whereby GMA affects task/job performance directly and indirectly through job knowledge. For the current study, WM replaces GMA in the partially-mediated model (See Figure 1), and the effects of GMA on both job knowledge and task performance are controlled. As stated earlier, Daneman and Carpenter (1980) noted that individuals with greater WM abilities are able to better hold information in storage during distracting tasks, as well as update this information and process irrelevant information (i.e., the distracting task) better than individuals with lesser WM abilities. This aligns with the dual functionality perspective of WM (Daneman & Carpenter, 1980). That is, WM consists of both a *capacity* function and a *processing* function. The working memory capacity (WMC) function allows individuals to maintain and recall information better, as well as maintain and recall *more* information in the presence of distractors.

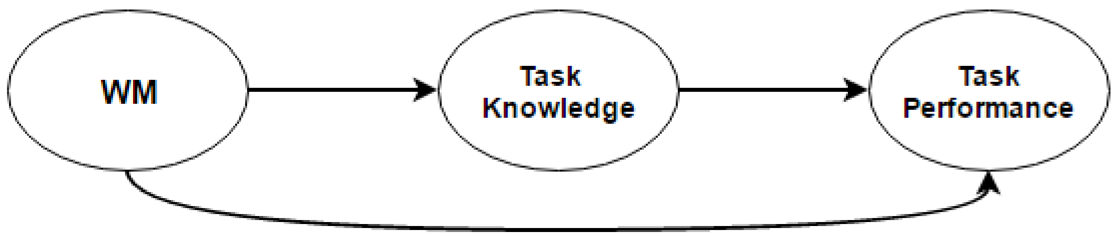


Figure 1. The direct and indirect effects of WM on task performance.

It is reasonable to posit that WMC affects learning, because people who maintain and recall more information in the presence of distractors probably process more information, and thus achieve greater levels of task knowledge. The working-memory-processing (WMP) function allows individuals to update, manipulate, and retrieve learned information, while inhibiting competing, distracting goals from the other task components they are required to complete. WMP is an important function in task performance, because the ability to retain and recall information while inhibiting distracting elements of a task itself should directly lead to better performance, even after controlling for accumulated knowledge on the task.

Bosco et al. (2015) made a similar claim when investigating the incremental validity of WM tasks over GMA. Specifically, they suggest that WM scores account for variance above GMA because typical measures of GMA tap into learned or acculturated knowledge, whereas WM measures assess more processing functions that are important for performance beyond learned knowledge. However, Carter et al. (2018) is the only study to provide empirical support for the proposed partial-mediation model. The current study was designed to replicate support for the model.

Domain Specificity of WM

Carter et al. (2018) only found support for the partial-mediating model when WM was measured using the RSPAN task. For the OSPAN assessment of WM there was no direct effect of WM on task performance. One possible explanation for the differential pattern of support is that the experimental task was highly loaded on verbal ability – participants were asked to create a properly-formatted APA-style reference page. The domain specificity explanation is that the more verbally-loaded (i.e., RSPAN) measure of WM was more reflective of formatting references than the more quantitatively-loaded measure (i.e., OSPAN).

When considering the task-switching involved with completing a task such as the RSPAN, it is intuitive that verbal ability plays an important role. For example, when a participant is presented with a sentence during the RSPAN, he/she must determine whether or not the sentence makes sense. Once a letter to be remembered is presented, s/he must hold that letter in WM while "solving" the next sentence. The higher his/her verbal ability, the fewer cognitive resources or attention s/he will need to solve the sentence, giving him/her more resources to focus on the letters that need to be recalled.

In terms of the three-embedded-components model, the letters to be recalled are in an activated LTM state for a shorter time, and therefore, these letters are less likely to decay. In contrast, someone with low verbal ability will need to spend more time and resources attending to the sentences, and will have to sacrifice more of his/her FoA processing the sentences while letters to needed to be recalled are in ALTM. Therefore, it is proposed that because WM is conceptualized and operationalized as simultaneous storage and processing of information, individuals differ in their WM capabilities based on the specific domains of the information (e.g., verbal or quantitative) that needs to be stored or processed.

To further examine the domain specificity of WM, a third purpose of this study was to test for WM domain specificity effects by including both the verbally-loaded citation task from Carter et al. (2018) and a new, quantitatively-loaded task.

Black/White Effects

Although GMA is a strong predictor of job performance, practitioners are hesitant to use cognitive ability measures because of large racial subgroup differences associated with these assessment tools (Herrnstein & Murray, 1994; Jensen, 1998; Roth, Bevier, Bobko, Switzer, & Tyler, 2001). Other racial subgroup differences exist, but because 94% of the U.S. population is either Black or White (Bureau of Labor Statistics, 2018), this discussion focuses on the Black-White differences associated with tests of GMA.

Any predictor that exhibits large subgroup differences is a concern because of the resulting adverse impact. That is, the use of cognitive ability scores for selection increases the probability of hiring successful employees, but simultaneously decreases the probability of hiring Blacks. This is labeled the “validity-diversity dilemma” (Pyburn, Ployhart, & Kravitz, 2008). As such, if g-loaded predictors are used in the selection process, an organization striving to increase diversity must seek a method to offset the adverse impact caused by the use of g-loaded predictors.

Subgroup Differences for WM. Although WM assessments correlate with GMA, there is no consensus on the magnitude of the relationship (Ackerman, Beier, & Boyle, 2005). The correlation between WM scores and GMA has ranged from 0.40 – 0.80 (Colom, Flores-Mendoza, & Rebollo, 2003; Fry & Hale, 1996; Jurden, 1995; Salthouse, Mitchell, Skovronek, &

Babcock, 1989; Tucker & Warr, 1996), with many estimates near the low end of this range. Ackerman et al. (2005) conducted a meta-analysis that examined both WM and STM tasks and their correlations with scores on GMA across 86 samples. They found that verbal WM tasks (r = 0.35), numerical WM tasks (r = 0.41), and spatial WM tasks (r = 0.59) had more moderate correlations with GMA; the average correlation of WM scores with GMA was r = 0.48.

Given that many studies have shown moderate correlations between WM and GMA, it is not surprising that prior research has shown smaller Black-White differences on WM than on GMA (Bosco et al., 2015). The current study was designed to provide further descriptive results about Black-White differences on assessments of WM. Such descriptive results are important because if future research replicates Bosco et al.'s (2015) finding that WM explains incremental variance in job performance, then smaller Black-White differences make WM assessments an obvious choice for helping to ameliorate the validity-diversity dilemma.

Moderation of the Partial Mediation Model. From a more scientific perspective, the critical issue is whether the proposed partial-mediation model is moderated by race (i.e., Black vs. Whites). Although test fairness research does not evaluate the proposed partial-mediation model, the results of this research will determine whether the model is moderated by race. Test fairness researchers examine slope differences for Black and White samples when using GMA to account for variance performance. A meta-analysis conducted by Hunter, Schmidt, and Hunter (1979) gathered 866 Black-White GMA–job performance validity pairs. The researchers found the regression lines for Blacks and Whites to be nearly identical in their analyses. Of the 866 pairs examined, only 8% (objective criteria) and 6% (subjective criteria) of the studies resulted in significant differential validity. They concluded that this proportion was just barely above the 5% expected due to chance. Others (e.g., Jensen, 1986; Jensen, 1998) have since reached similar conclusions, providing support that the GMA–performance relationship is not moderated by race.

Given that WM is moderately correlated with GMA, it is unlikely that differential prediction models using WM would find significant amounts of slope bias, thereby suggesting the effects the partial-mediating model of the effects of WM on task/job performance is not moderated by race. Furthermore, there is no logical/theoretical reason that support for the partial mediating model is moderated by race. Nevertheless, it is important to empirically test for moderation because other researchers have argued that slope bias is more prevalent than claimed due to the statistical nuances of detecting moderation (Aguinis, Beatty, Boik, & Pierce, 2005; Aguinis, Culpepper, & Pierce, 2010; Aguinis & Stone-Romero, 1997).

Overview

This study aimed to replicate Carter et al. (2018) in terms of attempting to show that WM explains variance in task performance beyond GMA, and provide support for the partial mediation model (See Figure 1). This study also extended the Carter et al. research through two important methodological changes. First, a quantitatively-oriented task was included with the verbally-loaded citation task and second, an equal number of Black and White participants were tested. The quantitative task was added to test the domain specificity hypothesis. An equal number of Blacks and Whites were sampled in order to provide further descriptive evidence that WM scores have smaller race effects than assessments of GMA, and to test whether participant race moderates support for the proposed partial-mediation model.

Chapter 2

Literature Review

General cognitive aptitude has been of interest to organizational researchers and practitioners to account for variance in organizational outcomes ever since cognitive ability measures were used for selection by the U.S. in World War I (Thomas & Scroggins, 2006). Cognitive aptitude, which has also been used interchangeably in the I-O literature as general mental ability, intelligence, cognitive ability, and psychometric GMA, has established itself as one of the best predictors of several organizational outcomes over the past few decades (Schmidt & Hunter, 2004). One specific cognitive aptitude that has also gathered support for its predictive validity in the educational and cognitive psychology realms is working memory (WM).

Genesis for this Study

Bosco, Allen, and Singh (2015) conducted four studies that tested the incremental validity of WM scores above two commonly-used measures of Gf – the Wonderlic Personnel Test (WPT) and Raven's Advanced Progressive Matrices (RAPM). In their first study, the authors gave undergraduates three tests of executive attention (EA), with two of the tests being the OSPAN and RSPAN. Additionally, participants were given the WPT and a management simulation game. The researchers found the composite of EA explained an additional 10% of variance in simulation performance after accounting for GMA, for a total of 42% variance explained.

In their second study, undergraduates were given the same predictor measures (i.e., the EA composite with two WM tasks) and a managerial simulation. In addition, in-role performance scales were given to their academic advisors. The results of Study 2 showed EA composite scores affected simulation performance, but the WPT scores did not. However, neither measure affected advisor-rated in-role performance.

The researchers' third study administered the same predictors, as well as a management-simulation task to employees of a large financial organization. In addition, employees' supervisors were given a task-performance scale as a measure of the employees' job performance. EA composite scores explained an additional 5% variance in simulation performance, and an additional 7% variance in supervisor-rated task performance above WPT scores.

Their final study replicated study 3 without the management simulation game and by changing the assessment of GMA to the RAPM. They found EA composite scores only accounted for an additional 2% (p = .09) variance in job performance after accounting for the RAPM. Bosco et al. (2015) concluded over the four studies that EA explains task performance above GMA because the underlying processes that drive both EA and task performance are independent of learned or acculturated knowledge which are explained by GMA. The current study aimed to replicate the findings of Bosco et al. (2015) in addition to extending their study by assessing task knowledge as a potential mediator of the WM → performance relationship.

More recently, Carter et al. (2018) examined the validity of scores on two WM measures over GMA for explaining task knowledge and task performance. The authors suggested that after controlling for GMA, WM had both a direct and indirect effect on task performance through task knowledge (i.e., partial mediation). Undergraduate students were given two common WM measures, RSPAN and OSPAN, followed by a progressive matrices task as a measure of GMA. Participants were then trained on how to properly format an APA-reference page, followed by a test that assessed their task knowledge of APA-reference-page formatting. Finally, to measure task performance, participants were given citations to format into an APA-reference page.

After controlling for GMA, results indicated that the RSPAN had both direct and indirect effects on task performance through task knowledge, whereas the OSPAN only had an indirect effect on task performance through task knowledge. The current study attempted to replicate the Carter et al. (2018) finding that WM explains incremental variance in task knowledge/performance over GMA, while also replicating the results that supported the partial-mediation model. In addition, the current study extended this research by adding knowledge tests and performance measures that better align with the quantitative OSPAN measure. Finally, this research examined Black-White race effects, similar to the study conducted by Bosco et al. (2015).

Working Memory

WM did not become a well-known concept in the study of human cognitive processes until Baddeley and Hitch (1974) developed their multicomponent model of WM. They considered WM to be short-term memory at work, and suggested that WM involves the ability to hold information in short-term memory while simultaneously completing some other cognitive task. They found that increasing irrelevant task demands decreased the amount of information and the quality of the information that could be recalled. After empirical investigation of the cognitive processes that occur while individuals are completing these dual tasks, Baddeley and Hitch developed the "multicomponent model of WM."

The multicomponent model is the most referenced model of WM in the literature to date (D'Esposito & Postle, 2015). However, recent advances in cognitive neuroscience have introduced models of WM that more accurately reflect how the WM processes unfold when individuals complete WM tasks. Some of these models are more abstract models of the processes of WM (similar to the multicomponent model) and others are more specific neural models which examine neural mechanisms that facilitate WM processes. The current study used as its framework a cognitive model of WM that matches the processes used to assess WM.

Activated LTM models. Activated LTM models are cognitive models that focus on the notion that information in the WM system is maintained with the help of LTM representations of the information (Lewis-Peacock & Postle, 2008). Specifically, these models all posit that when goal-relevant information is perceived (e.g., dialing a phone number), the information is attended to and remains in an activated state until it is no longer needed.

Oberauer's (2002) ALTM model consists of three embedded components: 1) The focus of attention (FoA), 2) The region of direct access (RDA), and 3) Activated long-term memory (ALTM). This model of WM suggests that when individuals are completing WM tasks, the item or information to which they are immediately attending is activated in the FoA. The FoA is most proximal to the individual and the immediate cognitive process the individual is completing

(Oberauer, 2009). The scope of the FoA is extremely limited in that the individual can directly focus on one particular piece of information.

The RDA consists of related information that is activated in order to assist in the cognitive processes carried out by the FoA. For example, when you are completing a math problem such as (7 x 2) + 9, your FoA activates whichever particular number you are focused on at any given time, while your RDA may be updating your representation from seven to 14 to 23. The RDA is not as proximal to the immediate cognitive task, but still contributes greatly to the task at hand by allowing you to keep activated the numbers you have already calculated (i.e., fourteen in this example) and enabling you to update that representation with the number that is being newly attended to (i.e., nine in this example).

Both of these components are embedded within the ALTM. The ALTM consists of more distally-activated information drawing on LTM representations. When you rehearse and maintain information "in the back of your head" while completing a simultaneous cognitive task, the information is activated in LTM. The ALTM also consists of activated information that assists in the immediate cognitive task. For example, when solving the previous math problem, (7 x 2) + 9, you draw on long-term memory by determining which operation to complete first (e.g., remembering the PEMDAS rule for order of operations). Therefore, in solving this simple math problem, all three components are activated, and information is moved to different levels of activation based on the task at hand.

Oberauer (2009) further suggests that when individuals are maintaining information in memory, they create cognitive bindings or associations between target pieces of information and with other representations in LTM in order to better maintain the target information. This is termed relational integration (Robin & Holyoak, 1995). Oberauer concludes that individuals' limit in working memory capacity is not due to an inability to store accumulating items of information, but an inability to properly connect information into a common schema. A graphical representation of the three-embedded-components model adapted from Oberaur (2002) is depicted in Figure 2.

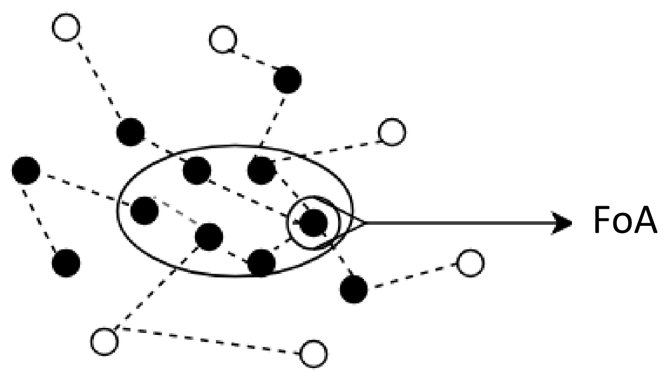


Figure 2. Oberauer's (2002) visual of the three-embedded-components model.

In Figure 2, the black circle within the white circle reflects the item in the FoA. The black circles within the oval are within the RDA. The nodes outside the RDA are in ALTM. Representations or pieces of information are linked together via associations (i.e., dotted lines). Activated representations are indicated by black nodes; non-activated representations are indicated by white nodes. The movement of information between levels of activation can be further clarified by considering how WM is typically measured.

General Mental Ability (GMA)

Charles Spearman (1904) set the framework for the study of human intelligence with his research involving young children's performance across a wide array of cognitive ability tests. In order to describe the positive manifold among cognitive ability tests, Spearman coined the "general factor," or psychometric GMA. This GMA is meant to represent the interrelationships across specific cognitive ability tests, and more generally speaking, to represent intelligence.

Spearman referred to aptitude for the lower level tasks (e.g., spatial ability, verbal reasoning, and psychomotor ability) as specific abilities or “s” abilities, which were considered hierarchically beneath GMA.

Through the years researchers sought to better understand GMA and the structure of cognitive abilities. Ultimately, Carroll’s (1993) model of psychometric GMA became the most accepted model, but due to the overlap between Carroll’s model and prior research by Horn and Catell (1966), Carroll’s model is typically referred to as the Cattell-Horn-Carroll (CHC) model (McGrew, 2005).

In the CHC model (see Figure 3), human cognitive abilities are conceptualized as three levels or strata: (1) general cognitive ability is the third level, or GMA, (2) “broad abilities” representing second-level abilities of GMA, and (3) “specific abilities” are even more granular, first-level abilities that represent information-processing capabilities that underlie broad abilities.

The second-order broad abilities include fluid reasoning (Gf), quantitative knowledge (Gq), reading and writing (Grw), comprehension knowledge (Gc), short-term memory (Gsm), long-term storage and retrieval (Glr), visual and auditory processing (Gv, Ga), and processing speed (Gs).

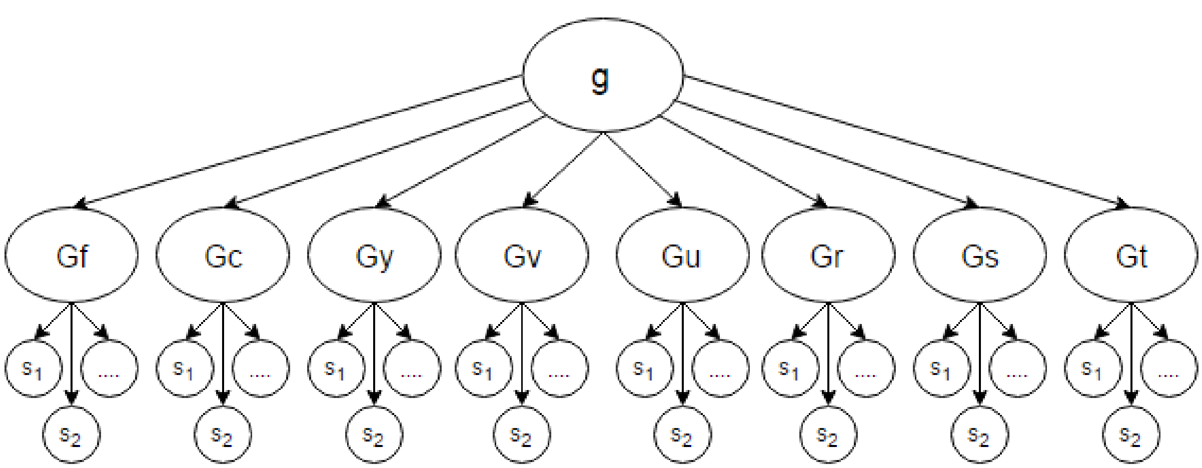


Figure 3. Carroll's (1993) three-stratum theory. Broad abilities are fluid intelligence (Gf), crystallized intelligence (Gc), general memory and learning (Gy), broad visual perception (Gv), broad auditory perception (Gu), broad retrieval ability (Gr), broad cognitive speediness (Gs), and processing speed (Gt). There are over 80 specific abilities (e.g., s_1, s_2, etc.) hierarchically below the broad abilities.

Is WM Part of the GMA Hierarchy?

Earlier versions of the CHC model (e.g., Figure 3) do not include any reference to the construct of WM. The main reason for this is that a very small number of the 400 studies that Carroll analyzed included any measure of WM. There is some debate as to whether WM is a specific ability of GMA, a higher-level processing construct closely related to GMA, or separate from GMA altogether.

Some researchers, such as Schneider and McGrew (2012) have considered WM to be a specific ability (i.e., first-order ability) of STM (i.e., a second-order ability) in the CHC model (Schneider & McGrew, 2012) due to the fact that the construct of WM is said to reflect STM capabilities (i.e., rather than long-term storage). Other researchers (e.g., Kane et al., 2007), argue that WM is not a specific ability, but rather reflects cognitive processes related to all cognitive abilities. Bosco et al. (2015) make the claim that WM is separate from GMA and how psychologists measure it. The researchers claim GMA reflects learned or acculturated knowledge that is affected by environmental and sociocultural factors, whereas WM does not. Regardless of the stance one takes with regard to WM and the GMA hierarchy, measures of GMA that psychologists use for selection will typically be moderately correlated with GMA.

Many researchers use fluid reasoning (Gf) or comprehension knowledge (Gc) as an operational definition of GMA rather than using a complete assessment of GMA (e.g., Roberts, Goff, Anjoul, Kyllonen, Pallier, & Stankov, 2000) due to typically high correlations between these specific abilities and overall GMA. For example, the Armed Services Vocational Aptitude

Battery (ASVAB) is a popular measure of aptitude consisting of ten subscales, seven of which measure comprehension or learned knowledge (Gc). A list of the subscales measured by the ASVAB can be found in Appendix A. Another commonly used assessment of GMA is the Wonderlic Personnel Test (WPT) – a measure that can be administered in 12 minutes or less (Wonderlic, 2018). This assessment tool includes items that evaluate one's ability to engage in abstract reasoning or novel problem solving (Gf).

Because these two specific abilities (i.e., Gc and Gf) are highly saturated with overall GMA, they are often used as operational definitions of general cognitive ability. Although some have debated whether WM is a lower-level ability in the hierarchy of GMA, closely related to GMA, or separate from GMA, it still does not represent GMA as well as Gc and Gf. Therefore, because WM is not as saturated with GMA, it would probably not be used as an operational definition of GMA or contributing factor to a composite that represents GMA.

Since WM is not used to create composites of GMA or to serve as the operational definition of GMA, WM may have the ability to explain incremental variance above typically-used measures of GMA. Despite past efforts that have failed to show the incremental validity of other specific abilities, WM has both a lower g-loading and is associated with processes important for affecting relevant organizational outcomes. Indeed, rehearsing, maintaining, updating, task-switching, and inhibiting are not only processes uncaptured by current measures of GMA, but they are also processes important for affecting learning and performance.

Incremental Validity of WM

The accumulated evidence regarding the validity of GMA for explaining job performance is so compelling (Colquitt, LePine, Noe, 2000; Hunter, 1986; Ree & Carretta, 1998; Ree & Earles, 1991; Roth, Huffcutt, & Bobko, 2003) that Schmidt and Hunter (2004) to conclude that

GMA accounts for variance in performance "better than any other ability, trait, or disposition and better than job experience" (p. 162). GMA also affects training performance. Hunter (1986) concluded that superior ability to acquire and retain/access information manifests as high scores on training knowledge assessments and overall training success. These higher scores on training assessments manifest as higher levels of on-the-job task performance and overall job performance (e.g., supervisor ratings).

Selection research has evolved to the point that research examining newer antecedents of job performance usually check the incremental validity of the new measure's scores over some GMA score (e.g., Oh, Le, Whitman, Kim, Yoo, Hwang, & Kim, 2014; Ohme & Zacher, 2015; Van Iddekinge, Aguinis, Mackey, & DeOrtentiis, 2017), making GMA the standard by which to judge the utility of a new predictor. Therefore, when considering WM as a potential antecedent of the learning of task knowledge and performance, it is worth considering whether WM can affect these outcomes even after controlling for GMA.

Learning outcomes. Research has shown that WM scores affect critical learning outcomes. Specifically, WM has been positively associated with learning computational skills (Wilson & Swanson, 2001) and knowledge of mathematical problem-solving (Alloway & Passolunghi, 2010; Friso-van den Bos, Van der Ven, Kroesbergen, & Van Luit, 2013; Gersten, Jordan, & Flojo, 2005; Swanson, 2004). WM has also accounted for variance in the ability to: a) remember lengthy instructions (Gathercole, Lamont, & Alloway, 2006), b) learn a second language (DeKeyser & Koeth, 2011; Linck & Weiss, 2015; Wen & Skehan, 2011), and c) process visually-presented material (Logie, Gilhooly, & Wynn, 1994).

In summary, evidence clearly supports that WM can affect learning criteria, but the research above does not address whether WM accounts for variance learning outcomes over and above GMA. Luo, Thompson, and Detterman (2006) found that WM scores accounted for variance in school achievement beyond that of fluid reasoning (i.e., a specific ability highly reflective of GMA). In a sample of undergraduate students, Krumm, Ziegler, and Buehner (2008) used WM scores to account for incremental variance in language learning above GMA. Interestingly, Alloway and Alloway (2010) examined school-aged children across six years and concluded that WM affects overall learning *more* than GMA. This is consistent with Alloway's (2009) findings that WM affected learning more than GMA for children with disabilities. Finally, as previously mentioned, Carter et al. (2018) found WM to affect knowledge of a newly learned verbally-loaded task after controlling for fluid intelligence. In conclusion, research has supported the premise that WM scores can account for additional variance in learning outcomes above GMA.

Task/job performance. Few studies have compared the predictive validity of both WM and GMA for task/job performance. Wolfe, Alderton, Larson, and Held (1995) found that adding a WM measure to the ASVAB accounted for additional variance in task and simulation performance across 19 military technical schools examined. Specifically, their criteria across the schools consisted of task performance in the form of tracking simulator performance, a typing performance test, an air-traffic-control-operator performance test, an aviation-electrician performance test, and fire-control-person-radar tests, among others. Furthermore, Held and Wolfe (1997) found WM scores explained an additional 5% - 10% variance in three different measures of task performance: basic Air Force modules that simulate real tower operations, advanced Air Force modules that simulate real tower operations, and Operations Specialist ratings which are based on performance on a job-like positioning and targeting task.

As detailed earlier, Bosco et al. (2015) used WM scores to explain 2% – 10% of additional variance in supervisor-rated performance and simulation performance. This study aimed to extend Bosco et al.'s findings by examining task knowledge as a partial mediator of the WM – performance relationship.

Partial Mediating Model of WM

In a landmark study, Hunter (1986) proposed and empirically tested one explanation for the effect of GMA on performance. Specifically, he suggested that GMA affects job performance so well because GMA affects job knowledge well, and job knowledge correlates highly with job performance. That is, people of higher intelligence are better able to learn information relevant for the job, and thus apply this to execute the job effectively. His results indicated a large indirect effect of GMA on job performance through job knowledge, as expected. This mediating role of knowledge in the GMA–performance relationship has since been supported by numerous researchers (Borman, Hanson, Oppler, & Pulakos, 1993; Borman, White, Pulakos, & Oppler, 1991; Schmidt, 2002; Schmidt & Hunter, 1992).

Hunter (1986) suggested that GMA has a direct effect on job performance, such that cognitive ability allows people to perform job responsibilities more effectively, even when the job requires the use of skills or knowledge they have not yet learned from training or past experience. In other words, GMA has a direct effect on job performance after controlling for job knowledge. Hunter's (1986) findings supported this theory, and he concluded that more intelligent individuals "are faster at cognitive operations on the job, are better able to prioritize between conflicting rules, are better able to adapt old procedures to altered situations, are better able to innovate to meet unexpected problems, and are better able to learn procedures quickly as the job changes over time"(p. 354). This direct effect has also been empirically supported by others (e.g., Borman et al., 1991; Schmidt, Hunter, & Outerbridge, 1986), suggesting task/job knowledge partially mediates the relationship between GMA and performance.

Considering WM as a cognitive ability that past research has supported as an antecedent of both learning and performance outcomes, it was expected that WM has the same direct and indirect effects on task performance through task knowledge. As detailed earlier, Carter et al. (2018) was the only study to test and support the partial mediation model using WM instead of GMA providing support for the dual functionality of WMC and WMP.

Dual functionality of WM. Starting with Baddeley and Hitch (1974), researchers have considered WM to serve two functions: capacity and processing. Working memory capacity (WMC) refers to the amount of accurate information that can be held in WM while performing some sort of simultaneous task. Working memory processing (WMP) refers to the ability to update information, task-switch, and inhibit distracting stimuli while successfully prioritizing the cognitive processes required to achieve a goal. Both of these functions are intertwined such that individuals with better WMP are able to free up resources to store more accurate information in WMC; higher WMC individuals are less taxed by information in WMC, and can therefore allocate resources that allow for a more efficient WMP.

The dual functionality perspective of WM provides the underlying logic for the proposed partial-mediating model for WM. WMC reflects the amount of information that can be maintained and recalled in the presence of distractors. It is intuitive that the WMC function would be important for the acquisition of any task knowledge, because acquiring knowledge requires individuals to remember presented information long enough for it to be recalled when prompted. This model can then be extended to propose that WMC would affect task performance because of the well-known relationship between knowledge and performance (Hunter, 1986).

This results in a mediated model whereby WM affects task knowledge and task knowledge affects task performance due to the successful application of this learned knowledge.

WMP reflects the ability to manipulate learned information, update representations, and process cognitive tasks while maintaining recently-presented information in memory. It is suggested that processes associated with WMP are important for task performance, independent of task knowledge. That is, even when two individuals have the same amount of task knowledge, the person with more efficient WM will perform better on the task because of the function of WMP. Therefore, when accounting for both WMC and WMP, this study hypothesized that WM would have both a direct effect and indirect effect on task performance, partially mediated by task knowledge (Figure 1).

Domain Specificity of WM

The results of Carter et al. (2018) that the partial-mediating model was only supported using the RSPAN may have been due to the fact that their performance task was verbally loaded. They suggested verbal WM affects verbal knowledge and task performance more than quantitative WM because WM may not be domain-general (i.e., the type of WM measure may be relevant for the predictive accuracy of WM scores). As stated earlier, this study was designed to test this hypothesis by using measures of both verbal and quantitative task knowledge and performance.

Some research has supported the domain-specificity argument. This literature mostly involves examining the correlation between WM tasks of different domains as well as the differential validity for WM in affecting outcomes of different domains. Daneman and Tardif (1987) found that verbal WM affected performance on a verbal-ability measure, while spatial WM did not. In line with these findings, Morrell and Park (1993) also found that a verbal, complex span task correlated strongly with scores on performance measures of text comprehension, but spatial WM tasks did not. Spatial WM affected performance on model-assembling task (i.e., a spatial task), but verbal WM tasks did not.

Shah and Miyake (1996) gave participants both spatial and verbal WM span tasks and measures of spatial and verbal ability. They found that spatial WM correlated with spatial ability but not with verbal ability, and verbal WM correlated with verbal but not spatial ability. They concluded that different resources are needed to complete WM tasks that tax a domain in which participants have a low ability than are needed for a domain in which they have a high ability. These findings were later supported by Friedman and Miyake (2000).

Domain specificity is further supported by the findings of Alloway, Gathercole, and Pickering (2006) when examining several verbal and visuospatial WM and STM tasks with children between the ages of four and eleven. They found domain specificity for the "storage" aspects (measured by STM tasks) of WM such that verbal STM loaded higher onto verbal WM, and visuospatial STM loaded higher onto visuospatial WM.

Other researchers have pushed back on the domain specificity of WM, claiming that WM is more of a general system that facilitates processing across a wide range of domains. Kane et al. (2007) acknowledge there are domain-specific processes associated with WM, but that these domain-specific studies ignore the key domain-general processes associated with WM. They further claim that the conclusions of other researchers were based on studies using the typical university student population where the range of general cognitive ability is restricted. That is, many of these studies use a population of students who were selected based on cognitive ability (e.g., SAT scores), and therefore the variation obtained on cognitive ability tasks would be more attributable to domain-specific processes than more general processes. Further, the authors point out that previous studies typically only compare two or three WM tasks, disallowing the study of the complex processes of WM.

To support their point, Kane et al. conducted a study that examined how six different WM tasks and six different STM tasks loaded onto latent-factor models in both a university sample and a public sample. Their findings indicated that 70 to 85% of the variance in spatial and verbal WM was shared between the two, and that the best model suggested verbal and spatial WM both load onto a higher-order WM factor. However, the researchers found the verbal and spatial STM tasks, which they considered to reflect the storage component of WM, to only share 40% indicating that domain-specific processes are associated with those STM tasks. The authors concluded by acknowledging that domain-specific storage mechanisms play a role in WM tasks, but that domain-general processes of WM are responsible for the moderate-to-high correlation between WM and GMA (Kane et al., 2007).

These findings are not particularly surprising, as the more cognitive ability measures that are administered or the more cognitive processes surveyed across tasks, the more these scores will reflect a higher-order general cognitive ability factor. The conclusion of Kane et al. (2007) that domain-general processes (e.g., attentional control) are responsible for the correlations with GMA are important for this study, as one of the goals of this study was to examine the incremental validity of WM above GMA. Extending their conclusion, it would seem intuitive to suggest that the domain-specific variance associated with WM is important for dissociating WM from GMA and accounting for variance not accounted for by GMA. Although the domain-specificity debate has not yet been fully reconciled, this study adopted the theoretical perspective that performance on WM tasks is dependent on the domain content of the distracting stimuli.

Although a more applied perspective is taken in the discussion, two important considerations must be made when reviewing the literature involving the comparison of a measure against GMA. The first, which has been discussed in detail, is the predictive validity of WM for affecting task knowledge and performance. The second consideration is the extent to which WM has differential validity based on race and whether or not there are subgroup mean differences on WM scores.

Black/White Differences

Even since Charles Spearman (1904) first proposed a theory of general intelligence, it was apparent there were racial-subgroup differences on various tests across various cognitive abilities. After reviewing studies that examined the cognitive abilities of Black and White Americans, Spearman (1927) pointed out the IQ differences between the two racial groups, even when education, environment, and social class were controlled. Since then, differences in IQ have been compared for various different demographic subgroups.

Numerous studies have found that Asian-Americans tend to score between 0.20 and 0.33 standard deviations (SD) higher than Whites on measures of GMA (Hough, Oswald, & Ployhart, 2001; Jensen, 1998). Hispanic-Americans' scores on tests of GMA are about 0.58 to 0.83 SD lower than Whites, but typically higher than Blacks (Roth et al., 2003).

The most noteworthy subgroup difference associated with tests of GMA is the Black-White race difference. On standardized tests of GMA (i.e., with a mean of 100 and SD of 15), Blacks tend to have mean scores around 85, one SD lower than the mean of 100 for Whites, (Jensen, 1998; Roth, Bevier, Bobko, Switzer, & Tyler, 2001). The one SD difference between Blacks and Whites has been reproduced in hundreds of studies over decades of research (Herrnstein & Murray, 1994). Jensen (1998) conducted a meta-analysis of 156 studies and placed the mean standardized difference at 1.08 SD. He concluded the true Black-White standardized difference to be between 0.72 and 1.44 SD (i.e., one SD on either side of the mean). This range is supported by numerous meta-analytic estimates. On the low end of this estimate, Schmitt, Clause, and Pulakos (1996) found a 0.83 SD difference between Blacks and Whites. However, most researchers have concluded the difference to be higher, such as 1.08 (Hernstein & Murray, 1994) or 1.10 (Roth et al., 2001). Although the exact estimate is sometimes disputed, a significant mean difference between the Black and White populations on GMA is accepted.

Validity-Diversity Dilemma

The validity-diversity dilemma is a tradeoff that exists for organizations during the selection process. Organizations often have to balance the validity of their selection measures to predict job performance with the amount of diversity that is present among the chosen candidates for a job. Because many of the selection measures with more predictive accuracy tend to exhibit larger Black-White differences, relying on these selection measures often results in the selection of a larger proportion of Whites than Blacks. For example, if an organization has a specific cut score of 100 when using a standardized scores on a measure of GMA, many more of the White applicants would likely reach that cut score than Blacks, resulting in differential selection rates. These differential selection rates for certain groups result in adverse impact (Equal Employment Opportunity Commission, 1978), and can be problematic for organizations financially, politically, and legally (Pyburn et al., 2008).

Along these lines, when examining whether WM scores are useful in accounting for incremental variance in learning and performance, it is also important to consider the extent to which race plays a role in the relationship between WM, knowledge, and performance.

Moderation of the Partial-Mediation Model

The mean race differences on cognitive ability have led researchers to suggest GMA has differential validity for Blacks and Whites such that the assessment tools do not predict outcomes for Blacks as strongly as they do for Whites. However, decades of research has supported the conclusion that there is typically no differential validity between Blacks and Whites when using GMA to explain various outcomes (Jensen, 1998). Although mean race differences on GMAdo exist, these differences are also reflected in outcomes such as educational achievement (Jencks & Phillips, 1998), job knowledge (McKay & McDaniel, 2003; McKay & McDaniel, 2006), training performance (McKay & McDaniel, 2006), and both objective measures of job performance (Roth, Huffcuft, & Bobko, 2003) and ratings (Kraiger & Ford, 1985; Waldman & Avolio, 1991).

The few studies that have suggested differential validity for GMA have done so with small samples, which is problematic when detecting differential validity between groups (Hunter & Hunter, 1984). As Hunter et al. (1979) explain, range restriction, non-normality, and non-independence are all biases that can lead to the false detection of differential prediction (i.e., a Type I error). For a review of the arguments and vast empirical evidence refuting differential validity of GMA, see Jensen (1998, Chapter 11).

Because WM is moderately correlated with GMA, it is expected that there would not be differential prediction between Blacks and Whites using WM scores. More specifically, no theory was found that has suggested WM would affect any performance outcomes more or less for Blacks compared to Whites. Considering the paths proposed in Figure 1, the theory of how WM processes are responsible for both learning and task performance should hold, regardless of any racial-group classification. Not only did the authors not find studies examining differential validity of WM for affecting performance outcomes, but even studies examining *mean* race differences on WM are scarce. This study was designed to contribute to the literature by testing for the differential validity of WM and examining mean race differences associated with WM.

Racial Differences on Specific Cognitive Abilities

Spearman's Hypothesis. Models by Spearman (1927) and Carroll (1993) suggest each specific ability has a different loading on the GMA factor, and some of these specific ability loadings are relatively small. Spearman suggested that the magnitude of mean race differences on any given specific ability is dependent upon the factor loading of that specific ability on GMA.

Jensen (1998) later examined these factor loadings across 149 cognitive tests (to test "Spearman's hypothesis") and found that some of the specific abilities exhibited smaller or larger race differences based on how much the ability loads onto GMA. For example, Jensen found larger Black-White differences on tests of spatial visualization (e.g., object rotation) than typically found with other cognitive tests. Further, smaller race differences were observed on tests of short-term memory (STM), such as Digit Span and Coding. Others have also supported Spearman's hypothesis with meta-analyses (e.g., Repko, 2011; Te Nijenhuis & Dragt, 2010).

Dahlke and Sackett (2017) further provided evidence in support of this hypothesis when examining mean race differences exhibited by different predictors such as specific cognitive abilities, biodata, integrity tests, interviews, personality, emotional stability, and many others. They found that the predictors' g-loadings had a correlation of 0.84 with the Black-White standardized differences on these measures (i.e., cohen's *d*). This provided further support for the common theory that predictors exhibit race differences based on their cognitive saturation.

Working memory and subgroup differences. Based on the overwhelming support of Spearman's hypothesis in the literature, subgroup differences on WM are expected to be observed to the extent that it is g-loaded. So, if there are large g-loadings, Whites would score higher than Blacks, based on past empirical findings. That raises the question – how g-loaded is WM?

There is no consistent agreement as to how much WM correlates with GMA (Ackerman, Beier, & Boyle, 2005). One reason for this is that many meta-analytic studies claiming to examine WM – GMA correlations (e.g., Conway, Macnamara, & Engel, 2013) are actually examining some tasks that are not WM tasks and tasks that do not measure GMA, but rather measure fluid reasoning, arithmetic, comprehension knowledge, or some other broad ability. Furthermore, WM tests are administered less often than STM tests, which can be found on the WAIS. Hence, less literature available has assessed the WM – GMA correlation.

Cognitive ability correlations are also affected by range restriction in that often the civilian or college admissions samples used to assess correlations between GMA and other variables do not exhibit the same range of cognitive ability as the general population (e.g., Berry, Cullen & Meyer, 2014). GMA correlations are further affected by how the tests are administered (e.g., whether a time constraint is placed on test takers), and the content overlap of the tests, among other factors. The correlation between WM and GMA has been cited to range anywhere from 0.40 to 0.80 (Colom, Flores-Mendoza, & Rebollo, 2003; Fry & Hale, 1996; Jurden, 1995; Salthouse, Mitchell, Skovronek, & Babcock, 1989; Tucker & Warr, 1996), with many estimates at the low end of this range.

Ackerman et al. (2005) conducted a meta-analysis that examined both WM and STM tasks and their correlations with GMA across 86 samples. The found that verbal WM tasks (r = 0.35), numerical WM tasks (r = 0.41), and spatial WM tasks (r = 0.59) had modest correlation with tasks of GMA. The average correlation of WM tasks with tasks of GMA was 0.48.

Unsurprisingly, spatial WM had the highest g-loading. As alluded to earlier, Jensen (1998) found large Black-White mean differences in spatial visualization, and therefore would suggest spatial content on a WM task to result in larger subgroup differences.

Some research has examined whether the closely-related construct of STM exhibits smaller racial subgroup differences than measures of GMA. Mayfield and Reynolds (1997) gave fourteen different STM tasks to over 1,000 Black and White children between 5 and 19 years of age to examine the subgroup differences associated with the scores on these measures. They found the highest Cohen's *d* exhibited in favor of Whites was 0.06, which is much lower than the typically reported *d* for general cognitive ability. Verive and McDaniel (1996) conducted a meta-analysis that analyzed 31 Black-White effect sizes for STM reported between the '50s and '90s. With their total sample size of 27,973, they found the average effect size to be 0.42 in favor of Whites, a value that is still lower than typically reported among effect sizes for GMA.

Bosco et al. (2015) was the only study found that directly compared the Black-White effect sizes of WM scores with those of GMA. In their first three studies, they found WM scores (d = 0.67, 0.72, 0.53) to exhibit smaller race differences than the WPT (d = 1.01, 1.16, 1.48). Additionally, the same WM scores (d = 0.86) exhibited smaller race differences than the RAPM (d = 1.20) in their fourth study.

Given that many studies have shown moderate correlations WM and GMA, the author expected to find smaller Black-White subgroup differences associated with WM compared with those of GMA. This proposition was further supported by evidence for smaller subgroup differences using STM and the study conducted by Bosco et al. (2015). Smaller Black-White differences would make the use of WM tasks more appealing for I-O researchers and practitioners, as smaller Black-White differences would help address the validity-diversity dilemma, assuming that WM has predictive validity, as discussed earlier. That said, the fourth purpose of this study was to examine the Black-White differences associated with WM measures compared to those of GMA.

The Current Study

This study was a laboratory study that examined the use of WM (i.e., RSPAN and OSPAN) as an antecedent of task knowledge and task performance compared to that of GMA, which was represented by fluid reasoning or Gf (i.e., a progressive matrices task). Gf was used instead of a full measure of GMA because a large cognitive test battery would have been required to obtain a full assessment of GMA. Since Gf is the second-order factor in the GMA hierarchy that is most saturated with GMA (Carroll, 1993), Gf was used as the operational definition of GMA.

GMA is typically not estimated using any measure of WM, and therefore it was hypothesized that the WM scores would capture variance associated with important cognitive processes that are not captured by GMA. Further, past research has provided evidence that WM scores are able to account for variance in both knowledge and performance outcomes. Based on these arguments, the following hypotheses were developed:

Hypothesis 1: After controlling for fluid intelligence (Gf), WM will account for unique variance in task knowledge.

Hypothesis 2: After controlling for Gf, WM will account for unique variance in task performance.

The dual functionality perspective of WM suggests that WM has two functions: WMC and WMP. In this study, WMC enabled individuals to store higher quality and quantity of information in the presence of distracting stimuli, which in turn enabled them to better learn and recall knowledge presented in a training video (i.e., task knowledge). WMP enabled individuals to effectively perform tasks while updating and manipulating stored representations, task-switching, and inhibiting distracting goals. In this study, WMP enables individuals to more effectively complete a task that applied their learned knowledge (i.e., task performance), even after controlling for task knowledge. Therefore, WM was expected to have both a direct and indirect effect on task performance through task knowledge.

Hypothesis 3: Controlling for Gf, task knowledge will partially mediate the relationship between WM and task performance.

It was argued that individuals in this study who have high verbal ability would be required to spend less time on the distracting verbal stimuli of the RSPAN, and will therefore get higher scores on the RSPAN. Due to the high verbal load of both measures, high verbal-ability were expected to score better on APA formatting knowledge tests and tasks. The same can be said of the OSPAN and the highly quantitatively-loaded knowledge test and task. Using this logic, RSPAN was expected to affect performance on the verbal test and task more than the OSPAN, while the OSPAN was presumed to affect performance on the quantitative test and task more than the RSPAN. Therefore, it was hypothesized:

Hypothesis 4: Because WM is domain specific, WM scores will better account for variance in knowledge and learning outcomes that align with the content in the WM measure.

Hypothesis 4a: Specifically, WM is domain specific such that verbal WM affects verbal task knowledge and performance more than quantitative WM.

Hypothesis 4b: WM is domain specific such that quantitative WM affects quantitative task knowledge and performance more than verbal WM.

Past research has shown that specific abilities exhibit race differences to the extent they are g-loaded (correlated with GMA). Earlier research has indicted that WM tends to have more moderate correlations with GMA than highly-reflective specific abilities such as Gf. This sample of this study consisted of half Black and half White participants, providing the opportunity to examine mean race differences associated with WM compared to those of Gf. Due to more moderate WM-GMA correlations, the following hypothesis emerged:

Hypothesis 5: WM scores will exhibit smaller subgroup differences than Gf scores.

Lastly, race was included as a moderator of the partial mediation model in order to examine whether or not there is differential validity for WM in affecting task knowledge or task performance. However, based on lack of evidence for differential prediction using GMA and the moderation correlation between WM and GMA, the model was not expected to be moderated by race.

Chapter 3

Method

I. Participants

Participants for the proposed study included 210 undergraduate and graduate students from a Southeastern university. The sample included 50% Black students (n = 105) and 50% White students (n = 105). Recruitment was conducted in three ways. First, an advertisement was posted on the psychology department bulletin boards and on the psychology department online psychology experiment system. Second, an announcement was sent out weekly to the graduate student email list that advertised the study and pointed out that both Black and White students are needed. Third, the researcher actively recruited from clubs and organizations on campus that have larger Black student representation. Specifically, the researcher reached out to club leaders and requested permission to present the study at club meetings in order to garner interest. Participants were told the study lasts 90 minutes and were offered $15 for their participation. Those students recruited from the psychology department were given the option to count participation toward extra course credit points instead of receiving cash. Students were also told the top performers on the experimental task would be entered into a lottery for two prizes – a $50 gift card and $25 gift card.

II. Procedure

Participants signed up for the study by using the psychology experiment management system or emailing the researcher to reserve a timeslot, depending on how they were recruited. Participants were scheduled to complete the study in groups of one to eight people. After signing consent forms, each participant was assigned to his/her own computer. Prior to the experimental tasks, participants were given a measure of WM, Gf, and another measure of WM. The quantitative and verbal WM tasks were randomly counterbalanced before and after the measure of Gf to ensure there were no order effects. After the pre-task assessments, participants were asked to take an active break where they were prompted to close their eyes and think of a relaxing environment for five minutes in order to prevent fatigue.

Each group reporting to the testing location was randomly assigned to a task condition. As the end of the experiment approached, random assignment was suspended in order to balance the number of participants in the four cells of the design, a 2 (Race) x 2 (Task) factorial. Participants were administered their respective experimental tasks using the same procedure: (1) A pre-test assessed participants' baseline knowledge of either the verbal or quantitative task. (2) Either a six-and-a-half-minute training video or nine-and-a-half-minute training video was presented to participants that provided a learning opportunity where they could acquire task knowledge. (3) A post-test assessed participants' task knowledge of the verbal or quantitative task. (4) Participants were given a task that applied their newly-learned knowledge in order to measure task performance. The knowledge tests (i.e., pre-tests and post-tests) were counterbalanced so that participants were randomly given one of the knowledge tests before and after the training. Finally, participants were debriefed and thanked for their participation. All of the measures took about 90 minutes in total.

Throughout the recruiting process and laboratory protocol, the researcher did everything possible to not activate the stereotype threat that has been associated with race and intelligence measures. This was accomplished by simply calling the tasks “cognitive tasks”, “puzzle tasks”, and “learning tasks” and avoiding terms such as intelligence, ability, or IQ.

Verbal experimental task. The verbal experimental task was identical to the one administered by Carter et al. (2018). Participants were presented with a nine-minute, thirty-one second video that instructed them on how to create a properly-formatted APA-style reference page. The video provided a detailed demonstration highlighting APA formatting rules with narration, thereby providing a learning opportunity for the participants. The complete video can be found at https://www.youtube.com/watch?v=XNGHw6Gwt7Q.

After completing the post-training knowledge assessment, participants completed the performance task. Participants were given typed information about six journal articles and were asked to create a properly-formatted APA style reference page using their computer. Each journal article given to them included the journal title, article title, author name(s), date published, and page numbers. The exact handout given to them can be found in Appendix B. Participants were given 15 minutes to type up the reference page for all six articles.

At the end of the 15 minutes, participants were asked to save their work, and documents were printed by the researcher(s) for later scoring.

Quantitative experimental task. Participants were presented with a six-minute, thirty-five-second video instructing them on how to calculate the allocation of federal taxes given a starting gross annual salary. The Khan Academy© video provided a detailed demonstration of how to calculate taxable income, federal taxes, FICA (social security and Medicare) taxes, and net take-home pay. The video tutorial can be found at https://www.youtube.com/watch?v=DtCfOMl3qo0.

After completing the post-training knowledge assessment, participants completed the performance task. Specifically, participants were given information about the gross annual salary, personal exemptions, and standard deductions of three different people. For each person, the participants were asked to calculate the taxable income, federal taxes, social security taxes, Medicare taxes, FICA taxes contributed by employer, and net take-home pay. The exact handout given to the participants can be found in Appendix C. All participants were provided with both paper and a calculator to complete their calculations. Participants were given 15 minutes to calculate the desired information and write answers on the worksheet provided. At the end of the 15 minutes, participants were asked to hand in their work, and worksheets were collected by the researcher for later scoring.

III. Measures

APA-style knowledge tests. Each APA-style knowledge test consisted of 10 multiple-choice questions addressing the detailed rules and guidelines of creating an APA-style reference page https://owl.english.purdue.edu/owl/resource/583/03/. Each question presented participants with four or five response options. Participants were given eight minutes to complete each test. Tests were scored based on number of correct responses. The APA knowledge-test items can be found in Appendix D.

Quantitative knowledge tests. The pre- and post- quantitative knowledge tests each contained 10 multiple-choice questions with four or five response options. Participants were given 12 minutes to complete each test. Tests were scored based on number of correct responses. The quantitative knowledge test items can be found in Appendix E.

Pilot testing was conducted for both the APA and quantitative knowledge tests. Pilot testing allowed for an update to the APA knowledge test from Carter et al. (2018) in order to eliminate items that were too easy. Also, pilot testing provided time and difficulty estimates for both the APA knowledge tests and the newly-developed federal taxes quantitative-knowledge test. Some items (i.e., items assessing certain material of the subject matter) were more difficult than others. Therefore, for both the pretest and posttest, participants randomly received ten test items of the pool of twenty items. This randomization was done within item type such that participants got the same number of each type of item on both the pretest and the posttest.

Automated RSPAN. The automated reading span (RSPAN) task is a computer-based complex span task which assesses WM by presenting participants with information to be remembered as well as distracting information designed to interfere with recall. Participants were first given instruction for the RSPAN followed by three practice blocks. Throughout the practice blocks, participants were presented with shortened versions of the test blocks in an attempt to give participants familiarity with how the stimuli are presented. Furthermore, the practice blocks served to establish a baseline reaction time for each participant. Response latency established at baseline was used to determine the length of time the test stimuli were presented to participants.

For the test blocks, participants were presented with a short sentence such as: “Alice likes to trampoline to the grocery store every weekend”. The sentence remained on the screen for a short amount of time, after which the participants was asked to indicate whether or not the sentence makes sense semantically. In the above example, the participants should have correctly indicated the sentence does not make sense by clicking “False” on the subsequent screen. Once they chose an option, a target letter was presented to them, which they have been told to remember for later recall at the end of the block. After the target letter was briefly presented, the participants evaluated another sentence, followed by another target letter. This process continued until 3-7 letters were presented in each block. There were two sets of each block size (i.e., two sets of 3, 4, 5, 6, 7) for a total of ten blocks. Block sizes were not constant, as the unpredictability of the block size further taxed the WM system.

At the end of the block, participants were presented with a recall screen, where they indicated which letters have been presented, *in the order they were presented.* Participants were told they can hold the place of a letter they have forgotten by inputting a "BLANK" in its place. The participants were not given a time limit for the recall screen. After the recall screen, participants started a new block of sentences and letters until all ten blocks are completed. Ten blocks of an average of five letters for each block resulted in 50 total target letters. Participants earned one point for every letter they remembered in the correct order. Therefore, participants could have earned a total of 50 points on each WM task.

In order to ensure participants were properly focusing on both the recall and the successful performance on the sentences, participants were told that in order for their results to be counted, they must respond to the sentences with 85% accuracy or better. They were provided with feedback on the computer screen informing them of their rate of correct responses throughout the test blocks. It has been recommended by past researchers (e.g., Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005) to not use participants' data if they are not properly investing resources in both the distracting task and the letter recall. This study followed these guidelines. The internal consistency is high (e.g., Conway et al., 2005) for the RSPAN task. The total time to complete the RSPAN was about 12 minutes.

Automated OSPAN. The automated operation span (OSPAN) was different from the RSPAN in one way – the distracting stimuli were math operations rather than sentences. Therefore, participants were presented with three practice blocks followed by ten test blocks. During the test blocks, participants were presented with simple math operations (e.g., 10 X 2 – 3). They were asked to determine whether or not the subsequent number presented on screen reflects the correct answer. In the above example, if "17" were presented, participants should have correctly clicked "TRUE" on the screen. After their response to the math problem, they were presented with a target letter they are told to remember for later recall. They were then presented with another math problem, followed by another target letter, and so on. Like the RSPAN, the OSPAN consisted of ten blocks which presented them with 3-7 target letters. After all letters in the block are presented, participants reported their letters on the recall screen.

Scoring was identical to the RSPAN in that participants received a point for each letter remembered in the correct position. Therefore, they could have earned a total of 50 points for the OSPAN. Participants were again able to monitor their accuracy in solving the math problems on the screen. Data were not used for any participant that does not achieve an 85% success rate or better. The average time for the OSPAN was about 12 minutes. A screenshot of the OSPAN can be found in Appendix F.

Gf. The progressive matrices task (PMT) is a matrix task developed and published online by Mensa® to measure IQ. The PMT has the same design as Raven's Advanced Progressive Matrices (RAPM) task, which is a measure that has been empirically supported to measure fluid reasoning (Gf). Matrices tasks are widely used in both research and selection to measure aptitude (Pearson, 2018). Gf is considered the best indicator of GMA, and consistently has the highest factor loading onto GMA (Carroll, 1993). Therefore, the current study will be using Gf as the operational definition of GMA, as stated earlier.

Like other matrix tasks, the PMT presented participants with a matrix of patterns, shapes, and colors, with one box missing from the matrix. The participants were tasked with figuring out what the “rules” are of the matrix, and extend these rules or patterns to the final block to determine what the missing block looks like. The PMT consisted of 20 multiple-choice items, with eight options for each item. The 20 items got progressively harder as the participants moved through the task. Participants were scored based on number of items answered correctly and had 17 minutes to complete the PMT. An example item of the PMT can be found in Appendix G.

In order to verify that the PMT captures much of what the RAPM measures and is therefore an acceptable measure of Gf, a post-hoc study was conducted to see how highly scores on the PMT and RAPM were correlated. In a sample of 56 students, the RAPM and PMT had a correlation of 0.69 ($p < .01$), indicating they are strongly correlated and the use of the PMT to measure Gf is justified.

Task Development. The verbal and quantitative tasks were developed in an attempt to assess performance in applying relatively newly-learned knowledge. In the verbal domain, although undergraduates have been somewhat exposed to APA writing style, many do not have a working knowledge of how to format an APA-style reference page. So, this would be mostly new information that is able to be learned in a short period of time. For many students, learning APA style is also knowledge that is useful for them in the future. In a similar vein, calculating different federal taxes is knowledge that most students have not yet acquired. This is information that is both new and information that may be useful for them in the future.

Task knowledge items for both types of tasks were created based on *only* information that was directly presented in the training video. Therefore, no other domain specific knowledge (i.e., knowledge other than reading, using a calculator, and simply arithmetic) was required to answer these items correctly. The tasks themselves also only required participants to apply knowledge of the information presented in the videos.

Verbal task performance. To score the verbal task, participants could have earned eight points for each of the six references they were asked to cite, and an additional point for putting the references in the proper order. Therefore, participants could have earned a total of 49 points. Specifically, participants could have earned one point for each of the following for each of the six references: 1) Spelling everything correctly throughout the reference (e.g., author names, article titles), 2) Properly inserting commas in the correct places throughout the reference, 3) Properly placing periods throughout the reference, 4) Properly placing parentheses throughout the reference, 5) Properly capitalizing throughout the reference, 6) Properly using spacing throughout the reference, 7) Properly using italics throughout the reference, and 8) Properly ordering the items within each reference (e.g., date following author names). Participants' task performance was scored based on the accuracy of their formatting in comparison to the proper formatting per APA guidelines. The scoring rubric used to score each APA reference page can be found in Appendix H.

Quantitative task performance. Participants could have earned eleven points for each of the two profiles for which they were asked to calculate taxes, for a total of 22 possible points. Specifically, participants could have earned one point for correctly calculating the following: 1) Taxable income, 2) Employee's contribution to social security, 3) Employee's contribution to Medicare, 4) Employee's total FICA contribution, 5) Employer's total FICA contribution, 6)

Federal taxes (excluding FICA taxes), and 7) Net take-home pay. Because four of these calculations required the successful calculation of a previous step, an incorrect calculation early on could have resulted in all incorrect calculations. This may not have reflected a lack of understanding of the task, but rather a simple mistake. Therefore, in addition, participants were given a point for each calculation in numbers 4, 5, 6 and 7 which follow the rule to calculate them (e.g., employee's total FICA contribution is the sum of employee's social security and employee's Medicare contributions). This resulted in eleven points which could have been earned for each profile they calculated taxes for, summing to 22 total possible points. The scoring rubric can be found in Appendix I.

V. Analysis

Two-hundred participants were desired to reach the power to observe significant results in support of the partial mediation model. Fritz and MacKinnon (2007) estimated that using bootstrapping as the analysis, for a small-medium effect of WM on task knowledge path and a medium effect of task knowledge on task performance (i.e., similar to the results of Carter et al., 2018), between 118 and 122 participants are recommended to reach adequate power.

The task knowledge measure statistically controlled for pretest knowledge by regressing the posttest knowledge scores on the pretest knowledge scores and saving the deleted residuals. These deleted residuals represented task knowledge scores controlling for the variance in pretest knowledge. The same was done for the task performance measure. In the mediation models, Gf was statistically controlled for using this same technique.

To test hypothesis 1, hierarchical regression was conducted by entering Gf at the first stage, followed by the WM score which aligns with the domain of the task knowledge measure. Results were interpreted in terms of the percent variance incremented by the WM scores as well as the statistical significance of the change in R^2. Hypothesis 2 was tested in the same way, but with the corresponding measure of task performance.

To test the partial mediation model (hypothesis 3), bootstrapping was conducted in R using the “mediate” functions. More specifically, ten thousand simulations with percentile bootstrapping were used to test the direct and indirect effects of WM on task performance after controlling for pretest and Gf, as well as create 95% confidence intervals.

To test hypothesis 4, two partial mediation models were compared using the same bootstrapping method just mentioned for each of the task types. In other words, to test domain specificity, the model with RSPAN affecting *verbal* task knowledge and performance were compared to that of OSPAN affecting *verbal* task knowledge and performance. The same was done for the RSPAN vs. the OSPAN in affecting *quantitative* task knowledge and performance. Direct and indirect effects and beta paths were compared between the RSPAN and OSPAN models within each of the task domains.

To test hypothesis 5, Cohen’s *d* effect sizes and confidence intervals were estimated to examine the size of the mean differences between the Black and White sample on tests of Gf compared to those of the RSPAN and OSPAN.

Chapter 4

Results

Descriptive Statistics and Correlations

Of the total sample of 210 participants, two were removed due to participants withdrawing from the study before completing the WM or PMT tasks. Two other participants were removed after being identified as outliers on the PMT task, having scored lower than four standard deviations below the mean. Two participants were removed due to failing to meet the 85% accuracy requirement on the WM tasks. Four participants were eliminated for scoring eight or above on the verbal pretest in order to mitigate ceiling effects for task knowledge. These removals resulted in a final sample of 200 participants. The descriptive statistics for the WM tasks, PMT, pretests, posttests, and tasks for the final sample are given in Table 1.

Table 1.
Descriptive Statistics

	Minimum	Maximum	Mean	Std. Deviation
V-pretest	1	7	3.60	1.96
Q-pretest	0	6	2.10	1.28
Gf	6	18	12.38	2.26
OSPAN	12	50	38.99	8.36
RSPAN	12	50	36.52	7.88
VTK	3	10	7.75	1.97
VTP	11	49	34.82	9.95
QTK	0	10	5.43	2.86
QTP	1	22	12.07	5.69

Notes. V-pretest = Verbal task knowledge before tutorial. Q-pretest = Quantitative task knowledge before tutorial. Gf = Fluid reasoning task. RSPAN = Reading Span. OSPAN = Operation Span. VTK = Verbal task knowledge after tutorial. VTP = Verbal task performance. QTK = Quantitative task knowledge after tutorial. QTP = Quantitative task performance. $^{**}p < .01$, $^{*}p < .05$ (one-tailed). $N = 200$ for predictors and $n = 100$ for criteria.

To test for order effects with the predictor measures (i.e., RSPAN, OSPAN, and PMT), descriptive statistics were computed for the predictor variables based on whether participants completed the RSPAN or OSPAN first. These are provided in Table 2. Results indicated that there were no differences between the descriptive statistics of the RSPAN, OSPAN, or PMT based on which WM task was completed first.

Table 2.
Descriptive statistics for predictors based on which WM task was completed first.

	RSPAN		**OSPAN**		**PMT**	
1st WM Task	Mean	SD	Mean	SD	Mean	SD
RSPAN	37.06	7.83	39.08	8.89	12.21	2.26
OSPAN	36.00	7.93	38.13	7.76	12.35	2.27

Notes. $N = 200$. Ninety-eight (49%) participants completed the RSPAN first. One-hundred two (51%) participants completed the OSPAN first.

The bivariate correlations among the measured variables are provided in Tables 3, 4, and 5. To mitigate possible range restriction on the WM tasks from using undergraduates from the same university, correlations between the WM tasks and the criteria were corrected for direct range restriction on WM in Tables 4 and 5. Specifically, Thorndike's (1947) 2nd formula was applied using standard deviations obtained by Unsworth, Heitz, Schrock, and Engle (2005) in a student and community-based sample of 252 participants. This sample more closely represents the population (i.e., participants from the general population) with a standard deviation scaled down to a sample size similar to the current sample. Range restriction corrections were not conducted for the PMT as no normative data (i.e., means and SDs) were found for the PMT.

No effects were found for age nor year in school on any of the predictor or criterion variables. A gender difference occurred for the scores on the verbal measures, with females scoring higher than males on the verbal pretest ($p < .01$), the verbal posttest ($p < .05$), and the verbal task ($p < .01$). Further, psychology majors/minors scored higher than non-psychology students on the verbal task ($p < .05$), but this effect was nonsignificant after controlling for race ($p = .19$)

Table 3.
Correlations among predictors administered to all participants (N = 200).

	1	2	3
1. Gf	__		
2. RSPAN	.23**	__	
3. OSPAN	.22**	.60**	__

Table 4.
Correlations among measured variables for verbal experimental tasks (n = 100)

	1	2	3	4	5	6
1. Gf	__					
2. RSPAN	.19*	__			.28*	.29**
3. OSPAN	.21*	.56**	__		.27*	.14
4. V-pretest	.09	.24**	.12	__		
5. VTK	.28**	.24**	.25**	.38**	__	
6. VTP	.30**	.25**	.13	.31**	.64**	__

Notes. V-pretest = Verbal task knowledge before tutorial. Gf = Fluid reasoning task. RSPAN = Reading Span. OSPAN = Operation Span. VTK = Verbal task knowledge after tutorial. VTP = Verbal task performance. Correlations below the diagonal are uncorrected; correlations above the diagonal are corrected for direct range restriction on WM. $^{**}p < .01$, $^{*}p < .05$ (one-tailed)

Table 5.
Correlations among measured variables for quantitative experimental tasks (n = 100).

	1	2	3	4	5	6
1. Gf	__					
2. RSPAN	.25**	__			.32**	.36**
3. OSPAN	.21*	.63**	__		.38**	.36**
4. Q-pretest	.16	-.01	-.10	__		
5. QTK	.37**	.28**	.35**	.11	__	
6. QTP	.49**	.32**	.33**	.12	.74**	__

Notes. Q-pretest = Quantitative task knowledge before tutorial. Gf = Fluid reasoning task. RSPAN = Reading Span. OSPAN = Operation Span. QTK = Quantitative task knowledge after tutorial. QTP = Quantitative task performance. Correlations below the diagonal are uncorrected; correlations above the diagonal are corrected for direct range restriction on WM. $^{**}p < .01$, $^{*}p < .05$ (one-tailed)

As can be seen from Table 3, the WM measures correlated strongly with each other (r = 0.60). Gf had a significant but more modest correlation with the RSPAN and OSPAN (r = 0.22 and r = 0.23, respectively). The RSPAN was the only predictor to correlate significantly with the verbal pretest (Table 4). As expected, Gf had the largest correlations with both verbal task knowledge and verbal task performance, although the correlations between the WM tasks and verbal task knowledge were comparable. There was a strong correlation between the verbal knowledge and performance measures (r = 0.64).

No predictors had a significant correlation with the quantitative pretest (Table 5). Once again, Gf had the largest correlations with quantitative task knowledge (r = 0.37) and performance (r = 0.49). WM had a correlation between r = 0.28 ($p < .01$) and r = 0.35 ($p < .01$) with the quantitative criteria. There was a strong correlation between the quantitative knowledge and performance measures (r = 0.74).

Creating Residuals for Hypothesis Tests

For both the quantitative and verbal tasks, posttest knowledge scores were regressed on pretest knowledge scores and deleted residuals were saved, creating a task knowledge score that controlled for baseline pretest knowledge. Pretest knowledge accounted for 14.4% ($p < .01$) of variance in posttest knowledge for the verbal content, and 1.2% ($p > .05$) of variance in posttest knowledge for the quantitative content.

Similarly, task performance scores were regressed on pretest knowledge scores and deleted residuals were saved. This created a task performance score that controlled for baseline pretest knowledge for both the verbal and quantitative tasks. Pretest knowledge accounted for 9.8% ($p < .01$) of variance in task performance on the verbal task, and 1.5% of ($p > .05$) of variance in task performance on the quantitative task.

Incremental Validity of WM over Gf

To test Hypotheses 1 and 2, hierarchical regression analyses were conducted to examine whether WM scores accounted for incremental variance in task knowledge and task performance above Gf. Eight separate hierarchical regression analyses were conducted to examine the whether or not WM accounted for unique variance in outcomes – all combinations of the factorial of 2 (WM measures) X 2 (task knowledge criteria) X 2 (task performance criteria). Gf was entered at the first step of each analysis, with a WM score entered at the second step. The dependent variables were verbal task knowledge, quantitative task knowledge, verbal task performance, and quantitative task performance. The results of these analyses are given in Table 6.

Hypotheses 1 and 2 – Incremental Validity

In Hypotheses 1 and 2, it was predicted that WM would add incremental variance above Gf for both task knowledge and task performance. Gf scores accounted for 6.9% ($p < .01$) of variance in verbal task knowledge; RSPAN scores accounted for an additional 1.3% ($p = .24$) of variance in verbal task knowledge; OSPAN scores accounted for an additional 3.0% ($p = .08$) of variance in verbal task knowledge. Gf scores accounted for 12.3% ($p < .01$) of variance in quantitative task knowledge. RSPAN scores accounted for an additional 4.0% ($p < .05$) of variance in quantitative task knowledge; OSPAN accounted for an additional 8.8% ($p < .01$) of variance in quantitative task knowledge.

Table 6.
Hierarchical Regression Analyses

	Verbal Task Knowledge				Verbal Task Performance			
	B	*SE B*	β	ΔR²	*B*	*SE B*	β	ΔR²
Step 1								
Gf	0.25	0.09	0.26	.069**	1.36	0.47	0.28	0.080**
Step 2								
Gf	0.22	0.09	0.24		1.24	0.47	0.26	
RSPAN	0.03	0.02	0.12	0.013	0.16	0.12	0.13	0.016
Step 1								
Gf	0.25	0.09	0.26	.069**	1.36	0.47	0.28	0.080**
Step 2								
Gf	0.21	0.09	0.23	.	1.33	0.48	0.28	
OSPAN	0.04	0.02	0.18	0.030	0.04	0.12	0.04	0.001

	Quantitative Task Knowledge				Quantitative Task Performance			
	B	*SE B*	β	ΔR²	*B*	*SE B*	β	ΔR²
Step 1								
Gf	0.42	0.11	0.35	0.123**	1.11	0.21	0.47	0.223**
Step 2								
Gf	0.36	0.11	0.30		0.98	0.21	0.42	
RSPAN	0.08	0.04	0.21	0.040*	0.16	0.07	0.22	0.047**
Step 1								
Gf	0.42	0.11	0.35	0.123**	1.11	0.21	0.47	0.223**
Step 2								
Gf	0.34	0.11	0.29		0.99	0.21	0.42	
OSPAN	0.10	0.03	0.30	0.088**	0.17	0.06	0.26	0.062**

Notes. Gf = Fluid reasoning (PMT) task. RSPAN = Reading Span. OSPAN = Operation Span.

For task performance, Gf scores accounted for 8.0% ($p < .01$) of variance in verbal task performance. The RSPAN scores accounted for an additional 1.6% ($p = .18$) of variance in verbal task performance, while the OSPAN scores accounted for an additional 0.1% ($p = .72$) of variance in verbal task performance.

Gf scores accounted for 22.3% ($p < .01$) of variance in quantitative task performance. The RSPAN scores accounted for an additional 4.7% ($p < .01$) of variance in quantitative task performance, while the OSPAN scores also accounted for an additional 6.2% ($p < .01$) of variance in quantitative task performance.

These results indicated that Hypotheses 1 and 2 were supported when accounting for variance in quantitative knowledge and performance, but not accounting for variance in verbal knowledge and performance. This was the case for both the RSPAN and the OSPAN scores. Although not significant, WM scores accounted for up to 3.0% variance in verbal task knowledge and up to 1.6% variance in verbal task performance. Failure to reach significance may be due to a smaller sample size (i.e., $n = 100$ per model) than the sample size ($N = 156$) in Carter et al. (2018).

Mediation Model Not Controlling for Gf

Before examining results with regard to Hypothesis 3 (i.e., the mediated model controlling for Gf), four partial mediated models were run without controlling for Gf. The resulting standardized path models and effects are provided in Figure 4. Without controlling for Gf, the OSPAN had indirect effects on both verbal ($p < .05$) and quantitative task performance ($p < .01$) through task knowledge, but no direct effect. The RSPAN had both direct ($p < .05$) and indirect ($p < .01$) effects on quantitative, but not on verbal task performance. The OSPAN had a significant total effect on quantitative task performance ($p < .01$), but not on verbal task performance. The RSPAN had significant total effects on both verbal ($p < .05$) and quantitative task performance ($p < .01$).

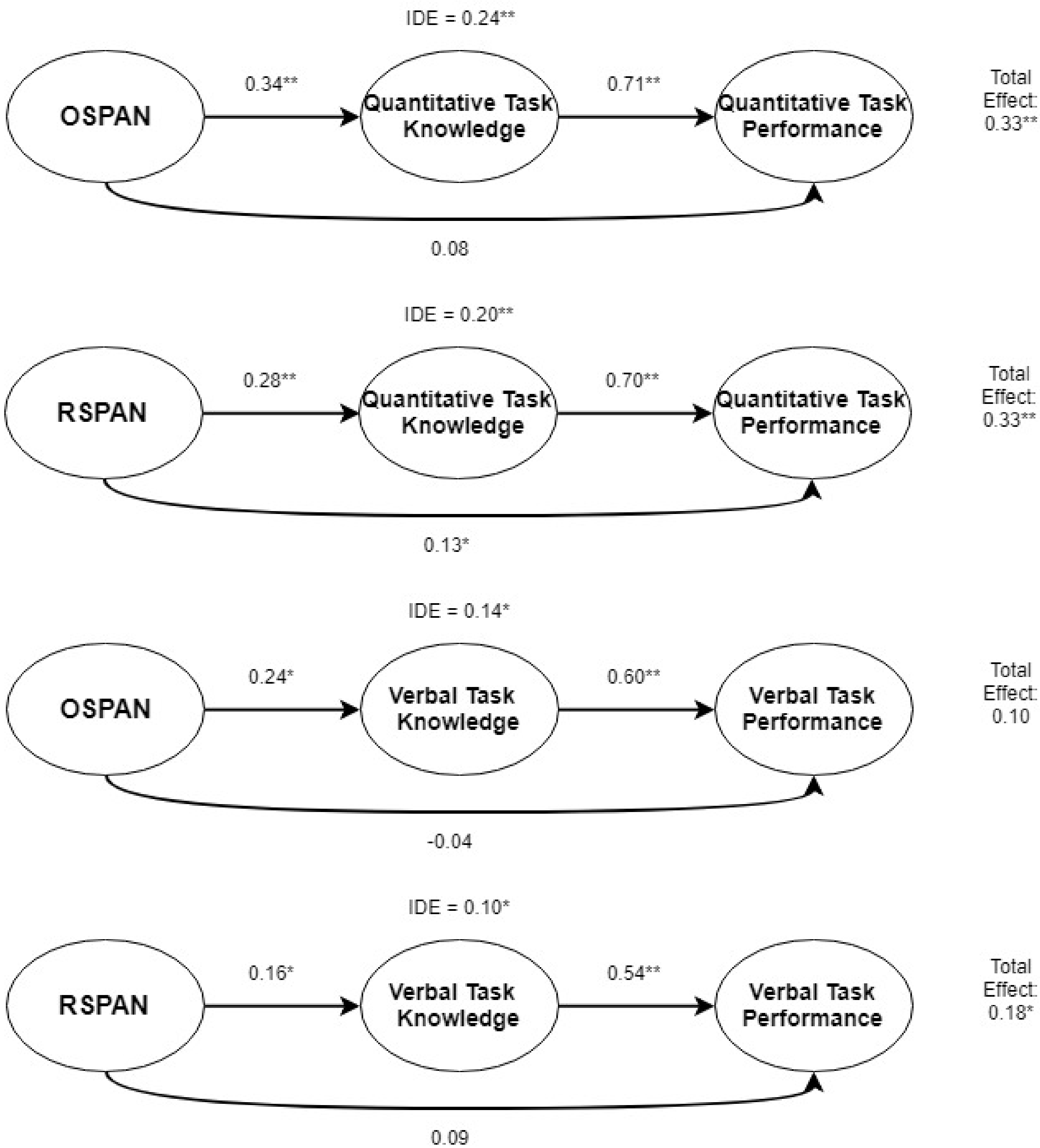


Figure 4. Partial mediation models, not controlling for Gf.

Hypothesis 3 – Mediation models controlling for Gf

Deleted residuals were created once again, resulting in scores of knowledge and performance that controlled both pretest knowledge and Gf. Both task knowledge scores and both task performance scores were the dependent variables in four separate hierarchical regression analyses. Percentile bootstrapping mediation was conducted separately for each combination of WM task (i.e., RSPAN and OSPAN) and criterion domain (i.e., verbal vs. quantitative), resulting in four partial mediation models. The resulting standardized path models and effects are shown in Figure 5.

After controlling for Gf, both the OSPAN and the RSPAN had only an indirect effect on quantitative task performance ($p < .01$ and $p < .05$, respectively); the direct effects were not significant. The OSPAN had only an indirect effect on verbal task performance ($p < .05$), but the RSPAN did not have significant a direct or indirect effect on verbal task performance. Both the RSPAN and OSPAN had significant total effects on quantitative task performance ($p < .05$ and $p < .01$, respectively), but neither the OSPAN nor the RSPAN had a significant total effect on verbal task performance. Therefore, Hypothesis 3 was partially supported (i.e., indirect effects were significant) when using the quantitative task, but not when using the verbal task.

Hypothesis 4 was not supported. WM scores did not account for variance in criteria based on the alignment between the WM task and task content (i.e., the domain specificity hypothesis).

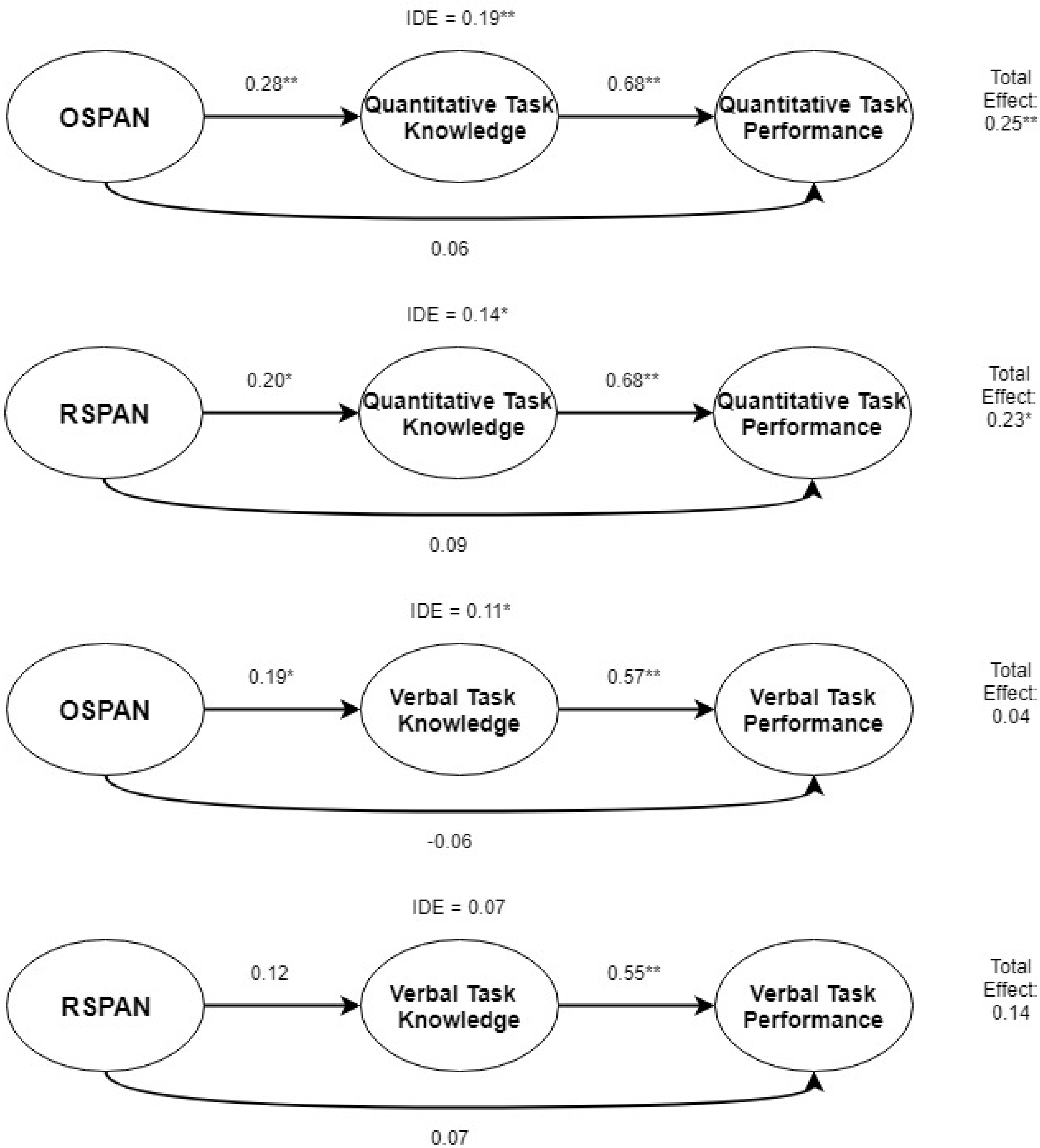


Figure 5. Partial mediation models, controlling for Gf.

Moderated Mediation Analyses

Participants' race was not expected to moderate the hypothesized path models. All three causal paths were tested for moderation by race in all four models (i.e., each WM/task domain combination). These were tested by entering a predictor x race interaction term into the models. No interaction terms were significant, meaning no paths were significantly moderated by race. In addition, bootstrapping moderated mediation was conducted, which tests for differences in the direct and indirect effects of WM on task performance based on race. Across all four models, WM did not have significantly different indirect, direct, or total effects on task performance based on race. The path coefficients for the interaction term as well as their *p* values are provided in Table 7.

Table 7.
PredictorXrace interaction term path coefficients and p values.

Path	Beta estimate	*p* value
RSPAN → VTK	0.37	0.07
OSPAN → VTK	-0.14	0.52
RSPAN → VTP	0.21	0.23
OSPAN → VTP	-0.02	0.90
VTK → VTP	0.20	0.23
RSPAN → QTK	-0.02	0.94
OSPAN → QTK	0.03	0.88
RSPAN → QTP	-0.01	0.97
OSPAN → QTP	-0.06	0.66
VTK → QTP	-0.14	0.36

Notes. Positive Beta estimates indicate the path is stronger for the Black sample.

Although no paths were significantly moderated, it is noteworthy that the RSPAN → VTK path was trending toward significant ($p = 0.07$). This meant that the path was nearly significantly stronger for the Black sample than it was for the White sample. This finding could simply be a statistical artifact due to sample size or sampling error. It may be in part due to the Black sample having somewhat larger variance on the RSPAN than the White sample, which can be seen in Table 8. However, Levene's test for homogeneity of variance between groups for the RSPAN was nonsignificant (p = 0.19).

Hypothesis 5 – Race group differences

Cohen's *d* was computed using the means and standard deviations of measured variables for Black and White participants. The descriptive statistics of measured variables for each race group can be found in Table 8 along with Cohen's *d* to compare race groups.

Table 8.
Descriptive statistics and Cohen's d estimates for measured variables.

	Black			White			Effect Size Estimates		
	N	Mean	SD	*n*	Mean	SD	*d*	Lower	Upper
V-pretest	50	3.18	1.80	50	4.02	2.04	0.44*	0.04	0.83
Q-pretest	50	2.06	1.06	50	2.14	1.49	0.06	-0.33	0.45
Gf	100	12.05	2.48	100	12.71	1.96	0.30*	0.02	0.57
RSPAN	100	35.34	8.49	100	37.70	7.07	0.30*	0.02	0.58
OSPAN	100	38.15	8.78	100	39.82	7.87	0.20	-0.08	0.48
VTK	50	7.44	1.94	50	8.06	1.97	0.32	-0.08	0.71
VTP	50	32.14	10.95	50	37.50	8.09	0.56*	0.15	0.95
QTK	50	4.70	2.58	50	6.16	2.97	0.53*	0.12	0.92
QTP	50	10.28	5.41	50	13.86	5.44	0.66**	0.25	1.06

Notes. V-pretest = Verbal task knowledge before tutorial. Q-pretest = Quantitative task knowledge before tutorial. Gf = Fluid reasoning task. RSPAN = Reading Span. OSPAN = Operation Span. VTK = Verbal task knowledge after tutorial. VTP = Verbal task performance. QTK = Quantitative task knowledge after tutorial. QTP = Quantitative task performance. *d* = Cohen's *d* standardized group difference estimate. Positive Cohen's *d* favors White participants.

As Table 8 indicates, White participants scored better than Black participants on the verbal pretest, Gf, RSPAN, verbal task performance, quantitative task knowledge, and quantitative task performance. Among the antecedent variables, Gf scores had the largest standardized group difference (d = 0.30; $p < .05$). The RSPAN scores had a comparable *d* of 0.30 ($p < .05$), while the OSPAN had a *d* of 0.20. The group difference for OSPAN scores was lower than Gf scores, but the confidence intervals overlapped, indicating the Cohen's *d* for OSPAN scores are not significantly lower than that of Gf. Although the OSPAN has a smaller absolute group difference, these findings do not statistically support Hypothesis 5, which stated that WM scores would exhibit smaller race group differences.

Chapter 5

Discussion

The current study investigated whether WM scores accounted for variance in task knowledge and task performance above Gf scores; in addition, racial subgroup differences of WM were compared to those of Gf. Results indicated that WM scores accounted for variance beyond Gf scores for quantitative knowledge and performance, but not for verbal knowledge and performance. Most of the effect of WM on task performance occurred through task knowledge, as indicated by significant indirect effects and nonsignificant direct effects on task performance. Lastly, WM scores exhibited similar race differences to those of Gf scores for the RSPAN and slightly smaller absolute race differences for the OSPAN scores.

Conceptual Implications

The hierarchical regression analyses revealed that WM scores explained significant additional variance in quantitative task knowledge and performance, and incremental variance for verbal knowledge and performance approached significance. These findings support the findings of Carter et al. (2018) in establishing WM as an antecedent of knowledge acquisition and application. More specifically, these findings indicate that even after controlling for Gf, specific WM processes play a role in learning newly-presented information for later recall. Intuitively, this newly-learned information assists in the completion of tasks which apply this knowledge. Perhaps most importantly, these findings contributed to understanding the process by which certain specific cognitive abilities may account for significant variance in these outcomes.

Hunter (1986) found that GMA is an antecedent that affects job knowledge and, in turn, job knowledge affects performance. Since then, this partially mediated model has been widely cited in the literature (e.g., Borman et al., 1993; Bormam et al., 1991; Schmidt, 2002). However, psychometric GMA or intelligence simply represent positive manifold among specific aptitudes; therefore, no widely-established specific process is available to explain how people high on GMA better acquire and apply knowledge. WM provides a more highly specified process that connects intelligence to performance. That is, relative to people with more limited WMs, people with better WM are able to store, update, and manipulate information more effectively in the presence of competing stimuli. This successful maintenance allows people to access this knowledge in the process of completing tasks that apply this knowledge.

It is noteworthy that WM scores accounted for notably more variance in quantitative tasks than that of the verbal tasks. This may have been the case because the quantitative knowledge experimental task was more complex and more difficult than the verbal knowledge experimental task, as indicated by the lower scores and higher variability on the quantitative knowledge assessment (Table 1). It could be argued that the higher scores and lower variability on the verbal tasks are a result of the fact that many of the participants (40%) were psychology majors/minors, and therefore they did better on the APA-oriented verbal tasks. However, follow-up analyses revealed that the means of both psychology- and non-psychology majors/minors were higher and less variable (i.e., in absolute value) on verbal task knowledge compared to quantitative task knowledge. This pattern of means and standard deviations holds when comparing psychology- and non-psychology majors/minors, both overall and within race (Table 9).

Table 9.
Descriptive statistics for criteria broken down by psychology major/minor and race.

	VTK		VTP		QTK		QTP	
	n	*M*(*SD*)	*n*	*M*(*SD*)	*n*	*M*(*SD*)	*n*	*M*(*SD*)
All								
Psych	40	7.93(1.89)	40	37.13(9.95)	38	5.89(2.78)	38	12.39(5.85)
Non-Psych	60	7.63(2.03)	60	33.28(9.73)	62	5.15(2.90)	62	11.87(5.64)
White								
Psych	26	8.35(1.72)	26	38.00(8.72)	23	5.57(2.91)	23	12.65(5.69)
Non-Psych	24	7.75(2.21)	24	36.96(7.50)	27	6.67(2.97)	27	14.89(5.10)
Black								
Psych	14	7.14(1.99)	14	35.50(12.11)	15	6.40(2.59)	15	12.00(6.24)
Non-Psych	36	7.56(1.93)	36	30.83(10.35)	35	3.97(2.24)	35	9.54(4.92)

Notes. VTK = Verbal task knowledge after tutorial. VTP = Verbal task performance. QTK = Quantitative task knowledge after tutorial. QTP = Quantitative task performance. Psych = self-reported psychology major or minor. Non-Psych = Self-reported non-psychology major/minor. *N* = 200.

In the quantitative video tutorial, participants were required to follow how the demonstrator was obtaining each value, what values were being input into the calculator, and retain information pertaining to how each of the calculations were related to each other. Because of the complexity of the information and the greater need to store, update, and manipulate information when acquiring necessary knowledge, WM was more essential in affecting quantitative task knowledge and performance.

In contrast, the verbal video tutorial only required participants to remember a repeated order of information (i.e., the order of information included in a reference). Participants may have also been able to simply use familiarity, judging whether or not a journal reference was correct based on whether it “just looked right” compared to previous reference examples presented. In this sense, the verbal tasks may have been less complex for most participants, rendering WM as less essential for the verbal tasks than the quantitative tasks.

This line of theory aligns well with past research that examined job complexity as a moderator of the cognitive ability→performance relationship. Specifically, it has been empirically supported that the validity coefficients for GMA in affecting knowledge and performance are higher for more complex jobs, moderate for moderately-complex jobs, and lower for less-complex jobs (Schmidt & Hunter, 1998). In a similar vein, as performing complex tasks relies more on cognitive functioning, one would expect WM to better account for variance in more complex tasks, especially considering the attentional and inhibitory processes associated with WM.

The results further imply that WM processes may be more important for the acquisition of knowledge than the application of this knowledge (i.e., task performance). In other words, much of the effect of WM on performance was mediated by task knowledge. However, this does not necessarily mean WMC is more important than WMP for affecting task performance. In this study, the measures of task knowledge and performance were strongly related in that the task applied only knowledge which they could have learned in the tutorial. That is, all of the information on the task knowledge assessments was presented in the video, and the task performance measure solely tested participants' application of this presented knowledge.

There were no factors that should have affected task performance greatly other than task knowledge. This means much of the variance in task performance was represented by task knowledge, as indicated by the large knowledge→performance standardized beta paths ranging from 0.55 to 0.68. After accounting for task knowledge and Gf, much less variance in task performance is left over to be explained by WMP. Thus, no significant direct effects of WM on performance were found.

The differences in difficulty between the quantitative and verbal tasks make it difficult to examine the results with regards to Hypothesis 4, which examined the role of domain specificity. This difficulty arose mainly because the OSPAN and RSPAN scores both accounted for more variance in the quantitative tasks than verbal tasks. No significant increments in variance were accounted for by WM scores for the verbal tasks. So based on Table 6, it is concluded that there was no evidence for domain specificity – verbal WM did not affect performance on the verbal tasks better than quantitative tasks and vice versa. These findings align with some researchers (e.g., Kane et al., 2007) who have claimed that WM is a domain-general construct and that WM affects outcomes of different content domains similarly.

Although past research cited in the literature review (e.g., Ackerman et al., 2005; Verive & McDaniel, 1996) suggests that WM may exhibit smaller race-group differences than Gf, findings for this study were difficult to interpret. This is because although the RSPAN (d = 0.30) and OSPAN (d = 0.20) scores had smaller race differences compared to commonly-found *d*'s of around 1.0 in the general population, Gf scores also exhibited a smaller Cohen's d of 0.30. This smaller race difference for Gf scores could be mainly due to range restriction on the cognitive ability of participants such that Black and White students in the same age group and enrolled at the same university did not differ as much on cognitive ability as the general population.

WM represents more processing components of cognitive ability, and focuses less on abstract reasoning and logic (Gf), so it was expected that WM would not display the same large race differences as Gf. Furthermore, as Spearman's (1927) hypothesis states, the closer one gets to measuring GMA the larger the race differences should be. Extending this theory, the modest

WM correlations with Gf provide an initial case that WM may exhibit smaller race differences when the general population or a more diverse workforce sample is used.

Practical Implications

Because of the high predictive validity of GMA in affecting job performance and the relatively little evidence for incremental prediction using specific abilities (Ree & Earles, 1991; Ree et al., 1994), organizational researchers have mostly accepted that solely measuring GMA is the best that can be done to tap into the variance accounted for by cognitive abilities. In turn, research on specific abilities has been mostly abandoned in recent years.

In conjunction with the Carter et al. (2018) study, this study further provides evidence for the effects of WM on task knowledge and performance. Although in some cases adding WM to the model only accounted for a small change in R^2 (e.g., OSPAN accounted for an additional 3.0% of quantitative task knowledge), these increments are meaningful when considering a large number of applicants over a long period of time.

These results provide evidence for WM as an antecedent of *task* performance in the moment. That is, WM is important for the successful completion of a laboratory task that requires some level of inhibitory and attentional processes. An important question is whether this task performance would translate to global job performance (e.g., supervisor ratings) in an organizational setting. It should be expected that WM would affect performance across a wide variety of tasks that are pertinent to job responsibilities. It is further expected that performance on job-relevant tasks accumulates in some way to make up overall job performance (e.g., supervisors rating employees as a result of their performance across important smaller job tasks).

From this perspective, although better performance on one small task may not have a large impact on job performance, the accumulation of performance across all job tasks should.

Thus, WM would probably be an important antecedent of *job* performance. This notion was supported by Bosco et al. (2015), who found initial evidence for the effect of WM on supervisor ratings. The accumulating evidence supporting WM as an antecedent of task knowledge and performance suggests that WM should be considered for inclusion in personnel selection batteries.

Although the WM battery takes time (~25-40 minutes) and requires applicants to be proctored, the possible incremental validity of its use may outweigh the costs. That is, after controlling for GMA, WM is able to account for variance in performance. This incremental variance especially outweighs the costs in the case of personnel-selection scenarios where it is the norm to require applicants to be assessed by a large battery of tests (e.g., many military selection and placement procedures). Small but significant increments of variance accounted for are also essential in jobs where human performance can impact critical societal outcomes.

On the other hand, just as general cognitive aptitude tests are not frequently essential for selection into lower level, less complex jobs, WM tests may not be cost-effective for selecting applicants to these jobs. Performance in these cases is typically not as dependent upon more complex cognitive processing; and therefore WM measures would not be a worthwhile expense of time and money. This argument was supported by the results of this study, where WM did not affect performance as much when the task was less difficult.

It is important to consider the race differences on any new selection measure, as race differences can lead to adverse impact. Although standardized race differences were similar between RSPAN and Gf scores in this study, OSPAN scores did exhibit a smaller absolute race difference (d = 0.20 vs. 0.30), which may provide initial evidence for a reduction in adverse impact through the use of WM. If future research establishes WM scores as having smaller race differences, WM scores may be preferable to Gf scores for accounting for variance in performance in selection contexts due to this reduced adverse impact. However, this study found no initial evidence for significantly smaller race differences for WM scores compared to those of Gf.

On the other hand, if WM scores are able to account for variance in knowledge and performance at least as well as Gf scores in selection contexts, future research may reveal that applicants perceive WM tests to be more face valid as a predictor of workplace performance than measures of Gf such as progressive matrices tasks. That is, it may be more accepted by applicants that memory skills are important for job performance as opposed to “puzzle solving” skills.

Limitations and Future Directions

A primary limitation of this study is the use of laboratory data rather than field data. Field data would have allowed for the examination of WM as an antecedent of *job* knowledge and *job* performance as they are typically operationalized in field studies. Not only did the current methodology use university students as participants, but it also required that *task* knowledge and *task* performance are examined rather than their real-world organizational counterparts. Field data would allow researchers to examine whether or not WM affects real-world job knowledge acquisition and job performance.

It does seem intuitive that the same processes that foster acquisition of task knowledge would be relevant for the acquisition of job knowledge over time, and the application of this task knowledge would be similar to the application of job knowledge for on-the-job performance. Still, field data would perhaps allow for a more realistic measure of performance which does not overlap with knowledge as much as the two measures of task performance used in this laboratory study. Future research should focus on using WM as an antecedent in a large selection context to examine incremental validity of WM scores for predicting job performance as well as both the direct and indirect effects of WM on performance through job knowledge.

Another limitation is related to the nature of the tasks themselves. The quantitative tasks were found to be more complex and difficult for the participants than the verbal tasks, making it difficult to interpret the domain-specificity hypothesis. Future research should examine this hypothesis with more standardized tests of job knowledge and performance that have similar levels of difficulty or complexity.

Sample size should be considered a limitation. Although data for 200 participants were used, only 100 participants were used for each mediation model (i.e., 50 per race in each model). This sample size is considerably lower than the sample size of 156 used by Carter et al. (2018) – the study on which the current study was based. This makes it difficult to detect effects in incremental validity, especially for a small effect, which is expected when trying to provide incremental variance above GMA scores.

Another limitation with the current study was the small Black-White race difference associated with the Gf scores. Typical race difference estimates in the general population are higher (Jensen, 1998; Roth et al., 2001), around $d = 1.0$. The observed small Cohen's *d* for Gf makes it difficult to interpret the small *d*'s for both the RSPAN and OSPAN scores comparative to Gf. Future research should further examine race differences on WM tasks, and especially compare race differences to those of more commonly-used measures of GMA, such as the Wonderlic or RAPM. It is hoped that such future studies will paint a clearer picture as to how useful WM is when included in selection batteries, including whether WM scores have smaller differences due to its representation of more processing components as opposed to higher level reasoning or logical components.

Furthermore, the students in the Black sample represented a more diverse population than the White sample. The Black sample was recruited campus-wide, and therefore represented a wide variety of academic majors with many involved in extracurricular organizations on campus. The White sample was recruited from Psychology courses and represented a smaller variety of academic majors (50% Psychology majors/minors). This may be one reason for the greater performance variation on both WM and Gf scores for the Black sample compared to the White sample. Future research in either laboratory studies or real-world studies should focus on recruitment from a more variable sample of Blacks and Whites to better eliminate any extraneous artifacts.

Finally, the partial mediation model is based on the functions of WMC and WMP. However, the current study had no way of teasing apart these two functions, thereby inhibiting the ability to determine that WMC was indeed responsible for the indirect effects on performance and WMP was responsible for the remaining direct effects. Researchers will need to find unique ways to measure WM in order to separate WMC and WMP, as these cognitive functions typically work together simultaneously (Daneman & Carpenter, 1980).

Conclusion

The current study aimed to further examine whether WM could help address the validity-diversity dilemma. Related to validity, this research added support for the Carter et al. (2018) study, showing WM scores can account for unique variance for two important outcomes: task knowledge and task performance. However, the mean race differences of WM scores still need to be better established against those of GMA scores in order to address the diversity side of the validity-diversity dilemma. Future research using WM in a selection battery is needed to provide real-world evidence of the utility of using WM scores to account for variance in performance while combating the presence of adverse impact.

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.

Aguinis, H. (1995). Statistical power problems with moderated multiple regression in management research. *Journal of Management, 21,* 1141-1158.

Aguinis, H., & Stone-Romero, E. F. (1997). Methodological artifacts in moderated multiple regression and their effects on statistical power. *Journal of Applied Psychology, 82,* 192-206.

Aguinis, H., Culpepper, S. A., & Pierce, C. A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology, 95*(4), 648-680. http://dx.doi.org/10.1037/a0018714.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuo-spatial short-term and working memory in children: Are they separable? *Child Development, 77*, 1698–1716.

Alloway, T.P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*, 25, 92-98.

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.

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences, 21*(1), 133-137.

Avis, J. M., Kudisch, J. D., Fortunato, V. J. (2002). Examining the incremental validity and adverse impact of cognitive ability and conscientiousness on job performance. *Journal of Business and Psychology, 17*(1), 87-105.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (pp. 47-89). San Diego, CA: Academic Press.

Berry, C. M., Cullen, M. J., & Meyer, J. M. (2014). Racial/ethnic subgroup differences in cognitive ability test range restriction: implications for differential validity. *Journal of Applied Psychology, 99*(1), 21–37.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory performance ratings. *Journal of Applied Psychology, 76*, 863– 872.

Borman, W. C., Hanson, M. A., Oppler, S. H., Pulakos, E. D., & White, L. A. (1993). Role of early supervisory experience in supervisor performance. *Journal of Applied Psychology, 78*(3), 443-449.

Bosco, F., Allen, D. G., & Singh, K. (2015). Executive attention: An alternative perspective on general mental ability, performance, and subgroup differences. *Personnel Psychology*, *68*(4), 859–898.

Bureau of Labor Statistics (2018, August). *Labor force characteristics by race and ethnicity, 2017*. Retrieved from https://www.bls.gov/opub/reports/race-and-ethnicity/2017/home.htm.

Carey, N. (1994). Computer predictors of mechanical job performance: Marine Corps findings. *Military Psychology, 6,* 1-30.

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

Carter, D., Hauenstein, N. M. A., & Geller, E. S. (2018). *Examining the incremental validity of working memory for predicting learning and task performance: A partial mediation model.* Paper presented at the annual meeting of the Society for Industrial and

Organizational Psychologists, Chicago, IL.

Colom, R., Flores-Mendoza, C., & Rebollo, I. (2003). Working memory and intelligence. *Personality and Individual Differences, 34,* 33-39.

Colquitt, J. A., LePine, J. A., & Noe, R. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology, 85*, 678 –707.

Conway, A. R., 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,* 769 –786.

Conway, A. R. A., Macnamara, B. N., & Engel de Abreu, P. M. J. (2013). Working memory and intelligence: An overview, in T. Alloway (Ed.)*Working memory: The new intelligence*. London: Psychology Press.

D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology,* 66, 115-142.

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology.* Advance online publication. doi:10.1037/apl0000235

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

Daneman, M., & Tardif, T. (1987) Working memory and reading skill re-examined. In: Coltheart, M., Ed., *Attention and Performance*, Erlbaum, London, Vol. 7, 491-508.

DeKeyser, R. M., & Koeth, J. (2011). Cognitive aptitudes for second language learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 395-

406). London: Routledge.

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology, 92,* 616 – 627.

Equal Employment Opportunity Commission (1978). *Employment tests and selection procedures*. Retrieved from https://www.eeoc.gov/policy/docs/factemployment_procedures.html.

Friso-van den Bos, I., Van der Ven, S. H. G., Kroesbergen, E. H., & Van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review, 10,* 29-44.

Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233-239.

Fry, A. F., Hale, S. (1996). Processing speed, working memory, and fluid intelligence. *Psychological Science, 7,* 237-241.

Gathercole, S. E., Lamont, E., & Alloway, T. P. (2006). Working memory in the classroom. In Pickering, S. (Ed.), *Working memory and education* (pp. 219–240). Oxford, UK: Elsevier.

Gersten, R., Jordan, N.C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*, 293–304.

Held, J. D., & Wolfe, J. H. (1997). Validities of unit-weighted composites of the Armed Services Vocational Aptitude Battery (ASVAB) and Enhanced Computer Administered Tests (ECAT). *Military Psychology*, *9*, 77-84.

Herrnstein, R. A., & Murray, C. (1994). *The bell curve.* New York: Grove Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253-270.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection, and amelioration of adverse impact in personnel selection procedures: Issues, evidence, and lessons learned. *International Journal of Selection and Assessment, 9,* 152-194.

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459 – 473.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86,* 721-735.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96,* 72-98.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitude, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*(3), 340-362.

Jencks, C., & Phillips, M. (1998). *The black-white test score gap*. Washington, DC: Brookings Institution Press.

Jensen, A. R. (1986). G: Artifact or reality? *Journal of Vocational Behavior, 29*, 301-331.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Jensen, A. R. (2002). Psychometric g: Definition and substantiation. In R. J. Sternberg & E. L. Grigorenko (Eds.), *General factor of intelligence: How general is it?* (pp. 39 –54). Mahwah, NJ: Erlbaum.

Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology, 87,* 93–102

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory*. New York, NY: Oxford Press.

Kaufman, S. B. (2009). *Beyond general intelligence: The dual-process theory of human intelligence* (Unpublished Ph.D. dissertation). Yale University, New Haven, CT.

Kraiger, K., & Ford, J. K. (1985). A meta-analysis of ratee race effects on performance ratings. *Journal of Applied Psychology, 70*, 56-65.

Krumm, S., Ziegler, M., & Buehner, M. (2008). Reasoning and working memory as predictors of school grades. *Learning and Individual Differences, 18*(2), 248–257.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *Journal of Neuroscience,* 28, 8765–8771.

Linck, J., & Weiss, D.J. (2015) Working memory predicts second language classroom learning. *Sage Open*, 5(4), doi: 10.1177/2158244015607352.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994) Counting on working memory in arithmetic problem solving. *Memory and Cognition*, *22*, 395–410.

Luo, D., Thompson, L.A., & Detterman, D. K. (2006). The criterion validity of tasks of basic cognitive processes. *Intelligence*, 34, 79–120.

Mayfield, J. W., & Reynolds, C. R. (1997). Black–White differences in memory test performance among children and adolescents. *Archives of Clinical Neuropsychology, 12*(2), 111–122.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79,* 493–505.

McGrew, K. S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 136–181). New York, NY: Guilford Press.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43,* 335-354.

McKay, P. F., & McDaniel, M. A. (2003, April). *A Re-examination of Black-White differences in job performance.* Manuscript presented at the annual Society for Industrial and Organizational Psychology Conference, Orlando, FL.

McKay, P. F., & McDaniel, M.A. (2006). A reexamination of black–white differences in work performance: More data, more moderators. *Journal of Applied Psychology, 91*(3), 538–554.

Morales, M., & Ree, M. J. (1992, June). *Intelligence predicts academic and work sample training performance.* Paper presented at the annual meeting of the American Psychological Society, San Diego, CA.

Morrell, R. W., & Park, D. C. (1993). The effects of age, illustrations, and task variables on the performance of procedural assembly tasks. *Psychology and Aging, 8,* 389 –399.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 411-421.

Oh, I. S., Le, H., Whitman, D. S., Kim, K., Yoo, T. Y., Hwang, J.-O., & Kim, C. S. (2014). The incremental validity of honesty–humility over cognitive ability and the Big Five personality traits. *Human Performance, 27*, 206–224.

Ohme, M., & Zacher, H. (2015). Job performance ratings: The relative importance of mental ability, conscientiousness, and career adaptability. *Journal of Vocational Behavior, 87*, 161-170.

Pyburn, K. M., Jr., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61*(1), 143-151.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*, 321-332.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, *79*, 518-524.

Ree, M. J., & Carretta, T. R. (1998). General cognitive ability and occupational performance. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology, 13,* 159-171.

Repko, J. (2011). Spearman's hypothesis tested with Raven's Progressive Matrices: A psychometric meta-analysis. *Unpublished Thesis.*

Roberts, R. D., Goff, G. N., Anjoul, F., Kyllonen, P. C., Pallier, G., & Stankov, L. (2000). The Armed Services Vocational Aptitude Battery (ASVAB): Little more than acculturated learning (Gc)!? *Learning and individual differences, 12*(1), 81–103.

Roth, P. L., BeVier, C., Bobko, P., Switzer, F., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297–330.

Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal of Applied Psychology, 88*(4), 694–706.

Salthouse, T. A., Mitchell, D. R. D, Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology Learning: Memory & Cognition, 15,* 507-516.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology*, *71*(3), 432-439.

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

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, *15*(1-2), 187-211.

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

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*, 3rd ed., (pp. 99–144). New York, NY: Guilford Press.

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.

Spearman, C. (1904). General intelligence, objectively determined and measured. *The American Journal of Psychology*, *15*(2), 201-293.

Spearman, C. (1927). *The abilities of man*. Caldwell, NJ: Blackburn Press.

Stanhope, D., & Surface, E. (2014). Examining the incremental validity and relative importance of specific cognitive abilities in a training context. *Journal of Personnel Psychology*, *13*(3), 146-156.

Swanson, H. L. (2004). Working memory and phonological processing as predictors of children's mathematical problem solving at different ages. *Memory & Cognition, 32,* 648–661.

Te Nijenhuis, J., & Dragt, J. (2010). Causes of group differences studied with the method of correlated vectors: A psychometric meta-analysis of Spearman's hypothesis. *Unpublished manuscript,* University of Amsterdam, the Netherlands.

Thomas, S. L., & Scroggins, W. A. (2006). Psychological testing in personnel selection: Contemporary issues in cognitive ability and personality testing. *Journal of Business Inquiry*, *5*, 28-38.

Tucker, P., & Warr, P. (1996). Intelligence, elementary cognitive components, and cognitive styles as predictors of complex task performance. *Personality and Individual Differences, 21,* 91–102.

Van Iddekinge, C. H., Aguinis, H., Mackey, J. D., & DeOrtentiis, P. S. (2017). A meta-analysis of the interactive, additive, and relative effects of cognitive ability and motivation on performance. *Journal of Management.* Advance online publication. doi:10.1177/0149206317702220.

Verive, J. M., & McDaniel, M. A. (1996). Short-term memory tests in personnel selection: Low adverse impact and high validity. *Intelligence, 23*(1), 15-32.

Waldman, D. A., & Avolio, B.J. (1991). Race effects in performance evaluations: controlling for ability, education, and experience. *Journal of Applied Psychology, 76,* 897-901.

Wen, Z., & Skehan, P. (2011). A new perspective on foreign language aptitude research: Building and supporting a case for “working memory as language aptitude.” *Revista Ilha do Desterro*, *60*, 15–43.

Wilson, K. M., & Swanson, H. L. (2001). Are mathematics disabilities due to a domain-general or a domain-specific working memory deficit? *Journal of Learning Disabilities*, *34*(3), 237–248.

Wolfe, J. H., Alderton, D. L., Larson, G. E., & Held, J. D. (1995). Incremental validity of enhanced computer-administered testing (ECAT). San Diego, CA: Navy Personnel Research and Development Center.

Wonderlic (2018, October). Retrieved from https://www.wonderlic.com/about-wonderlic.

Appendices

Appendix A – Subtests of the Armed Services Vocational Aptitude Battery (ASVAB). Adapted from official-asvab.com/docs/asvab_fact_sheet.pdf.

Subtest	Description	Domain
General Science (GS)	Knowledge of physical and biological sciences	Science/Technical
Arithmetic Reasoning (AR)	Ability to solve arithmetic word problems	Math
Word Knowledge (WK)	Ability to select the correct meaning of a word presented in context and to identify best synonym for a given word	Verbal
Paragraph Comprehension (PC)	Ability to obtain information from written passages	Verbal
Mathematics Knowledge (MK)	Knowledge of high school mathematics skills	Math
Electronics Information (EI)	Knowledge of electricity and electronics	Science/Technical
Auto Information (AI)	Knowledge of automobile technology	Science/Technical
Shop Information (SI)	Knowledge of tools and shop terminology and practices	Science/Technical
Mechanical Comprehension (MC)	Knowledge of mechanical and physical principles	Science/Technical
Assembling Objects (AO)	Ability to determine how an object will look when its parts are put together	Spatial

Appendix B – Journal Article Information for Verbal Task Performance Measure

Please use the following information about six **journal articles** and create a properly formatted APA style reference page. Based on the tutorial you just watched, try to be as accurate as possible including proper *formatting, punctuation, capitalization, spelling, italics, grammar, and spacing*. You will have 15 minutes to complete this task.

1. Journal: European Journal of Marketing
Pages of article in journal: Pages 1245-1283
Volume of journal: Volume 41
Published: November 17, 2007
Author(s): Douglas Brownlie
Title of article: Toward effective poster presentations: An annotated bibliography

2. Journal: Group Processes & Intergroup Relations
Pages of article in journal: Pages 28-45
Volume of journal: Volume 16
Published: January 10, 2013
Author(s): Emery Cecile, Thomas S. Calvard, Meghan E. Pierce
Title of article: Leadership as an emergent group process: A social network study of personality and leadership

3. Journal: Psychological Review
Pages of article in journal: Pages 369-389
Volume of journal: Volume 94
Published: June 16, 1987
Author(s): Kay Deaux, Brenda Major
Title of article: Putting gender into context: An interactive model of gender-related behavior

4. Journal: Academy of Management Review
Pages of article in journal: Pages 889-913
Volume of journal: Volume 31
Published: September 24, 2006
Author(s): Piers Morgan, Cornelius J. Konig
Title of article: Integrating theories of motivation

5. Journal: Journal of Personality and Social Psychology
Pages of article in journal: Pages 1173-1182
Volume of journal: Volume 51
Published: March 26, 1986
Author(s): Reuben M. Baron, David A. Kenny
Title of article: The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations

6. Journal: Journal of Individual Differences
Pages of article in journal: Pages 185 – 189
Volume of journal: Volume 35

Published: January 1, 2014
Author(s): Matthias Ziegler, Christoph J. Kemper, and Peter Kruyen
Title of article: Short Scales – Five Misunderstandings and Ways to Overcome Them

Appendix C – Federal Tax Items for Quantitative Task Performance Measure

Rate	Individuals	Married Filing Jointly
10%	Up to $9,525	Up to $19,050
12%	$9,526 to $38,700	$19,051 to $77,400
22%	38,701 to $82,500	$77,401 to $165,000
24%	$82,501 to $157,500	$165,001 to $315,000
32%	$157,501 to $200,000	$315,001 to $400,000
35%	$200,001 to $500,000	$400,001 to $600,000
37%	over $500,000	over $600,000

Based on the information given for the people below, calculate and write in the amount of money that goes toward each of the following. Use the knowledge that you learned in the tutorial presented. Use the scratch paper and calculator provided, as needed. You will have 15 minutes to complete this task.

For John, assume that the social security tax rate is 6.0% and the Medicare tax rate is 1.5%.

	John
Gross Annual Income	$76,500
Standard Deduction	$8,200
Personal Exemption	$5,800
Taxable Income	
Employee's contribution to Social Security	
Employee's contribution to Medicare	
Employee's total FICA contribution	
Employer's total FICA contribution	
Federal taxes (excluding FICA taxes)	
Net take-home pay	

For Paul and Betty, assume that the social security tax rate is 5.0% and the Medicare tax rate is 2.0%.

	Paul and Betty (filing jointly)
Gross Annual Income	$236,700
Standard Deduction	$24,900
Personal Exemption	$18,900
Taxable Income	
Employee's contribution to Social Security	
Employee's contribution to Medicare	
Employee's total FICA contribution	
Employer's total FICA contribution	
Federal taxes (excluding FICA taxes)	
Net take-home pay	

Appendix D – APA Pretest & Posttest items for Verbal Task Knowledge Measure

1. How are entries ordered in a Reference list?
A. Title, author's name, date of publication, journal, volume, pages.
B. Author's name, date of publication, title, journal, volume, pages.
C. Date of publication, author's name, title, journal, pages, volume.
D. Author's name, title, date of publication, journal, pages, volume.

2. Which of the following is the correct APA Reference list entry?
A. Reese, G. (2000). ***Database Programming with JDBC and Java*****. Beijing: O'Reilly Media.**
B. Reese, George. *Database programming with JDBC and Java*. Beijing: O'Reilly Media. 2000.
C. Reese, G. (2000). *Database programming with JDBC and Java*. O'Reilly Media, Beijing, China.
D. Reese, G. (July 2000). *Database programming with JDBC and Java*. O'Reilly Media, Beijing, China.

3. Below is a reference page listing for a book with two authors. How should the title of the book be formatted?
Giorgis, C., & Glazer, J. I. (2009). Literature for young children: Supporting emergent literacy, ages 0 - 8 (6th ed). Boston, MA: Pearson Education.
A. The book title should be in quotations.
B. The book title should be in italics.
C. The book title should have all important words capitalized.
D. The book title should not have "Supporting" capitalized.
E. The book title should be listed without special formatting.
F. The book title should be underlined.

4. Select the answer that provides the correct reference page format for a journal article published online and in print.
A. Recent work on epistemic value. *American Philosophical Quarterly*, *44* (2), 85-110. Retrieved from http://www.jstor.org/stable/20464361.
B. Pritchard (2007). Recent work on epistemic value. Retrieved from http://www.jstor.org/stable/20464361.
C. Pritchard, D. (2007). Recent work on epistemic value. ***American Philosophical Quarterly*****,** ***44*** **(2), 85-110. Retrieved from http://www.jstor.org/stable/20464361.**
D. Pritchard, D. (2007, April). *American Philosophical Quarterly*, *44* (2), 85-110. Retrieved from http://www.jstor.org/stable/20464361.

5. In the following example of a reference page listing of a journal article retrieved from an electronic database, which item(s) of information are missing?
Florian. (2010). Challenges for interactivist-constructivist robotics. *New Ideas in Psychology*. 350-353. doi: 10.1016/j.newideapsych.2009.09.009
A. The DOI, the month, and the page number

B. The author's initials and the page number
C. **The author's initials, the journal volume and issue number**
D. The month, volume, and issue number

6. What is wrong about the following APA reference?
Bower, H. (2001). The Gender Identity Disorder in the DSM-IV Classification: A Critical Evaluation. *Australian and New Zealand Journal of Psychiatry, 35*, 1-8.

A. The title of the article should come before the date
B. Only the first word of the article title should be capitalized.
C. Only the first word of the journal title should be capitalized.
D. The title of the article should be italicized, not the journal.
E. The (2) in parentheses should not be used unless it is a book.
F. Only one of the author's initials should be included.

7. Which one of the following references is properly formatted?
A. Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, *35*, 127-145.
B. Alibali, M., W. (1999). How Children Change Their Minds: Strategy Change Can Be Gradual or Abrupt. *Developmental Psychology*, *35*, 127-145.
C. Alibali, M. W. (1999). *How children change their minds: Strategy change can be gradual or abrupt*. Developmental Psychology, 35, 127-145.
D. Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology, 35*, p. 127-145.

8. What is wrong with the following APA style journal reference?
Swan, J. T., Rail, D. Q., & Bushcombe, M. N. (1995). "Students and the problem of entitlement." *Social Work and Research*, 68(2), 127-137.
A. The article's title should not be in quotations.
B. The journal title should not be in italics.
C. The important words in the article title should be capitalized.
D. The date should include the month the article was published.

9. What type of citation is this?
Austin, J. H. (1998). *Zen and the brain: Toward an understanding of meditation and consciousness*. Cambridge, MA: MIT Press.
A. Journal article
B. Magazine article
C. Book chapter
D. **Book**
E. Newspaper article
F. Academic Publication

10. What is wrong with the following APA journal reference?
Carroll, L., Gilroy, P. J., & Ryan, J. (2002). Counseling transgendered, transsexual, and gender-variant clients. *Journal of Counseling and Development, 80*. 131-129.

A. The pages numbers should be listed as “p. 1-20.”
B. There should be a period after the journal title
D. The word “and” should be between the authors’ names instead of “&”
E. **There should be a comma after the journal number**
F. The article title is not formatted correctly.

11. When typing up a journal reference, how should the journal title be formatted?
A. Written in italics, only important words capitalized, followed by a period.
B. Written in bold, only first word capitalized, followed by a comma.
C. No special formatting, only first word capitalized, followed by a period.
D. Written in italics, only important words capitalized, followed by a comma.
E. Written in bold, only first word capitalized, followed by a comma.

12. What is wrong about the following APA reference?
Jacoby, W. G. (1994). Public Attitudes Toward Government Spending. *American Journal of Political Science*, *38*(2), 336-361.

A. The title of the article should come before the date
B. Only the first word of the article title should be capitalized.
C. Only the first word of the journal title should be capitalized.
D. The title of the article should be italicized, not the journal.
E. The (2) in parentheses should not be used unless it is a book.
F. Only one of the author’s initials should be included.

13. Which of the following is a properly formatted reference for the reference page?
A. Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war: An historical review. *American Political Science Review*. *97*(01), 75-98. doi: 10.1017/S0003055403000534.
B. Fearon, J. D., and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war: an historical review. *American Political Science Review*, *97*(01). 75-98. doi: 10.1017/S0003055403000534.
C. Fearon, J. D., & Laitin, D. D. (June 2003). Ethnicity, insurgency, and civil war: An historical review. *American political science review, 97*(01), 75-98, doi: 10.1017/S0003055403000534.
D. Fearon, J. D., and Laitin, D. D. (2003). Ethnicity, insurgency, and civil war: An historical review. *American Political Science Review*. *97*(01), 75-98. doi: 10.1017/S0003055403000534.
E. Fearon, J. D., & Laitin, D. D. (2003). Ethnicity, insurgency, and civil war: An historical review. *American Political Science Review*, *97*(01), 75-98. doi: 10.1017/S0003055403000534.

14. What is the correct order of items to be included in a reference?
A. Author, article title, year published, journal, volume, pages.
B. Author, year published, article title, journal, volume, pages.
C. Author, journal, year published, article title, page numbers
D. Article title, author, year published, page numbers, journal.

E. Author, year published, journal, article title, pages.
F. Author, year published, article title, journal, pages, volume.

15. What is wrong with the following APA style journal reference?
Eagly, A. H., & Carli, L. L. (1981). Sex of researchers and sex-typed communications as determinants of sex differences in influenceability: A meta-analysis of social influences. *Psychological Bulletin, 90*. 1-20.

A. The pages numbers should be listed as "p. 1-20."
B. There should be a period after the journal title
C. Influenceability should be capitalized since it is before the colon
D. The word "and" should be between the authors' names instead of "&"
E. There should be a comma after the volume number
F. The article title is not formatted correctly.

16. What is wrong with the following APA style journal reference?
Tidwell, L. C., & Walther, J. B. (2002). Computer-mediated communication effects on disclosure, impressions, and interpersonal evaluations: getting to know one another a bit at a time. *Human Communication Research*, *28,* 317-348.

A. Something that is not capitalized should be.
B. The journal title should not be in italics.
C. The date should include the month the article was published.
D. The initials and names of the authors are not formatted correctly.
E. There should be a period after the volume number.
F. None of the above – the reference is correct as is.

17. How should the names of the authors be formatted in an APA reference?
A. B. Harrison, J. F. Wood, V. Smith, & N. Westbrook (2004).
B. Harrison B., Wood J. F., Smith V., & Westbrook N. (2004).
C. Harrison, B., Wood, J. F., Smith, V., & Westbrook, N. (2004).
D. Harrison, B., Wood, J. F., Smith, V., and Westbrook, N. (2004).
E. Harrison, B. Wood, J. F. Smith, V. & Westbrook, N. (2004).
F. Harrison, B.; Wood, J. F.; Smith, V.; & Westbrook, N. (2004).

18. When should periods appear in an APA reference page?
A. After authors' initials, after article title, and after page numbers.
B. After authors' initials, after year published, after article title, and after page numbers.
C. After authors' initials, after year published, after article title, after journal title, and after page numbers.
D. After authors' initials, after year published, and after page numbers.
E. After authors' initials, after article title, after volume numbers, and after page numbers.

19. When should commas appear in an APA reference page?
A. Between authors, after journal title, and after volume number.
B. After article title, after journal title, and after volume number.

C. Between authors, after article title, after journal title.
D. Between authors, and after journal title.
E. After journal title, after volume number, and after year published.

20. Which of the below should you NOT see in an APA journal reference?
A. An italicized journal title with important words capitalized.
B. An article title with only one capitalized word.
C. A comma after journal title.
D. The issue number of the journal's volume.
E. A period after volume number.
F. A capitalized word after a colon.

Appendix E – Federal Taxes Knowledge Tests for Quantitative Task Knowledge

Rate	Individuals	Married Filing Jointly
10%	Up to $9,525	Up to $19,050
12%	$9,526 to $38,700	$19,051 to $77,400
22%	38,701 to $82,500	$77,401 to $165,000
24%	$82,501 to $157,500	$165,001 to $315,000
32%	$157,501 to $200,000	$315,001 to $400,000
35%	$200,001 to $500,000	$400,001 to $600,000
37%	over $500,000	over $600,000

1. Given the tax brackets for 2018 shown above, how much would a single person pay for federal taxes (excluding FICA taxes) if he/she has $151,800 of taxable income?
A. $33,428
B. $36,432
C. $29,336
D. $30,722
E. $34,662

2. Stacy has a gross annual salary of $65,000 and a taxable income of $55,000. Assume that Social Security deductions are at 6.0% and Medicare deductions are at 1.5%. How much money will Stacy have deducted to pay for Social Security?
A. $3,900
B. $3,300
C. $4,875
D. $4,125
E. $2,925

3. Kevin has a gross annual salary of $45,000 and a taxable income of $38,000. Assume that Social Security deductions are at 6.0% and Medicare deductions are at 1.5%. Between his own and his employer's contributions, how much total FICA taxes will be paid on Kevin's behalf?
A. $3,375
B. $5,700
C. $2,850
D. $6,750

4. Which of the following is true?

A. Standard deductions reduce taxable income, while personal exemptions are added to taxable income
B. If an individual pays $4,200/year on FICA taxes, his her employer pays an additional $2,100
C. Social Security and Medicare taxes are both FICA taxes that are taken out of your annual taxable salary
D. Standard deductions and personal exemptions both reduce the amount of taxable income a person has

5. Based on the tax brackets for 2018 shown above, what is the formula for calculating the federal taxes (excluding FICA taxes) for someone who has a taxable income of $80,000?
A. Federal taxes = 9525(.10) + (38700)(.12) + 80000(.22)
B. Federal taxes = (80000 – 38700)(.22) + (38700 – 9525)(.12) + 9525(.10)
C. Federal taxes = (80000 – 9525)(.10) + (80000 – 38700)(.12) + 80000(.22)
D. Federal taxes = 9525(.10) + (38700 + 9525)(.12) + (80000 + 38700)(.22)

6. Which of the following is true?
A. With all else being equal, a person making $82,400 gross annual salary and in the 22% tax bracket would receive more net take-home pay than a person making $83,000 and in the 24% tax bracket
B. A person who has $100,000 of taxable income will have ***at least*** $24,000 (24%) in federal taxes
C. A couple filing jointly who have a gross annual income of $23,000, a standard deduction of $3,000 and a personal exemption of $2,000 will be taxed at a rate of 12%
D. A couple filing jointly who have $332,000 of taxable income will have only $17,000 taxed at the 32% rate and the rest of the taxable income will be taxed at a lower rate

7. George has a gross annual salary of $77,900 and a taxable income of $68,200. George pays $4,200 in Social Security taxes. George's employer is contributing a total of $5,624 toward FICA taxes for George. How much is George paying for Medicare taxes?
A. $9,824
B. $1,424
C. $5,624
D. $2,812

8. Henry has a gross annual salary of $84,000 and a taxable income of $66,000. Assume that Social Security deductions are at 7.0% and Medicare deductions are at 2.5%. How much money will Henry have deducted to pay for Medicare?
A. $1,650
B. $7,980
C. $5,880
D. $4,620
E. $2,100

9. Michelle has a gross annual salary of $105,000 and a taxable income of $83,000. Assume that Social Security deductions are at 7.0% and Medicare deductions are at 2.5%. Between her own

and her employer's contributions, how much in total FICA taxes will be paid on Michelle's behalf?
A. $7,885
B. $19,950
C. $9,975
D. $15,770
E. $13,565

10. Betty has a gross annual salary of $43,500 and a taxable income of $32,700. Betty pays $2,600 in Social Security taxes. Her employer is contributing a total of $3,470 toward FICA taxes for Betty. How much is Betty paying for Medicare taxes?
A. $870
B. $3,470
C. $6,070
D. $2,600
E. $10,800

11. Given the tax brackets for 2018 shown above, how much would a single person pay for federal taxes (excluding FICA taxes) if he/she has $98,500 of taxable income?
A. $17,610
B. $23,640
C. $13,549
D. $17,930
E. $19,662

12. Travis has a gross annual salary of $94,000 and a taxable income of $82,000. Assume that Social Security deductions are at 6.0% and Medicare deductions are at 1.5%. How much money will Stacy have deducted to pay for Social Security?
A. $7,050
B. $5,640
C. $4,920
D. $6,150
E. $4,230

13. Kim has a gross annual salary of $38,000 and a taxable income of $29,000. Assume that Social Security deductions are at 5.5% and Medicare deductions are at 1.5%. Between his own and his employer's contributions, how much total FICA taxes will be paid on Kevin's behalf?
A. $4,060
B. $5,320
C. $2,660
D. $2,030
E. $3,040

14. Which of the following is true?
A. If an individual pays $5,600/year on FICA taxes, his her employer pays an additional $2,800

B. Because of standard deductions and personal exemptions, taxable income can be more than gross income in certain situations
C. Standard deductions and personal exemptions both reduce the amount of taxable income a person has
D. Social Security and Medicare taxes are both types of FICA taxes that are taken out of your annual taxable salary

15. Based on the tax brackets for 2018 shown above, what is the formula for calculating the federal taxes (excluding FICA taxes) for someone who has a taxable income of $81,500?
A. Federal taxes = 9525(.10) + (38700)(.12) + 81500(.22)
B. Federal taxes = (81500 – 9525)(.10) + (81500 – 38700)(.12) + 81500(.22)
C. Federal taxes = 9525(.10) + (38700 + 9525)(.12) + (81500 + 38700)(.22)
D. **B. Federal taxes = (81500 – 38700)(.22) + (38700 – 9525)(.12) + 9525(.10)**

16. Which of the following is true?
A. A person who has $200,000 of taxable income will have ***at least*** $70,000 (35%) in federal taxes
B. A couple filing jointly who have $197,000 of taxable income will have only $32,000 taxed at the 24% rate and the rest of the taxable income will be taxed at a lower rate
C. With all else being equal, a person making $38,600 taxable income and in the 12% tax bracket would receive more net take-home pay than a person making $38,800 taxable income and in the 22% tax bracket
D. A person who has a gross annual income of $14,000, a standard deduction of $3,000 and a personal exemption of $2,000 will be taxed at a rate of 12%

17. Paul has a gross annual salary of $56,200 and a taxable income of $47,800. George pays $3,250 in Social Security taxes. George's employer is contributing a total of $4,270 toward FICA taxes for George. How much is George paying for Medicare taxes?
A. $7,520
B. $2,135
C. $1,020
D. $1,625

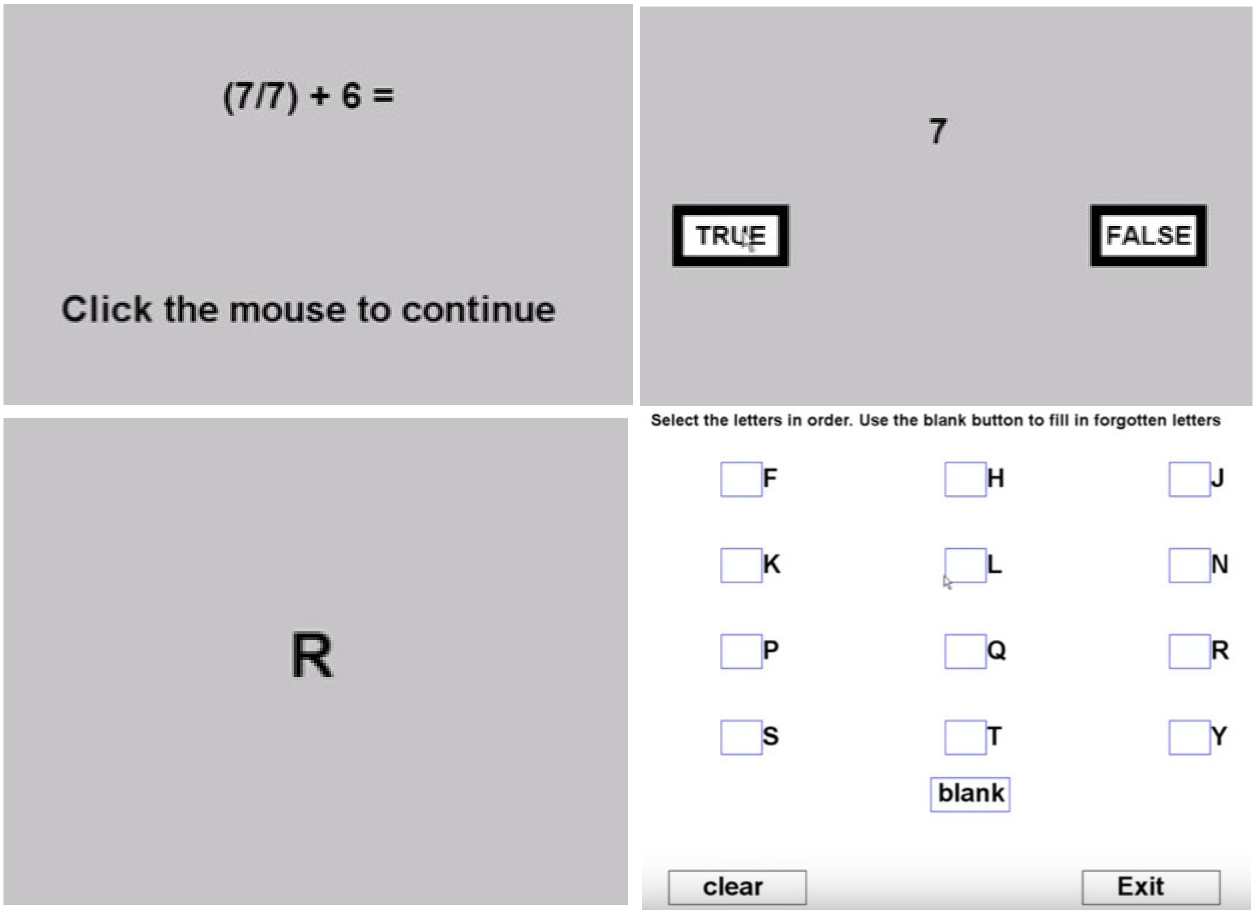
18. Bob has a gross annual salary of $91,000 and a taxable income of $81,700. Assume that Social Security deductions are at 6.5% and Medicare deductions are at 1.5%. How much money will Henry have deducted to pay for Medicare?
A. $1,226
B. $4,550
C. $4,085
D. $2,655
E. $1,365

19. Katie has a gross annual salary of $112,000 and a taxable income of $101,400. Assume that Social Security deductions are at 5.5% and Medicare deductions are at 2.0%. Between her own and her employer's contributions, how much in total FICA taxes will be paid on Michelle's behalf?
A. $7,605
B. $16,800
C. $15,210
D. $8,400
E. $11,840

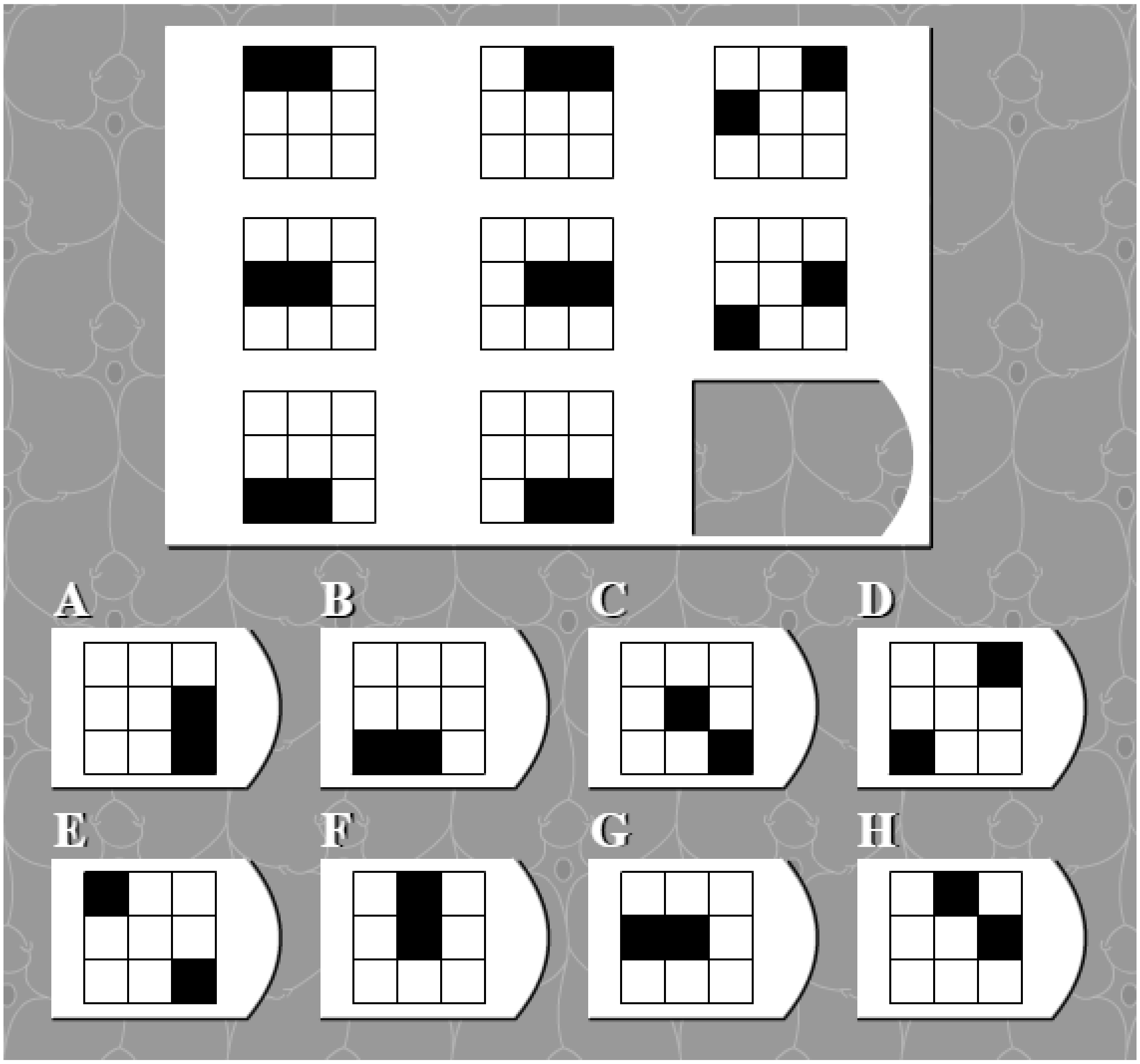
20. Susie has a gross annual salary of $41,600 and a taxable income of $35,500. Betty pays $2,280 in Social Security taxes. Her employer is contributing a total of $3,204 toward FICA taxes for Betty. How much is Betty paying for Medicare taxes?
A. $924
B. $1,848
C. $5,484
D. $4,560
E. $1,328

Appendix F – OSPAN

OSPAN. Example of one item from the OSPAN. Letters to be remembered were presented in blocks of three to seven.



Appendix G – A sample item from the Progressive matrices task (PMT)



Appendix H – Scoring Rubric for the APA Reference Page Task

Reference #	Criterion	Points scored
1	Participant spells everything correctly throughout the reference	
1	Participant includes periods in all the correct places.	
1	Participant places commas in the correct places throughout the reference	
1	Participant places parentheses in the correct places throughout the reference	
1	Participant properly capitalizes as needed throughout the reference	
1	Participant puts spaces in the correct places throughout the reference	
1	Participant puts necessary information in italics	
1	Participant places each item of information in the proper order in the entire reference	
2	Participant spells everything correctly throughout the reference	
2	Participant includes periods in all the correct places.	
2	Participant places commas in the correct places throughout the reference	
2	Participant places parentheses in the correct places throughout the reference	
2	Participant properly capitalizes as needed throughout the reference	
2	Participant puts spaces in the correct places throughout the reference	
2	Participant puts necessary information in italics	

2	Participant places each item of information in the proper order in the entire reference	
3	Participant spells everything correctly throughout the reference	
3	Participant includes periods in all the correct places.	
3	Participant places commas in the correct places throughout the reference	
3	Participant places parentheses in the correct places throughout the reference	
3	Participant properly capitalizes as needed throughout the reference	
3	Participant puts spaces in the correct places throughout the reference	
3	Participant puts necessary information in italics	
3	Participant places each item of information in the proper order in the entire reference	
4	Participant spells everything correctly throughout the reference	
4	Participant includes periods in all the correct places.	
4	Participant places commas in the correct places throughout the reference	
4	Participant places parentheses in the correct places throughout the reference	
4	Participant properly capitalizes as needed throughout the reference	
4	Participant puts spaces in the correct places throughout the reference	
4	Participant puts necessary information in italics	

4	Participant places each item of information in the proper order in the entire reference	
5	Participant spells everything correctly throughout the reference	
5	Participant includes periods in all the correct places.	
5	Participant places commas in the correct places throughout the reference	
5	Participant places parentheses in the correct places throughout the reference	
5	Participant properly capitalizes as needed throughout the reference	
5	Participant puts spaces in the correct places throughout the reference	
5	Participant puts necessary information in italics	
5	Participant places each item of information in the proper order in the entire reference	
6	Participant spells everything correctly throughout the reference	
6	Participant includes periods in all the correct places.	
6	Participant places commas in the correct places throughout the reference	
6	Participant places parentheses in the correct places throughout the reference	
6	Participant properly capitalizes as needed throughout the reference	
6	Participant puts spaces in the correct places throughout the reference	
6	Participant puts necessary information in italics	

6	Participant places each item of information in the proper order in the entire reference	
	Participant puts all references in the correct order on the reference page	
Total Score:	Sum of the points in the boxes to the right.	

Appendix I - Scoring Rubric for the Federal Tax Calculations Task

	John	Points	Paul and Betty (filing jointly)	Points
Gross Annual Income	$76,500		$236,700	
Standard Deduction	$8,200		$24,900	
Personal Exemption	$5,800		$18,900	
Taxable Income	$62,500		$192,900	
Employee's contribution to Social Security	$4,590		$11,835	
Employee's contribution to Medicare	$1,147.50		$4,734	
Employee's total FICA contribution	$5,737.50		$16,569	
Does Employee's total FICA = Social Security + Medicare?				
Employer's total FICA contribution	$5,737.50		$16,569	
Does Employer's total FICA equal employee's total FICA?				
Federal taxes (excluding FICA taxes)	$9,689.50		$34,875	
Does Federal taxes equal the proper calculations, but using gross income?	$12,769.50		$45,387	
Net take-home pay	$61,073		$185,256	
Does net take-home pay = gross annual income – total FICA – federal taxes?				
Total points (add up columns 3 and 5) **/22**				