Intelligence 50 (2015) 186–195

Contents lists available at ScienceDirect

Intelligence

Wonderlic, working memory capacity, and fluid intelligence ☆

Kenny L. Hicks *, Tyler L. Harrison, Randall W. Engle

Georgia Institute of Technology, United States

ARTICLE INFO

Article history:
Received 26 February 2014
Received in revised form 14 March 2015
Accepted 23 March 2015
Available online xxxx

Keywords:
Wonderlic
Working memory capacity
Fluid intelligence
Personnel selection
General mental abilities

ABSTRACT

Despite the widespread popularity of the Wonderlic Personnel Test, evidence of its validity as a measure of intelligence and personnel selection is limited. The present study sought to better understand the Wonderlic by investigating its relationship to multiple measures of working memory capacity and fluid intelligence. Our results show that Wonderlic has no direct relationship to fluid intelligence once its commonality to working memory capacity is accounted for. Further, we found that the Wonderlic was a significant predictor of working memory capacity for subjects with low fluid intelligence, but failed to discriminate as well among subjects with high fluid intelligence. These results suggest that the predictive power of the Wonderlic could depend on the characteristics of the sample it is administered to, whereas the relationship between fluid intelligence and working memory capacity is robust and invariant to the cognitive capabilities of the sample.

1. Introduction

Despite numerous validation studies and over 200 million administrations in commercial and research domains, the validity of the Wonderlic is largely unknown. Its claim to quickly and easily measure intelligence might appeal to researchers and test-makers alike, but it is important that this claim first be substantiated with supporting research.

The present work seeks to understand this instrument in relation to working memory capacity and novel reasoning. Because of the predictive power of working memory capacity, researchers have begun to use it as a vehicle to understand the core mechanisms involved in reasoning and other constructs of interest (Oberauer et al., 2007). Using measures grounded in a solid theoretical framework such as working memory capacity and fluid intelligence, we can make better inferences about what drives performance on the Wonderlic.

2. The Wonderlic

Eldon Wonderlic's early research interests explored predictors of job performance. His first experiment was a large-scale, exploratory study including indicators of personality, intelligence, and supervisor ratings. The results showed that only the indicator of intelligence, the Otis Self-administering Test of Mental Ability was a significant predictor of job performance (Stevens & Wonderlic, 1934). Further evidence for the validity of the Otis Test for personnel selection was found when Wonderlic conducted a follow-up study in which the number of questions missed and omitted on the Otis Test was found to be significantly correlated with the job performance of office managers (Stevens & Wonderlic, 1934).

Despite these findings, the Otis Test drew criticism from researchers for having poor psychometric properties (Hovland & Wonderlic, 1939; Stevens & Wonderlic, 1934). For instance, although the Otis Test claimed to be a power test, item level analyses on several parallel forms of the test found that items were not ranked properly from easiest to most difficult (Hovland & Wonderlic, 1939). Further, the validity coefficients

☆ This research was supported by the Office of Naval Research Grant N00014121040.We would like to thank Zach Shipstead for helpful comments on the data analysis and to members of the Attention and Working Memory Lab at Georgia Tech for their help running subjects and collecting the data for the current study. Further, we thank Kasey Harris for helpful suggestions on revisions to the manuscript.

* Corresponding author at: School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332, United States.
E-mail address: khicks6@gatech.edu (K.L. Hicks).

http://dx.doi.org/10.1016/j.intell.2015.03.005
0160-2896/

for the Otis Test were found to be severely restricted due to an excess of easy items. This issue is reflected by the finding that the Otis Test is poor at discriminating at the extremes, particularly at the higher end of cognitive performance (Cattell, 1931). The Wonderlic test was adapted from the Otis Test and shortened into a 12-minute measure by selecting a subset of Otis items based on their psychometric properties. This new instrument was marketed specifically for personnel selection. However, early research on the Wonderlic Personnel Test found that it also suffered from an inability to discriminate at the upper end of the distribution (Buckley, 1957; Wonderlic & Hovland, 1939), a point that we will return to later.

3. Validation

Due to a growing concern for the lack of standards for testing in 1950, the APA appointed the Committee on Test Standards, led by Lee J. Cronbach. The group was tasked with developing test standards for psychological and educational measurement. Their report included recommendations for determining test validity (Cronbach & Meehl, 1955). Guided by this work, we consider two issues with the Wonderlic. First, the best demonstration of validity should be criterion validity, how well it predicts real world job performance. In addition, there should be a theoretical account for the underlying mechanisms responsible for performance on the Wonderlic.

Regarding the issue of criterion validity, results supporting the validity of the Wonderlic are almost exclusively demonstrated by their correlation with other intelligence measures (i.e. construct validity) rather than how well the test predicts actual job performance (i.e. criterion validity). When interpreting these correlations, it is important to consider that many of the intelligence tests used in the 20s and 30s such as the Army Alpha and Beta, Otis Self-administering Test, Wonderlic Personnel Test, and the Wechsler–Bellevue, originated from the Stanford–Binet. Therefore, correlations between the Wonderlic and other test batteries developed during this time will be inflated due to domain-specific overlap in item content.

A single piece of evidence is frequently cited as ample reliability and validity evidence for the Wonderlic. In this study, the Wonderlic and the Wechsler Adult Intelligence Scale — Revised (WAIS-R) were administered to 120 community subjects (Dodrill, 1981). The results of this study found that the Wonderlic and the WAIS-R were almost perfectly correlated ($r = .91$–$.93$). Despite this evidence, earlier work comparing the Wonderlic to the WAIS failed to find such a strong correlation ($r = .65$), suggesting that the two tests are related, but far from isomorphic (Buckley, 1957). This finding is also supported by other researchers who have failed to find correlations between the Wonderlic and WAIS at the magnitude of Dodrill's earlier work (Edinger, Shipley, Watkins, & Hammett, 1985).

The validation research mentioned thus far has only compared performance on the Wonderlic to the WAIS. As we have mentioned before this is problematic because the items are inherently related. Further, discrepancies in correlations reported so far may have arisen due to differences in the composition of the studies' samples, or it may be the case that the Wonderlic fails to systematically predict certain aspects of intelligence. The Cattell–Horn model of *g* specifies two distinct aspects of intelligence that drive cognitive performance. The first is referred to as crystallized intelligence (gC), which taps general knowledge and education such as knowledge of vocabulary definitions or state capitals. The second, fluid intelligence (gF), taps the ability to derive logical solutions to novel problems (Carroll, 1982; Cattell, 1963).

To date only two studies have reported correlations between the Wonderlic and distinct gC and gF subscales. Bell, Matthews, Lassiter, and Leverett (2002) assessed the Wonderlic and the Kaufman Adult and Adolescent Intelligence Test (KAIT), and found the Wonderlic to be a significant predictor of both crystallized and fluid abilities. However, more recent work correlating the Wonderlic and the Woodcock–Johnson — Revised (Matthews & Lassiter, 2007) demonstrated that the Wonderlic was related to gC, but not gF. From a theoretical perspective, this result suggests that while the Wonderlic is a reliable predictor of learned knowledge, it has failed to reliably predict gF, the ability to learn and adapt in situations that require novel reasoning.

Despite researchers submitting the Wonderlic to rigorous validation studies, little is known about the specific cognitive mechanisms responsible for individual differences on the test (Bosco & Allen, 2011; Culbertson, Huffcutt, & Goebl, 2013). This question is difficult to assess, as the Wonderlic was not developed from an underlying theory of cognitive performance. This and the proprietary nature of the test make revealing specific cognitive mechanisms, or processes, difficult. This is, in part, a result of ability testing beginning outside of psychological theory (Cronbach & Meehl, 1955; Anastasi, 1967; Sternberg, 1982; Sternberg & Kaufman, 2011; Embretson & Reise, 2000).

Working memory capacity captures specific aptitudes beyond gF (Bosco & Allen, 2011; Hambrick, Oswald, Darowski, Rench, & Brou, 2010; König, 2005). Recent work in this area challenges the idea that the Wonderlic is the best indicator of performance in the laboratory or on the job. For instance, previous research shows that working memory capacity predicts performance on air traffic control simulations (as cited in Ackerman, Beier, & Boyle, 2007), SAT performance (Engle, Kane, & Tuholski, 1999; Turner & Engle, 1989), academic performance, job performance, and in multi-tasks designed to simulate high-stakes work environments (Hambrick et al., 2010). In contrast, research finds that Wonderlic fails to predict academic performance (Chamorro-Premuzic & Furnham, 2008; Furnham, Chamorro-Premuzic, & McDougall, 2002; McKelvie, 1994) and has an inconsistent relationship to predictors of job performance such as customer service or sales volume (Barrick, Mount, & Strauss, 1993; Frei & McDaniel, 1998; Hogan & Hogan, 1995; Rode, Arthaud-Day, Mooney, Near, & Baldwin, 2008).

In addition to inconsistent findings in traditional job settings, research on the Wonderlic and NFL performance does not support the validity of the instrument in sports settings, despite widespread use in this field. Research in this area finds that the Wonderlic does not predict future NFL performance, selection decisions during the draft, or the number of games started (Lyons, Hoffman, & Michel, 2009). Conversely, experimental research has demonstrated that tactical decision making in sports is dependent on working memory capacity (Furley & Memmert, 2012). Additional work shows that working memory is also critical for coordinating activities in groups. For instance, Furley & Memmert (2013) found that attention control guides decision making in tasks that simulate the role of football quarterback. As the number of interactions the quarterback had with other players increased, so did the attentional demands of the task.

To date, the majority of research investigating working memory capacity focuses on developing a theory to specify its role in cognition, and researchers have made substantial progress toward this goal. Numerous studies have shown that working memory capacity plays an integral role in situations that require subjects to maintain goal-relevant information in the focus of attention and accurately retrieve information from long-term memory (Cowan, 2001; Engle et al., 1999; Kane et al., 2004; Shipstead, Redick, Hicks, & Engle, 2012; Unsworth & Engle, 2007). The role of working memory capacity is particularly important in situations when it is difficult to control and direct attention, due to an introduction of task interference, irrelevant distractions, or pressure to perform quickly while remaining accurate.

Research has shown that working memory capacity adds almost twice as much unique variance in multi-tasking performance as do general mental abilities (Hambrick et al., 2010). Further, recent studies demonstrate that working memory capacity is the best predictor of multi-tasking (e.g., Damos, 1993; Hambrick & Meinz, 2011; Hambrick et al., 2010; Stankov, Fogarty, & Watt, 1989). Working memory capacity remained predictive of multi-tasking performance even after controlling for individual differences in intelligence (König, 2005).

In a study examining intelligence and working memory capacity in the workplace, Bosco and Allen (2011) administered a working memory battery comprised of the Complex Span and a flanker task. The authors found that the Wonderlic and the working memory battery were equally predictive of performance on a job simulation task. However, when supervisor ratings and actual job performance were assessed, working memory capacity was the best predictor.

Bosco and Allen also assess the Wonderlic and adverse impact in their 2011 study. In addition to being a better predictor of real-world job metrics, working memory capacity was less vulnerable to adverse impact when compared to the Wonderlic. When the authors evaluated performance between African-Americans and Whites, they found major group differences on the Wonderlic. Tests of cognitive ability have the potential for adverse impact if performance differs across race, religion, national origin, age, or gender, and could lead to test bias in high-stakes testing situations and might disqualify applicants from employment (i.e. pre-employment testing). The saturation of *g* in measures of cognitive ability has been considered as a reason for race-group differences on these measures (Jensen, 1980, 1984, 1998; Hunter & Hunter, 1984). Test developers and researchers interested in tests of general mental ability should minimize adverse impact by developing and administering tasks that are less dependent on previous knowledge, such as measures of working memory capacity. Further, employers and practitioners who use the Wonderlic for selection purposes should be aware of adverse impact potential.

In a follow-up study, Bosco and Allen found that the Wonderlic was associated with a racial difference score that was 50% greater than the composite score of working memory capacity. Additionally, the authors found that working memory capacity was a significant predictor of a task simulating the work environment, whereas the Wonderlic was not. Finally, a third experiment showed that the working memory battery predicted an additional 7.2% of the variance on a measure of job simulation, as well as an additional 5.2% of the variance in supervisor ratings of job performance above and beyond the Wonderlic. Analogous to previous findings, Wonderlic performance was associated with ethnic differences in supervisor ratings. Scores on the working memory battery were not.

4. Aims of the present study

Our first aim in the present study was to determine if the Wonderlic is related to a factor of fluid intelligence. Although demonstrating a relationship between the Wonderlic and fluid intelligence lends evidence for the test's validity, it does not provide a complete picture of what the Wonderlic represents. Therefore, our second aim was to test the relationship between the Wonderlic and various tasks of working memory capacity. We chose to include measures that reflect multiple paradigms, carrying different theoretical considerations. For example, the Visual Arrays tasks measure the number of discrete slots or "chunks," that one can maintain in short-term memory, while the Complex Span tasks have traditionally been used to measure executive attention.

An added advantage of using fluid intelligence and working memory capacity is that both are predictive across a host of real-world tasks, and it is the chief interest of our lab to understand the theoretical mechanisms that drive these cognitive abilities (Engle, 2002). Working memory capacity and fluid intelligence have demonstrated their predictive power outside the laboratory. For instance, both working memory capacity and fluid intelligence predict scores on the ASVAB (Alderton, Wolfe, & Larson, 1997; Gottfredson, 1997), making them powerful tools of selection.

5. Participants

We collected data from 63 Georgia Tech students and 71 community members in Atlanta. All participants completed three measures of the Complex Span, three measures of fluid intelligence, three change detection tasks, and the Wonderlic Personnel Test — Revised. Less than 2% of the current data was missing due to computer and/or experimenter errors. In order to maximize power, we handled these data by imputing the missing values using the Expectation–Maximization algorithm in EQS 6.1 (Bentler & Wu, 2005).

To test whether or not the Wonderlic discriminates for both high and low ability subjects, we divided subjects into groups of high and low fluid intelligence to determine if working memory capacity and the Wonderlic shared overlapping variance within each group.

6. Procedure

We examined the psychometric properties of the Wonderlic assessment by analyzing its correlation with factors of working memory capacity and fluid intelligence. All tasks with the exception of the Wonderlic were conducted using E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). The Wonderlic was scored by the Wonderlic Company.

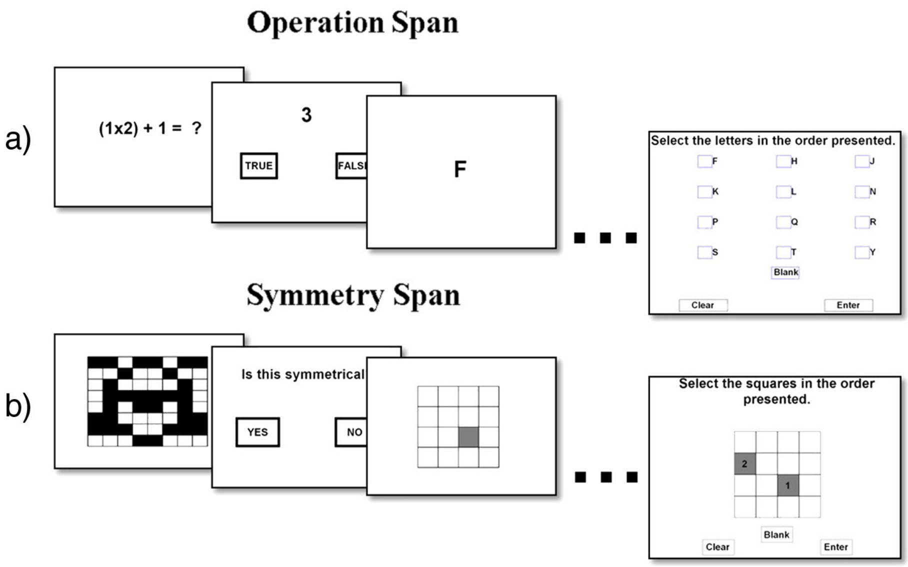

Fig. 1. Examples of the (a) operation span where to-be-recalled items are interleaved with simple math equations that must be solved, (b) and Symmetry span where subjects judge the symmetry of a picture shown in an 8 × 8 grid (Harrison et al., 2013).

6.1. Task descriptions

6.1.1. Automated Operation span (Unsworth et al., 2005)

The subject first completed a practice procedure in which they answer a series of simple math operations ($1 \times 2 + 1 = ?$); after the math practice, subjects' maximum time allotted to solve the math problems on the real trials is calculated by their mean reaction time plus 2.5 standard deviations. Subjects also performed a practice procedure where they were presented with two letters and were required to recall them in the order they were presented. After practice, subjects were presented with a list of 15 trials of three to seven randomized letters interleaved with simple math operations. An example is shown in Fig. 1a.

6.1.2. Symmetry span (Unsworth, Redick, Heitz, Broadway, & Engle, 2009)

This task required the subject to judge whether a picture is symmetrical at the vertical axis while remembering two to five specific locations highlighted on a 4 × 4 grid. An example is presented in Fig. 1b.

6.1.3. Reading span (Unsworth et al., 2009)

Subjects were asked to judge whether a sentence made logical sense. After the sentence judgment, subjects were presented with three to seven letters to recall in their proper serial order.

6.1.4. Raven's Advanced Progressive Matrices (Raven, 1990)

Subjects were presented with a 3 × 3 matrix of figures in which the lower right part of the matrix was missing. Subjects had to select a figure from one of eight answer choices that logically completed the matrix. Subjects had 10 min to complete 18 problems.

6.1.5. Letter Sets (Ekstrom, French, Harman, & Dermen, 1976)

On each trial, five four-letter strings were presented. Four of the sets followed a specific rule. The test-taker needed to discern this rule and decide which string did not follow it.

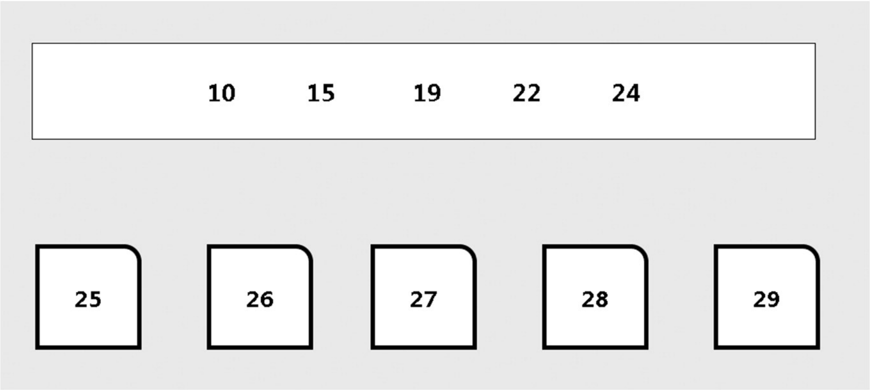

Fig. 2. Example of the Number Series task. Subjects are shown a series of numbers that follow a logical rule. The subject's task is to determine the next number in the sequence (e.g., 25) and choose it from a set of available options.
Adapted from Thurstone (1938).

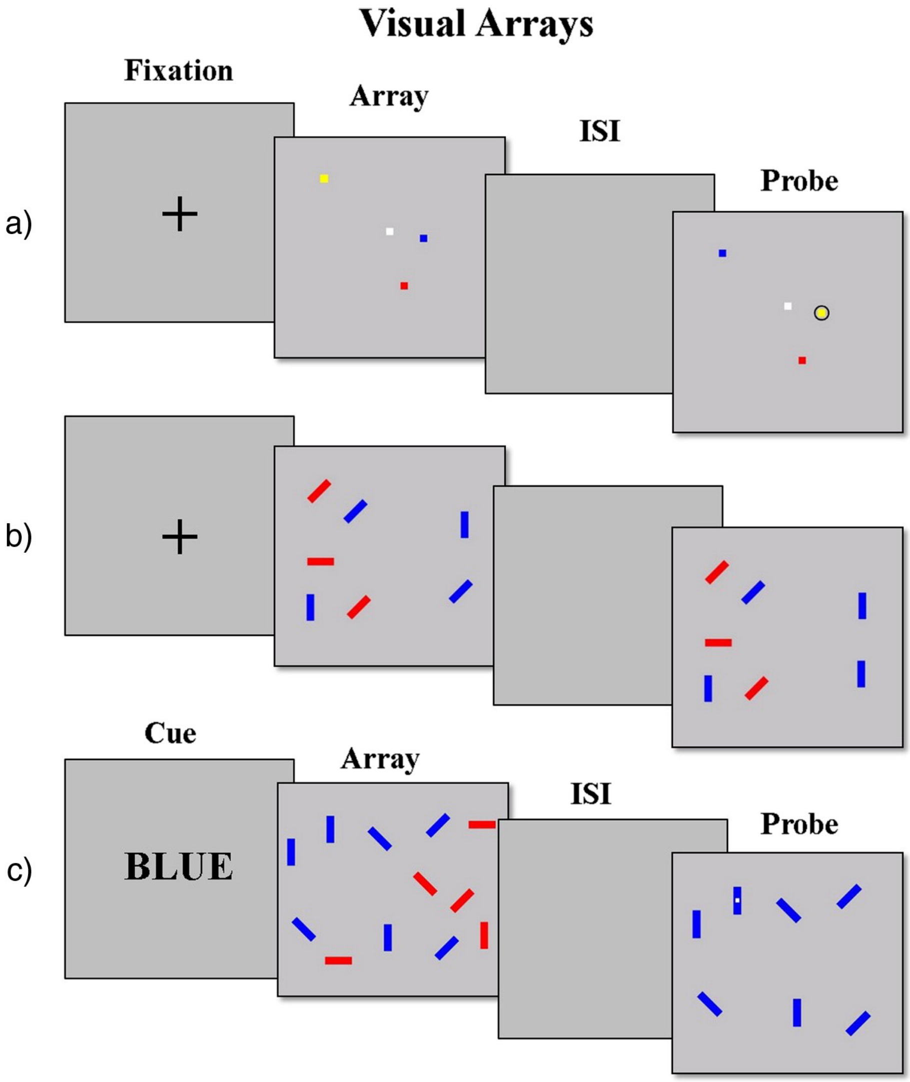

Fig. 3. The Visual Arrays tasks (Shipstead et al., 2014). In (a) the subject determines whether an encircled box has changed color. (b) The subject indicates whether the position of the box has changed (i.e., changed orientation from vertical to horizontal). (c) The subject decides whether the box containing a white dot has changed position.

Seven minutes were given to complete 30 problems. The dependent variable was the number of correct responses.

6.1.6. Number Series (Thurstone, 1938)

Subjects were presented with a series of numbers and asked to select one out of the five choices that represented the next logical number in the sequence. Subjects had 5 min to complete 15 problems (see Fig. 2).

6.1.7. Visual Arrays 1 (color judgment, Fig. 3a) (Shipstead, Lindsey, Marshall, & Engle, 2013)

The subject is shown either 4, 6, or 8 colored squares for 250 ms, followed by a 900 ms inter stimulus-interval (ISI). After the ISI, the array of squares reappeared with one square circled. The subject must indicate whether the square is the same color as its original presentation. The color of each square was randomized at the beginning of each trial. All possible color

Table 1
Descriptive statistics.

Variable	Min	Max	M	SD	Sk	Kurtosis
Wonderlic	6	39	23.88	7.811	.029	−.905
Ravens	0	16	9.48	3.874	−.348	−.658
LS	0	18	10.37	3.752	−.333	−.367
NS	1	15	8.95	3.428	−.370	−.644
Ospan	10	75	56.19	14.133	−.941	.540
Sspan	0	42	27.11	9.327	−.439	−.455
Rspan	0	75	50.99	16.509	−.837	.157
VA1	−.57	5.90	3.65	1.273	−.969	1.039
VA2	−.65	5.47	2.69	1.371	−.430	−.469
VA4	−1.08	5.35	2.02	1.255	−.143	−.221

Note: Wonderlic = Wonderlic Personnel Test − Revised; Raven = Raven's Advanced Progressive Matrices; LS = Letter Sets; NS = Number Series task; Ospan = Automated Operation span; Sspan = Symmetry span; Rspan = Reading span; VA1 = Visual Arrays, color judgement; VA2 = Visual Arrays, orientation judgement; VA4 = Visual Arrays, selective orientation judgement.

Table 2
Correlations among variable.

	Wonderlic	Raven	LS	NS	Ospan	Sspan	Rspan	VA1	VA2	VA4
Wonderlic	–									
Raven	0.62	–								
LS	0.60	0.63	–							
NS	0.68	0.63	0.71	–						
Ospan	0.46	0.39	0.40	0.60	–					
Sspan	0.45	0.60	0.48	0.50	0.50	–				
Rspan	0.52	0.52	0.50	0.46	0.62	0.56	–			
VA1	0.47	0.51	0.48	0.48	0.32	0.39	0.41	–		
VA2	0.34	0.51	0.35	0.36	0.21	0.40	0.32	0.58	–	
VA4	0.49	0.52	0.55	0.57	0.38	0.45	0.44	0.67	0.64	–

Note. Wonderlic = Wonderlic Personnel Test — Revised; Raven = Raven's Advanced Progressive Matrices; LS = Letter Sets; NS = Number Series task; Ospan = Automated Operation span; Sspan = Symmetry span; Rspan = Reading span; VA1 = Visual Arrays, color judgement; VA2 = Visual Arrays, orientation judgement; VA4 = Visual Arrays, selective orientation judgement. All correlations are significant at (p = 0.05).

options included green, purple, yellow, blue, white, and black. The dependent variable of interest was a Cowan's *k* (cf. Cowan et al., 2005).

6.1.8. Visual Arrays 2 (orientation judgment, Fig. 3b)

Subjects were presented with either 5 or 7 rectangles in various orientations. All rectangles were either red or blue and remained the same color within each trial. The subject was asked to indicate whether any of the squares had changed in orientation from the first time they were presented.

6.1.9. Visual Arrays 4 (selective orientation judgment, Fig. 3c)

This task was very similar to the Visual Arrays 2 task. Subjects were presented with 5 or 7 red and blue squares in various orientations. The difference in this task is that subjects were cued at the beginning of each trial. This cue indicated which color square the subject should attend to. In the example presented in Fig. 3c the subject is told to attend to the blue squares and is asked to judge whether the square with a white dot has changed orientation.

6.1.10. Wonderlic Personnel Test — Revised (Wonderlic, 2007)

The test taker was given 12 min to solve a 50-item reasoning test. Potential test items included numerical (An item costs $5. How much would four of those items cost? Answer: $20), verbal (i.e., analogies or word comparisons), and spatial reasoning items (i.e., matrix reasoning problems where the subject is asked to provide a missing element in a matrix comprised of different geometric shapes).

7. Results

When judging how well a given model fits the data, researchers judge the overall agreement of multiple fit indices. Several fit statistics are reported for the models included here. The standardized root mean square residual (SRMR) evaluates how well the specified model represents the raw variance–covariance matrix, while the root mean square error of approximation (RMSEA) tests how accurately the model reproduces the correlation matrix. The cut-off values for these indices vary in the psychometric literature. For instance, some authors suggest values ≤ 0.08 for the SRMR and values ≤0.10 for the RMSEA to indicate acceptable model fit (Kline, 2011; MacCallum, Browne, & Sugawara, 1996). However, Hu and Bentler suggest the more stringent cut-off values of ≤0.06 for SRMR and ≤0.08 for RMSEA to indicate acceptable fit (1999). Both the comparative fit index (CFI) and non-normed fit index (NNFI) evaluate the specified model in comparison to a null model. The NNFI corrects for model complexity while the CFI is less affected by sample size. Acceptable fit for these indices is also variable in the literature. Kline (2011) suggests 0.90 or above, while Hu and Bentler (1999) recommend values of 0.95 and above. Another fit measure, the chi square, is also a goodness of fit test but is highly influenced by sample size as well. Therefore, the statistic that is typically reported is the chi square divided by degrees of freedom (chi square/df) where values less than 3 are considered favorable. See Kline (2011) for a more in depth discussion regarding fit indices. For model comparisons, an additional index is reported to demonstrate the model fit between models (the AIC) (Jöreskog, 1993).

Table 3
Fit statistics for confirmatory factor analysis.

χ^2	df	χ^2/df	RMSEA	SRMR	NNFI	CFI
65	24	2.72	0.11	0.05	0.90	0.94

Descriptive statistics for all tasks are found in Table 1. We submitted all indicators of fluid intelligence, Complex Span, and Visual Arrays to a confirmatory factor analysis (CFA) (see Fig. 4). Model fit was acceptable (CFI = 0.94, NNFI = 0.90, SMR = 0.05, RMSEA = 0.11[1]) (see Table 3).

Next, we conducted a structural equation model to reveal the relationship between the Wonderlic and factors of working memory capacity and fluid intelligence. We created latent factors for the Complex Span and Visual Arrays, and loaded them onto a common factor in order to maximize the power and interpretability of the results.[2] Although previous research has shown that the Visual Arrays and the Complex Span tasks are separable aspects of working memory capacity (Shipstead et al., 2012), the current model loaded both latent constructs onto a single factor as a way to explore the predictive power of their shared variance. Both Complex Span and Visual Arrays

[1] The RMSEA tends to over-reject true population models when the sample size is less than 250 subjects (i.e. there is a substantial increase in the probability of a Type II error) (Hu & Bentler, 1999). In our case, Hu and Bentler argue that less emphasis be placed on the RMSEA due to a limited sample size.

[2] Due to the model containing a second order factor, the variance of the common factor was set to 1. Additionally, since the Wonderlic is only a single indicator, we estimated the error variance to be .10 in order to identify the model.

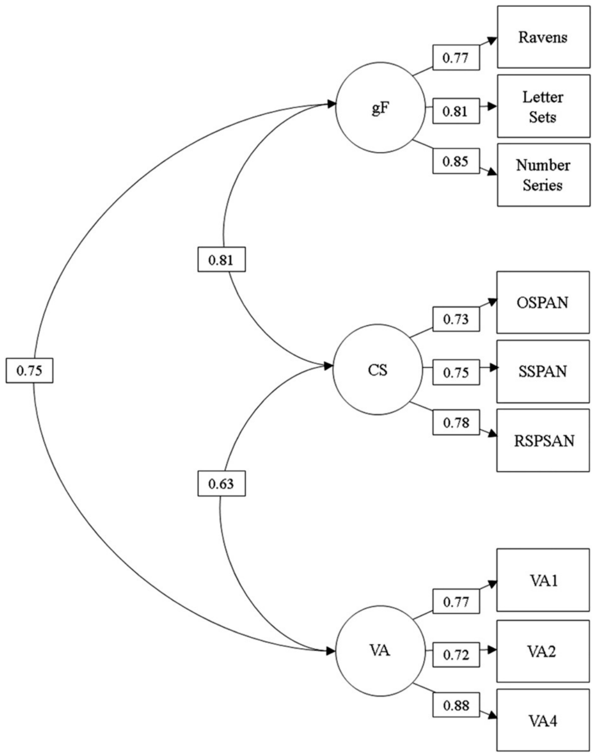

Fig. 4. Confirmatory factor analysis (CFA) of fluid intelligence = a factor comprised of the Raven, Letter Sets, Number Series; Complex Span = a factor of the Operation, Reading, and Symmetry span; Visual Arrays = a factor comprised of all three Visual Arrays tasks. Note: gF = fluid intelligence, CS = Complex Span, VA = Visual Arrays.

factors loaded highly on the common factor (>.80 CS, >.70 VA, respectively). Since measurement error cannot be derived from a single indicator, we estimated a factor of the Wonderlic by setting the relationship between the latent factor and the manifest variable to 1 as recommended by Kline (2011).

Model 1 tested whether the Wonderlic, noted as general mental ability (GMA), mediated the relationship between working memory capacity and gF (Fig. 5). Overall model fit was poor (see Table 4). In particular, χ^2/df > 3, RMSEA > 0.10, SRMR > 0.12.

Model 2 specified a direct link between working memory capacity and fluid intelligence (r = 0.91), shown in Fig. 6. Overall model fit was acceptable (see Table 4). Further, after the direct link between working memory capacity was added, the Wonderlic was no longer a significant predictor of fluid intelligence. Model fit was acceptable. Adding a direct path

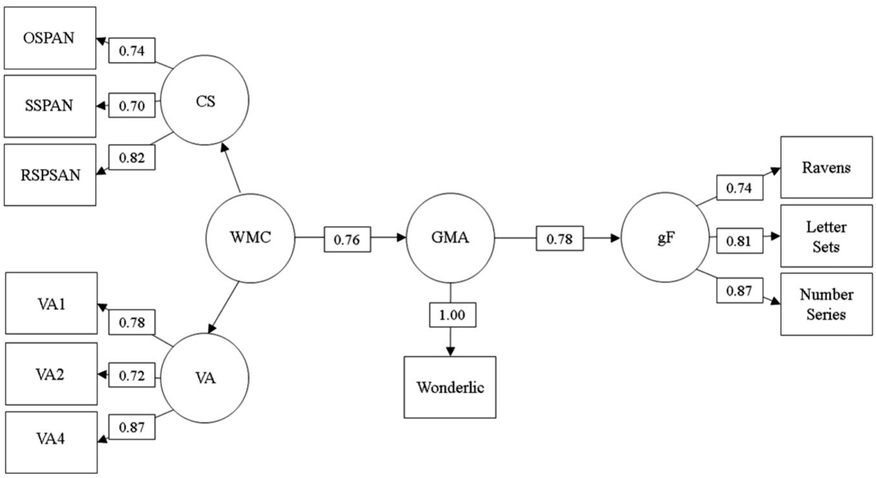

Fig. 5. Structural equation model testing whether Wonderlic (GMA) mediates the relationship between working memory capacity and fluid intelligence. Complex Span = a factor comprised of the Operation, Reading, and Symmetry span. Visual Arrays = a factor comprised of all three Visual Arrays tasks. WMC = composed of both latent factors of the Visual Arrays and Complex Span. Note: gF = fluid intelligence, CS = Complex Span, VA = Visual Arrays.

Table 4
Fit statistics for working memory capacity mediation model.

	χ^2	df	χ^2/df	RMSEA	SRMR	NNFI	CFI	AIC
Model 1	119	32	3.7	0.14	0.12	0.99	0.99	55.62
Model 2	72	31	2.3	0.10	0.05	0.99	0.99	10.11

from WMC to gF lead to a substantial decrease in the AIC and chi-square, indicating that Model 2 is preferred over Model 1.

Next, we explored whether the predictive power of the Wonderlic was dependent on the ability level of the subjects under investigation. We tested how the Wonderlic correlated with measures of working memory capacity when subjects were split into groups of high and low fluid intelligence. Analyzing group differences is highly contested in the intelligence literature. The proper classification of subjects to their respective group(s) is not straightforward. A common technique for splitting groups is based on splitting factor scores on a latent construct such as fluid intelligence (for more robust methods see Molenaar, Dolan, Wicherts, & van der Maas, 2010). Factor scores can be computed in several ways, and although they are highly correlated, the various methods may fail to rank-order subjects in the same way (see DiStefano, Zhu, & Mindrila, 2009 for a review of these methods). However, we chose the Anderson–Rubin method as it preserves the underlying variance–covariance matrix.

Factor scores were derived by submitting all measures of fluid intelligence, Complex Span, and Visual Arrays, to a confirmatory factor analysis specifying three factors using principal axis factoring and a Promax rotation as the factors were all correlated.

We split the current sample into subjects with high and low fluid intelligence ("high gF" and "low gF") by ranking each subject on the factor with the highest fluid intelligence loadings and performing a median split based on the Anderson–Rubin factor scores. We then correlated the Wonderlic to the Anderson–Rubin factor scores generated for the Complex Span and Visual Arrays factors for each group. Our results for the "high gF" group showed a Pearson correlation of .27 and .26 for the Complex Span and Visual Arrays respectively (n = 67). For the "low gF" group, we observed a Pearson correlation of .44 and .33 for the Complex Span and Visual Arrays (n = 67).

The results showed that the Wonderlic predicted Complex Span and Visual Arrays performance in both the high and low fluid intelligence groups. However, correlations using the Anderson–Rubin method were lower for subjects with high ability compared to subjects with low fluid intelligence. A more direct test of this relationship would be to fit the same structural equation model to both high and low ability groups. Unfortunately in the current study, the sample size is too small to fit two separate models. However, we encourage future researchers to replicate and extend our findings with larger sample sizes.

8. Discussion

The present study sought to better understand the Wonderlic by investigating its relationship to multiple measures of working memory capacity and fluid intelligence. Our results show that the Wonderlic was not a sufficient mediator of the relationship between WMC and gF. Further, the Wonderlic failed to predict gF after accounting for the relationship between WMC and gF. Further, we found that the Wonderlic was a significant predictor of working memory capacity for subjects with low fluid intelligence, but failed to discriminate as well among subjects with high fluid intelligence. These results suggest that the predictive power of the Wonderlic is dependent on the characteristics of the sample it is administered to, a finding in line with previous claims that measures of specific abilities such as working memory capacity are more robust predictors of cognitive performance. For instance, a recent study found that the relationship between the Wonderlic and performance on a CAPTCHA job simulation task was mediated by working memory capacity (Stermac-Stein, 2014). This is in line with previous research showing that working memory capacity accounts for job performance above and beyond the Wonderlic (Perlow, Jattuso, & De Wayne Moore, 1997).

While the Wonderlic is a widely used personnel selection tool, its relationship to intelligence has remained unclear. Results

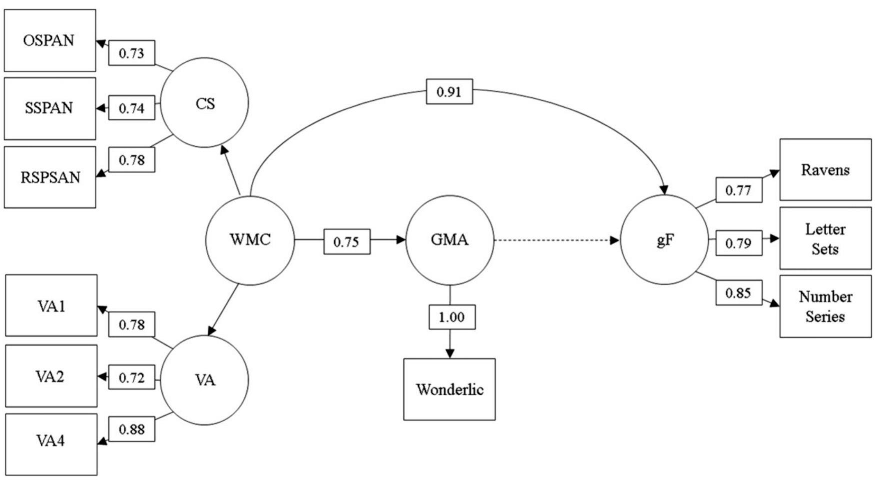

Fig. 6. Structural equation model testing whether WMC has a direct relationship to fluid intelligence above and beyond the Wonderlic (GMA). Complex Span = a factor comprised of the Operation, Reading, and Symmetry span. Visual Arrays = a factor comprised of all three Visual Arrays tasks. WMC = composed of both latent factors of the Visual Arrays and Complex Span. Note: gF = fluid intelligence, CS = Complex Span, VA = Visual Arrays.

in the present study suggest that the predictive validity of the Wonderlic is dependent on the characteristics of the sample it is administered to, whereas the relationship between fluid intelligence and working memory capacity is robust and invariant to the samples' cognitive capabilities.

Using cognitive measures that are grounded in a solid theoretical framework allowed us to draw more concrete inferences about mechanisms driving performance on the Wonderlic. An additional strength of the current study was our inclusion of multiple indicators of each construct. A key limitation worth noting is that our statistical analyses and theoretical interpretations are essentially limited because the Wonderlic provides only a single score, derived from multiple subtests. Due to the proprietary nature of the test, it was not possible to assess the unique contribution of any of these subtests to other dependent measures of interest since we were only provided with a single aggregate score. Tasks on working memory capacity, however, are freely available and open source.

Instead of focusing on tests of "general mental ability," such as the Wonderlic, organizations should consider using other constructs such as working memory capacity. Future work in this area should further validate its relationship to job performance along with other measures that have less emphasis on crystallized intelligence (Matthews & Lassiter, 2007). Researchers have found that working memory capacity has less potential for adverse impact than traditional measures of intelligence or personnel selection such as the Wonderlic (Bosco & Allen, 2011; Hough, Oswald, & Ployhart, 2001; Verive & McDaniel, 1996). Working memory capacity is grounded in a solid theoretical perspective and will lend a greater understanding of individual differences, particularly in the study of individuals with higher intelligence.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working Memory and Intelligence: The Same or Different Constructs? *Psychological Bulletin*, *131*(1), 30–60. http://dx.doi.org/10.1037/0033-2909.131.1.30.

Alderton, D. L., Wolfe, J. H., & Larson, G. E. (1997). The ECAT battery. *Military Psychology*, *9*(1), 5–37. http://dx.doi.org/10.1207/s15327876mp0901_1.

Anastasi, A. (1967). Psychology, psychologists, and psychological testing. *American Psychologist*, *22*(4), 297–306. http://dx.doi.org/10.1037/h0024521.

Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediating effects of goal setting. *Journal of Applied Psychology*, *78*(5), 715–722. http://dx.doi.org/10.1037/0021-9010.78.5.715.

Bell, N. L., Matthews, T. D., Lassiter, K. S., & Leverett, J. P. (2002). Validity of the wonderlic personnel test as a measure of fluid or crystallized intelligence: Implications for career assessment. *North American Journal of Psychology*, *4*(1), 113–120.

Bentler, P. M., & Wu, E. J. C. (2005). *EQS 6.1 for Windows.* Encino, CA: Multivariate Software.

Bosco, F. A., & Allen, D. G. (2011). *Executive attention as a predictor of employee performance* (Academy of Management Proceedings No. 1). *Academy of Management Journal*. http://dx.doi.org/10.5465/AMBPP.2011.65869703.

Buckley, W. D. (1957). *Designing A validity study of the Wonderlic Personnel Test (Unpublished doctoral dissertation).* Richmond, VA: University of Richmond.

Carroll, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29–120). Cambridge, MA: Cambridge University Press.

Cattell, P. (1931). Why Otis' 'IQ' cannot be equivalent to the Stanford-Binet IQ. *Journal of Educational Psychology*, *22*(8), 599–603. http://dx.doi.org/10.1037/h0074143.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1–22. http://dx.doi.org/10.1037/h0046743.

Chamorro-Premuzic, T., & Furnham, A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences*, *44*(7), 1596–1603. http://dx.doi.org/10.1016/j.paid.2008.01.003.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185. http://dx.doi.org/10.1017/S0140525X01003922.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, *51*(1), 42–100. http://dx.doi.org/10.1016/j.cogpsych.2004.12.001.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281–302. http://dx.doi.org/10.1037/h0040957.

Culbertson, S. S., Huffcutt, A. I., & Goebl, A. P. (2013). Introduction and empirical assessment of executive functioning as a predictor of job performance. *Psych Journal*, *2*(2), 75–85. http://dx.doi.org/10.1002/pchj.20.

Damos, D. L. (1993). Using meta-analysis to compare the predictive validity of single- and multiple-task measures to flight performance. *Human Factors*, *35*(4), 615–628.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research & Evaluation*, *14*(20), 1–11.

Dodrill, C. B. (1981). An economical method for the evaluation of general intelligence in adults. *Journal of Consulting and Clinical Psychology*, *49*(5), 668–673. http://dx.doi.org/10.1037/0022-006X.49.5.668.

Edinger, J. D., Shipley, R. H., Watkins, C. E., & Hammett, E. B. (1985). Validity of the Wonderlic Personnel Test as a brief IQ measure in psychiatric patients. *Journal of Consulting and Clinical Psychology*, *53*(6), 937–939. http://dx.doi.org/10.1037/0022-006X.53.6.937.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factorreferenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, *11*(1), 19–23. http://dx.doi.org/10.1111/1467-8721.00160.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. http://dx.doi.org/10.1037/0096-3445.128.3.309.

Frei, R.L., & McDaniel, M.A. (1998). Validity of customer service measures in personnel selection: A review of criterion and construct evidence. *Human Performance*, *11*(1), 1–27.

Furley, P. A., & Memmert, D. (2012). Working memory capacity a controlled attention in tactical decision making. *Journal of Sport & Exercise Psychology*, *34*(3), 322–344.

Furley, P., & Memmert, D. (2013). "Whom should I pass to?" the more options the more attentional guidance from working. *PLoS ONE*, *8*(5), e62278. http://dx.doi.org/10.1371/journal.pone.0062278.

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2002). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences*, *14*(1), 47–64. http://dx.doi.org/10.1016/j.lindif.2003.08.002.

Gottfredson, L. S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, *24*(1), 79–132. http://dx.doi.org/10.1016/S0160-2896(97)90014-3.

Hambrick, D. Z., & Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science*, *20*(5), 275–279. http://dx.doi.org/10.1177/0963721411422061.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*(8), 1149–1167. http://dx.doi.org/10.1002/acp.1624.

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychological Science*, *24*, 2409–2419. http://dx.doi.org/10.1177/0956797613492984.

Hogan, R., & Hogan, J. (1995). *Hogan Personality Inventory manual* (2nd ed.). Tulsa, OK: Hogan Assessment Systems.

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*(1-2), 152–194. http://dx.doi.org/10.1111/1468-2389.00171.

Hovland, C. I., & Wonderlic, E. F. (1939). A critical analysis of the Otis self-administering test of mental ability—higher form. *Journal of Applied Psychology*, *23*(3), 367.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. http://dx.doi.org/10.1080/10705519909540118.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, *96*(1), 72.

Jensen, A. R. (1980). *Bias in mental testing.* New York, NY: Free Press.

Jensen, A. R. (1984). Test bias: Concepts and criticisms. In C. R. Reynolds, & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 507–586). New York, NY: Plenum.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT, US: Praeger Publishers/Greenwood Publishing Group.

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, CA: Sage.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The Generality of Working Memory Capacity: A Latent-Variable Approach to Verbal and Visuospatial Memory Span and Reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189–217. http://dx.doi.org/10.1037/0096-3445.133.2.189.

Kline, R. B. (2011). *Principles and practice of structural equation modeling.* New York: Guilford.

König, C. J., Bühner, M., & Mürling, G. (2005). Working Memory, Fluid Intelligence, and Attention Are Predictors of Multitasking Performance, but Polychronicity and Extraversion Are not. *Human Performance*, *18*(3), 243–266. http://dx.doi.org/10.1207/s15327043hup1803_3.

Lyons, B. D., Hoffman, B. J., & Michel, J. W. (2009). Not much more than g? An examination of the impact of intelligence on NFL performance. *Human Performance*, *22*(3), 225–245. http://dx.doi.org/10.1080/08959280902970401.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. http://dx.doi.org/10.1037/1082-989X.1.2.130.

Matthews, T. D., & Lassiter, K. S. (2007). What does the Wonderlic Personnel Test measure? *Psychological Reports*, *100*(3, Pt 1), 707–712. http://dx.doi.org/10.2466/PR0.100.3.707-712.

McKelvie, S. J. (1994). Validity and reliability findings for an experimental short form of the Wonderlic Personnel Test in an academic setting. *Psychological Reports*, *75*(2), 907–910. http://dx.doi.org/10.2466/pr0.1994.75.2.907.

Molenaar, D., Dolan, C. V., Wicherts, J. M., & van der Maas, H. J. (2010). Modeling differentiation of cognitive abilities within the higher-order factor model using moderated factor analysis. *Intelligence*, *38*(6), 611–624. http://dx.doi.org/10.1016/j.intell.2010.09.002.

Oberauer, K., Süß, H. -M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York: Oxford University.

Perlow, R., Jattuso, M., & Moore, D. W. (1997). Role of verbal working memory in complex skill acquisition. *Human Performance*, *10*(3), 283–302. http://dx.doi.org/10.1207/s15327043hup1003_4.

Raven, J. C. (1990). *Advanced progressive matrices.* Oxford, England: Oxford Psychological Press.

Rode, J. C., Arthaud-Day, M. L., Mooney, C. H., Near, J. P., & Baldwin, T. T. (2008). Ability and personality predictors of salary, perceived job success, and perceived career success in the initial career stage. *International Journal of Selection and Assessment*, *16*(3), 292–299. http://dx.doi.org/10.1111/j.1468-2389.2008.00435.x.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide.* Pittsburgh, PA: Psychology Software Tools.

Shipstead, Z., Lindsey, D. B., Marshall, R. L., & Engle, R. W. (2014). The mechanisms of working memory capacity: Primary memory, secondary memory, and attention control. *Journal of Memory and Language*, *72*, 116–141. http://dx.doi.org/10.1016/j.jml.2014.01.004.

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory*, *20*(6), 608–628. http://dx.doi.org/10.1080/09658211.2012.691519.

Stankov, L., Fogarty, G., & Watt, C. (1989). Competing tasks: Predictors of managerial potential. *Personality and Individual Differences*, *10*(3), 295–302. http://dx.doi.org/10.1016/0191-8869(89)90102-5.

Stermac-Stein J. (2014). *Exploring the relationship between general intelligence, executive function, and performance* (Unpublished doctoral dissertation). The University of Guelph, Ontario, Canada.

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

Stevens, S. N., & Wonderlic, E. F. (1934). The relationship of the number of questions missed on the Otis mental tests and ability to handle office detail. *Journal of Applied Psychology*, *18*(3), 364–368. http://dx.doi.org/10.1037/h0072351.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, *1*(9), 1–121.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. http://dx.doi.org/10.1016/0749-596X(89)90040-5.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132. http://dx.doi.org/10.1037/0033-295X.114.1.104.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. http://dx.doi.org/10.3758/BF03192720.

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*(6), 635–654. http://dx.doi.org/10.1080/09658210902998047.

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. http://dx.doi.org/10.1016/S0160-2896(96)80003-1.

Wonderlic, E. F. (2007). *Wonderlic Personnel Test-Revised: Manual.* Los Angeles, CA: Western Psychological Services.

Wonderlic, E. F., & Hovland, C. I. (1939). The Personnel Test: a restandardized abridgment of the Otis S-A test for business and industrial use. *Journal of Applied Psychology*, *23*(6), 685–702. http://dx.doi.org/10.1037/h0056432.