Journal of Applied Research in Memory and Cognition xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Applied Research in Memory and Cognition

journal homepage: www.elsevier.com/locate/jarmac

Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL)☆

Kenny L. Hicks, Jeffrey L. Foster, and Randall W. Engle*

Georgia Institute of Technology, United States

The Complex Span paradigm is one of the most influential and widely used instruments for measuring working memory capacity (WMC). We report the results of four experiments designed to explore the feasibility of obtaining valid estimates of WMC online. We explored the relationships between the Complex Span tasks and fluid intelligence (gF) in the lab and on the web using a new platform called the Online Working Memory Lab (the OWL). The OWL is universally accessible across all computer operating systems and functions in both local and remote contexts, allowing researchers to sample more diverse subjects from practically anywhere. Experiments 1 and 2 showed that the Complex Span failed to predict gF when the to-be-remembered stimuli were letters and the tests were taken online. We increased the predictive validity of the test battery in Experiments 3 and 4 by replacing the letters with memory stimuli that were more difficult to write down in an unproctored setting. This work describes our most recent attempts to measure working memory capacity in the wild.

Keywords: Online testing, Working memory capacity, Fluid intelligence, Individual differences

Psychometric and individual differences research on the construct of working memory capacity (WMC) has exploded since the late 90s when Engle and colleagues found that it predicted fluid intelligence above and beyond short-term memory (Engle, Tuholski, Laughlin, & Conway, 1999). Since then, large-scale studies examining the latent correlation between these two constructs have repeatedly demonstrated that WMC is the single best predictor of fluid intelligence (Oberauer et al., 2007).

Different research teams have various theoretical accounts of what cognitive mechanisms are responsible for the domain-general nature of WMC in higher-order cognition. Some researchers have emphasized the importance of short-term memory limitations (Cowan, 2010), while others have concentrated their research toward understanding the relationship between inhibition and WMC (Hasher et al., 2007). The theory put forward by Engle and colleagues has a primary focus on the importance of attention control and the ability to avoid interference, while remaining vigilant and proactive when engaging in goal-relevant behavior.

In line with the attention theory put forward by Engle and colleagues, researchers have also found that the Complex Span tasks are powerful predictors of controlled attention. For instance, measures of the Complex Span predict the subjects' ability to look away from an attention orienting stimulus in the anti-saccade task (Kane, Bleckley, Conway, & Engle, 2001; Unsworth, Schrock, & Engle, 2004), the ability to inhibit automatic processing in the Stroop task (Kane & Engle, 2003), and the ability to focus their attention like a spotlight when taking the flanker (Heitz & Engle, 2007). Measures of Complex Span predict individual differences in reading comprehension, scholastic performance, note-taking, and the ability to follow instructions. These measures also predict the ability to focus attention and prevent mind-wandering when it becomes disruptive to performance (Kane et al., 2007; Kane & Engle, 2003; Kane & McVay, 2012).

In short, researchers have found that individual differences in WMC are reliable across the lifespan, and in many ways mimic the same pattern of results found for fluid intelligence, but not

Authors Note

This research was supported by the Office of Naval Research Grant N00014121040.

☆ Please note that this paper was handled by the former editorial team of JARMAC.

* Correspondence concerning this article should be addressed to Randall W. Engle, School of Psychology, Georgia Institute of Technology, 654 Cherry Street, Atlanta, GA 30332, United States. Contact: randall.engle@gatech.edu

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

always. For instance, Hambrick et al. (2010) found that the Complex Span predicted individual differences in multi-tasking even after controlling for fluid intelligence. Whereas fluid intelligence requires each test item to be novel to the subject, individual differences on measures of the Complex Span persist, even after multiple exposures. For most tests, a situation in which an answer key or some other test material is stolen or leaked, the test is likely to be completely compromised and no longer valid. This same situation would have very little impact on subjects' performance on the Complex Span as long as the stimuli are randomized. The inherent nature of the WMC tests make them resilient instruments of individual differences and may turn out to be more culture fair than measures of fluid intelligence altogether. A recent investigation by Bosco et al. (2015) revealed a substantial difference in how well tests of intelligence properly classify white subjects versus black subjects. Although this finding was not new in relation to tests of intelligence, the finding that measures of attention and WMC demonstrated nearly three times less adverse impact than traditional tests of fluid intelligence highlight the social and political advantages of using the most culture-fair measure of the subject's cognitive ability.

Researchers have also found individual differences in WMC among experts. For instance, Lopez et al. (2012) conducted a study in which highly experienced pilots were subjected to 35 h of sleep deprivation. The results showed that the Operation Span accounted for 42% of the errors made by expert pilots on an advanced flight simulator. Meinz and Hambrick (2010) found a similar pattern of data for expert pianists. When subjects were asked to sight-read a new piece of music, individual differences in WMC were predictive of their performance, above and beyond the number of hours practiced.

Traditional Experimentation With the Complex Span Paradigm

Our lab has spent a considerable amount of time developing measures of the Complex Span and making them freely available to other researchers. Over the years, these tasks have been downloaded by over 1000 research teams and translated into more than 10 languages. The majority of the Complex Span tasks we have written are programmed in the E-Prime framework, although some have been written with other experiment-development packages such as Inquisit and MEL.

General Method

Overview

Recently, our lab developed an online battery of Complex Span tasks called the Online Working Memory Lab (the OWL), which is capable of both laboratory and network deployment. The OWL is accessible on Mac/Linux/and the Windows operating system. In fact, tasks included in the OWL can be accessed by any device with a web browser (such as Android, iPhone, and the iPad). The following experiments lay out our most recent attempts to measure working memory capacity in the wild.

Materials and Procedures

The first iteration of the OWL consisted of four Complex Span tasks: Operation Span, Reading Span, Symmetry Span, and the Running Span. We assessed the construct validity of the OWL by examining the correlations among the WMC and fluid intelligence (gF) measures performed in the laboratory, under the supervision of a proctor, compared to subjects' performance on the web. Descriptions of each task can be found below.

The Automated Operation Span (Unsworth, Heitz, Schrock, & Engle, 2005) requires the subject to first complete 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 with memory required is calculated by their mean time to answer the math question plus 2.5 standard deviations. Subjects also perform a practice procedure where they are presented with two letters and asked to recall them in the order they were presented. After the practice phase, subjects are presented with the real trials that combine the math and letter procedures of the experiment. Subjects are presented with a list of 15 trials of 3–7 randomized letters interleaved with simple math operations. After each list is complete, subjects are required to recall the letters in the order presented. An example is shown in Figure 1.

The Symmetry Span (Unsworth, Redick, Heitz, Broadway, & Engle, 2009) task (shown in Figure 1) is a spatial version of the complex span which requires the subject to judge whether a picture is symmetrical around a vertical axis while remembering 2–5 specific locations highlighted on a 4×4 grid.

Reading Span (Unsworth et al., 2009) procedure is identical to the Operation Span except that in the place of math operations, subjects must judge if a sentence makes sense such as, "Tom ran to the store with rainbow shoes." The to-be-remembered items in the reading span are letters.

Running Span (Broadway & Engle, 2010) does not require the subject to make any judgments between presentations of the to-be-remembered stimulus, instead, subjects are shown varying set sizes of letters and are required to only recall a subset of those letters. For instance, the subject may see the letters J, L, K, P, T and only be required to recall the last 3 letters in the order they were presented (K, P, T).

Mental Counters (Alderton, Wolfe, & Larson, 1997) presents three lines to the subject. At the beginning of the experiment each line is given a value of five. Squares appear rapidly and in random order above and below the line, changing its value. When a square appears above the line, the value of that line *increases* by one; when it appears below the line the value *decreases* by one. The subject must keep track of the value of all three lines. In Figure 2, a square appears above the third line. In this example, the subject should add 1 to the current value of the third line ($5 + 1 = 6$). At the end of each trial the subject must indicate the final number value of each line.

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

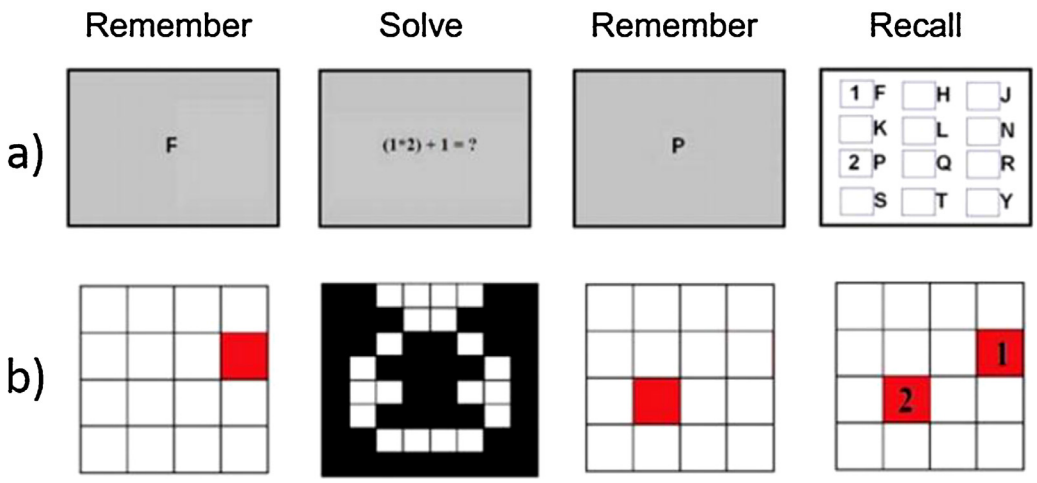

Figure 1. Examples of the Operation Span (a) and Symmetry Span (b).

Figure 2. The Mental Counters task. Subjects have to keep track of the three values or "counters" that change rapidly and in random order.

Fluid Intelligence[1]**. Letter Sets**. In this task, subjects were presented with four groups of letters, three of which were constructed using the same rule while the fourth was composed using a different rule. Subjects were asked to identify the grouping that was created with a different rule than the other three. (e.g., AAAA BBBB CCCC ***DDED***). Subjects were given 5 min to complete 15 questions.

Number Series. This task showed subjects a string of numbers that followed a particular rule. Subjects were asked to determine the next number in the sequence. For instance, an example of a practice item may be: 1 2 3 4 5?. In this example the subject is expected to determine the rule, which is to that each new number in the sequence is simply one more than the previous number. The rules became increasingly complex throughout the task. Subjects were given 4 min and 30 s to complete 15 questions.

Paper-Folding required each subject to imagine mentally folding a piece of paper and putting a hole through it. They were then asked to imagine where each hole on the piece of paper would be if they then unfolded the piece of paper. Paper-Folding contained a total of 20 questions. Subjects were given a 6 min time limit to complete the task.

Matrix Reasoning showed subjects a 3×3 matrix of patterns that were constructed using a common rule. On each trial the bottom right picture of the matrix was missing and each subject was asked to select the next picture in the pattern by selecting the correct answer from several alternatives. The Matrix Reasoning test contained a total of 18 questions, with a 10 min time limit.

Online Performance on the OWL Battery

Experiment 1 (E1). The purpose of our first experiment was to compare performance in the laboratory to online performance on the OWL. We asked 58 subjects who had previously completed E-Prime versions of the Complex Span and fluid intelligence measures in our lab to take them a second time online using the OWL. Last, all subjects returned to our lab to complete two additional tests in the laboratory that were not administered in the original laboratory test battery (Reading Span and Mental Counters). After subjects agreed to participate, we sent them an email link to the task manager which tracked their subject ID and directed them through the battery automatically. Subjects were compensated $30.00 for their participation. A screenshot of the task manager is located in Appendix A. It contains a queue that displays the tests the subject still needs to complete, an approximate administration time for those tasks, and a deadline for completion. This allowed us to conduct the entire cognitive battery without a proctor as the task manager guided the subject through each task with a simple mouse click.

[1] The measures of gF conducted on the web were new items created by our lab for this project. We accomplished this by solving the rule of the original problem and using that same rule to build a new item.

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

Table 1
Correlations Among Composite Variables

	WMCz_Online	WMCz_Inlab	gFz_Online	gFz_Inlab
WMCz_Online	–			
WMCz_Inlab	0.57	–		
gFz_Online	0.45	0.67	–	
gFz_Inlab	0.31	0.66	0.74	–

Note: WMCz_Online and WMCz_Inlab are composite variables of WMC tasks including Operation Span, Reading Span, Running Span, Symmetry Span, and Mental Counters. gFOnline and gFInlab are composites of fluid intelligence tasks performed online and in lab respectively.

Method (E1). The purpose of the first experiment was to compare subjects' performance in the laboratory to their behavior on the internet. The descriptive statistics for the measures of WMC are presented in experiment 1. Our primary interest in E1 was to compare and contrast subjects' performance in the lab to their behavior online. In order to test this question, we created two composite scores for the working memory capacity tasks (separate scores for "in lab" and "online" performance), as well as two composite scores for the measures of gF ("in lab" and "online").

Our main interest was to determine whether the WMC measures conducted on the web were predictive of fluid intelligence performance both in-lab and online. The results of our first analysis can be found in Table 1. First, the fluid intelligence tests conducted online correlated very highly with the same measures taken in the lab ($r = .74$).[2] Additionally, the WMC measures taken in the lab predicted fluid intelligence online and in the lab. However, our newly developed online measures of WMC were less predictive of fluid intelligence on the web ($r = 45$), and in the lab ($r = .31$) (see in Appendix A). We expected the relationship between our online tests of WMC and fluid intelligence to be stronger and had no clear explanation for this pattern of results. That is, until we conducted the second experiment, which provided more insight into these results.

Experiment 2. The primary of our second experiment was to further explore the validity of our WMC measures when they are performed online. We recruited a sample of young adults from the United States aged 18–35. All that the OWL required was for each subject to have access to a web browser and a United States IP address.

Method (E2). We recruited 100 online subjects using Amazon's Mechanical Turk. Each subject used the OWL to complete the same eight measures presented in E1. All subjects received $10.00 for their participation. Twelve subjects failed to complete the entire battery of tests within a 24-h period. These subjects were dropped from the analysis leaving an $n = 88$. Descriptive statistics for each test can be found in Table 3 in Appendix A.

Results (E2). Our first Confirmatory Factor Analysis (CFA) loaded all five memory tasks on to a single factor labeled WMC which we correlated with a factor of fluid intelligence (see Figure 3). The WMC factor only predicted roughly 18% of the variance in fluid intelligence ($r = .43$), which is much lower than what we expected to find.

The fit indices of the first model indicated that the fit was poor. In order to determine which parameters of the model were not specified correctly we analyzed the standardized residual correlation matrix, which includes any additional correlations between variables that were not captured by the model. We observed a clear, but unexpected pattern of data. All three of the verbal measures of WMC had negative standardized residual correlations to gF that were not shared by the spatial measures of WMC. This pattern of data suggested that subjects with low gF had high verbal WMC scores, but low spatial WMC scores. Overall, our analysis showed that only the spatial measures of WMC were significant predictors of gF.

Last, we attempted to extract a common WMC factor and a verbal specific factor in order to better understand subjects' performance on the verbal tests of WMC. This analysis is presented

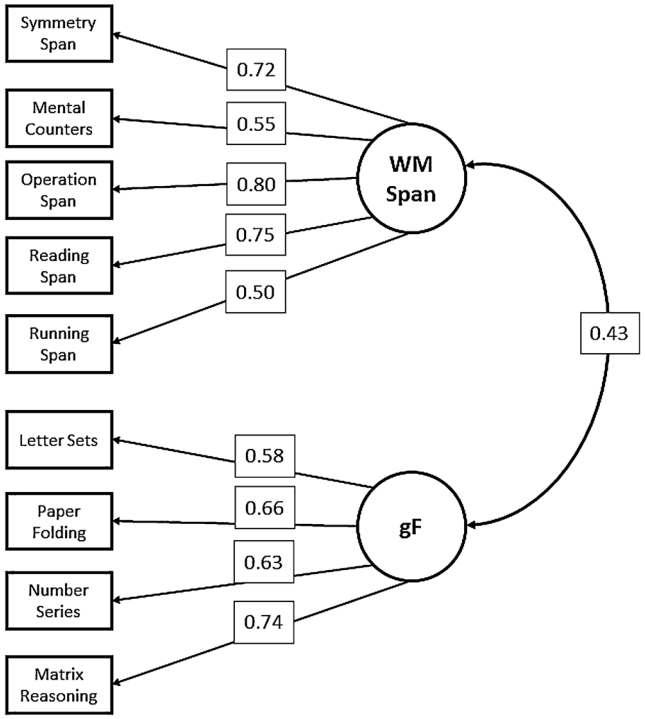

Figure 3. Confirmatory Factor Analysis for all measures of working memory capacity (WMC) predicting fluid intelligence (gF). WMC = working memory capacity; gF = fluid intelligence.

[2] The gF measures conducted in-lab were original items, whereas the gF measures online were newly created. Thus, this correlation can be viewed as cross-validation evidence for our newly created items. The original and new items correlated as high as the expected reliability.

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

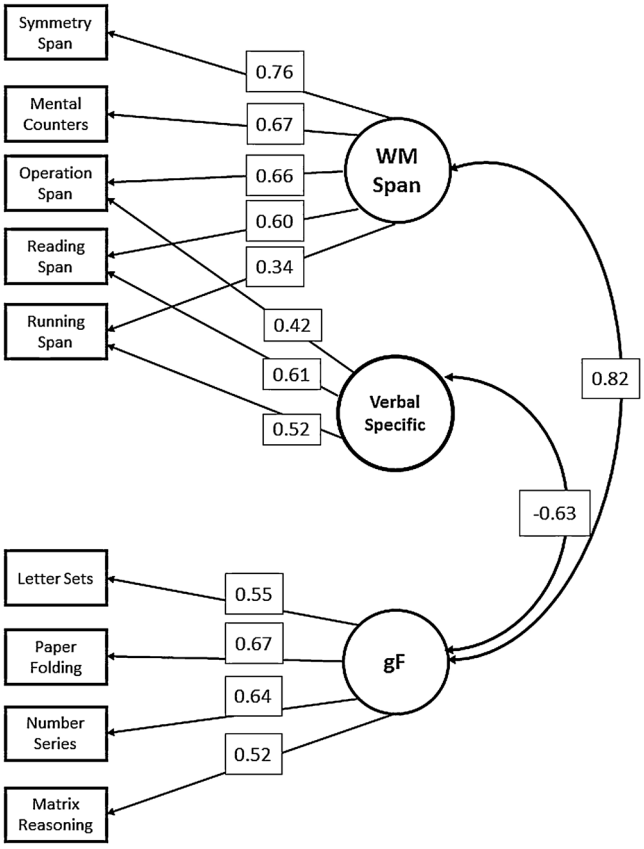

Figure 4. Confirmatory Factor Analysis of a common working memory capacity factor and a specific verbal factor predicting fluid intelligence (gF).

in Figure 4. We created a domain general factor of WMC by loading all of the tasks onto a single factor and a specific verbal WMC factor by loading all of the verbal WMC tests onto a second factor. The correlation between the domain-general WMC construct and gF was substantial. The estimate was very similar to previous research findings that have assessed their relationship in the lab. However, the relationship between the verbal specific factor and fluid intelligence was negative, which surprised us at first. This pattern of evidence suggests that subjects with lower cognitive ability had higher verbal WMC performance on the web. What could explain this result? Our hypothesis is that lower ability subjects were more likely to cheat by writing down the to-be-remembered stimuli instead of recalling them as instructed (Table 2).

Table 2
Fit Statistics for Working Memory Capacity CFA Models

Model	χ^2	df	χ^2/df	RMSEA	SRMR	NNFI	CFI
Single Factor	81	26	3.1	0.16	0.13	0.70	0.79
Verbal_Specific	33	22	1.5	0.07	0.05	0.93	0.96

Note: The only model approaching acceptable fit is the last model, specifying a unique verbal factor (see Figure 4).

Re-analysis of Experiment 1

As Experiment 2 was in progress, we realized that subjects could easily cheat our verbal tests of WMC by simply writing down the to-be-remembered stimuli, which were letters. Therefore, we asked approximately 50 subjects in E2 if they had written down any of the to-be- remembered letters during the online experiment, and 10% (5–50) reported that they had indeed written down letters at some point in the assessment. Based on this information, we thought that subjects might have cheated in a similar manner during the online battery of E1. We conducted a re-analysis of E1 to test this hypothesis.

After separating the measures of working memory capacity into verbal and spatial composites we assessed their relationship to fluid intelligence in the lab and online. The results of Table 3 show the same pattern of data found in E2. Although the spatial tasks were highly predictive of fluid intelligence across both in-lab and online environments, the verbal measures of working memory capacity taken online failed to predict fluid intelligence in the lab and were only weakly predictive of fluid intelligence online. Therefore, the evidence suggests that subjects in our first experiment had likely written down the to-be-remembered stimuli in the same way that we observed in the second experiment.

Despite strong correlations between spatial measures of WMC and gF in the lab and online, we found that the verbal WMC tests had little to no correlation to fluid intelligence when conducted on the internet. This finding was in complete contrast to what our lab and many others have found when assessing the relationship between verbal tests of WMC and novel reasoning in the lab (Chuderski, 2014; Engle et al., 1999; Kane et al., 2004; Kyllonen & Christal, 1990; Oberauer et al., 2003).

Improving Validity on the OWL Battery

Experiment 3 (E3). The goal of our third experiment was to decrease the incidence of cheating on the web-based versions of

Table 3
Correlations Among Variables

	WMC_VerbalOnline	WMC_SpatialOnline	gF_Online	gF_Inlab
WMC_VerbalOnline	–			
WMC_SpatialOnline	0.66	–		
gF_Online	0.35	0.67	–	
gF_Inlab	0.21	0.56	0.74	–

Note: WMC_VerbalOnline is a composite score of verbal WMC tasks including Operation Span, Reading Span, Running Span. WMC_SpatialOnline is a composite of Symmetry Span and Mental Counters. gF_Online and gF_Inlab are composites of fluid intelligence tasks performed online and in lab respectively. The only correlation that was not significant was the correlation between WMC_VerbalOnline and gF_InLab (n.s.).

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

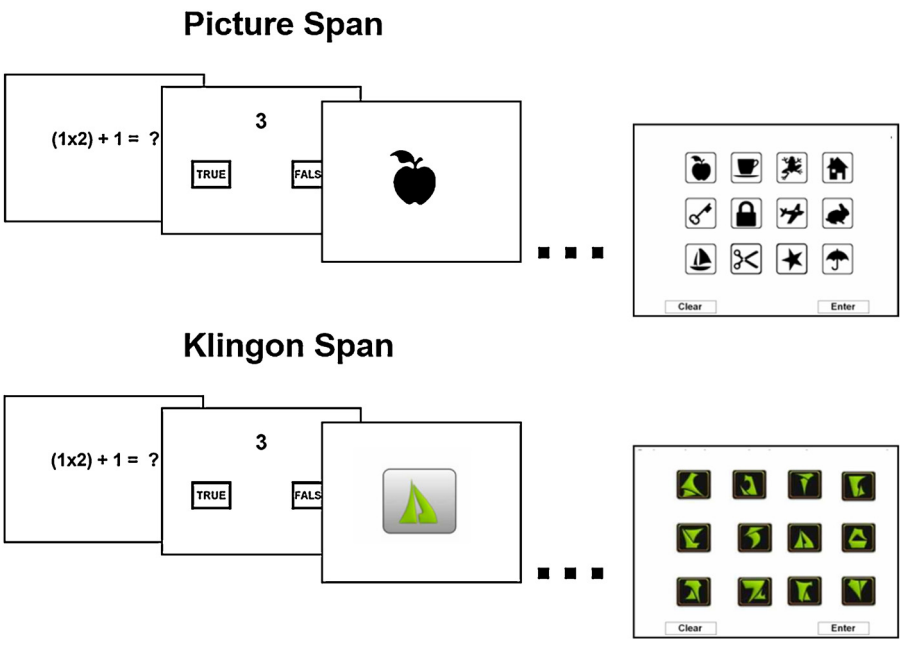

Figure 5. Images of the newly developed Picture and Klingon Span tasks.

the WMC tests by replacing the to-be-remembered stimuli from letters to pictures of objects that could be easily verbalized in one task, and Klingon characters that were much more difficult to rehearse in another task. To preview our results, we found that replacing the letter stimuli in the Operation Span with pictures of common objects (e.g., an umbrella, or frog), or Klingon characters, lead to a substantial increase in the correlation between the online measures of WMC and gF.

Method for E3. As in the first experiment, experiment three was conducted in order to compare subjects' online performance on the two newly developed tests to their behavior in the lab. Therefore, our first study included 112 subjects aged 18–35 who we selected from our lab's database to complete the OWL online. All subjects had previously participated in research at Georgia Tech's Attention and WMC lab. Out of the 112 subjects, 10 subjects were not usable due to an excessive amount of missing data, leaving us with a final sample of 102 subjects. A small number of these subjects also contained missing data. We imputed the missing data for these subjects using the EM (Estimated Maximization) procedure in EQS.

Tasks. *Measures of WMC*

Operation Span (Unsworth et al., 2005) – See E1 for description.

Picture Span[3] – this task was modeled after the Operation Span, the only difference was that the to-be-remembered stimuli consisted of pictures of ordinary objects such as a frog, an umbrella, or plane. Each object was interleaved with simple arithmetic problems. Each subject received 15 trials of the Picture Span, each size set (2, 3, 4, 5, 7) a total of 3 times each. The task was randomized for each subject.

Klingon Span – the Klingon Span was also based on an alteration to the traditional Operation Span (which presented words or letters). The to-be-remembered stimuli were replaced from letters to Klingon symbols. We chose items that to be as distinct from one another as possible. Figure 1 shows the stimuli chosen. Subjects received three trials of each of the four set-sizes (2, 3, 4, 5). This resulted in a total of 12 trials. Although this altered version contained spatial stimuli as the to-be-remembered stimuli, it also required subjects to solve the same arithmetic between each item presentation. Therefore, the interleaving task was verbal in nature and required the subject to solve math problems (Figures 5–7).

Measures of Fluid Intelligence

Number Series – See E1 for description.

Paper Folding – See E1 for description.

Form Boards – in the Form Boards task the subject is asked to create an object, such as a square, from 4 out of 5 stimuli presented beneath the final shape. The correct four stimuli will create the object perfectly, but if a wrong stimulus is chosen the final shape would not be created correctly. Each subject was administered 10 test items.

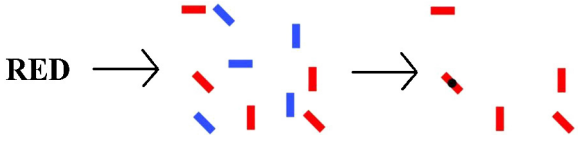

Figure 6. Picture of the Visual Arrays task.

[3] Stimuli chosen for the Picture Span task were simple, single word objects that could be displayed in black and white.

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

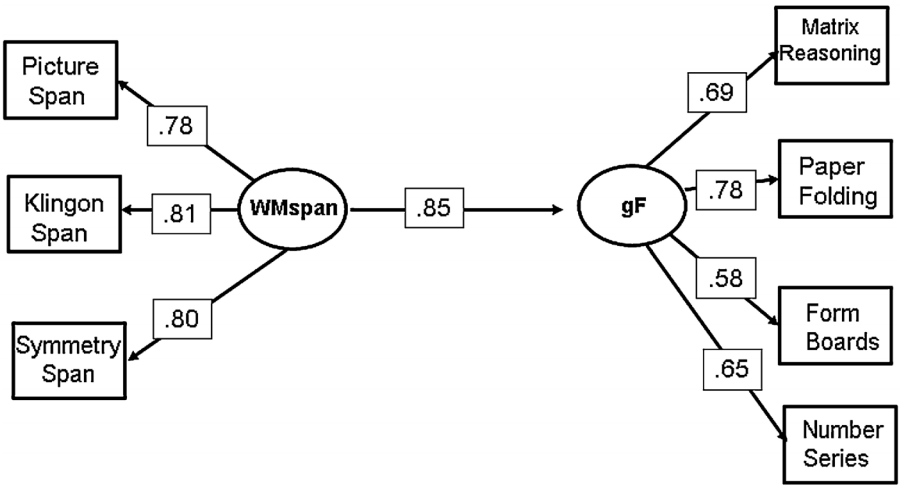

Figure 7. Confirmatory Factor Analysis.

Matrix Reasoning – See E1 for description.

Results (E3). Our initial analyses were geared toward understanding the relationship between subjects' performance on WMC and gF measures in an online environment compared to their performance in the laboratory.

In order to investigate the relationship between WMC and gF across both the online and laboratory contexts we created two *z*-score composites of each construct, one for performance in the lab and one for performance on the web. As we predicted, performance on both the online and in lab measures of WMC and gF were highly correlated with one another (see Table 4). Specifically, gF performance in the lab was a strong predictor of WMC performance on the web ($r = .71$). gF performance on the web was also a strong predictor of WMC performance in the lab ($r = .61$). This evidence suggests that we successfully increased the validity of our web-based measures of WMC. Subjects' performance on gF in the lab and on the web was high enough to meet test re-test standards ($r = .81$), as was performance on WMC in the lab and on the web ($r = .75$).

Replication of Results and Introduction of Attention Measures

Experiment 4 (E4). Our fourth and final experiment had two aims. First, we wanted to replicate the results of the previous experiment in a random sample of online subjects from anywhere within the United States. Second, we wanted to investigate the feasibility of administering measures of attention control and the focus of attention in an online environment.

Method (E4). We added two additional measures to the OWL, a measure of attention control (anti-saccade) and a measure of the Visual Arrays. We initially recruited 140 subjects from Amazon's crowd-sourcing service, Mechanical Turk, to take the OWL battery online. However, only 112 subjects completed the entire battery. Therefore, the results reported for Experiment 4 are the data for 112 subjects aged 18–35. The only additional constraint for participating in the experiment was that the subject must have a United States IP address. The tasks administered in Experiment 4 were identical to Experiment 3 with the exception of two additional measures.

Additional Tasks in Experiment 4

Anti-saccade – the subject was asked to fixate their attention on a cross appearing in the middle of the computer screen at the beginning of each trial. After the fixation disappeared an asterisk was display on either the left or right side of the computer screen. After the asterisk disappeared, one of three letters appeared on the opposite side of the screen (B, P, or R) and was quickly masked. The subject's job was to look away from the asterisk, toward the opposite side of the screen in order to identify which of the three letters appeared.

Visual Arrays – the subject was shown a display of 10 or 14 rectangles in various orientations on each trial. The subject shown the word RED or BLUE at the start of each trial in order to cue them to pay attention to either the red or blue stimuli.

Table 4
Correlations Among Composite Variables of WMC and gF

	gFz_InLab	gFz_Online	WMCz_InLab	WMCz_Online
gFz_InLab	1			
gFz_Online	0.81	1		
WMCz_InLab	0.64	0.61	1	
WMCz_Online	0.71	0.73	0.75	1

Note: gF variables: gFz_InLab = composite fluid intelligence variable (in lab administration), gFz_Online (taken online). WM variables: WMCz_InLab = composite working memory capacity variable (in lab administration), WMCz_Online (taken online).

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

Table 5
Correlations Among New OWL Measures and Composites of gF and WMC

	AntiSacc	VisualArrays	gFz	WMCz
AntiSacc	1			
VisualArrays	0.35	1		
gFz	0.39	0.49	1	
WMCz	0.38	0.54	0.61	1

Note: AntiSacc = Antisaccade, Visual Arrays. gFz = gF composite. WMCz = working memory capacity composite.

After the first display of blue and red rectangles disappeared the subject was shown a second display with only blue or red rectangles (depending on the trial). The subject then saw a black circle inside one of the rectangles and were asked to determine whether or not the orientation had changed from the first display. We computed a *k* value for each subject by using the single probe correction equation (Cowan et al., 2005). The calculation is $k = N^*$ (hits + correct rejections − 1). These values were calculated for each set size. Last, we averaged each set size in order to obtain the final *k* value.

Results (E4). The results of Experiment 4 replicated our findings in Experiment 3. Specifically, the modified versions of the Operation Span were significantly correlated with all measures of gF as well as our newly developed tests of attention control (anti-saccade) and our new measure of the Visual Arrays. There was an exception with the Anti-Saccade task, which did not significantly predict Matrix Reasoning at the bivariate level ($p = .066$). However, both tasks significantly predicted composite measures of WMC and gF. These findings provide further support for the validity of the OWL in online contexts and extend the OWL's measurement capability to constructs related to, but separate from WMC (Table 5).

In order to maximize the power and interpretability of our data, we combined the online performance[4] of the experimental groups from experiments 3 and 4. The current model consists of 214 subjects who completed all three of the WMC measures and all four of the gF tasks online.

Our analysis included specifying a Structural Equation Model in which WMC predicted gF. This model was a good fit to our data. WMC predicted 72.3% of the variance in gF (this value is obtained by squaring the loading on the single-headed arrow going from WMC to gF). As we stated previously, numerous studies have shown that WMC and gF share a substantial amount of overlapping variance. After modifying the Operation Span to make the stimuli more difficult to write down we found the same pattern of results that researchers have demonstrated in much more controlled environments.

Despite the results of E2 in which we found poor model fit when we included verbal measures of WMC with easily written down stimuli, both our newly created measures were highly loaded onto the WMC factor even though the Picture Span stimuli could be rehearsed verbally (Table 6).

Table 6
Fit Statistics for Confirmatory Factor Analysis (Figure 4)

χ^2	df	χ^2/df	RMSEA	SRMR	NFFI	CFI
24	13	1.8	0.06	0.03	0.97	0.98

Discussion

Although WMC has proved to be one of the best predictors of intelligence, its measurement has remained restricted to the laboratory. In the current article we report four experiments investigating the predictive validity of working memory capacity with a flexible online battery that can be administered both locally and online. The findings were clear. Our online measures predicted fluid intelligence both in and outside the lab. Correlations for all measures in this battery are consistent with evidence observed in traditional lab settings.

We have also identified a potential threat to the validity of measures of verbal WMC when measured online. Our data revealed that low ability subjects were more likely to attempt to "game" the verbal tasks by writing down the to-be-remembered stimuli when they could be easily copied (e.g., verbal stimuli such as letters) in remote administrations when no proctor was present. We addressed this issue in experiments 3 and 4 by replacing the letter stimuli previously used in the Operation Span task with stimuli that were more difficult to write down (e.g., Pictures and Klingon characters). The current article also investigated the validity of measuring the Anti-saccade and the Visual Arrays on the web. Findings reported here provide further support for the validity of the OWL and extend its measurement capability to other constructs that contribute to our understanding of working memory capacity.

The ability to measure higher-order cognition over the web offers clear advantages to researchers. Researchers have raised serious concerns regarding the representation of minorities in samples recruited for most of the social sciences (Henrich, Heine, & Norenzayan, 2010). This issue is particularly relevant to the study of issues such as adverse impact, in which the goal is to determine how accurately a measure reflects the true ability of subjects across different races. This question cannot be properly addressed if the researcher fails to acknowledge and control for the social differences between whites and minorities, such as the fact that minorities have substantially reduced access to the research lab.

Conclusion

The OWL offers a completely new method of subject recruitment which researchers can use to overcome many of the sampling issues facing researchers today. In addition, researchers can also take advantage of new technological advances in crowd-sourcing recruitment, such as Amazon's Mechanical Turk. These technologies allow researchers to access a wide range of subjects and offer tailored recruitment

[4] We estimated two factor scores for each group separately and tested for homogeneity of variance using Levene's test. The results showed that the groups did not differ on the working memory capacity or fluid intelligence factor.

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

procedures. Paired with the OWL battery, for example, researchers can recruit from a database that more closely resembles the general population, relying less on convenience samples.

Our research on the feasibility of obtaining valid estimates of WMC on the web has revealed several important findings. First, our previous work in this area showed clear evidence that measures of spatial WMC were valid when taken on the web. However, we also found that subjects were likely to cheat on our verbal measures of WMC (Operation Span, Reading Span and Running Span) when they were performed online versus in the lab. In the current article we successfully modified the verbal Complex Span tasks by creating new to-be-remembered stimuli which were more difficult to write down. The results of two experiments using these new measures demonstrated a substantial increase in the correlation between WMC and gF when subjects performed the OWL online. In addition, we extended the measurement capability of the OWL to include a measure of attention control (the anti-saccade) as well as a measure of the Visual Arrays. This work contributes to a growing literature on online recruitment and measurement.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

Appendix A.

Experiment 1

See Table A1.

Table A1
Descriptive Statistics for Online/In Lab Comparison

Variables	Min	Max	Mean	SD	Skew	Kurtosis
OSpan_Online	0	75	50.4	26.1	−0.93	−0.76
OSpan_Inlab	3	75	53.4	16.7	−1.1	0.84
RunSpan_Online	0	75	45.9	21.9	−0.54	−0.79
RunSpan_Inlab	10	81	54.3	18	−0.51	−0.46
ReadSpan_Online	0	75	44.1	26.3	−0.41	−1.36
ReadSpan_Inlab	3	75	44.5	20.2	−0.4	−0.85
SymSpan_Online	0	42	22.8	12.7	−0.18	−1.22
SymSpan_Inlab	4	42	24.9	9.9	−0.34	−0.39
MentalCount	0	32	16.7	10.4	−0.3	−1.32
MentalCount	1	31	17.9	8.2	−0.48	−0.5

Note: WM tasks: OSpan = Operation Span, ReadSpan = Reading Span, RunSpan = Running Span, SymSpan = Symmetry Span, MentalCount = Mental Counters. gF tasks: LS = Letter Sets, PaperFold = Paper Folding, NS = Number Series, MatrixReas = Matrix Reasoning.

Experiment 2

See Tables A2–A4.

Table A2
Descriptive Statistics for Online Tasks

Variable	Min	Max	Mean	SD	Skew	Kurtosis
OSpan	2	75	55.7	18.95	−1.34	−1.34
ReadSpan	5	75	55.93	16.83	−1.06	−1.06
RunSpan	32	100	73.07	15.35	−0.22	−0.22
SymSpan	1	42	26.02	9.38	−0.46	−0.46
MenCount	6	32	24.99	5.02	−1.28	−1.28
LettSets	2	23	12.76	3.87	−0.15	−0.15
PaperFold	2	19	10.19	4.07	0.22	0.22
NumSeries	3	14	8.13	2.54	0.1	0.1
MatrixReas	2	17	10.09	3.39	−0.26	−0.26

Note: Partial scores are reported. WM tasks: OSpan = Operation Span, ReadSpan = Reading Span, RunSpan = Running Span, SymSpan = Symmetry Span. gF tasks: MenCount = Mental Counters, LettSets = Letter Sets, PaperFold = Paper Folding, NumSeries = Number Series, MatrixReas = Matrix Reasoning.

Table A3
Correlations Among Variables

	OSpan	ReadSpan	RunSpan	SymSpan	MentalCount	LS	PaperFold	NS	MatrixReas
OSpan	–								
ReadSpan	0.65	–							
RunSpan	0.41	0.54	–						
SymSpan	0.57	0.48	0.27	–					
MentalCount	0.39	0.35	0.2	0.49	–				
LS	0.1	0.08	0.04	0.32	0.3	–			
PaperFold	0.15	0.15	−0.12	0.4	0.46	0.52	–		
NS	0.17	0.06	0.02	0.4	0.4	0.39	0.28	–	
MatrixReas	0.24	0.05	−0.03	0.41	0.46	0.33	0.5	0.55	–

Note: WM tasks: OSpan = Operation Span, ReadSpan = Reading Span, RunSpan = Running Span, SymSpan = Symmetry Span, MenCount = Mental Counters. gF tasks: LS = Letter Sets, PaperFold = Paper Folding, NS = Number Series, MatrixReas = Matrix Reasoning.

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

Table A4
Correlation Matrix for E4

	Online						In Lab						
	PicSpan	Klingon	SymSpan	MatrixReas	PaperFold	FormBoard	NumSeries	Ospan	SymSpan	RotSpan	Ravens	NumSeries	LettSets
Online													
PicSpan	1												
Klingon	0.63	1											
SymSpan	0.62	0.72	1										
MatrixReas	0.49	0.54	0.56	1									
PaperFold	0.50	0.62	0.57	0.60	1								
FormBoard	0.40	0.49	0.45	0.45	0.58	1							
NumSeries	0.55	0.47	0.54	0.47	0.52	0.46	1						
In Lab													
Ospan	0.71	0.47	0.44	0.40	0.30	0.29	0.41	1					
SymSpan	0.60	0.51	0.58	0.55	0.47	0.30	0.45	0.66	1				
RotSpan	0.61	0.57	0.57	0.57	0.60	0.35	0.54	0.56	0.66	1			
Ravens	0.54	0.59	0.59	0.63	0.70	0.53	0.60	0.41	0.53	0.61	1		
NumSeries	0.57	0.44	0.45	0.44	0.52	0.38	0.68	0.49	0.51	0.57	0.61	1	
LettSets	0.55	0.54	0.48	0.49	0.46	0.46	0.63	0.38	0.39	0.48	0.55	0.60	1

Experiment 3

See Tables A5 and A6.

Table A5
Descriptives for E3 (Online Administration)

	N	Range	Minimum	Max	Mean	SD	Skew	Kurtosis
MatrixReas	102	10	0	10	5.9	2.2	−0.85	0.23
NumSeries	102	10	0	10	6.2	2.6	−0.4	−0.77
PaperFold	102	10	0	10	5.3	2.7	0.06	−1.05
FormBoards	102	10	0	10	5.9	2.5	−0.35	−0.75
OSpan	102	75	0	75	47.4	24.4	−0.81	−0.85
SymSpan	102	39	0	39	22.2	11.4	−0.46	−1.02
PictureSpan	102	71	1	72	44.1	20.5	−0.63	−0.86
Klingon	102	69	0	69	21.3	18.6	0.85	−0.24

Note: Descriptives from online tasks. gF tasks: MatrixReas = Matrix Reasoning, NumSeries = Number Series, PaperFold = Paper Folding, FormBoard = Form Boards. WM tasks: Ospan = Operation Span, SymSpan = Symmetry Span, PictureSpan = Pictures Span, Klingon = Klingon Span.

Table A6
Descriptives for E3 (in Lab Administration)

	N	Range	Minimum	Max	Mean	SD	Skew	Kurtosis
Ravens	102	17	0	17	9.17	4.1	−0.367	−0.88
NumberSeries	102	15	0	15	9.2	3.6	−0.357	−0.59
LettSets	102	22	4	26	15.84	4.8	−0.185	−0.25
OSpan	102	61	14	75	54.9	15	−0.86	0.04
SymSpan	102	41	0	41	26.6	9.7	−0.87	0.37
RotSpan	102	41	0	41	24.8	10.7	−0.57	−0.52

Note: Descriptives from in tasks with in lab administration. gF tasks: Ravens = Ravens Advanced Progressive Matrices, NumSeries = Number Series, LettSets = Letter Sets. WM tasks: OSpan = Operation Span, SymSpan = Symmetry Span, RotSpan = Rotation Span.

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010

Experiment 4

See Tables A7 and A8.

Table A7
Descriptives for E4

	N	Min	Max	Mean	Std	Skew	Kurtosis
PictureSpan	112	0	75	47.09	22.02	−0.74	−0.69
Klingon	112	0	74	26.74	19.78	0.45	−0.92
SymSpan	112	0	42	28.23	10.14	−0.61	−0.28
MatrixReas	112	2	9	6.17	1.55	−0.44	−0.2
PaperFold	112	0	10	5.67	2.62	−0.31	−0.76
FormBoard	112	0	10	5.88	2.28	−0.54	−0.24
NumSeries	112	1	10	6.63	2.09	−0.43	−0.45
AntiSacc	112	0.2	0.93	0.5	0.15	0.64	0.04
VisualArrays	112	−1.65	4.5	2.4	1.28	−0.47	−0.04

Note: WM tasks: Picture Span, Klingon = Klingon Span, SymSpan = Symmetry Span. gF tasks: MatrixReas = Matrix Reasoning, PaperFold = Paper Folding, FormBoard = Form Boards, NumSeries = Number Series. AntiSacc = Antisaccade, Visual Arrays.

Table A8
Correlation Matrix for E4

	PictureSpan	Klingon	SymSpan	MatrixReas	PaperFold	FormBoard	NumSeries	AntiSacc	VisArrays
PicSpan	1								
Klingon	0.66	1							
SymSpan	0.56	0.61	1						
MatrixReas	0.47	0.24	0.43	1					
PaperFold	0.55	0.45	0.54	0.51	1				
FormBords	0.29	0.24	0.28	0.33	0.32	1			
NumSeries	0.45	0.36	0.35	0.37	0.4	0.47	1		
AntiSacc	0.32	0.42	0.22	0.18	0.35	0.23	0.36	1	
VisArrays	0.42	0.53	0.47	0.31	0.49	0.37	0.26	0.36	1

Note: WM tasks: Picture Span, Klingon = Klingon Span, SymSpan = Symmetry Span. gF tasks: MatrixReas = Matrix Reasoning, PaperFold = Paper Folding, FormBoard = Form Boards, NumSeries = Number Series. AntiSacc = Antisaccade, Visual Arrays.

See Figure A1.

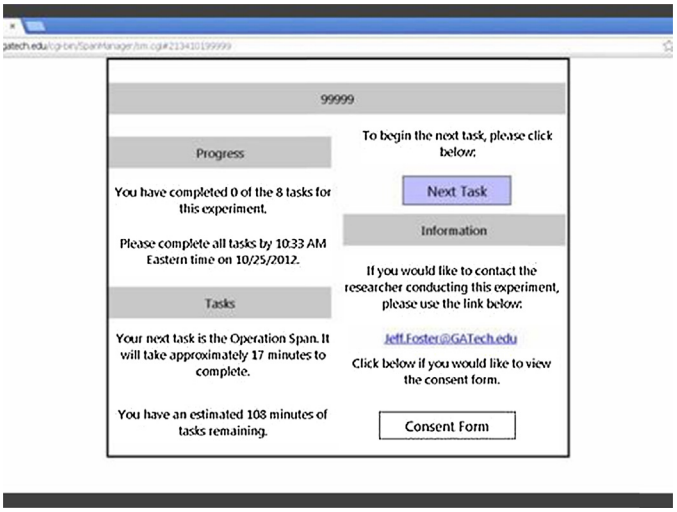

Figure A1. The task manager for our online tasks.

References

Alderton, D. L., Wolfe, J. H., & Larson, J. E. (1997). The Enhanced Computer Administered Test (ECAT) battery. *Journal of Military Psychology*, *9*, 5–37.

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.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*, 563–570.

Chuderski, A. (2014). The relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory & cognition*, *42*(3), 448–463.

Cowan, N. (2010). The magical mystery four how is working memory capacity limited, and why? *Current directions in psychological science*, *19*(1), 51–57.

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

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid

intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309.
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.
Hasher, L., Lustig, C., & Zacks, R. T. (2007). Inhibitory mechanisms and the control of attention. *Variation in working memory*, *19*, 227–249.
Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*(2), 217.
Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83.
Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General*, *130*(2), 169.
Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in working memory*, *1*, 21–48.
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.
Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47.
Kane, M. J., & McVay, J. C. (2012). What mind wandering reveals about executive-control abilities and failures. *Current Directions in Psychological Science*, *21*(5), 348–354.
Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence*, *14*(4), 389–433.
Lopez, N., Previc, F. H., Fischer, J., Heitz, R. P., & Engle, R. W. (2012). Effects of sleep deprivation on cognitive performance by United States Air Force pilots. *Journal of Applied Research in Memory and Cognition*, *1*(1), 27–33.
Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill the role of working memory capacity. *Psychological Science*,
Oberauer, K., Süß, H. M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. *Variation in working memory*, 49–75.
Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence*, *31*, 167–193.
Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1302.
Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.
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*, 635–654.

Received 1 February 2016;
received in revised form 2 July 2016;
accepted 5 July 2016
Available online xxx

Please cite this article in press as: Hicks, K. L., et al. Measuring Working Memory Capacity on the Web With the Online Working Memory Lab (the OWL). *Journal of Applied Research in Memory and Cognition* (2016), http://dx.doi.org/10.1016/j.jarmac.2016.07.010