Original Article

Measuring Working Memory Capacity With Automated Complex Span Tasks

Thomas S. Redick[1], James M. Broadway[1], Matt E. Meier[2], Princy S. Kuriakose[1], Nash Unsworth[3], Michael J. Kane[2], and Randall W. Engle[1]

[1]Georgia Institute of Technology, Atlanta, GA, USA [2]University of North Carolina Greensboro, Greensboro, NC, USA, [3]University of Oregon, Eugene, OR, USA

Abstract. Individual differences in working memory capacity are related to a variety of behaviors both within and outside of the lab. Recently developed automated complex span tasks have contributed to increasing our knowledge concerning working memory capacity by making valid and reliable assessments freely available for use by researchers. Combining the samples from three testing locations yielded data from over 6,000 young adult participants who performed at least one of three such tasks (Operation, Symmetry, and Reading Span). Normative data are presented here for researchers interested in applying cutoffs for their own applications, and information on the validity and reliability of the tasks is also reported. In addition, the data were analyzed as a function of sex and college status. While automated complex span tasks are just one way to measure working memory capacity, the use of a standardized procedure for administration and scoring greatly facilitates comparison across studies.

Keywords: working memory capacity, individual differences, validity, reliability

Throughout the psychological literature, working memory capacity (WMC) is a critical construct for cognitive functioning. Numerous studies showed that WMC is strongly related to intelligence (Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süß, 2005) and executive functions (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). WMC is seen as a core cognitive deficit in theories of aging, schizophrenia, ADD, Alzheimer's disease, and reading disability (Engle & Kane, 2004; Kane, Conway, Hambrick, & Engle, 2007). Furthermore, individual differences in WMC have been implicated in social psychology phenomena such as stereotype threat, emotion regulation, and intrusive thought suppression. WMC has also been studied in applied research on multitasking, driving under distraction, and fatigue in medical students and pilots (Engle, 2010).

Therefore, the proper measurement of individual differences in WMC is critical. To facilitate accurate and reliable measurement, Engle and colleagues created and made freely available automated versions of three of the most widely used WMC measures (Operation, Symmetry, and Reading Span), which take into account psychometric and theoretical considerations known to influence scores on these tasks. In the present article, we highlight the broad applicability of automated complex span tasks (CSTs) and present new analyses of data collected at three testing locations over the past 8 years. We begin with some background on the use of CSTs as WMC measures.

CSTs as WMC Measures

Simple span tasks such as Digit Span and Corsi Blocks, in which subjects serially report a series of items presented, have been widely used in standardized intelligence test batteries. The Reading Span (Daneman & Carpenter, 1980) combined the storage aspect of a simple span task (remember a series of words in order) with an interleaved processing task (reading a sentence) – hence the label "complex span task." Daneman and Carpenter found that the number of words recalled in the CST (Reading Span) – but not in the simple span task (Word Span) – predicted performance in reading comprehension and pronoun reference criteria measures. This finding supported the view that CSTs measure a dynamic *working* memory system that involves both the storage and processing of information, in contrast to simple span tasks, which measure a short-term memory capacity that involves storage only. Meta-analyses showed

DOI: 10.1027/1015-5759/a000123
© 2012 Hogrefe Publishing

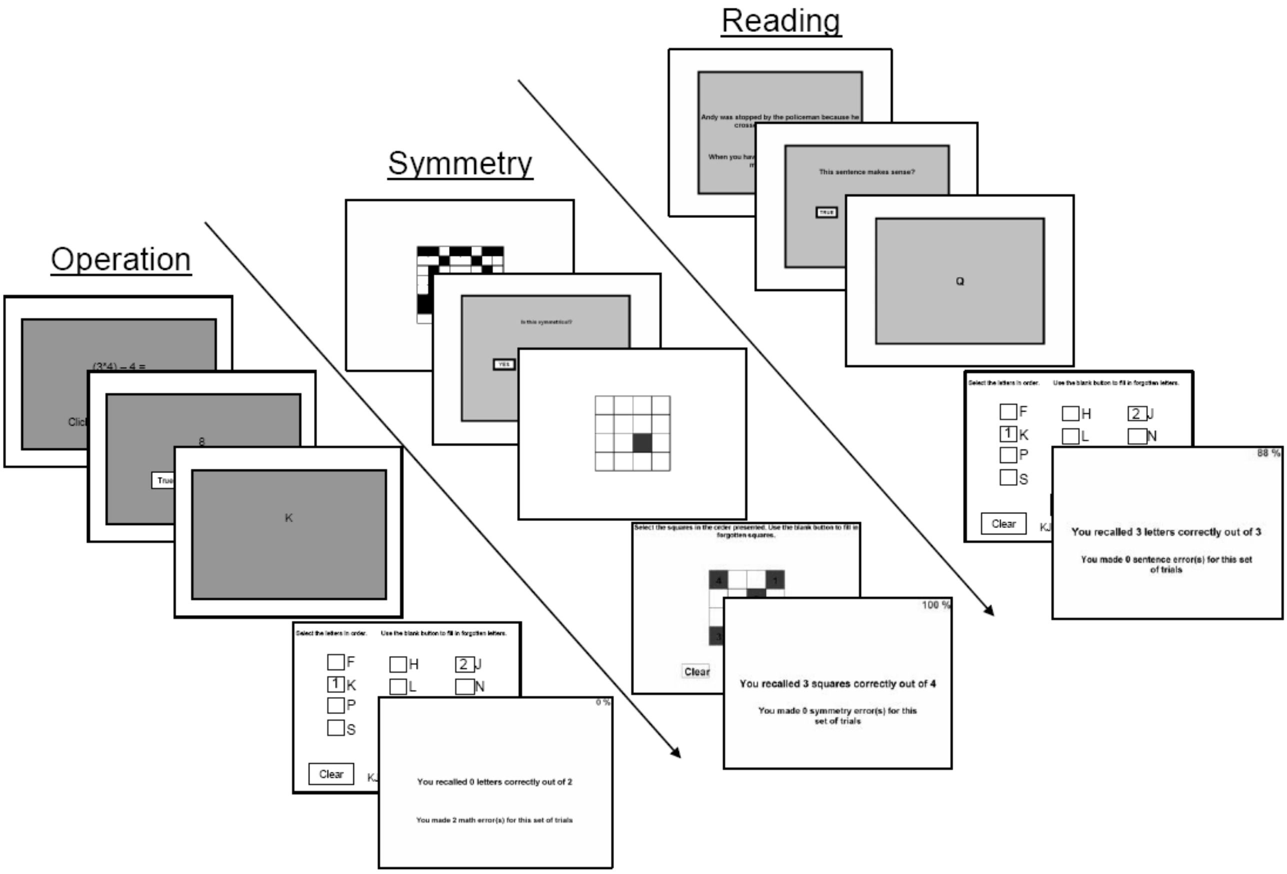

Figure 1. Example screenshots from Operation, Symmetry, and Reading Span. The first three slides show one processing-storage sequence, and the last two slides show the recall and feedback screens at the end of each trial.

that CST performance is strongly related to reading comprehension (Daneman & Merikle, 1996) and fluid intelligence (Ackerman, Beier, & Boyle, 2005; Kane et al., 2005; Unsworth & Engle, 2007), and that effects of cognitive aging are larger for CSTs than for simple span tasks (Bopp & Verhaeghen, 2005).

After the introduction of Reading Span, subsequent research explored potential associations of the processing component in CSTs with criterion-related abilities. Turner and Engle (1989) showed with Operation Span that the processing content of the CST (viz., math operations) need not be similar to the ability criterion in order to measure WMC. Later, Shah and Miyake (1996) developed Symmetry Span, in which participants made symmetry judgments and remembered spatial locations. Shah and Miyake demonstrated a verbal-spatial distinction between Symmetry and Reading Span in the prediction of spatial abilities, consistent with Baddeley's (1986) working memory model. However, Kane et al. (2004) demonstrated that the storage components of Reading, Operation, and Symmetry Span account for similar variance in verbal and spatial ability tests, leading to the view that individual differences in WMC are predominantly domain-general.

Individual Differences in Working Memory Capacity

To account for the broad predictive utility of CSTs, Engle and colleagues proposed that performance on Operation, Symmetry, and Reading Span primarily reflect individual differences in *executive attention* (Engle & Kane, 2004; Kane et al., 2007). Executive attention includes both memory and attention abilities (Unsworth & Engle, 2007), and reflects the ability to temporarily maintain goal-relevant information in primary memory and to retrieve information from secondary memory. The ability to maintain and retrieve information is especially important in situations with high interference, such as the CSTs, where attention must switch between mental representations, and where information needs to be remembered and then quickly forgotten again across trials. In support of this notion, numerous studies showed that performance on CSTs predicts performance on tasks requiring attentional control, such as Stroop, antisaccade, flankers, attentional blink, and go/no-go tasks, although these tasks do not have an obvious memory component (see Kane et al., 2007, for a review). Critically, low scorers on CSTs per-

© 2012 Hogrefe Publishing

form worse in attention tasks in the interference conditions, but not the control conditions, indicating that WMC does not reflect a general deficit.

Automated Complex Span Tasks

Unsworth, Heitz, Schrock, and Engle (2005) reported on the development of an automated version of Operation Span. Automated versions of Symmetry and Reading Span were described in Unsworth, Redick, Heitz, Broadway, and Engle (2009). Compared to "traditional" CSTs, the automated CSTs are quickly administered, completely computerized and mouse-driven, and automatically scored. In addition, the automated CSTs generate a random combination of trials and list lengths at each administration, use to-be-remembered items that are distinct from the processing task, and present feedback on processing and storage accuracy at the end of each trial.

For each trial in automated Operation Span, participants first see an arithmetic equation, then indicate whether a presented answer is correct, and finally see a letter to remember for later recall (Figure 1). After three to seven such processing-and-storage presentations, a recall grid is presented, and participants must click on the letters they saw during the trial in correct serial order. Reading and Symmetry Span are similarly structured except for obvious differences in content (Figure 1). For all automated CSTs, there are three practice conditions before proceeding to the real trials: (a) storage-task only, (b) processing-task only, and (c) processing-and-storage tasks interleaved. An upper bound on processing time during the processing-and storage task trials is based on the participant's performance during the processing-task only condition – the participant's mean plus 2.5 *SD*s. This method of establishing individualized time limits was determined after extensive pilot testing and was motivated by a concern with possible processing-storage tradeoffs if the participant was allowed to take as much time as desired during the processing task. Consistent with such concerns, research has shown that CSTs with unlimited processing times do not predict higher-order cognition compared to CSTs in which processing decision-times are constrained (Friedman & Miyake, 2004; St. Clair-Thompson, 2007).

For all of the automated CSTs, the final screen of the program displays five scores: (a) absolute storage score, which is the sum of all trials in which all items were recalled in the correct serial order; (b) partial storage score, which is the sum of items recalled in the correct serial position, regardless of whether the entire trial was recalled correctly; (c) processing errors, which are the total number of errors made on the processing task; (d) speed errors, which are the number of processing problems that were not answered before the individualized time limit; and (e) accuracy errors, which are the number of processing problems that were answered incorrectly (note that processing errors = speed errors + accuracy errors).

Although the automated CSTs provide storage scores based on an absolute-scoring and partial-credit scoring method, research indicates that the psychometric properties of partial-credit scoring are better. For the traditional CSTs, partial-credit scores have higher internal consistencies (Conway et al., 2005; Friedman & Miyake, 2005) and stronger relationships with reading comprehension (Friedman & Miyake, 2005) and matrix reasoning (Unsworth & Engle, 2007) compared to absolute-scoring scores. The absolute-scoring method makes less sense from a test-theory perspective, because information is discarded that could be used to distinguish among individuals' performance. We advocate the use of the partial scores based on analyses of the traditional CSTs: This paper represents the first attempt to compare the two scoring methods provided in the automated CSTs.

The purpose of the remainder of the article is twofold. First, we review previously published evidence for the test-retest reliability, construct validity, and criterion-related validity of the automated CSTs. Second, we report new analyses of data from over 6000 participants to examine the internal consistency and convergent validity of the automated CSTs, along with normative descriptive statistics. Although most researchers solely use performance on the storage aspect of CSTs to measure WMC, we also report normative data for processing errors (for detailed analyses of processing performance on the automated CSTs, see Unsworth, Redick et al., 2009).

Test-Retest Reliability

We begin by presenting information from previously published studies. Test-retest reliabilities for the automated CSTs from Unsworth et al. (2005) and Unsworth, Redick et al. (2009) are presented in Table 1. Note that the test-retest reliabilities based on the absolute scores are lower than the partial scores. In addition, although partial scores were significantly higher at time 2 relative to time 1 for each task (all *p*s < .05), the increase was only 2–3 items, indicating relatively small practice effects on the automated CSTs. Importantly, as indicated by the high test-retest reliabilities, the rank-ordering of individuals was stable across test sessions.

Table 1. Test-retest reliabilities

	Operation ($N = 78$)	Symmetry ($N = 138$)	Reading ($N = 138$)
Absolute score	.77	.62	.76
Partial score	.83	.77	.82

 © 2012 Hogrefe Publishing

Construct and Criterion-Related Validity

The automated and traditional versions of the CSTs correlate strongly with each other (Unsworth et al., 2005). The automated CSTs are also highly correlated with other commonly used working memory measures such as Running Letter Span (Broadway & Engle, 2010) and Letter-Number Sequencing (Shelton, Elliott, Hill, Calamia, & Gouvier, 2009). The correlation between the automated CSTs and the *n*-back task is small (Jaeggi, Studer-Leuthi et al., 2010; Unsworth, 2010b; Unsworth, Miller et al., 2009), although this is consistent with previous research (Jaeggi, Buschkuehl, Perrig, & Meier, 2010; Kane et al., 2007) showing little overlap between performance on traditional CSTs and *n*-back tasks.

Numerous studies have used the automated CSTs to predict performance in other domains. Table 2 shows studies in which automated CSTs were used to predict either performance on Raven Progressive Matrices (fluid intelligence) or a Vocabulary test (crystallized intelligence). As can be seen, the zero-order correlations are consistently higher with Raven than with Vocabulary. In addition, using a multidimensional scaling approach, Tucker-Drob and Salthouse (2009) showed that all three automated CSTs centrally loaded with highly g-loaded reference variables such as Raven, Letter Sets, and Shipley Abstraction. As has been found repeatedly with the traditional CSTs, performance on automated CSTs is predictive of higher-order cognitive abilities. Importantly, like traditional CSTs, the automated CSTs are also predictive of low-level attention abilities, including sustained attention (McVay & Kane, 2009), selective attention (Redick & Engle, 2006), and response inhibition (Unsworth, Spillers, & Brewer, 2009). Finally, CST performance declines with increasing age in adulthood for both traditional (McCabe et al., 2010) and automated (Salthouse, Pink, & Tucker-Drob, 2008) CSTs.

Table 2. Automated CST correlations from previous studies

Source	N	Raven	Vocab
Broadway & Engle (2010) E1	89	.44	
Broadway & Engle (2010) E2a	151	.53	
Broadway & Engle (2010) E2b	143	.52	
Jaeggi, Studer-Leuthi et al. (2010)	104	.24	
Shelton, Elliott, Hill, Calamia, & Gouvier (2009)	174	.29	
Unsworth, Heitz, Schrock, & Engle (2005)	252	.38	
Unsworth, Brewer, & Spillers (2009)	173	.26	
Unsworth, Redick et al. (2009)	138	.51	
Unsworth, Spillers, & Brewer (2009)	155	.23	
Unsworth & Spillers (2010)	181	.20	
Unsworth (2010b)	165	.32	.09
Unsworth (2010a)	161		.12
Unsworth & Brewer (2010)	177		.12
Unsworth, Spillers, & Brewer (2011)	156		.15

Note. Correlations from studies that administered multiple automated CSTs were averaged together.

Normative Sample

In order to answer other psychometric questions about automated Operation, Symmetry, and Reading Span, we combined the data of three research sites that used the identical versions of the tasks from 2004 to 2009. Although many of these participants' data were analyzed for individual studies, the present analyses are new. All participants at the University of Georgia (UGA) and the University of North Carolina Greensboro (UNCG) were students participating for credit. Participants at the Georgia Institute of Technology (GT) included: (a) GT students participating for either credit or pay; (b) students enrolled at other area colleges and technical schools participating for pay; or (c) nonstudents from the community participating for pay. All participants were young adults between the ages of 17 and 35 who completed at least one automated CST. An advantage of the aggregate sample is that our participants have a wide range of cognitive abilities, operationally defined here as "college status." According to *Princeton Review*, incoming freshmen in the 2008 class had the following verbal/quantitative/total SAT scores: GT (644/690/1334); UGA (616/618/1234); UNCG (522/523/1045).

In addition, our inclusion of nonstudents from the Atlanta area allowed a more adequate representation of the lower end of the cognitive ability spectrum, which is useful for generalization to the population of all young adults (not just college students). For example, although many theoretical research studies use only college students as participants, clinical and medical studies are more likely to use nonstudents from the general population. With the current dataset (N = 6,274), we were able to address typical performance, internal consistency, and relationships among the automated CSTs.

The normative data are useful for a variety of reasons. First, many WMC researchers (including ourselves) select participants who score in the upper or lower quartile of our distribution of automated CST performance and then compare these individuals on another task of interest. The extreme-groups approach was crucial to the development of the executive-attention theory of WMC, because the need to manipulate multiple variables of interest within the attention tasks does not fit easily within a regression-based approach. For example, Kane and Engle (2003) compared high- and low-WMC individuals on accuracy and response times to different Stroop task conditions by manipulating the frequency of incongruent trials, the order of task conditions, and the presence of feedback. While the extreme-groups method has its limitations (Preacher, Rucker, MacCallum, & Nicewander, 2005), it is analogous to aging studies that sample young and older adults and compare

© 2012 Hogrefe Publishing

their performance on a task of interest. Thus, one use of the normative data is that researchers can compare their samples to the normative data and conduct extreme-groups studies without sampling the middle part of the distribution. In addition, the normative data may be helpful for individuals who want to use the automated CSTs in clinical assessments. For example, automated Operation and Symmetry Span are part of a consensus panel's cognitive battery for schizophrenia research (Barch et al., 2009).

Descriptive Statistics

Descriptive statistics are presented in Table 3. Although the storage scores are normally distributed (skewness < 2 and kurtosis < 4; Kline, 1998), the skewness and kurtosis for the processing accuracy variables are very high (see also Unsworth, Redick et al., 2009). There is often a floor effect for processing errors for college students because the processing decisions are intended to be a distraction for the storage task. Given the high skewness and kurtosis, processing accuracy analyses should be conducted after correcting for the deviation from normality (e.g., Unsworth, Redick et al., 2009, used an arcsin transformation). Table 4 shows the storage scores for each task as a function of the percentile in the dataset. These percentiles are provided for median-, tercile-, and quartile-based assignment of group (high/low).

We also examined gender effects on automated CSTs. Gender information was not available for $N = 469$ participants. Examining the partial storage scores, males remembered more items than females on all three tasks: Operation, $t(5815) = 3.42$, $p < .01$, $d = .09$; Symmetry, $t(5589) = 9.78$, $p < .01$, $d = .26$; Reading, $t(5112) = 2.24$, $p = .03$, $d = .06$. The male advantage was only 1–2 items for each of the tests, and as the effect sizes indicate, the magnitude of the gender effects were small, especially for Operation and Reading Span. Males made more processing errors than females on two of the tasks: Operation, $t(5815) = 3.21$, $p < .01$, $d = .08$; Symmetry, $t(5589) = 0.94$, $p = .35$, $d = .03$; Reading, $t(5112) = 7.01$, $p < .01$, $d = .19$. The female advantage was less than one error for each test, and again the magnitude of the gender effects was small.

Performance was also examined based on the school that the participants were attending (Table 5). Consistent with the SAT scores reported above, the automated CST scores of the GT students were slightly higher than those of the UGA students, and both GT and UGA students scored higher than the UNCG students. The scores of the UNCG students were most similar to the group of non-GT participants, which represented a mix of Atlanta area college students and nonstudents.

Table 3. Descriptive statistics

Measure	Mean	*SD*	Skew	Kurtosis
Operation span (*N* = 6,236)				
Absolute score	42.04	17.67	–0.28	–0.59
Partial score	57.36	13.65	–1.30	1.74
Processing errors	6.58	5.27	3.34	20.73
Speed errors	1.48	2.38	12.21	305.13
Accuracy errors	5.11	4.49	3.41	18.59
Symmetry span (*N* = 6,018)				
Absolute score	18.76	9.62	0.20	–0.64
Partial score	27.87	8.26	–0.61	–0.03
Processing errors	3.41	4.02	2.94	13.25
Speed errors	0.67	1.54	11.55	239.69
Accuracy errors	2.74	3.55	3.03	13.23
Reading span (*N* = 5,537)				
Absolute score	36.51	18.83	–0.01	–0.86
Partial score	53.81	15.09	–0.97	0.56
Processing errors	5.08	4.70	4.04	30.98
Speed errors	1.41	1.67	2.41	11.11
Accuracy errors	3.66	4.21	5.23	48.22

Table 4. Percentiles for storage scores

Measure	5	25	33.3	50	66.6	75	95
Operation span (*N* = 6,236)							
Absolute score	10	30	34	43	51	55	68
Partial score	29	51	55	61	65	67	73
Symmetry span (*N* = 6,018)							
Absolute score	4	12	14	19	23	25	35
Partial score	13	23	25	29	33	34	39
Reading span (*N* = 5,537)							
Absolute score	6	22	27	37	46	51	68
Partial score	24	46	50	57	63	65	73

Table 5. Partial scores as a function of college status

	Overall	GT	UGA	UNCG	nonGT
Operation	57.35 (13.66)	62.46 (9.79)	61.16 (10.87)	52.02 (15.20)	53.31 (14.93)
	N = 6,236	*N* = 1,245	*N* = 2,010	*N* = 1,511	*N* = 1,470
Symmetry	27.87 (8.26)	31.29 (6.71)	29.67 (7.32)	26.42 (8.24)	24.31 (8.79)
	N = 6,018	*N* = 1,245	*N* = 1,786	*N* = 1,512	*N* = 1,475
Reading	53.81 (15.09)	59.78 (11.19)	58.11 (12.87)	46.77 (15.46)	49.58 (16.58)
	N = 5,537	*N* = 1,037	*N* = 2,000	*N* = 1,504	*N* = 996

 © 2012 Hogrefe Publishing

Internal Consistency

The internal consistency of the storage scores was calculated in two different ways with the same methods used by Engle, Tuholski, Laughlin, and Conway (1999) and Kane et al. (2004) with traditional CSTs. First, because there are three presentations of each of the five list lengths in the CSTs, each of the 15 trials can be identified as the first, second, or third instance of a particular list length. The first instance of all list lengths can be used to create a score based on those five trials, then the same for the second and third instances of each list length, and then calculate a Cronbach's α based on each third of the test. Second, using Kane et al.'s method, we calculated the proportion of letters correctly recalled for each of the 15 trials and then calculated Cronbach's α across the 15 items. As Table 6 shows, the reliabilities for each task were well above the recommended level of .70 (Nunnally, 1978) when using the partial scores; internal consistencies were lower for the absolute scores.

Table 6. Cronbach's α for automated CSTs

Method	Operation (N = 6,077)	Symmetry (N = 5,871)	Reading (N = 5,389)
Engle, Tuholski, Laughlin, & Conway (1999)	.86/.80	.81/.73	.88/.83
Kane et al. (2004)	.84/.75	.76/.63	.86/.78

Note. Reliabilities are presented for partial/absolute scores, respectively.

Convergent Validity

Crosstask correlations (Table 7) were examined for participants who completed all three automated CSTs (N = 5,316). All correlations were r = .52 or higher, with the highest correlation observed between the two nonspatial automated CSTs. Note that correlations among the absolute scores were lower by .05–.07. Also presented in Table 7 are the correlations as a function of the college status of the sample. The Operation and Reading Span correlations did not differ much as a function of college status, but the correlations of Operation and Reading Span with Symmetry Span increased as the ability level of the subsample decreased. For example, the Reading-Symmetry Span correlation was r = .36 in the GT sample, whereas the correlation was r = .59 in the non-GT sample tested at the same site. Correlations were also examined separately for males (N = 1,897) and females (N = 2,992) who completed all three CSTs. The correlation between Operation and Reading Span did not differ between the sexes (.68/.67 for males/females), but the correlations with Symmetry Span were slightly higher for males than females. Operation and Symmetry Span were correlated at r = .57 and .48 for males and females, respectively; Reading and Symmetry Span were correlated at r = .59 and .48 for males and females, respectively.

Table 7. Correlations among the partial storage scores on the automated CSTs

	1. Operation	2. Symmetry
1. Operation	–	
2. Symmetry	.52/.36/.42/.51/.55	–
3. Reading	.68/.61/.61/.63/.68	.53/.36/.42/.51/.59

Note. N = 5,316/1,037/1,781/1,506/992 for the overall, GT, UGA, UNCG, and non-GT samples, respectively.

Discussion

The data indicate that automated CSTs have desirable psychometric properties as evidenced by (1) high test-retest reliability, (2) high internal consistency, (3) convergent and discriminant construct validity, and (4) criterion-related validity. The normative data indicate substantial variability in automated CST performance. In addition, the automated CSTs show extremely small or no gender effects, in contrast to researchers who claim male advantages for WMC based on other types of measures (e.g., Lynn & Irwing, 2008).

The analyses also show that automated CST performance is consistent with the expected cognitive ability level of the population from which the sample was obtained. The correlations among the automated CSTs also varied as a function of the sample type, with lower correlations obtained in the samples with higher mean performance. The goal of these analyses was not to provide data specific to these particular institutions, but rather to clearly demonstrate the role that the sample plays in WMC studies using automated CSTs. For example, a researcher interested in the domain-general or domain-specific nature of WMC may arrive at a different conclusion if the participants are all GT students as opposed to community nonstudents.

Our results confirm previous findings with traditional CSTs that partial scoring is superior to absolute scoring, based on (1) higher test-retest correlations, (2) higher internal consistencies, and (3) higher correlations among the three CSTs. As stated previously, the partial scoring method picks up the same variance as the absolute method, plus additional variance due to the partial credit. Unless there is a strong theoretical reason to use absolute scores, we recommend using partial scores as the method that is more reliable and sensitive to individual differences in WMC.

Automated CSTs are just one way to measure WMC, and a variety of measures should be used to eliminate the influence of specific method-variance (Lewandowsky, Oberauer, Yang, & Ecker, 2010). However, the standardized administration and scoring of the automated CSTs, together with the large corpus of data available, are beneficial to researchers interested in measuring WMC without creating their own task. Consistent use of validated, reliable, standardized measures also supports the generalization of results from a particular study by allowing for easier tests of replication across samples in different research labs. Finally, we feel that making the computerized tasks freely available for download from our website (in E-Prime or Inquisit) can increase the diversity of research on the relationship of individual differences in WMC with other constructs. In conclusion, the automated CSTs are valid and reliable tools to further our understanding about the nature of WMC and about why individual differences in WMC are related to a variety of behaviors in and out of the lab.

Acknowledgments

During the writing of the manuscript several of the coauthors were being supported by a grant from the Office of Naval Research (Award No. N000140910129). We thank Chad Schrock and Rich Heitz for their programming work in creating and piloting the automated complex span tasks. We thank the hundreds of undergraduate and graduate students who helped administer these tasks as research assistants in our labs. The tasks are available for download at http://www.psychology.gatech.edu/renglelab

References

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

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

Barch, D. M., Berman, M. G., Engle, R., Jones, J. H., Jonides, J., MacDonald, A., III, . . . Sponheim, S. R. (2009). CNTRICS final task selection: Working memory. *Schizophrenia Bulletin, 35,* 136–152.

Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 60B*, 223–233.

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.

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

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

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

Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology, 51*, S17–S26.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145–199). New York: Elsevier.

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

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language, 51*, 136–158.

Friedman, N. P., & Miyake, A. (2005). Comparison of four scoring methods for the reading span test. *Behavior Research Methods, 37*, 581–590.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory, 18*, 394–412.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning – Implications for training and transfer. *Intelligence, 38*, 625–635.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working-memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–48). New York: Oxford University Press.

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,* 47–70.

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

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 visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133,* 189–217.

Kline, R. B. (1998). *Principles and practice of structural equation modeling*. NY: Guilford.

Lewandowsky, S., Oberauer, K., Yang, L.-X., & Ecker, U. K. H. (2010). A working memory test battery for MatLab. *Behavior Research Methods, 42*, 571–585.

Lynn, R., & Irwing, P. (2008). Sex differences in mental arithmetic, digit span, and g defined as working memory capacity. *Intelligence, 36*, 226–235.

McCabe, D. P., Roediger, III, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24*, 222–243.

McVay, J. C., & Kane, M. J. (2009). Conducting the train of thought: Working memory capacity, goal neglect, and mind wandering in an executive-control task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 196–204.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. (2005). Working memory and intelligence – Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 61–65.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme-groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*, 178–192.

Redick, T. S., & Engle, R. W. (2006). Working memory capacity

 © 2012 Hogrefe Publishing

and Attention Network Test performance. *Applied Cognitive Psychology, 20,* 713–721.
Salthouse, T. A., Pink, J. E., & Tucker-Drob, E. M. (2008). Contextual analysis of fluid intelligence. *Intelligence, 36*, 464–486.
Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*, 4–27.
Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence, 37*, 283–293.
St. Clair-Thompson, H. L. (2007). The influence of strategies upon relationships between working memory and cognitive skills. *Memory, 15*, 353–365.
Tucker-Drob, E. M., & Salthouse, T. A. (2009). Confirmatory factor analysis and multidimensional scaling for construct validation of cognitive abilities. *International Journal of Behavioral Development, 33*, 277–285.
Turner, M., & Engle, R. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.
Unsworth, N. (2010a). Interference control, working memory capacity, and cognitive abilities: A latent variable analysis. *Intelligence, 38*, 255–267.
Unsworth, N. (2010b). On the division of working memory and long-term memory and their relation to intelligence: A latent-variable approach. *Acta Psychologica, 134*, 16–28.
Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language, 62*, 19–34.
Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity-fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin and Review, 16*, 931–937.
Unsworth, N., & Engle, R. W. (2007). On the division of short-term and working memory: An examination of simple and complex spans and their relation to higher-order abilities. *Psychological Bulletin, 133*, 1038–1066.
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., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J., Campbell, W., & Goodie, A. S. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences, 30*, 194–200.
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.
Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language, 62*, 392–406.
Unsworth, N., Spillers, G. J., & Brewer, G. A. (2009). Examining the relations among working memory capacity, attention control, and fluid intelligence from a dual-component framework. *Psychology Science Quarterly, 4*, 388–402.
Unsworth, N., Spillers, G. J., & Brewer, G. A. (2011). Variation in verbal fluency: A latent variable analysis of switching, clustering, and overall performance. *Quarterly Journal of Experimental Psychology, 64*, 447–466.

Published online: February 15, 2012

Thomas S. Redick

Indiana University
Purdue University Columbus
4601 Central Avenue
Columbus, IN 47203
USA
E-mail tsredick@iupuc.edu

© 2012 Hogrefe Publishing