



## Assessing working memory capacity in a non-native language

Christopher A. Sanchez<sup>a,\*</sup>, Jennifer Wiley<sup>b</sup>, Timothy K. Miura<sup>c</sup>, Gregory J.H. Colflesh<sup>d</sup>, Travis R. Ricks<sup>e</sup>, Melinda S. Jensen<sup>f</sup>, Andrew R.A. Conway<sup>g</sup>

<sup>a</sup> Arizona State University, United States

<sup>b</sup> University of Illinois at Chicago, United States

<sup>c</sup> Loyola University, Chicago, United States

<sup>d</sup> University of Illinois at Chicago, United States

<sup>e</sup> University of Illinois at Chicago, United States

<sup>f</sup> University of Illinois at Urbana-Champaign, United States

<sup>g</sup> Princeton University, United States

### ARTICLE INFO

#### Article history:

Received 4 January 2010

Received in revised form 25 March 2010

Accepted 1 April 2010

#### Keywords:

Working memory capacity

Bilingual

Individual differences

### ABSTRACT

The present studies directly test the usefulness of two English-language working memory capacity (WMC) assessments with two samples of students whose native language was not English. Participants completed two widely used complex span tasks, Reading Span (RSpan) and Operation Span (OSpan), in English. To determine whether the well-established relationship between WMC and Raven's Advanced Progressive Matrices (RAPM) would be observed when span tasks were not given in the native language of the participant, span scores were regressed against performance on the RAPM. Results indicated that while OSpan was a reliable and valid predictor of RAPM in non-native-English speakers, RSpan administered in English was not.

© 2010 Elsevier Inc. All rights reserved.

Given the increasing population of individuals who speak more than one language, a topic of growing interest is the effect of bilingualism on cognition and the mechanisms that underlie effective bilingual memory and language processing. Many recent studies have explored how bilingualism affects executive functioning, flexibility and attentional control in individuals (Bialystok, 2007; Colzato et al., 2008; Ransdell, Barbier, & Nilt, 2006). Importantly, a methodological point that has not been subject to much scrutiny in previous studies on bilinguals is the language in which various cognitive tests are performed.

Specifically, little attention has been given to the linguistic background of participants in many cognitive experiments at English-speaking universities. In these settings, assessments of executive functioning and general cognitive abilities are routinely administered only in English. Although many cognitive tasks have been shown to be reliable and valid measures for native speakers, it is imperative to determine whether these same tasks are reliable and valid predictors of ability when given in a non-native language. This point seems especially relevant for tasks with a large verbal component, such as one of the most popular working memory span tasks (e.g., Reading Span). The present study directly tests the usefulness of English-language working memory

capacity assessments with a population whose native language is not English.

Working memory capacity (WMC) has emerged in the past 30+ years as a powerful predictor of performance across a myriad of tasks (Conway et al., 2005). WMC predicts performance in dichotic listening, Stroop, and anti-saccade tasks (Colflesh, & Conway, 2007; Kane et al., 2001; Kane, & Engle, 2003), as well as reading comprehension (Daneman, & Carpenter, 1980; Long, & Chong, 2001; Payne, Kalibatseva, & Jungers, 2009), metacomprehension skill (Griffin, Wiley, & Thiede, 2008), memory for baseball games (Hambrick, & Engle, 2002) and learning in science (Sanchez, & Wiley, 2006, 2009). Further, WMC also predicts reasoning, problem solving and decision making (Ash, & Wiley, 2006; Copeland, & Radvansky, 2004; Ricks, Turley-Ames & Wiley, 2007; Shamosh et al., 2008).

WMC is typically assessed using complex span tasks with two distinct components (1) a processing component, and (2) a concurrent storage component in which the participant attempts to remember unrelated information for later recall (see Conway et al., 2005 for a review). As these tasks require participants to process and store information concurrently, complex span tasks are considered measures of a domain-general cognitive ability which some have argued is essentially the ability to control attention (Kane et al., 2001), the ability to focus attention (Cowan, 2005), or the ability to resist interference (Bunting, 2006). This construct goes beyond just the basic capacity of the short-term store as assessed by simple word or digit-span tests which do not require concurrent processing (Daneman, & Merikle, 1996; Engle et al., 1999).

\* Corresponding author. Department of Applied Psychology, Arizona State University, 7271 E. Sonoran Arroyo Mall Mesa, AZ 85212, United States. Tel.: +1 480 727 1589; fax: +1 480 727 1538.

E-mail address: [c.sanchez@asu.edu](mailto:c.sanchez@asu.edu) (C.A. Sanchez).

The original reading span task (RSpan; Daneman, & Carpenter, 1980) required participants to read sets of 2–6 sentences out loud while also trying to remember the last word of each sentence for later recall. While this original RSpan measure was found to correlate well with measures of reading comprehension and VSAT performance (Daneman, & Carpenter, 1980), some have argued that this correlation could be a result of domain-specific skills in verbal ability supporting better memory for the sentences that were read (Baddeley, 2003; Daneman, & Merikle, 1996; Ericsson, & Kintsch, 1995). In an effort to eliminate the potential influence of verbal proficiency, newer RSpan tasks require that participants recall unrelated letters instead of the last words of the sentences.

Alternatively, several parallel measures that are less verbal, but share the same basic components as the RSpan task, have been developed including the operation span task (OSpan, Turner, & Engle, 1989). OSpan involves making mathematical judgments for number strings rather than sensibility judgments for sentences. Despite the difference in the processing component, previous research has found a high correlation between RSpan and OSpan tasks with native-English speakers ( $r = .60$ , Conway et al., 2005). Further, in native-English populations, both OSpan and RSpan have been shown to correlate at .30 with performance on the Raven's Advanced Progressive Matrices which is a prototypical measure of fluid intelligence (Unsworth, & Engle, 2005). While some have suggested that the constructs of fluid intelligence ( $G_f$ ) and WMC are nearly synonymous (Kyllonen, & Christal, 1990), more recent research suggests instead that these constructs are indeed separate (but highly correlated) entities that share variance due to the need to control or focus attention (Unsworth, & Engle, 2005). Given the strong and well-documented correlation between WMC and  $G_f$ , a standard way to assess the validity of a WMC assessment as a measure of controlled attention is to examine the extent to which it correlates with RAPM. If the task is truly tapping the WMC construct, then it should reveal a moderate to strong correlation with RAPM.

The present studies are the first to examine the reliability and predictive validity of a set of English-language span tasks with a population whose native language is not English. The individuals in the first study are American college students who report Spanish as their native language, but have spoken English for ten or more years. To further generalize the effects, a second experiment was conducted using another sample of American college students who were non-native-English speakers, where native language was free to vary. These participants reported a variety of native languages (including Spanish).

To assess the predictive validity of these standard span tasks within these populations, the analysis of interest is whether span scores will correlate with RAPM scores. Results for these non-native-English speakers are contrasted with the results of native-English speakers from the same college population. If administration of these tasks in a non-native language fundamentally changes the nature of the construct measured by these tasks, then one might expect a divergence in the correlations between span measures and RAPM across the two linguistic groups. In the most extreme case, neither span task might predict RAPM performance in the non-native population. On the other hand, if both English-language RSpan and OSpan are reliable and valid measures of the WMC construct for native-English and non-native individuals, the usual relationship between each of these measures and RAPM across language groups should be seen.

## 1. Experiment 1

### 2. Method

A total of 134 participants were recruited from a large public Midwestern university in the United States. Half of these participants were native-English speakers, and the remaining 67 participants were

native-Spanish speakers who reported having spoken fluent English for at least 10 years ( $M = 14.25$  years,  $SD = 2.82$ ). The average age (in years) that the bilinguals began speaking English was 5.33 years old ( $SD = 2.45$ ). All participants received course credit for their participation.

All participants completed OSpan, RSpan, and RAPM with breaks between each task. The order of the span tasks was varied across participants. At the end of the session all participants completed a Language History Questionnaire in which they reported their native language and any additional languages they spoke or read, the number of years they had spoken English fluently, and the age at which they began speaking English.

### 2.1. Operation Span (OSpan)

The OSpan task (Turner, & Engle, 1989) requires participants to verify simple mathematical strings while also trying to remember unrelated words. Standards of administration and proportional scoring followed the recommendations in Conway, Kane, Bunting, Hambrick, Wilhelm and Engle (2005). For each trial, participants were required to read aloud and verify the correctness of a single math problem, and then immediately read and remember the word following the equation. For example,

$$IS(8 / 2) - 1 = 1? \text{ BEAR}$$

As soon as the participant read the word, the next operation–word string was presented. The operation–word strings were presented in sets of two to five trials. At the end of each set, the participant was prompted to recall all words in the set, in the correct order. Three trials of each set size were presented, in a random order. In order to ensure that participants were attending to the processing task, an 85% accuracy criterion on the math operations was required. Reliability for the OSpan task was calculated following Unsworth, Heitz, Schrock and Engle (2005). The proportion score for the first presentation of every set size was averaged together to form a sub-score. The same procedure was followed for the second and third presentation of each set size, yielding 3 sub-scores. Cronbach's  $\alpha$  was computed using these 3 sub-scores. As shown in Table 1, while reliability was slightly higher for native English speakers in OSpan, the reliability for bilinguals remains high and similar to previous values (Conway et al., 2005 reports that reliability estimates for span tasks are typically in the .7–.9 range).

### 2.2. Reading Span (RSpan)

The RSpan task (Kane et al., 2004) is identical to the OSpan task except that participants read sentences, make sensibility judgments, and then remember a set of unrelated letters, rather than verifying math equations and remembering words. Administration and scoring was identical to OSpan. Reliability estimates were computed as above, and as shown in Table 1, RSpan appears to be a highly reliable measure in both language groups.

### 2.3. Raven's Advanced Progressive Matrices (RAPM)

RAPM (Raven et al., 1998) is considered a measure of fluid intelligence. The test consists of 36 individual items presented in ascending order of difficulty (i.e., the easiest item is presented first). Each item consists of a matrix of geometric patterns with the bottom-right corner missing. The participant selects, from 8 alternatives, the option that correctly completes the matrix. The task was administered according to standard instructions with a 40-minute time limit. A participant's score was the total number correct. Reliability of the RAPM was determined by computing Cronbach's  $\alpha$  from the subscores

**Table 1**  
Descriptive statistics for all tasks according to language group in Experiment 1.

Measure	Group	Mean (sd)	Median	Cronbach's $\alpha$	Skew	Kurtosis	25% Dist	75% Dist
OSpan	Native English	.63(.16)	.61	.83	.17	−.76	.52	.74
	Bilingual	.54(.07)	.53	.71	.07	.10	.45	.64
	Overall sample	.59(.15)	.58	.79	.20	−.29	.48	.68
RSpan	Native English	.70(.14)	.71	.75	−.05	−.45	.61	.80
	Bilingual	.60(.16)	.61	.75	−.16	−.49	.48	.71
	Overall sample	.65(.16)	.66	.78	−.21	−.29	.54	.75
RAPM	Native English	21.43(4.47)	21.00	.72	.16	−.10	18.00	25.00
	Bilingual	20.58(4.25)	21.00	.74	.34	−.01	17.00	23.00
	Overall sample	21.01(4.36)	21.00	.73	.25	−.12	18.00	24.00

for even and odd items. RAPM was also highly reliable across both language groups.

### 3. Results and discussion

#### 3.1. Descriptive statistics

Before moving to the main analysis, descriptive statistics for each task for the full sample and by language group are considered. As shown in Table 1, the ranges for span scores and RAPM in the two language groups were similar, which is critical for the target correlational analyses. Although bilingual individuals had significantly lower raw scores on both the OSpan ( $t(132) = 3.49, p < .01$ ) and RSpan ( $t(132) = 4.15, p < .01$ ) tasks, there was no difference between language groups in terms of average performance on the RAPM ( $t(132) = 1.13, p > .05$ ).

A final preliminary analysis indicated that the usual high correlation between RSpan and OSpan performance was obtained (see Table 2) and did not differ by language group. The length of time speaking fluent English or age of fluency in the bilingual sample was not correlated with span scores.

#### 3.2. Correlations between span scores and RAPM

Correlations between span scores and RAPM were used to test whether English-language WMC tasks are valid predictors of  $Gf$  in both language groups. As shown in Table 2, across language groups both OSpan and RSpan were significantly correlated with RAPM.

There was no difference ( $z < 1, p > .05$ ) in the magnitude of the correlation between OSpan and RAPM for the bilingual and native-English groups. Interestingly, the pattern of results did differ for RSpan when the language groups were considered separately ( $z = 1.96, p < .05$ ). The relationship for native-English speakers was

**Table 2**  
Correlations overall and by language group in Experiment 1.

Group	Measure	OSpan	RSpan	RAPM
Native English	OSpan	–	–	–
	RSpan	.69**	–	–
	RAPM	.42**	.50**	–
Bilingual	OSpan	–	–	–
	RSpan	.66**	–	–
	RAPM	.41**	.20	–
	Years speaking	.10	.09	.10
	Age fluency	−.18	−.29	.03
Overall	OSpan	–	–	–
	RSpan	.70**	–	–
	RAPM	.42**	.35**	–

\*\*  $p < .01$ .

again significant, but not for bilinguals. (The number of years speaking English or age of fluency also did not correlate with performance on the RAPM.)

In an effort to further establish the differential predictive validity of RAPM by these span tasks, the span measures were entered in sequential fashion into hierarchical regressions with RSpan in the first block and OSpan in the second. The logic behind these regression analyses is that if RSpan is as adequate a predictor of RAPM as OSpan, the addition of OSpan should not significantly increase the overall amount of variance predicted. However, if RSpan is failing to catch a significant portion of the variance in the RAPM that is usually associated with complex span tasks, the addition of OSpan should provide a significantly better fit of the data, as OSpan is providing a more valid measure of the WMC construct. As scores on both complex span tasks were depressed in the bilingual group, it is of interest whether these lower scores (likely a result of the verbal component in each span task; albeit less in the OSpan task) represent a decrease in the sensitivity of these tasks to capture the essence of the relationship between the WMC construct and RAPM.

For the native-English speakers, results from the first block of analysis indicate that RSpan ( $\beta = .50, p < .01$ ) is a significant predictor of RAPM ( $R^2 = .23, F(1,65) = 21.07, p < .01$ ). The addition of OSpan in the second block of analyses did not produce a significant change in overall variance accounted for ( $R^2$  change = .01,  $p > .05$ ). This suggests that RSpan and OSpan scores in this sample both capture the same portion of variance in the RAPM, and that these measures are tapping the same construct of WMC in native-English speakers.

For bilingual speakers, results from the first block of analysis indicate that RSpan ( $\beta = .20, p > .05$ ) is not significantly predictive of RAPM ( $R^2 = .03, F(1,65) = 2.77, p > .05$ ). The addition of OSpan in the second block ( $\beta = .48, p < .01$ ) did allow the model to account for a significant overall portion of variance in the RAPM ( $R^2 = .15, F(2,64) = 6.68, p < .01$ ) that was not identified previously ( $R^2$  change = .13,  $F(1,64) = 10.20, p < .01$ ). These results suggest that RSpan and OSpan did not explain the same variance in RAPM in this bilingual sample. When OSpan was entered into the regression model, it was able to capture a significant amount of variance not already accounted for by RSpan. This suggests that for non-native-English speakers OSpan remains an accurate assessment of the WMC construct, whereas RSpan does not seem to be a good measure of this construct.

Although the correlation between the two complex span tasks in the bilingual group was similar to that of the native group, it is important to note that equivalent correlations across groups does not necessarily mean that the correlations reflect the same underlying construct across the groups, especially when tasks are multiply determined (i.e., performance is dependent upon multiple mechanisms). One way to determine if the correlations reflect the same underlying mechanism is to examine how the two tasks correlate with a third measure. That is exactly the approach we have taken here. The fact that RSpan and OSpan account for different variance in RAPM in the bilingual

group, but account for the same variance in RAPM in the native group, suggests that the nature of the correlation is different across the two groups. As an analogy, two measures of verbal short-term memory may be correlated at similar levels for children and for adults, however this does not necessarily mean that these correlations reflect the same underlying mechanism in each of the two groups.

Furthermore, the amount of RAPM variance accounted for by OSpan in bilinguals is consistent in magnitude with prior research (Conway et al., 2005), and there was no initial difference between language groups in RAPM performance, so it can be reasonably concluded that even while OSpan scores were depressed in this particular sample of bilinguals, the variance that was captured by the OSpan task is indeed the same portion of variance that represents the idealized relationship between WMC and RAPM, namely the controlled attention construct. Thus, even though scores on the span tasks were lower for bilinguals, again likely due to the verbal component inherent in both complex tasks, this verbal portion of the span tasks is not critical to the estimation of the relationship between WMC and measures of complex reasoning.

#### 4. Experiment 2

In order to further establish the utility of OSpan as a measure of the WMC construct in bilingual populations, a second experiment was conducted with a separate group of non-native-English speakers. In this second study, participants were not limited to native-Spanish speakers, but could speak any native language other than English. This second experiment was designed to examine the relationships between these cognitive measures in an English-speaking university subject pool setting, where individuals typically complete English-language assessments regardless of their language history (which is often unknown to experimenters unless participants are explicitly queried).

#### 5. Methods

119 participants (60 Native-English, and 59 non-native-English speakers) at a large Midwestern public university participated in this experiment for course credit. Non-native speakers were self-identified, and native languages spoken included several Asian dialects (e.g., Mandarin, Korean, Gujarati), eastern European dialects (e.g., Polish, Ukrainian, Albanian), Mediterranean/Middle Eastern languages (e.g., Greek, Hebrew, Arabic), in addition to Spanish. Unlike Experiment 1, participants were not required to have spoken English for more than 10 years. Unfortunately, data were missing for 16 of the non-native English speakers on their age of fluency in English, but were available for the remainder.

All materials, methods and procedures were identical to Experiment 1.

### 6. Results and discussion

#### 6.1. Descriptive statistics

All descriptive statistics and reliabilities are presented in Table 3. There was no difference in absolute performance levels across language groups for either RAPM or OSpan ( $t(117) < 1$ ). There was also no significant difference across language groups for RSpan ( $t(117) = 1.62, p > .05$ ).

#### 6.2. Relationships between tasks

Just as in Experiment 1, OSpan and RSpan were significantly correlated in both the native-English ( $r(59) = .60, p < .01$ ) and non-native-English language groups ( $r(60) = .76, p < .01$ ). OSpan was again significantly correlated to RAPM performance in both the native-English ( $r(59) = .39, p < .01$ ) and non-native-English groups ( $r(60) = .43, p < .01$ ). However, RSpan was only significantly correlated with RAPM in the native-English group ( $r(59) = .39, p < .01$ ), and not in the non-native-English group ( $r(60) = .23, p > .05$ ), although this difference in magnitude was not reliable ( $z < 1, p > .05$ ). Number of years speaking English also did not correlate with RAPM, or performance on either span task, in non-native speakers ( $r(43) < .04, p > .05$ ).

To further articulate this disconnect between RSpan and RAPM in the different language groups, hierarchical linear regressions were again conducted. In the native-English group, RSpan ( $\beta = .39$ ) significantly predicted RAPM performance in the first block of the analysis ( $R^2 = .14, F(1,58) = 10.66, p < .01$ ). The addition of OSpan in the second block did not predict a significant additional portion of variance ( $R^2$  change = .04,  $p > .05$ ). Thus, as in Experiment 1, it appears that these tasks are capturing the same construct.

In the non-native-English group, the first block of analysis indicated that RSpan ( $\beta = .23$ ) was not a significant predictor of RAPM performance ( $R^2 = .03, F(1,57) = 3.06, p > .05$ ). However, just as in Experiment 1, addition of OSpan in the second block accounted for a significant portion of additional variance ( $R^2$  change = .10,  $p < .05$ ), and produced an overall significant model ( $R^2 = .12, F(2,56) = 5.00, p < .01$ ). This again suggests that RSpan appears to be non-diagnostic for assessing individual differences in WMC, while OSpan still remains a reliable, valid predictor of this construct in this non-native-English population.

#### 7. General discussion

This set of results indicates that while RSpan and OSpan administered in English are both reliable measures for non-native-speaker samples, only OSpan is a valid predictor of RAPM. Thus, it appears that completing RSpan in a non-native language does not permit an accurate assessment of WMC.

**Table 3**  
Descriptive statistics for all tasks according to language group in Experiment 2.

Measure	Group	Mean (sd)	Median	Cronbach's $\alpha$	Skew	Kurtosis	25% dist	75% dist
OSpan	Native English	.61(.14)	.60	.80	.25	-.16	.50	.71
	Non-Native English	.62(.12)	.61	.64	.30	.62	.54	.68
	Overall sample	.62(.13)	.61	.74	.25	.13	.53	.69
RSpan	Native English	.66(.13)	.66	.70	-.18	.04	.60	.74
	Non-Native English	.62(.15)	.62	.63	-.52	1.13	.53	.71
	Overall sample	.64(.14)	.64	.67	-.42	.79	.56	.73
RAPM	Native English	19.12(6.53)	19.00	.90	.02	-.99	13.25	24.75
	Non-Native English	18.63(5.26)	19.00	.83	-.08	.40	16.00	21.00
	Overall sample	18.87(5.91)	19.00	.87	.01	-.49	15.00	22.00

It is likely that the weak relationship between RSpan and RAPM is a direct consequence of the linguistic nature of the processing task which prohibits an accurate assessment of the WMC construct. This result is consistent with the original intuitions of Turner and Engle (1989) who created the OSpan task specifically to be less sensitive to influences of reading or language ability. The current result also coincides with other research that has demonstrated issues with accurately assessing the WMC construct among populations with low verbal ability when tasks involve a significant language processing component (Nation et al., 1999). Finally, this result might also be related to findings that when bilinguals complete standardized tests in their non-native language, their performance tends to be lower, thus underestimating their true ability (Garcia, & Pearson, 1993; Lee, 1986; Mestre, 1986). Importantly, while linguistic information might prohibit the ability of RSpan to predict complex cognitive performance, this verbal information did not prohibit OSpan from still capturing the portion of RAPM variance commonly associated with WMC in native samples.

Given the current research literature that has focused on a possible “bilingual advantage” in executive functioning, at first it seems a bit puzzling that the bilingual participants did not outperform the monolingual controls in the present studies. There are a couple of possibilities why this may have occurred. The first is that the bilingual advantage has been difficult to replicate consistently (Bialystok, et al., 2004). A second, more incisive, point is that the bilingual advantage may depend on the type of task that is used (Colzato, et al., 2008). Where advantages have been consistently found has been on tasks such as the Simon task and the Wisconsin Card Sort task, which may require different executive functions than the complex span tasks used here. From this perspective, the present findings are consistent with the results of Colzato et al. (2008) who found that while bilinguals did exhibit an advantage on reactive inhibition tasks (i.e., inhibition as a result of more dedicated processing to relevant stimulus), they did not differ from monolinguals on tasks requiring other forms of active inhibition (i.e., inhibition as a result of suppressing irrelevant information). The buildup of proactive interference (or the requirement of active inhibition) is characteristic of both the OSpan and RSpan tasks (Bunting, 2006), but there is little need for reactive inhibition in these tasks. If this reactive inhibition process is critical for the bilingual advantage observed on previous tasks (Bialystok, 2001), the lack of this requirement in the complex span tasks might also explain the lack of a facilitative effect found across both experiments.

Bilinguals represent a portion of the population that is growing at a rapid rate, and non-native-English speakers are increasingly represented in studies of cognitive and social information processing at English-speaking universities. This study demonstrates that the failure to recognize the linguistic status of individuals can be problematic, and some English-language tasks will not be valid measures of cognitive constructs in those populations. Yet, even though OSpan was administered in English, the present results suggest that it still taps the same domain-general construct in non-native-English speakers enrolled in an English-speaking university. Thus, this set of studies provides an important demonstration that English-language OSpan can be used to effectively assess the WMC construct in a non-native-English speaker population.

## Acknowledgements

This research was supported by Grant R305H030170 from the Institute of Education Sciences, U.S. Department of Education, and Grant 0347887 from the National Science Foundation to Jennifer Wiley. The opinions expressed are those of the authors and do not represent views of these institutions. The authors thank Ivan K. Ash, Thomas D. Griffin, Patrick J. Cushen, Brigette Rubas, Melissa Bland, Jimmy Wong, Mital Patel, Yekaterina Zlatorkylova and Gerald

Ontiveros for assistance with data collection and comments on this manuscript.

## References

- Ash, I. K., & Wiley, J. (2006). The nature of restructuring in insight: an individual differences approach. *Psychonomic Bulletin and Review*, 13(1), 66–73.
- Baddeley, A. D. (2003). Working memory and language: an overview. *Journal of Communication Disorders*, 36, 189–208.
- Bialystok, E. (2001). *Bilingualism in Development: Language, Literacy, and Cognition*. New York: NY: Cambridge University Press.
- Bialystok, E. (2007). Cognitive effects of bilingualism: how linguistic experience leads to cognitive change. *The International Journal of Bilingual Education and Bilingualism*, 10(3), 210–223.
- Bialystok, E., Craik, F. I. M., Klein, R., & Vliswanathan, M. (2004). *Psychology and Aging*, 19(2), 290–303.
- Bunting, M. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 183–196.
- Colflesh, G. J. H., & Conway, A. R. A. (2007). Individual differences in working memory capacity and divided attention in dichotic listening. *Psychonomic Bulletin and Review*, 14(4), 699–703.
- Colzato, L. S., Bajo, M. T., van den Wildenberg, W., Paolieri, D., Nieuwenhuis, S., La Heij, W., et al. (2008). How does bilingualism improve executive control? A comparison of active and reactive inhibition mechanisms. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 34(2), 302–312.
- 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.
- Copeland, D. E., & Radvansky, G. A. (2004). Working memory and syllogistic reasoning. *The Quarterly Journal of Experimental Psychology*, 57A(8), 1437–1457.
- Cowan, N. (2005). *Working Memory Capacity*. New York: Psychology Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory capacity and reading. *Journal of Verbal Learning and Verbal Behavior*, 19(4), 450–466.
- Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychonomic Bulletin and Review*, 3(4), 422–433.
- 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.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245.
- Garcia, G. E., & Pearson, P. D. (1993). Assessment and diversity. *Review of Research in Education*, 20, 337–391.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition*, 36(1), 93–103.
- Hambrick, D. Z., & Engle, R. W. (2002). Effects of domain knowledge, working memory capacity, and age on cognitive performance: an investigation of the knowledge-is-power hypothesis. *Cognitive Psychology*, 44, 339–387.
- 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., Bleckley, K. M., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory capacity. *Journal of Experimental Psychology: General*, 130, 169–183.
- 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, 189–217.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14(4), 389–433.
- Lee, J. F. (1986). On the use of the recall task to measure L2 reading comprehension. *Studies in Second Language Acquisition*, 8, 201–211.
- Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: a paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 27, 1424–1429.
- Mestre, J. P. (1986). The problem with problems: Hispanic students and math. *Bilingual Journal*, 32, 15–19.
- Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology*, 73, 139–158.
- Payne, T. W., Kalibatsava, Z., & Jungers, M. K. (2009). Does domain experience compensate for working memory capacity in second language reading comprehension? *Learning and Individual Differences*, 19, 119–123.
- Ransdell, S., Barbier, M. -L., & Nilt, T. (2006). Metacognitions about language skill and working memory among monolingual and bilingual college students: when does multilingualism matter? *The International Journal of Bilingual Education and Bilingualism*, 9(6), 728–741.
- Raven, J. C., Raven, J. E., & Court, J. H. (1998). *Progressive Matrices*. Oxford: Oxford Psychologists Press.
- Ricks, T., Turley-Ames, K. J., & Wiley, J. (2007). Effects of working memory capacity on mental set due to domain knowledge. *Memory and Cognition*, 35, 1456–1462.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34(2), 344–355.
- Sanchez, C. A., & Wiley, J. (2009). To scroll or not to scroll: scrolling, working memory capacity and comprehending complex text. *Human Factors*, 51(5), 730–738.

- Shamosh, N. A., DeYoung, C. G., Green, A. E., Reis, D. L., Johnson, M. R., Conway, A. R. A., Engle, R. W., Braver, T. S., & Gray, J. R. (2008). Individual differences in delay discounting, Relay to intelligence, working memory, and anterior prefrontal cortex. *Psychological Science*, *19*(9), 904–911.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127–154.
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: examining the correlation between Operation Span and Raven. *Intelligence*, *33*, 67–81.
- 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.