Creation of Spatial Mental Models with Figural Stimuli: Validation of the Emoji-based Spatial Integration Task

Polly O’Rourke (porourke@casl.umd.edu)
Nick B. Pandža (npandza@umd.edu)
Susan Campbell (susanc@umd.edu)
Center for Advanced Study of Language
University of Maryland, 7005 52nd Avenue
College Park, MD 20742 USA

Abstract

The current study examined a new spatial integration (SI) task, based on figural rather than linguistic stimuli, to measure the construct of mental modeling ability. Previous tasks conflated linguistic ability with mental modeling ability by requiring sentence processing, which may have contributed to mixed findings with respect to the relationship between mental model ability and working memory capacity (WMC). The figural spatial integration task produced the canonical continuity effect, such that discontinuous items had lower accuracy than continuous items. Furthermore, WMC and visuospatial ability predicted SI task performance, and both were stronger predictors for the continuous condition. The interactions between predictors and task conditions suggest reliance on heuristics and/or rehearsal during performance of the more difficult discontinuous items.

Keywords: Spatial integration, mental modeling, working memory capacity, spatial manipulation.

Introduction

Mental models are abstract representations of a situation, derived from a narrative or some other form of input (Ehrlich & Johnson-Laird, 1982). Successful creation of mental models contributes to logical thinking (e.g., Bell & Johnson-Laird, 1998; Evans, Handley, Harper, & Johnson-Laird, 1999) and spatial and temporal reasoning (e.g., Baguley & Payne, 1999; Carreiras & Santamaria, 1997; Roberts, 2000). It is also strongly connected to the ability to comprehend written or spoken narratives (Bower & Morrow, 1990; de Vega, 1995; Radvansky & Copeland, 2004).

The experimental task most commonly used to assess spatial mental model ability is the Spatial Integration (SI) task (Copeland & Radvansky, 2007; Radvansky & Copeland, 2004). In this task, participants are presented with a sequence of three sentences (one at a time), each describing the spatial relation of two of four objects. Immediately following this presentation, participants select from an array the picture which represents the correct spatial arrangement of the four items. There are two conditions referring to how the spatial information is presented in the learning phase: continuous and discontinuous. In the continuous condition, the second screen includes one item from the first screen and the third screen includes one item from the second, enabling the participant to incrementally construct a mental model. In the discontinuous condition, the second screen and third screen are switched such that the second screen does not contain either of the items in the first screen but the third contains one item from each of the previous screens.

The use of sentence stimuli in this task is, however, problematic. First, task performance may reflect verbal rather than mental model abilities. Second, the processing demand associated with language comprehension may obfuscate the relationship between mental model ability and key underlying cognitive factors, like working memory capacity (WMC; Conway, Kane, Bunting, Hambrick, Wilhelm, & Engle, 2005). Extant findings support the problematic nature of using sentence stimuli in this task. While several studies have found no relationship between identification accuracy in the SI task and WMC (Radvansky & Copeland, 2001; 2004; Radvansky, Gibson & McNerney, 2014), O’Rourke and Bunting (in press) found that when controlling for reading comprehension ability, WMC predicted accuracy in the discontinuous condition. They also found that when participants performed the SI task in their second language, second language proficiency alone predicted performance. As operating in L2 is widely known to absorb available WM resources, this finding and the finding for L1 indicate that variability related to language processing may obfuscate the relationship between WM and mental model creation.

Copeland and Radvansky (2007), in their study of mental model ability in aging adults, a population which generally has reduced WMC (Myerson, Emery, White, & Hale, 2003), examined both verbal and figural versions of the task. They implemented the SI task with sentence stimuli describing the spatial configurations (Experiment 1), word stimuli appearing in the relevant spatial configurations (Experiment 2), and picture stimuli appearing in the configurations (Experiment 3). Aging adults did very poorly on both continuous and discontinuous conditions in Experiment 1, with performance on the discontinuous condition not differing significantly from chance. WMC (indexed by Operation Span; Turner & Engle, 1989) predicted identification accuracy in the older participants, but not the young adult group. Performance on the continuous condition was improved in Experiments 2 and 3. Only in Experiment 3, the figural version, did aging adults perform above chance in the discontinuous condition. WMC predicted performance in both age groups in Experiments 2 and 3. Furthermore, this finding suggests that the cognitive burden of language processing absorbs working memory resources required for successful performance of the SI task.

The goal of the current study was to validate a new figural version of the SI task and to examine the relationship between task performance and working memory capacity. Another potential source of variance in this task is spatial visualization ability, which reflects the ability to represent and manipulate parts of an image (Carroll, 1993). This ability may underpin performance of the figural version in particular as stimuli can be represented visually immediately, without the step of converting word/sentence stimuli into images. Spatial visualization ability will, therefore, also be included as a predictor in the analysis.

Method

Participants

A total of 161 (96 female) participants between the ages of 18 and 39 (M = 20.32, SD = 1.67) with normal or corrected-to-normal vision were tested and retained for analysis in the current experiment. They were paid for their participation. Two additional participants were excluded for exhibiting a pattern of not following study directions across multiple tasks.

Tasks

Spatial Integration Task

The SI task (adapted from Copeland & Radvansky, 2007) tests the ability to construct a mental model of the spatial arrangement of four items. In the learning phase, participants are presented with three screens, each containing two of four objects in particular spatial arrangements (see Figures 1 and 2). Items are presented in the continuous (see Figure 1) or discontinuous condition (in which the second and third screens in Figure 1 would be switched; see Figure 2). After the learning phase, participants must select from eight diagrams the one that correctly represents the spatial arrangement of all four objects in relation to one another (see Figures 1 and 2 for correct arrangement for the example item). The three screens in the learning phase are presented for 2 seconds per screen, while the test screen remains available until the participant responds. The task stimuli consisted of 80 emoji downloaded from the Emojione database (emojione.com), representing 20 sets of four semantically related emoji (e.g., fruits, vegetables, vehicles). Each task item was composed of one of the sets of four. Each item appeared once per stimulus list. Two forms of the test were created such that items were matched across conditions; a particular set of emoji appeared in one stimulus list in the continuous condition and in the other stimulus list in the discontinuous condition. As participants must choose the correct answer from eight options, chance performance is about 12%.

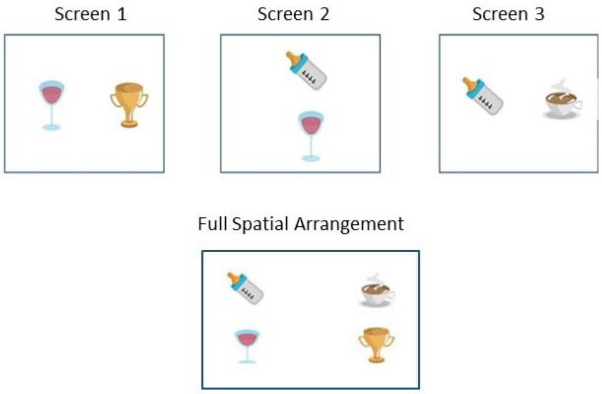

Figure 1. Example of continuous item from the SI task with the full spatial arrangement of the four items. Participants can build a partial model immediately.

Figure 2. Example of corresponding discontinuous item from the SI task with the full spatial arrangement of the four items. Participants must wait to integrate the model based on information that comes on the third screen.

Shapebuilder task

The Shapebuilder (SB) task is a complex visuospatial working memory measure (Atkins et al., 2014). In this task, participants are shown a series of shapes in a 4 x 4 grid, and they must recall the shape, color, and location of the series of the shapes in the correct presentation order. There are 26 items and the number of shapes in each item's sequence increases over the course of the task from two shapes (six items), to three shapes (nine items), and finally to four shapes (11 items). Points were earned for each shape in the sequence for which the location, shape, and color were correctly recalled. Partial credit was awarded if the location was correctly recalled. More points were awarded for longer sequences. See Atkins et al. (2014) for full scoring parameters and for WMC factor loadings alongside other working memory measures.

Paper Folding task

The Paper Folding (PF) task used in the current study to measure spatial visualization ability was a computerized test adapted from the ETS Kit of Factor-Referenced Tests (Ekstrom et al., 1976). Two forms were created with 20 items each. Items were ordered by increasing difficulty. As participants must choose the correct answer from five options, chance performance is 20%.

Procedure

The tasks pertinent to this study were administered as part of a larger battery of 18 behavioral tasks and surveys. The 18-task battery was administered in two sessions of three hours each with the opportunity for breaks between each task. Testing took place in a classroom-style computer lab. Written consent was obtained at the beginning of the first testing session. SI and PF were administered in session one and SB was administered in session two.

Results

Data from six participants in the SB task is missing due to study attrition as they did not return for the second session. As a result, the sample size for analyses including SB is 155. We ran Wilcoxon Signed Rank tests comparing forms, and found no significant differences; therefore forms for SI and PF were collapsed in the correlational analysis. See Table 1 and 2 for descriptive statistics and correlations among measures, respectively.

Table 1. Task performance – Descriptive statistics

Task	Mean	*SD*	Min	Max
SI_Con	.67	.26	.00	1.00
SI_Discon	.52	.22	.00	1.00
SB	1711.52	466.53	755.00	2325.00
PF	.74	.18	.15	1.00

Table 2. Correlations among measures. Numbers on the diagonal reflect average internal consistency.

Task	SI_Con	SI_Discon	SB	PF
SI_Con	(.75)			
SI_Discon	.50*	(.56)		
SB	.44*	.33*	(.73)	
PF	.60*	.37*	.43*	(.86)

* $p < .001$

We conducted a logistic multilevel model (MLM, or mixed-effects model) on the binary individual trial-level accuracy data in order to generalize across participants and items and account for the fact that particular items were present in both conditions (Baayen et al., 2008; Linck & Cunnings, 2015). Condition (continuous vs. discontinuous) was included as a fixed effect, nested within-item, as each item appeared in both conditions across the two forms.

The model predicts estimated log-odds (*b*) of a correct response on the SI task, from which we can derive the change in odds and probability of accurate performance on the task. The independent variables included to explain variance in subject and item performance were Condition (Continuous, discontinuous), z-scored SB and PF, and the two-way interactions of Condition with SB and PF. Results of this modeling procedure are shown in Table 3, with the model baseline being the discontinuous condition.

Table 3. Logistic MLM results for SI item accuracy

Fixed Effects	**Estimate (*b*)**	**Odds (exp(*b*))**	***SE***	***p*-value**
Intercept	0.10	1.11	0.13	.43
Continuous	0.18	2.17	0.08	<.001*
SB	0.22	1.25	0.09	.01*
PF	0.30	1.35	0.09	<.001*
Con × SB	0.16	1.18	0.10	.09^
Con × PF	0.42	1.53	0.09	<.001*
Random Effects		**Variance**	***SD***	
Intercepts | Subject		0.47	0.69	
Intercepts | Item		0.22	0.47	

* $p < .05$, ^ $p < .10$

The results show a main effect of Condition, confirming the effect of continuity for the SI task while controlling for the effects of SB and PF.

The results further show visuospatial WMC and spatial visualization (as measured by the SB and PF tasks, respectively) contribute independent variance to SI accuracy on both discontinuous and continuous items. Since SB and PF are z-scored and on the same scale, the sizes of the estimates can be directly compared. The effect for PF is slightly stronger than for SB on discontinuous items ($b_{PF\,=}\ 0.30 > b_{SB\,=}\ 0.22$) and the effect for PF is almost twice the size of the effect for SB on continuous items ($b_{PF} + b_{Con×PF\,=}\ 0.72 > b_{SB} + b_{Con×SB\,=}\ 0.38$).

The effect for SB is positive: as SB scores increase, so do the odds of a correct response on discontinuous items on SI. There is a marginal interaction of continuous × SB, suggesting that the effect of SB may be even stronger for continuous items than discontinuous (see Figure 3).

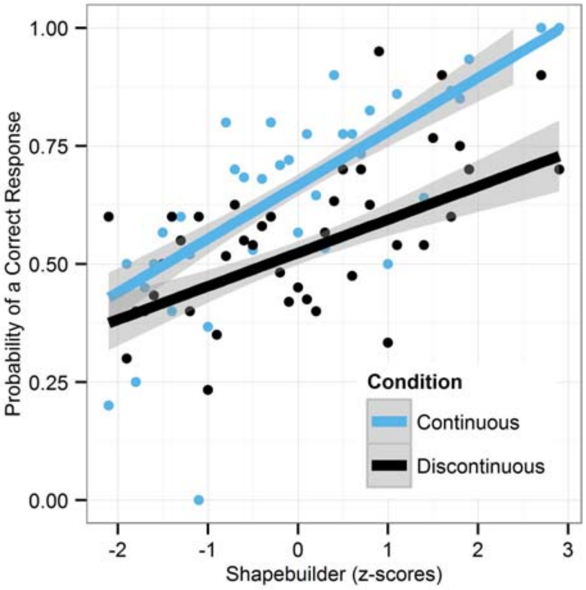

Figure 3. Depiction of SI accuracy regressed on Shapebuilder z-scores split by Condition. Shaded area around line represents 1 SE.

The effect for PF is also positive: as PF scores increase, so do the odds of a correct response on discontinuous SI items. Finally, there is a significant interaction of continuous × PF, indicating that, for each standard deviation increase on PF performance in our sample, the odds of a correct response are higher on continuous items than discontinuous items on the SI task (see Figure 4).

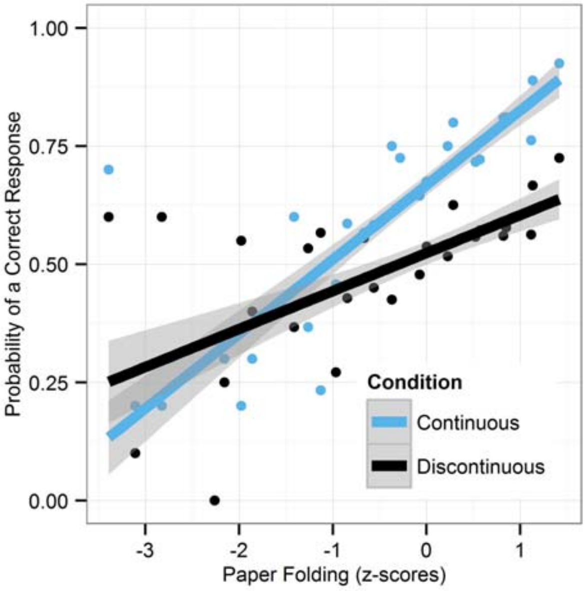

Figure 4. Depiction of SI accuracy regressed on Paper Folding z-scores split by Condition. Shaded area around line represents 1 SE.

Discussion

The results show that young adults perform similarly on this figural, emoji-based version of the SI task to previously tested text-based versions of the task. Our mean accuracies of 67% for continuous items and 52% for discontinuous (both of which are far above chance performance of 12%) are consistent with Copeland & Radvansky (2007)'s findings using the standard, sentence based task (68% for continuous condition and 47% for discontinuous), and their figure task (73% for continuous and 53% for discontinuous). The fact that accuracy levels and continuity effects for our figural version of SI mirror extant findings provides evidence that this new version of the task performs similarly to the standard task. One limitation of this study is that we did not compare performance on our figure version to a standard text based version of the SI task.

Our emoji-based SI task extends the findings of Copeland and Radvansky (2007) in regards to the utility of a non-linguistic SI task and has several advantages over their instantiation. In our version of the SI task, the learning phase for each item was experimenter-paced such that participants saw each of the three screens for two seconds. When examined by Copeland, Radvansky and colleagues (Copeland & Radvansky, 2007; Radvansky & Copeland, 2001; 2004; Radvansky et al., 2014) the learning phase was self-paced such that participants had as long as they wanted for each training screen and reading times for each screen were dependent variables. The two-second time limit for the present task was ased on pilot data such that it represented a window within which most participants advanced to the next screen. While accuracy was similar to previous results with a self-paced learning phase (Copeland & Radvansky, 2007), our experimenter-paced version of the task is easier to administer remotely, without a proctor, due to a more consistent task duration.

Another key methodological difference is that our SI task included pictures of real-world objects (e.g., coffee, glass) across different semantic sets (e.g., vegetables, vehicles), allowing greater generalizability of the items than simple geometric shapes (e.g., red square, green star). The more complex nature of the images could also have led to a lower probability of verbal rehearsal strategies, especially since the same four colors were used in every trial in the Copeland and Radvansky (2007) task; however, future research will be needed to test this claim.

Finally, the current study had greater power than Copeland and Radvansky (2007), in that there were more trials (20 versus 8 total, 10 versus 4 by condition), more participants (161 versus 60, the latter split between two groups), and all trials were used for more powerful statistical analyses thanks to a multilevel model design.

The examination of the cognitive underpinnings of mental model ability showed that both WMC and spatial visualization ability predict performance on the SI task. Spatial visualization ability emerged as a slightly stronger effect than WMC. Interestingly, both predictors accounted for more variance in the Continuous condition than the discontinuous condition.

In contrast to other studies using text based versions of the task (Copeland & Radvansky, 2007, Exp. 1; Radvansky & Copeland, 2001; 2004; Radvansky et al., 2014), the current study found evidence for WMC as a predictor of performance on the SI task in young adults. As previously noted, lack of effects using the sentence-based SI task could be due in part to the processing demands associated with converting linguistic representations into spatial representations. Furthermore, the interim spatial representations must be maintained while the next sentence is parsed into spatial information. While in figural form, the SI task is demanding on WM resources, eliminating the sentence as a conveyor of spatial information may have resulted in an increase in resources available for mental model creation.

Another possible reason for lack of effect in previous studies, particularly for the discontinuous condition, is that WMC may be a weaker predictor in the discontinuous condition, as shown by the marginal two-way interaction in the current study. This may be due to the "choke" factor whereby high WMC individuals start performing like low WMC individuals when they are under pressure (Sattizahn, Moser, & Beilock, 2016; Wang & Shah, 2014), such that WMC no longer predicts performance.

Strategy use may be another factor reducing the effect of individual WMC on performance. Wang and Shah (2014) note that when heuristics are available, people with high and low WM spans may perform similarly. It may be that all participants develop strategies in order to reduce the cognitive effort (Shah & Oppenheimer, 2008) involved in determining the correct response in the discontinuous condition and, therefore, WMC would no longer predict performance. For example, in the discontinuous condition, after seeing the third screen in the learning phase, a participant may only partially incorporate the spatial arrangement such that he/she knows the positioning of two of the four images (e.g., top two images in the square). This information, though incomplete, may be enough to select the correct answer in the test phase. It may be possible in future iterations of the task to reduce the utility of heuristics via changes to the design. Specifically, the options in the test phase could be modified such that strategy use would be less likely to lead to a correct response.

Another possibility is that participants were more likely to engage in rehearsal in the discontinuous task. Rehearsal is a means of maintaining information in a short-term memory store, without any WM or executive involvement (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002). As such, if participants were more likely to use rehearsal in order to remember the spatial configurations in the discontinuous condition, then WMC would be a less effective predictor of performance. There are many strategies for preventing rehearsal during cognitive task performance (Cowan, 2008). One example is adding a secondary processing task (e.g., counting backwards) for participants to perform during the period in which information needs to be retained. Given that the retention interval is fairly brief in this task, it is not clear that a secondary processing task would be effective. Another option might be further reducing the time for which the screens in the learning phase are presented to a less comfortable pace. This adjustment would likely have consequences, however, for the WMC demand. Additional testing is necessary to determine how researchers can prevent rehearsal in this task.

While WMC accounted for a significant amount of variability in SI performance, spatial visualization ability emerged as a stronger predictor. It is, perhaps, unsurprising that spatial visualization ability would predict performance on a spatial reasoning task like SI, particularly our figural version. Spatial visualization ability interacted with task condition such that its utility as a predictor was better in the continuous condition. This pattern is, of course, similar to the pattern observed for WMC, but with spatial visualization ability the effect was significant.

This finding supports the account that in the discontinuous condition, participants were more likely to not create true spatial mental models but rather to use heuristics or rehearsal in order to determine the correct answer at test. The case for rehearsal is particularly strong in that reduced role of spatial visualization ability in the discontinuous condition suggests that participants may not be creating visual representations. If that is the case, then verbal rehearsal is one way to perform the task. While the example of a strategy described above requires some level of visuospatial representation, there may be strategies, other than rehearsal, that do not.

Hitherto unexamined effects of individual variability in spatial visualization ability may have been another factor contributing to the mixed findings in the literature with respect to the contribution of WMC. In the previous text-based versions of the SI task, perhaps individuals with poorer spatial visualization ability had more difficulty transitioning from text-based representations to full visual representations, regardless of WMC, and therefore were unable to create a spatial mental model of the four items.

In conclusion, the current study validated a figure version of the SI task such that results from this task show performance levels and continuity effects consistent with previous studies. Given that this task is not sensitive to individual variability in language processing, or even native language, it can be used as a more pure measure of spatial mental model ability. This conclusion is supported by the finding that WMC predicted task performance in the figure version of the task, and language-related variability may have obscured this relationship in previous studies. Furthermore, we present evidence that spatial visualization ability is a significant predictor of task performance, and that while both WMC and spatial visualization ability predicted performance in both conditions, there was evidence suggesting the effect was stronger in the continuous condition. Future research will determine the source of this interaction.

References

Atkins, S. M., Sprenger, A. M., Colflesh, G. J., Briner, T. L., Buchanan, J. B., Chavis, S. E., ... & Harbison, J. I. (2014). Measuring Working Memory Is All Fun and Games. *Experimental Psychology*, 61, 417-438.

Baayen, H., Davidson, D., and Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *J. Mem. Lang.* 59, 390–412. doi: 10.1016/j.jml.2007.12.005

Baguley, T., & Payne, S.J. (1999). Recognition memory for sentences from spatial descriptions: A test of the episodic construction trace hypothesis. *Memory & Cognition*, 27, 962-973.

Bell, V.A., & Johnson-Laird, P.N. (1998). A model theory of modal reasoning. *Cognitive Science*, 22, 25-51.

Bower, G.H., & Morrow, D.G. (1990). Mental models in narrative comprehension. *Science*, 247, 44-48.

Carreiras, M., & Santamaria, C. (1997). Reasoning about relations: Spatial and nonspatial problems. *Thinking and Reasoning*, 3, 191-208.

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

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 & Review, 12(5), 769-786.

Conway, A. R., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*(2), 163-183.

de Vega, M. (1995). Backward updating of mental models during continuous reading of narratives. Journal of Experimental Psychology: *Learning, Memory & Cognition*, 21, 373-385.

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

Ehrlich, K., & Johnson-Laird, P. N. (1982). Spatial descriptions and referential continuity. *Journal of Verbal Learning and Verbal Behavior*, 21(3), 296-306.

Evans, J.St.B.T., Handley, S.J., Harper, C.N.J., & Johnson-Laird, P.N. (1999). Reasoning about necessity and possibility: A test of the mental model theory of deduction. Journal of *Experimental Psychology: Learning, Memory, & Cognition*, 25, 1495-1513.

Linck, J. A., and Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, 65, 185–207. doi: 10.1111/lang.12117

Myerson, J., Emery, L., White, D. A., & Hale, S. (2003). Effects of age, domain, and processing demands on memory span: Evidence for differential decline. *Aging, Neuropsychology, and Cognition*, *10*(1), 20-27.

O'Rourke, P. & Bunting, M. (in press). The Cognitive Underpinnings of Mental Model Construction in L1 and L2. *Quarterly Journal of Experimental Psychology*.

Radvansky, G.A., & Copeland, D.E. (2001). Working memory and situation model updating. *Memory & Cognition*, 28, 1073-1080.

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *The American journal of psychology*, 191-213.

Radvansky, G.A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review*, 14, 1027-1042.

Radvansky, G. A., Gibson, B. S., & MCNERNEY, M. W. (2014). Working memory, situation models, and synesthesia. *The American journal of psychology*, *127*(3), 325-342.

Roberts, M.J. (2000). Strategies in relational inference. *Thinking and Reasoning*, 6, 1-26.

Sattizahn, J. R., Moser, J. S., & Beilock, S. L. (2016). A Closer Look at Who "Chokes Under Pressure". *Journal of Applied Research in Memory and Cognition*, 5(4), 470-477.

Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: an effort-reduction framework. *Psychological Bulletin*, 134(2), 207-222.

Turner, M.L. & Engle, R.W. (1989). Is working memory capacity task independent? Journal of Memory and Language, 28, 127-154.

Wang, Z., & Shah, P. (2014). The effect of pressure on high-and low-working-memory students: An elaboration of the choking under pressure hypothesis. *British Journal of Educational Psychology*, *84*(2), 226-238.

Acknowledgements

This research is based upon work supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Maryland, College Park and/or any agency or entity of the United States Government. Nothing in this article is intended to be and shall not be treated or construed as an endorsement or recommendation by the University of Maryland, United States Government, or the author of the product, process, or service that is the subject of this article. Correspondence concerning this article should be addressed to Polly O'Rourke, Center for Advanced Study of Language, University of Maryland, 7005 52nd Avenue, College Park, MD 20742 (e-mail: porourke@casl.umd.edu). Special thanks to Meredith Mislevy Hughes and Valerie Karuzis.