BRIEF REPORT

Is working memory capacity a causal factor in fluid intelligence?

Alexander P. Burgoyne[1] · **David Z. Hambrick**[1] · **Erik M. Altmann**[1]

Abstract

It is well established that measures of reasoning ability and of working memory capacity (WMC) correlate positively. However, the question of what explains this relationship remains open. The purpose of this study was to investigate the *capacity hypothesis,* which ascribes causality to WMC. This hypothesis holds that people high in WMC are more successful in capacity-demanding cognitive tasks than people lower in WMC because they can temporarily maintain more information in the form of sub-goals, hypotheses, and partial solutions. Accordingly, this hypothesis predicts that the correlation between WMC and reasoning performance should increase as the capacity demands of the reasoning items increase. We tested this prediction using items from Raven's Progressive Matrices and two measures of WMC, complex span and the *k* estimate from the Visual Arrays task. Neither WMC measure showed the effect predicted by the capacity hypothesis. Furthermore, the results cannot be attributed to restriction of range in performance on the individual reasoning items. This finding adds to existing evidence calling into question the capacity hypothesis, and, more generally, the view that WMC has a causal influence on fluid intelligence.

Keywords Fluid intelligence · Reasoning ability · Raven's matrices · Working memory capacity

Introduction

Research has established that scores on tests of *fluid intelligence* (Gf), such as Raven's Progressive Matrices, correlate positively and moderately with estimates of *working memory capacity* (WMC) from tasks requiring simultaneous storage and processing of information (see Conway & Kovacs, 2013, for a review). However, the question of what, exactly, accounts for this correlation remains open. One possibility is that the causal arrow goes from WMC to Gf. More specifically, the ability to reason may be constrained by the amount of information a person can temporarily hold in an active state. As Unsworth, Fukuda, Awh, and Vogel (2014) stated:

> "Individuals with large capacities can simultaneously maintain more information in WM than individuals with smaller capacities. In terms of gF, this means that high capacity individuals can simultaneously attend to multiple goals, sub-goals, hypotheses, and partial solutions

for problems which they are working on allowing them to better solve the problem than low capacity individuals who cannot maintain/store as much information" (p. 3).

While this *capacity hypothesis* of the WMC-Gf correlation remains popular, nearly all measures of cognitive ability correlate positively and moderately with each other. This finding of "positive manifold" among measures of cognitive ability is, in fact, one of the most replicated findings in psychology (see Jensen, 1998). Thus, by itself, the finding that measures of WMC correlate with measures of Gf is insufficient to establish causation; additional tests of the hypothesis are required.

One such test is to evaluate whether the correlation between WMC and Gf increases as a function of the presumed information-processing demands of reasoning items. The capacity hypothesis implies that items that draw more heavily on WMC should be better at discriminating between high- and low-capacity reasoners. Accordingly, one approach to testing the capacity hypothesis is to analyze different items on tests of Gf in terms of differences in the extent to which they load WMC.

Carpenter, Just, and Shell (1990) developed one measure of the extent to which different items in Raven's Progressive Matrices load WMC. The measure is the number of *rule tokens* required to solve an item. The concept of a rule token is illustrated in Fig. 1 with two Raven's-like items. In the item on the left, only a single rule token is required to solve the

✉ Alexander P. Burgoyne
  burgoyn4@msu.edu

[1] Department of Psychology, Michigan State University, East
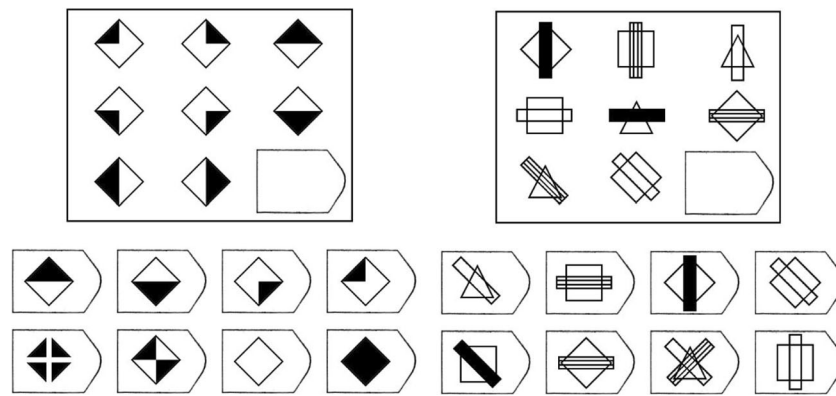 Lansing, MI 48824, USA

Fig. 1 Two Raven's-like items adapted from Carpenter et al. (1990). In the left panel, a single rule token is required to solve the problem; in the right panel, three rule tokens are required

problem: *addition* — an element from one column is added to another to produce the third. By contrast, in the item on the right, three rule tokens are required: (1) *distribution of three shapes* — each row contains one diamond, one square, and one triangle; (2) *distribution of three textures of the line* — each row has one dark line, one striped line, and one clear line; and (3) *constant-in-a-row* — the orientation of the line is the same within a row. Thus, rule tokens are patterns characterizing the relations between the figural elements of a Raven's item. In theory, as the number of rule tokens for items increases, the number of sub-goals, hypotheses, and partial solutions that the test taker must explore and hold in mind should also increase. Accordingly, for a given item, the number of rule tokens can be interpreted as an index of capacity demand.

This study builds on previous work that examined Raven's performance at the item level. Using an analytical approach introduced by Salthouse (1993) and Carpenter et al.'s (1990) classification of Raven's items, Unsworth and Engle (2005) tested whether capacity accounts for the WMC-Gf relationship. Participants completed a working-memory span task (operation span) to measure WMC and Raven's Progressive Matrices to measure Gf. Unsworth and Engle then examined whether the point-biserial correlation (r_{pb}) between WMC and Raven's item solution accuracy (i.e., incorrect or correct) increased as the number of rule tokens increased. Finding no support for this hypothesis, they concluded that the WMC-Gf relationship is not attributable to individual differences in capacity. As Unsworth and Engle (2005) stated, "The results of the present study strongly suggest that the number of goals or sub-results that can be held in memory does not account for the shared variance between working memory span measures and fluid intelligence" (p. 78).

Support for the capacity hypothesis has remained elusive. For instance, Wiley, Jarosz, Cushen, and Colflesh (2011) examined item-level correlations between WMC and Raven's accuracy, but found that the correlations did not increase as the number of rule tokens increased. Rather, they found that the WMC-Gf correlation was stronger for items that required

new combinations of rules. Wiley et al. (2011) suggested that these items place greater demands on executive functions (e.g., resistance to proactive interference from previous items), resulting in stronger WMC-Gf correlations. However, this finding has not been replicated (Harrison, Shipstead, & Engle, 2015; Little, Lewandowsky, & Craig, 2014).

Little et al. (2014) carried out the only study we know of to observe an increasing WMC-Gf correlation as the capacity demands of the Raven's items increased. Little et al. argued that the key difference between this study and previous studies was the strength of the overall WMC-Gf correlation in their sample ($r = .56$, compared to $r = .34$ in Unsworth & Engle (2005) and $r = .33$ in Wiley et al. (2011)), along with ceiling effects (i.e., very high accuracy rates) on early Raven's items. As Little et al. (2014) explained:

> "As the overall correlation between WMC and Raven's performance increases, the item-specific correlations can no longer be constant but must also increase across item difficulty. This is a necessary consequence of near ceiling performance on the early items which declines as the items become more difficult combined with a high overall correlation between Raven's and WMC" (p. 4).

This explanation suggests that restriction of range produced the pattern of results Little et al. (2014) observed, because it attenuated the WMC-Gf correlation for early items, leaving the later items' correlations stronger by comparison. Therefore, in the present study we examine and correct for the effects of restriction of range.

It should also be noted that working memory span tasks are not "process-pure" measures of capacity. That is, these measures capture more than just the amount of information a participant can temporarily maintain in an active state. For instance, in operation span, the participant must solve a series of arithmetic equations while attempting to remember a letter following each for later recall. Thus, operation span may be influenced not only by how many letters participants can hold in memory, but also by

their arithmetic skill and use of mnemonic strategies for remembering the letters, among other factors.

Arguably a more direct measure of capacity is the k estimate from the widely administered *visual arrays* task (Luck & Vogel, 1997). In this task, participants are shown two arrays of colored squares, one after another, with a brief delay in between. The second array is either identical to the first array, or the color of one square is different. Participants must determine whether the two arrays are the same or different. k reflects the number of colored squares (units of information) a participant can remember.

In the present study, we tested the capacity hypothesis by having participants complete the visual arrays task, two working memory span tasks, and Raven's Matrices. The empirical question is whether the different measures of WMC will converge on a null relationship between WMC and Gf, supporting and extending Unsworth and Engle's (2005) finding, or whether k and complex span will dissociate, perhaps because complex span is not process pure as an assessment of capacity.

Method

Participants

The participants were undergraduate students. In total, 311 participants contributed data to the study. Nearly all the participants were between the ages of 18 and 25 years; approximately 75% were female.

Procedure

Listed in order of administration, participants completed the following cognitive ability tests. Participants were given a short break between tests.

Visual arrays This task was modeled after Luck and Vogel's (1997) whole-display no-load task. As depicted in Fig. 2, a memory array of two to eight colored squares was displayed for 200 ms, followed by a blank display for 900 ms. This was followed by a test array, which was either identical to the memory array, or differed by the color of one of the squares. The test array was displayed until the participant used the keyboard to indicate whether the memory array and test array were the same or different. Following the participant's response, the next trial began.

Participants completed four practice trials and 80 test trials. Later test trials had larger set sizes. *Set size* refers to the number of squares in an array. There were 12 trials at each of set sizes 2, 3, 4, 5, and 6, and ten trials at each of set sizes 7 and 8. There were an equal number of *same* trials and *different* trials at each set size.

We used two scoring methods. The first calculated the percentage of correct responses (i.e., "visual arrays accuracy"); a participant who guessed randomly would earn a score of approximately 50%. The second method used Pashler's (1988) formula for estimating capacity:

$$k = N\left(\frac{h-f}{1-f}\right)$$

In this formula, k represents capacity, N is the relevant set size, h is the hit rate, and f is the false-alarm rate. The hit rate is computed as follows:

$$h = d + (1-d)f$$

Here, d represents the probability that the participant detects a change on *different* trials. The false alarm rate represents the proportion of trials in which a participant responded "different" on *same* trials. Participants with a false alarm rate of 100% were given an estimated capacity of zero for that set size. A separate estimate of capacity was computed for set sizes 2–8.

Operation span Participants must solve math equations, and remember a letter that follows each equation (Unsworth et al., 2005). After a series of equation-letter trials, participants must recall the letters in the order in which they were presented. There were three blocks of five sets of equation-letter trials. The measure was the number of letters recalled in the correct order.
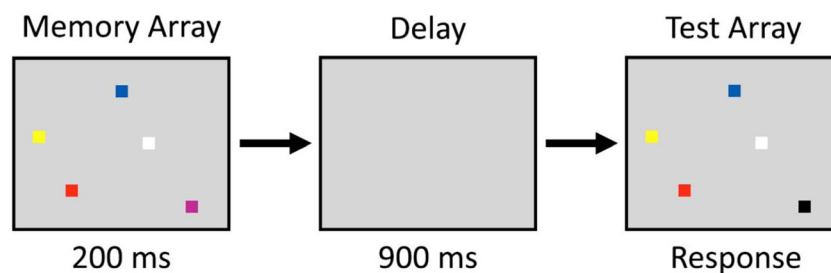

Fig. 2 Illustration of the visual arrays task. An array of colored squares is presented for 200 ms, followed by a 900-ms blank delay interval. Participants are then presented with a test array, and use the keyboard to indicate whether the test array and the memory array are the same or different

Symmetry span Participants must make symmetry judgements about abstract patterns, and remember the location of a red square that appears after each pattern (Oswald et al., 2015). After a series of symmetry-square trials, participants must recall the location of the red squares in the order in which they were presented. There were 12 sets of pattern-square trials. The measure was the number of square locations recalled in the correct order.

Raven's advanced progressive matrices Participants are presented with a set of patterns arranged in a 3 × 3 formation. The pattern in the lower right is missing, and participants must choose the alternative that best completes the set. Participants completed the 18 odd-numbered items from Raven's Advanced Progressive Matrices (Raven & Raven, 1998). The time limit was 10 min; the measure was the number correct.

Number series Participants are presented with a series of numbers following a pattern. Participants must select from four alternatives the number that logically completes the pattern. Participants completed 15 items from the test of primary mental abilities (Thurstone, 1938). The time limit was 4.5 min; the measure was the number correct.

Letter sets Participants are presented with five sets of four letters (e.g., ABCD) arranged in a row. Participants must choose the set that does not follow the same pattern as the other four. Participants completed 20 items from the ETS Kit of Factor-Referenced Tests (Ekstrom, French, Harmon, & Derman, 1976). The time limit was 5 min; the measure was the number correct.

Data screening

Six participants did not complete the visual arrays task and were excluded from analysis. Of the remaining 305 participants, 49 participants demonstrated chance-level performance or worse on visual arrays (visual arrays accuracy ≤ 50%) and were excluded. This left a usable sample size of 256 participants. There were no values more than 3.5 SDs from sample means.

Table 1 Descriptive statistics for cognitive ability measures

Measure	N	M	SD	Range	Skew	Kurtosis	α
Visual arrays (k)	256	3.28	0.75	.91 − 5.00	-0.53	0.12	.65
Operation span	255	36.36	18.34	0 − 75	0.11	-0.66	.81
Symmetry span	255	17.62	9.11	0 − 42	0.18	-0.53	.68
Raven's matrices	256	8.32	3.28	0 − 16	-0.04	-0.72	.69
Number series	256	8.43	2.47	2 − 15	0.05	-0.14	.68
Letter sets	255	9.67	2.88	2 − 18	0.11	-0.02	.66

Table 2 Correlation matrix

Measures	1.	2.	3.	4.	5.	6.
1. Visual arrays (k)	-					
2. Operation span	.04	-				
3. Symmetry span	.22	.48	-			
4. Raven's matrices	.26	.16	.25	-		
5. Number series	.27	.24	.26	.29	-	
6. Letter sets	.20	.27	.23	.29	.33	-

Note. Listwise n = 253. Correlations greater than .12 are statistically significant at $p < .05$

Measures of capacity

We averaged k estimates at set sizes 2–8 for each participant and refer to this variable as k. We also created a composite variable representing performance on working memory span tasks by averaging standardized (z) scores on operation span and symmetry span.

Classification of Raven's items

As our operational definition of capacity demand, we categorized Raven's items according to the number of rule tokens required to solve each problem using Carpenter et al.'s (1990)

Table 3 Descriptive statistics and k-solution accuracy correlation for Raven's items

Raven's item	Rule tokens	N	M	SD	r_{pb}
1	3	256	.75	.44	.13
3	2	256	.70	.46	.12
5	2	256	.83	.37	.13
7	1	256	.56	.50	.10
9	2	256	.71	.45	.16
11	1	256	.79	.41	.25
13	3	256	.36	.48	-.03
15	2	256	.65	.48	.07
17	2	256	.58	.49	.18
21	4	256	.33	.47	.21
23	4	255	.33	.47	.20
25	3	255	.33	.47	.07
27	2	252	.25	.44	.09
29	3	248	.15	.36	-.01
31	4	236	.19	.40	.09
33	2	224	.21	.41	.06
35	4	203	.15	.36	-.03

Note. Items 11, 15, 21, and 25 were not classified by Carpenter et al. (1990) because they contained small grids or crosshatching that could not be properly displayed on the monitors they had available to them at the time. We classified these items using Carpenter et al.'s (1990) framework. For each item, incorrect = 0 and correct = 1.

Table 4 Average *k*-solution accuracy correlation grouped by number of rule tokens

Number of rule tokens	Mean *n*	Accuracy *M*	Accuracy *SD*	Mean *k*-accuracy r_{pb}	*p* value for *k*-accuracy r_{pb}
1	256	.67	.36	.18	.004
2	251	.57	.24	.12	.058
3	254	.40	.25	.04	.526
4	238	.28	.29	.13	.045

Note. Accuracy *M* and *SD* are reported for the full sample (*n* = 256)

analysis. Of the 18 items, two required only one rule token, seven items required two rule tokens, four items required three rule tokens, and four items required four rule tokens. (Item #19 could not be classified using Carpenter et al.'s (1990) framework and was excluded from analysis.) Because Raven's is a timed test, some items were not attempted. Unattempted items were excluded from analysis.

Results

Descriptive statistics are presented in Table 1. *k* varied substantially across participants (*M* = 3.28, *SD* = 0.75), as did performance on Raven's (*M* = 8.32, *SD* = 3.28) and the other reasoning tests. Correlations are presented in Table 2. Correlations among the Gf measures (avg. *r* = .30), and between the WMC measures and the Gf measures (avg. *r* = .24), were in the expected range (e.g., Ackerman, Beier, & Boyle, 2005).

Visual arrays (*k*) and Raven's accuracy

To reiterate, the capacity hypothesis of the WMC-Gf relationship predicts that the correlation between WMC and solution accuracy

on Raven's items should increase as a function of the capacity demands of the items (i.e., number of rule tokens). To test this hypothesis, for each Raven's item, we first computed the point-biserial correlation (r_{pb}) between *k* and solution accuracy (i.e., incorrect = 0 or correct = 1). The results are shown in Table 3.

Next, we grouped Raven's items by number of rule tokens, and calculated the sample size-weighted average correlation between *k* and solution accuracy for each group of items (see Table 4 and Fig. 3). We tested for differences between all possible pairs of dependent correlations using Steiger's (1980) formula. None of the correlations differed significantly from one another (all *p*s > .05). The largest difference was one rule token (*r* = .18) versus three rule tokens (*r* = .04), *z* = 1.877, *p* = .060, but this difference was in the direction opposite to that predicted by the capacity hypothesis. Thus, contrary to the capacity hypothesis, the average correlation between *k* and Raven's accuracy did not increase as a function of the capacity demands of the Raven's items.

However, range restriction may have attenuated the *k*-accuracy correlation for items with more rule tokens. *SD*s tended to be smaller for items with two to four rule tokens than for items with one rule token (see Table 4). Therefore, we corrected each of the correlations in Table 4 for explicit range restriction using Pearson's (1903) formula (see, e.g., Wiberg & Sundström, 2009). The "unrestricted" standard deviation for each group of rule tokens was set to .36, the same value obtained for items requiring one rule token. The corrected *k*-accuracy correlations were r_c = .18 for items requiring one rule token, r_c = .18 for two rule tokens, r_c = .06 for three rule tokens, and r_c = .16 for four rule tokens. None of these correlations are significantly different to one another (all *p*s > .10).

We also examined the relationship between number of rule tokens and solution accuracy, to confirm that number of rule tokens is related to the difficulty of the items. Items requiring more rule tokens did, in fact, have lower average accuracy rates, $F(3, 765) = 145.67$, $p < .001$. A polynomial trend
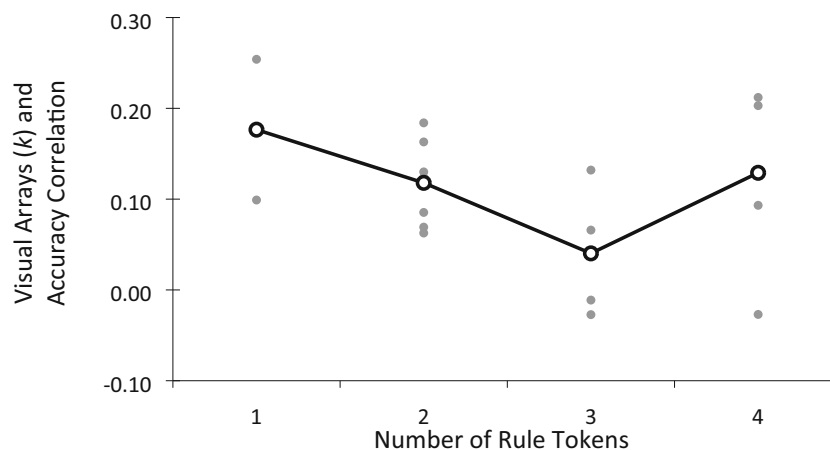

Fig. 3 The open circles (connected by the solid line) represent the average point-biserial correlations between visual arrays (*k*) and solution accuracy for Raven's items, grouped by number of rule tokens. The filled circles represent individual point-biserial correlations between visual arrays (*k*) and accuracy on each Raven's item

Table 5 Average correlation between working memory span composite and solution accuracy for Raven's items by the number of required rule tokens

Number of rule tokens	Mean n	Mean working memory span-accuracy r_{pb}	p value for working memory span-accuracy r_{pb}
1	256	.17	.006
2	251	.09	.155
3	254	.08	.204
4	238	.06	.357

analysis indicated a large linear relationship, $F(1, 255) = 321.70$, $p < .001$, partial $\eta^2 = .56$, and a small cubic relationship, $F(1, 255) = 5.47$, $p = .020$, partial $\eta^2 = .02$. The large linear relationship in particular weighs against an interpretation of our results which holds that number of rule tokens failed to moderate the relationship between k and solution accuracy simply because number of rule tokens bears no relationship to performance.

Working memory span tasks and Raven's accuracy

Finally, to replicate Unsworth and Engle's (2005) findings, we tested whether the correlation between working memory span performance and accuracy on Raven's items differed as a function of the number of rule tokens. Consistent with the preceding results, none of the correlations differed significantly from one another (all ps > .12; see Table 5 and Fig. 4). In fact, the trend in correlations across number of rule tokens is the opposite of what is predicted by the capacity hypothesis.

We then corrected each of the correlations in Table 5 for restriction of range using the same approach as before. Setting the "unrestricted" standard deviation to .36 for all rule token groups resulted in corrected working memory span-accuracy

correlations of $r_c = .17$ for items requiring one rule token, $r_c = .13$ for two rule tokens, $r_c = .11$ for three rule tokens, and $r_c = .07$ for four rule tokens. None of these correlations differ significantly from one another (all ps > .16).

Discussion

That there is a positive relationship between measures of WMC and of Gf is beyond dispute, but the question of what accounts for this relationship remains open. According to the capacity hypothesis (Unsworth et al., 2014), having a higher level of WMC facilitates problem solving by allowing individuals to simultaneously maintain more sub-goals, hypotheses, and partial solutions in an active state. This hypothesis leads to the prediction that the correlation between WMC and Gf should increase as a function of the capacity demands of the Gf items.

We tested this prediction using items from Raven's Progressive Matrices, which can be classified according to the number of rule tokens that are required to solve each item. Items that required more rule tokens had lower accuracy rates, supporting the argument that the number of rule tokens indicates the demands of the reasoning items. Contrary to the capacity hypothesis, however, the relationship between an estimate of WMC from the visual arrays paradigm (k) and solution accuracy on Raven's items did not increase with number of rule tokens. Furthermore, replicating the results of Unsworth and Engle (2005), the relationship between performance on working memory span tasks and accuracy on Raven's items did not differ as a function of capacity demands, either. Thus, our results suggest that no matter how WMC is measured, it is not a causal factor underlying variation in Gf.

If capacity does not account for the WMC-Gf correlation, what does? There are a number of possibilities. One is the ability to control attention (Engle, 2018): individuals who
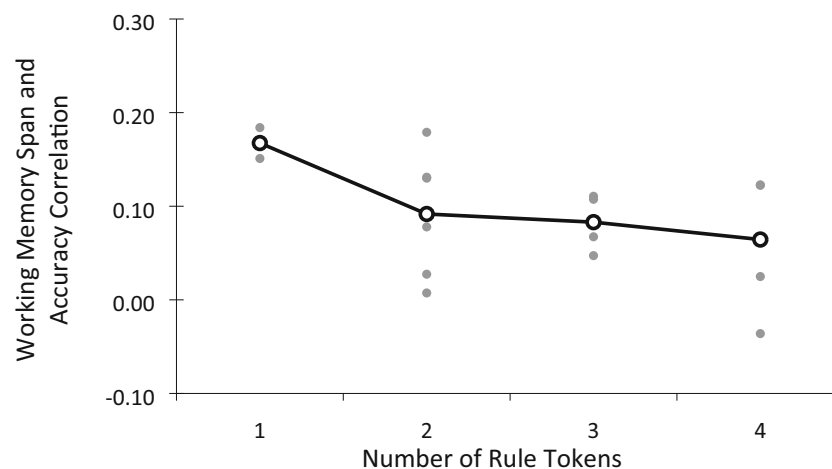

Fig. 4 The open circles (connected by the solid line) represent the average point-biserial correlations between the working memory span composite and solution accuracy for Raven's items, grouped by the

number of required rule tokens. The filled circles represent individual point-biserial correlations between the working memory span composite and accuracy for each Raven's item

are better able to control attention to maintain relevant information or disengage from irrelevant information may perform better on tests of both Gf and WMC (see also Stanovich & Toplak's, 2012, concept of *cognitive decoupling*). In the context of problem solving, attentional control might allow participants to flexibly shift from testing one hypothesis to another. In tests of WMC, attentional control might allow participants to resist the effects of distractors and proactive interference on memoranda.

Indeed, one limitation of the present study is ambiguity about the construct of *capacity* and its relationship to attentional control (Shipstead et al., 2015). In general, there is disagreement in the literature as to what WMC tasks measure. Capacity is often defined as the number of units of information that can be held in an active state (see Cowan, 2017, for a review), but some researchers have argued that measures of capacity should be interpreted as indicators of executive attention (e.g., Engle, 2018). In the present investigation, we tested a claim that invoked a "capacity as units of information" model of WMC. Whether capacity estimates really index capacity, or control of attention, or some combination of the two, our results suggest that the contribution of these factors to reasoning performance does not increase as the capacity demands of the Raven's items increase.

Conclusion

The capacity hypothesis argues that WMC constrains Gf. Item-level analyses using an estimate of capacity (k) and Raven's Advanced Progressive Matrices provided no support for this hypothesis. Taken together with the results of Unsworth and Engle (2005), our replication of their results, and Salthouse's (1993) results, there appears to be little evidence that WMC is a causal factor underlying individual differences in Gf. What accounts for the WMC-Gf relationship remains an open question.

Open Practices Statement The data and materials are available to researchers upon request. This work was not preregistered.

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.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404-431.

Conway, A. R., & Kovacs, K. (2013). Individual differences in intelligence and working memory: A review of latent variable models. In *Psychology of Learning and Motivation* (Vol. 58, pp. 233-270). Academic Press.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review, 24*, 1158-1170.

Ekstrom, R. B., French, J. W., Harmon, H. H., & Derman, D. (1976). ETS Kit of Factor-Referenced Cognitive Tests (Educational Testing Service, Princeton, NJ).

Engle, R. W. (2018). Working memory and executive attention: A revisit. *Perspectives on Psychological Science, 13*, 190-193.

Harrison, T. L., Shipstead, Z., & Engle, R. W. (2015). Why is working memory capacity related to matrix reasoning tasks? *Memory & Cognition, 43*, 389-396.

Jensen, A. R. (1998). The g factor: The science of mental ability (Vol. 648). Westport, CT: Praeger.

Little, D. R., Lewandowsky, S., & Craig, S. (2014). Working memory capacity and fluid abilities: The more difficult the item, the more more is better. *Frontiers in Psychology, 5*, 239.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*, 279-281.

Oswald, F. L., McAbee, S. T., Redick, T. S., & Hambrick, D. Z. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods, 47*, 1343-1355.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44*, 369-378.

Pearson, K. (1903). Mathematical contributions to the theory of evolution: II. On the influence of natural selection on the variability and correlation of organs. Royal Society Philosophical Transactions, 200(Series A), 1-66.

Raven, J., & Raven, J. C. (1998). Court, JH (1998). *Manual for Raven's progressive matrices and vocabulary scales*.

Salthouse, T. A. (1993). Influence of working memory on adult age differences in matrix reasoning. *British Journal of Psychology, 84*, 171-199.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2015). Working memory capacity and the scope and control of attention. *Attention, Perception, & Psychophysics, 77*, 1863-1880.

Stanovich, K. E., & Toplak, M. E. (2012). Defining features versus incidental correlates of Type 1 and Type 2 processing. *Mind & Society, 11*, 3-13.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245-251.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*.

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., Fukuda, K., Awh, E., & Vogel, E. K. (2014). Working memory and fluid intelligence: Capacity, attention control, and secondary memory retrieval. *Cognitive Psychology, 71*, 1-26.

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.

Wiberg, M., & Sundström, A. (2009). A comparison of two approaches to correction of restriction of range in correlation analysis. *Practical Assessment, Research & Evaluation, 14*, 2-9.

Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 256-263.