Intelligence 77 (2019) 101387

Contents lists available at ScienceDirect

Intelligence

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

Working memory capacity and strategy use on the RAPM

Andrew F. Jarosz[a,*], Megan J. Raden[a], Jennifer Wiley[b]

[a] Department of Psychology, Mississippi State University, P.O. Box 6161, Mississippi State, MS 39762, United States of America
[b] Department of Psychology, University of Illinois at Chicago, 1007 W. Harrison St. MC 285, Chicago, IL 60607, United States of America

ARTICLE INFO

Keywords:
Working memory capacity
Fluid intelligence
Strategy
Raven's Advanced Progressive Matrices
Individual differences

ABSTRACT

Despite many studies showing that high working memory capacity (WMC) individuals perform better on analytic reasoning and problem-solving tasks, the cognitive mechanisms underlying these relationships are still under debate. The present work explored the link between WMC and performance on a popular test of fluid intelligence (gF), the Raven's Advanced Progressive Matrices (RAPM; Raven, Raven, & Court, 1998), with the goal of assessing whether strategies might play a mediating role in the WMC and gF relationship. Using think-aloud protocols to assess strategies, it was determined that individual differences in strategy use on the RAPM partially mediated the relationship between WMC and performance. In addition, evidence suggested that participants decreased their use of constructive matching strategies as item difficulty increased. Finally, think-aloud protocols provided evidence for a third, hybrid strategy: isolate-and-eliminate. This new strategy goes beyond constructive matching and response elimination, utilizing aspects of each.

1. Introduction

Fluid intelligence (gF) is generally described as the ability to adapt to new situations and to perceive relations (Cattell, 1963). It is central to the construct of intelligence (Marshalek, Lohman, & Snow, 1983), relating highly to reasoning ability and loading quite strongly onto general intelligence factors (Carroll, 1993). However, there remains considerable debate about the mechanisms underlying the construct of fluid intelligence.

One finding that has provided a focus for many researchers is the correlation between working memory capacity (WMC) and gF, with the two constructs generally correlating in the range of 0.75–0.85 when estimates are derived from multiple measures of each construct (Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süß, 2005), though correlations between individual WMC and gF tasks tend to be lower (e.g., a large factor analytic study found individual WMC task correlations with a gF task to range from $r = 0.25$ to $r = 0.47$; Kane et al., 2004). WMC, originally defined as the ability simultaneously store and process information (Baddeley & Hitch, 1974), has been thought to drive performance on gF tasks via a variety of proposed mechanisms. For example, one such mechanism is a limit in storage capacity (Carpenter, Just, & Shell, 1990; Mulholland, Pellegrino, & Glaser, 1980). That is, higher WMC may allow reasoners to better maintain the rules and goals necessary to solve a problem. Alternatively, WMC may grant an increased ability to learn rules while solving (Verguts & De Boeck, 2002) and reapply those rules on later problems. Other work has focused on the attentional aspect of WMC (Engle, 2002), suggesting that WMC helps to deal with interference or distraction while solving (Engle & Kane, 2004; Jarosz & Wiley, 2012; Unsworth & Engle, 2005; Wiley, Jarosz, Cushen, & Colflesh, 2011), whether that be distraction from prior problems or information within the problem itself. Another possibility is that the strategies that individuals use to complete gF tasks may be related to WMC (Dunlosky & Kane, 2007). While progress has certainly been made, the fundamental question of why performance on WMC tasks predicts performance on gF tasks remains unclear. No single explanation of the relationship seems capable of explaining all of the constructs' shared variance, leaving many questions unanswered.

The main question for this study is testing whether strategies may serve as a mediator in the WMC-gF relationship. There is evidence that WMC can be related to the strategies an individual uses to complete cognitive tasks. One perspective, the strategy affordance hypothesis (Bailey, Dunlosky, & Kane, 2008), suggests that strategies will only mediate the relationship between WMC and a task if the strategies used in the target task and WMC task are identical. However, other work suggests WMC can be a limiting factor in what strategies individuals can use effectively. Sometimes this can be seen on the WMC tasks themselves (Dunlosky & Kane, 2007; Engle, Cantor, & Carullo, 1992; Imbo, Duverne, & Lemaire, 2007; Kaakinen & Hyönä, 2007; McNamara & Scott, 2001; Turley-Ames & Whitfield, 2003), but this also extends to

* Corresponding author.
E-mail addresses: afj62@msstate.edu (A.F. Jarosz), mjr302@msstate.edu (M.J. Raden), jwiley@uic.edu (J. Wiley).

https://doi.org/10.1016/j.intell.2019.101387
Received 29 December 2017; Received in revised form 9 August 2019; Accepted 5 September 2019
0160-2896/

other tasks. For example, work examining choking-under-pressure has demonstrated that high-WMC individuals are more likely to see their performance suffer in high-pressure scenarios, as a result of being unable to successfully use their normal, more effective strategies (Beilock & DeCaro, 2007). Research using a similar manipulation demonstrated the same deficit in performance on a gF task, the Raven's Advanced Progressive Matrices (RAPM; Raven, Raven, & Court, 1998), during high-pressure situations (presumably for similar reasons, although this work did not directly assess strategies; Gimmig, Huguet, Caverni, & Cury, 2006). Indeed, the high-pressure scenario eliminated the normal correlation between WMC and the RAPM. WMC has also been demonstrated to predict strategy differences in complex category learning (DeCaro, Carlson, Thomas, & Beilock, 2009; DeCaro, Thomas, & Beilock, 2008). These results suggest a place for strategies in the WMC and gF relationship.

Given the variety of roles that WMC is thought to play in reasoning, understanding the role of strategies on gF tasks could represent an important step in understanding the WMC-gF relationship. Additionally, strategies hold the benefit of being teachable. While the idea of training WMC as a whole is hotly debated (e.g., Shipstead, Redick, & Engle, 2012), altering the way an individual approaches a reasoning problem is a realistic possibility. Indeed, some work suggests that improvements in intelligence test performance after training can be largely explained by accounting for changes in the strategies implemented (Hayes, Petrov, & Sederberg, 2015). Early work on visual analogies implicated the use of distinct verbal-analytic strategies versus visual-holistic strategies by dissociating performance on different types of test items using multiple algorithms (Hunt, 1974) or differential effects of verbal overshadowing (DeShon, Chan, & Weissbein, 1995). Alternatively, work by Bethell-Fox, Lohman, and Snow (1984) identified two primary approaches to solution for visual analogies using eye tracking, with better solvers using a *constructive matching* strategy (involving the generation of the solution prior to searching for it in the response bank) and poorer solvers tending to use a *response elimination* strategy (where participants check each potential response in the response bank against the solution, until reaching the correct answer). Better solvers tended to favor the former strategy, while poorer solvers favored the latter.

Recent work by Gonthier and Thomassin (2015) attempted to shed some clarity on the intersection of WMC, strategy use, and performance on the RAPM. In particular, they focused on comparing a model in which strategies explain (or partially explain) the relationship between WMC and RAPM performance to a model in which strategies are considered to be "noise," with strategy differences being largely irrelevant. In their first study, participants completed a version of the RAPM intended to bias them towards a constructive matching strategy by withholding the response bank for 15 s and explicitly instructing them to use this strategy. This manipulation improved the performance of low-WMC individuals and reduced the WMC-RAPM correlation (as would be expected if strategy variation played a mediational role), but it did not eliminate it.

In a second study (Gonthier & Thomassin, 2015), participants self-reported their strategy use via a four-item questionnaire administered after completing the RAPM, endorsing descriptions of strategies on a Likert scale. Two of these questions assessed use of the constructive matching strategy, while the other two assessed use of a response elimination strategy (with reliabilities between items of 0.21 and 0.31 for each strategy). They found that composites of these self-report retrospective measures fully mediated the WMC-RAPM relationship. In contrast, later work came to quite the opposite conclusion (Jastrzębski, Ciechanowska, & Chuderski, 2018). When Jastrzębski and colleagues replicated their study using larger samples sizes and additional survey questions, they failed to find any mediation by strategy use. Based on the lack of any mediation effect, the authors argued that strategies do not determine the relation between WMC and RAPM. However, an alternative possibility is that the methods that were used to assess strategy use did not allow for valid measurement of actual strategy use.

Notably, each of these recent studies relies on self-report strategy questionnaires with predetermined strategy descriptions, administered after the completion of the entire RAPM. This in turn brings up several concerns that may impact the interpretation of these recent findings. First, a retrospective strategy questionnaire relies entirely upon participants' assessment of their own strategies. This can be problematic as it is based on the assumption that participants have access to the strategies that they used for each solution in memory, and, as Ericsson and Simon (1984) note, subjects are unlikely to explicitly generate such representations while completing a task. As a result, it is possible that retrospective strategy judgments are made based upon subjects' reasoning about their likely processes, rather than veridical recall. Similarly, given the relative inability of individuals to monitor higher-order cognitive processes (e.g., individuals' predictions of their comprehension of texts only correlates around 0.27 with their actual test performance after reading; Maki, 1998; Thiede, Griffin, Wiley, & Redford, 2009), it is unclear if participants would necessarily have access to the strategies that they are using. This is particularly problematic when one is attempting to study differences due to WMC since it is low-WMC individuals who tend to have the poorest ability to monitor their own cognition (Griffin, Wiley, & Thiede, 2008). Additionally, these studies did not collect strategy data on an item-by-item basis, but only asked about strategy use in aggregate. By asking only at a single, retrospective time point it is unclear which problems the assessments are targeting. Participants' reports could rely upon the first strategy used, the most recent strategy used, an average of all strategies used, or the strategies used on problems that stick out in their memory (perhaps due to primacy, difficulty, or qualities of the images in the problem). As such, a participant's reliance on a particular strategy and how it relates to WMC and solution success for a particular problem is unclear based upon these measures.

A final problem with this method is that the strategies assessed by retrospective reports are predetermined by the experimenter. By providing a list, this method constrains the types of strategies that can be reported. It may also invite erroneous responding, due to perceived demand characteristics or social desirability. As opposed to a more open-ended response format, providing descriptions of strategies may encourage participants to report their use at a higher rate than they were actually employed, or across a broader range.

Thus, several features of self-report, retrospective surveys are problematic, as is the inability to map the strategies that are reported onto specific items. Though Jastrzębski and colleagues' measures were more reliable than Gonthier's, they still predetermined the strategies that individuals could report and did not provide trial-by-trial data.

Other work assessing the role of strategy on gF tasks has relied upon eye-tracking methods. In Bethell-Fox et al.'s (1984) work, eye tracking was used to initially identify the constructive matching and response elimination strategies in a four-figure geometric analogy task (A is to B as C is to?) by examining the time spent on different parts of problems, and the movements between them. Other work used those same methods to extend these findings to the RAPM (Vigneau, Caissie, & Bors, 2006), demonstrating that better solvers tended to rely more on constructive matching on the RAPM as well. Another study found similar results using an adaption of Markov chains in combination with principal component analysis to assess strategy based on participant scan paths (Hayes, Petrov, & Sederberg, 2011). Other eye-tracking work has provided initial evidence that WMC might predict strategies on tasks like the RAPM (Jarosz & Wiley, 2012).

Eye-tracking methods circumvent some of the issues with erroneous self-reports due to retrospective checklists. Participants' awareness of their own strategy is no longer a necessity, nor is there an issue with the predetermination of strategies being assessed. However, eye-tracking studies are fraught with their own difficulties. For example, much like in work utilizing strategy questionnaires, prior studies have averaged across problems to create a single strategy measure for each individual. Even the work employing more detailed scan path analyses (Hayes et al., 2011) has collapsed strategy measures across problems for each solver, losing the more nuanced item-level data. This is unfortunate, as item-level data would allow for a better assessment not only of how strategy use might change over time, but also of how item difficulty might interact with individual differences in WMC and strategy use.

Additionally, eye-tracking studies tend to rely on coarse measures, such as the overall number of toggles between the problem and response bank (Bethell-Fox et al., 1984; Jarosz & Wiley, 2012; Vigneau et al., 2006). These are generally interpreted in terms of constructive matching and response elimination strategies, with fewer toggles between the problem and response bank, as well as proportionally less time spent looking at the response bank, signifying higher constructive matching use. However, these measures do not provide a way to determine the precise strategy used on any individual item, and only give an overall sense of whether the solver is acting “more like” someone using constructive matching, or “more like” someone using response elimination. Such analyses also cannot account for multiple strategies that may result in the same pattern of eye movements. That is to say, two participants could be processing problems in a completely different manner but looking at problems in a similar way, and eye tracking would be unable to differentiate between them. Thus, it is still possible that eye-tracking studies may miss the presence of unexpected strategies. Indeed, work with verbal analogies (Embretson, Schneider, & Roth, 1986) utilizing multicomponent latent trait models has suggested not only strategies similar to constructive matching and response elimination, but also highlighted the possibility that people may use “partial strategies” that do not account for all information in a problem.

The present work aimed to avoid the issues brought about by past work employing strategy questionnaires and eye tracking by utilizing a think-aloud methodology to assess the different approaches that solvers use as they attempt to solve RAPM items. Verbal protocols were collected to provide trial-by-trial information on strategy use, while also avoiding predetermining the strategies participants can report (and the associated risk of demand characteristics). Verbal protocols also have the potential to illuminate strategy differences that may be difficult to identify using only eye tracking.

Using an open-ended, trial-by-trial think-aloud design, the present study allows exploration of how strategies interact with WMC and gF through a novel lens. Trial-level strategy data allows for a closer examination of strategy selection and implementation on the RAPM, illuminating differences that may be missed when using coarser strategy-use measures. This then allows for a more thorough test of how WMC relates to strategy use on the RAPM, avoiding reliance on participants' metacognition and retrospection, and paints a more detailed picture of participants' strategy use on each item. The analyses test for a potential mediating role of strategies on RAPM performance, in addition to an examination of how high and low-WMC individuals naturally vary in their strategy selection on the RAPM, and how different RAPM items may vary in the strategies that are used to solve them. The potential mediating role of strategies are also tested on other gF tasks, to explore whether any mediation by strategy use on the RAPM is specific to predicting RAPM performance.

Based on prior work, strategy use as assessed by verbal protocols was expected to relate to both WMC and gF, with higher performers and higher-WMC individuals preferring the constructive matching strategy. Given the increased precision in measuring strategy, both in the presence of trial-level data and in the avoidance of biasing strategy reports, it was also expected that strategy use would mediate the relationship between WMC and RAPM performance.

2. Method

2.1. Participants

Participants were 60 (42 female) undergraduates from a large, midwestern university with complete task data who participated as part of the Introduction to Psychology subject pool, and who were naïve concerning the RAPM. Protocol data were not available for 5 participants due to the corruption of video files. Those participants were not included in analyses involving coded protocol data, but the full sample of 60 was available for other analyses.

Prior strategy research has demonstrated that both WMC (Jarosz & Wiley, 2012) and RAPM performance (Vigneau et al., 2006) correlate with measures of strategy use with a medium-to-large effect size. Based on this information, Fritz and MacKinnon's (2007) work, which used simulated data to estimate power in mediation analyses, can be used to assess whether the present study provides enough power to find mediation effects. With paths to and from the mediator being medium and large, and utilizing bootstrapping to test for mediation, a sample size of 59 results in power of 0.80. Thus, the study's initial sample size was set at 60.

In response to reviewer concerns about sample size, supplemental data from an additional 25 think-aloud participants (16 female), from both a large midwestern and a large southern university, were added to the sample. These additional data included new participants who completed an abbreviated study including the RAPM and the automated symmetry span. The supplemental sample also included some data that had been collected previously but was not complete on all measures. Nine of these participants completed a slightly different subset of RAPM problems. Combined with the original sample, this resulted in a total sample of 80 participants with complete symmetry span and think-aloud RAPM data.

3. Measures

Two sets of measures were used to assess WMC and gF in these studies, with the intention of creating a factor for each construct.

3.1. Working memory capacity tasks

The four WMC tasks used were the automated operation span (Unsworth, Heitz, Schrock, & Engle, 2005), automated symmetry span (Unsworth, Redick, Heitz, Broadway, & Engle, 2009), automated running span (Broadway & Engle, 2010), and backwards digit span tasks (adapted from Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000).

3.1.1. Automated operation span

The automated operation span task (Unsworth et al., 2005) involves remembering letters while simultaneously processing equations. On each trial, participants must decide whether the answer to an equation is correct or incorrect by identifying whether the number shown on the screen after the problem is the correct answer. Following this, a letter to be remembered by the participant appears on the screen. At the end of the trial, there is a recognition test, with participants clicking the letters they saw in the correct serial order. The entire task is presented via computer, and is paced adaptively according to participant performance during practice trials. Between 2 and 7 items are presented in each trial, with three trials of each length. The task takes roughly 15 min to complete.

3.1.2. Automated symmetry span

The automated symmetry span task is an additional complex span task (Unsworth et al., 2009). It also involves simultaneous memory and processing. The procedure is identical to that of the automated operation span, with several exceptions. First, a symmetry judgment is utilized as the processing part of the task, and memory for the location of a square in a 4 by 4 grid is the memory component. Rather than clicking on letters, participants respond by clicking the appropriate spaces in a grid in the order they were presented. Between 2 and 5 items are presented in each trial, with three trials of each length, and the task takes roughly 15 min to complete.

3.1.3. Automated running span

The automated running span is based on a version used previously by Broadway and Engle (2010). Participants are presented with a series of letters, but are told to only remember the last few letters in the series. Thus, they must keep track of the last letters while forgetting the earlier letters. On each trial, participants see between 4 and 9 letters, and must remember the last 3 to 6 letters, with three trials of each length. Participants completed two blocks of trials on this task, taking approximately 10 min to complete.

3.1.4. Backwards digit span

The backwards digit span involves remembering a series of numbers, then repeating them back in reverse order (adapted from Oberauer et al., 2000). In this computerized version, participants saw a string of numbers between 2 and 8 numbers in length, and were asked to type back the numbers in reverse order. Participants had 500 ms to view each digit in the number string, and 15 s to type their answer once the numbers disappeared. To prevent "cheating" on this task, typing was constrained so that participants were forced to type the numbers in backwards order (and could not enter the numbers in forward order first and then revise them, which would eliminate the need for mental transformation). Numbers were presented in sets of increasing size, with two trials of each size. The task takes approximately 10 min to complete.

3.2. General fluid intelligence tasks

The main focus for this study was performance on the RAPM (Raven et al., 1998). In addition, three other tasks were used to obtain an independent measure of the gF construct: the series completion task (Kotovsky & Simon, 1973), the letter sets task (Ekstrom, French, Harman, & Dermen, 1976), and a visual analogies task (Lohman & Hagen, 2001; Form 6, Level H, Test 8 of the cognitive abilities test).

3.2.1. Series completion task

The series completion task (Kotovsky & Simon, 1973) involves finding the next letter in a series of letters following a specific pattern, as seen in the example series below:

p o n o n m n m l m l k ___

The task contains a total of 15 items, each containing between 6 and 15 letters in the series. Participants are asked to type in the next letter of the series using the keyboard. They have 15 min to complete the task.

3.2.2. Letter sets

The letter sets task (Ekstrom et al., 1976) is a categorization task that presents 5 sets of 4 letters, as seen in the example below:

1	2	3	4	5
NOPQ	**DEFL**	**ABCD**	**HIJK**	**UVWX**

Four of these sets follow a pattern of some sort, unrelated to the sounds or shapes of the letters. Participants are required to figure out the pattern that those sets follow, and identify the set that does not follow the pattern by typing in the number of that set. There are two blocks of problems with 15 problems each. Participants have 7 min to complete each block of problems.

3.2.3. Visual analogies

The visual analogies task (Lohman & Hagen, 2001) involves completing 25 visual analogies of the form A : B :: C : ?, as seen in Fig. 1. Thus, participants must identify the rules governing the relationship between images A and B, and apply that to image C to extrapolate what image D must look like. Participants choose their answer from among 5 possible response options by typing in the letter of the appropriate response. Participants have 10 min to complete the task.

3.2.4. Abbreviated Raven's Advanced Progressive Matrices

The RAPM (Raven et al., 1998) is a matrix reasoning task that normally includes 36 items in ascending order of difficulty. For these studies, only 12 items from the RAPM were used in order to fit into a shorter time frame. An example item can be seen in Fig. 2. Participants must select the figure that best completes the matrix at the top of the page, out of a set of 8 possible responses. Participants had 15 min to complete the 12 items. Items were chosen to match several criteria. They cover a range of difficulty (item order was used as a proxy for difficulty), include both visual-holistic and verbal-analytic items (DeShon et al., 1995), cover a variety of different rule sets including both novel and repeated rules (Wiley et al., 2011), and differ in the most common errors made on the problems according to the manual (Raven et al., 1998). Table 1 shows which items were selected, and where they fell in each category. Of the problems selected, 4 were considered to require a verbal-analytic approach, 5 a visual-holistic approach, and 3 could be solved by either of the two approaches. Half of the items require a novel rule set when presented, while the other half require rules seen in previous items. For error types, 5 items tend to result in incomplete-solution errors, 5 in arbitrary-line-of-reasoning errors, and 2 in overdetermined-choice errors. Participants completed two practice items (the first two items of Set 1 of the RAPM), then proceeded to complete the target items in order. Nine of the participants in the additional sample completed items 1, 4, 8, 11, 15, 18, 21, 23, 25, 30, 31, and 35. Analyses involving item types are performed without these participants.

3.3. Think-aloud instructions

Participants were asked to talk aloud about what they were doing while attempting to solve each RAPM problem using instructions from a prior problem-solving study (Ash, Jee, & Wiley, 2012). The exact directions were as follows:

> So that we understand what you are doing while you solve each problem, you will be asked to talk aloud while solving the problem. Be sure to keep talking through the problem-solving process. If you are reading the problem, please do so aloud. If you write anything on paper while solving, please verbalize what you are writing. If you stop talking during the session, I will remind you to keep talking. Think of it as though you are turning up the volume on your inner voice. While you are solving, you will not be given any feedback as to how close you are to solution or the accuracy of your solution. Remember it is very important that you keep talking aloud while solving the problem. Do you have any questions?

Following these instructions, participants completed two "warm-up" tasks: a long division problem and an anagram task (Ash et al., 2012). The instructions were intended to focus participants on simply reporting their thoughts and mental actions, as opposed to instructions in which participants are prompted to explain their thoughts, which are more likely to be reactive (Ericsson & Simon, 1980). Ultimately, performance on the selected RAPM items while thinking aloud was slightly lower than expected based on norming data (Raven et al., 1998), with participants on average solving 42% of items, rather than the expected 55%. Participants were also asked to talk aloud during a second gF task (letter sets) to test whether providing verbal protocols changes the nature of the task, such that the gF measures administered using think-aloud instructions might correlate with each other, but no longer with the other gF tasks.

3.4. Procedure

During an initial session, students completed the WMC tasks, beginning with the operation span, followed by symmetry span, running

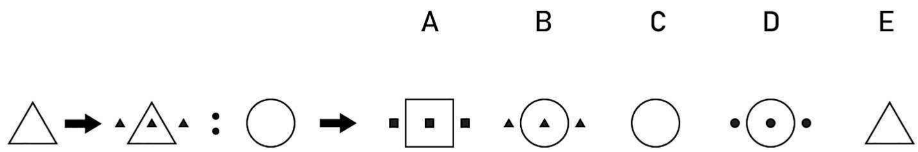

Fig. 1. An item similar to those in the visual analogies task.

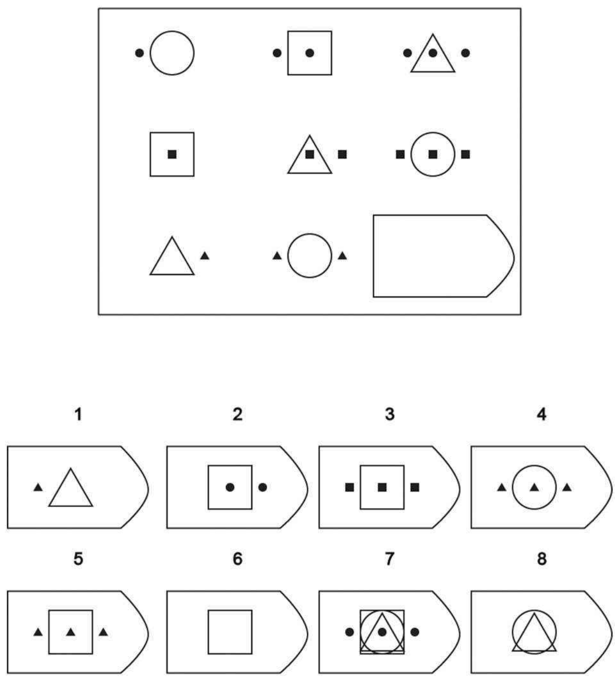

Fig. 2. An item similar to those in the Raven's Advanced Progressive Matrices.

Table 1
RAPM items by type.

Item	A/V/B	Novel/repeated	Rules	Error type
1	A	R	dist 3, dist 3, const	IS
4	A	N	prog, const	ALR
7	V	N	add	ODC
11	V	R	add	ALR
13	A	R	dist 3, dist 3, const	IS
16	V	N	add, sub	ALR
19	B	R	add	IS
22	V	R	add, sub	ODC
25	B	N	prog, change color, const	ALR
30	A	N	dist 2, dist 3, dist 3	IS
33	V	R	add, sub	IS
35	B	N	dist 2, dist 2, dist 3	ALR

Note: A = verbal-analytic, V = visual-holistic, B = both, R = repeated rule, N = novel rule, dist 3 = distribution of 3, const = constant, prog = progression, add = addition, sub = subtraction, dist 2 = distribution of 2, IS = incomplete solution, ALR = arbitrary lines of reasoning, ODC = overdetermined choices.

span, and backwards digit span. During a second session in the following week, the gF tasks were completed. Half the of participants received the tasks in the order of (think-aloud) RAPM, letter sets, (non-think-aloud) visual analogies, then series completion, and the other half in the order of visual analogies, series completion, the (think-aloud) RAPM, followed by letter sets. Think-aloud instructions and practice were always received immediately before the first think-aloud task, and in the case where think-aloud tasks came first, participants received instructions to cease thinking aloud before the final two tasks. Specifically, they were told, "*For the next two tasks, you will no longer need to think aloud. Just solve the problems like you normally would. You no longer need to verbalize everything that you are thinking or writing.*" All participants were videotaped during the think-aloud tasks.

3.5. Scoring

The WMC tasks were scored using the proportional scoring method, as the number of items recalled in the correct serial position (Broadway & Engle, 2010; Unsworth et al., 2005). Descriptive statistics for WMC tasks can be found in Table 2. Correlations between the WMC tasks can be found in Table 3. Finally, a principal component analysis identified a single WMC factor with all four tasks loading onto the factor (Backwards digit span: 0.76; Operation span: 0.67; Running span: 0.79; Symmetry span: 0.73; eigenvalue = 2.17, variance explained = 54.24%). This WMC factor significantly predicted performance on the RAPM at $r(58) = 0.38$, $p = .002$.

The RAPM, visual analogies, series completion, and letter sets tasks were scored by taking the sum of correctly answered items. Descriptive statistics for all gF tasks can be found in Table 2. All of the gF tasks significantly correlated with each other as shown in Table 3. To confirm that all four gF tasks were measuring the same construct, a principal component analysis was computed. A single gF factor was identified with all four tasks loading onto the factor (Figural analogies: 0.78; Letter sets: 0.82; Series completion: 0.78; RAPM: 0.63; eigenvalue = 2.27, variance explained = 56.92%). This is particularly important, as it demonstrates that collecting verbal protocols on the letter sets task and the RAPM did not harm their validity as measures of gF.

4. Results

All analyses were completed on the full sample of 85 participants (80 with think-aloud data), unless otherwise noted. Analyses using the initial sample of 60 participants with complete task data are provided in the Appendix and demonstrate the same pattern of results. Because the strategy codes do not represent different levels of a manipulated independent variable, but were determined post-hoc, chi-square analyses were used to compare solution rates across strategies on the different problem types. Analyses treating these strategy codes as dependent variables were performed using multilevel binary logistic regression.

4.1. Strategy identification

Table 4 outlines the coding scheme developed for the solution strategies that were used while completing the RAPM, based on a first pass of the verbal protocols. Three common strategies were identified during the RAPM. Two have been previously identified in the literature (Bethell-Fox et al., 1984; Vigneau et al., 2006): constructive matching and response elimination. To use a constructive matching strategy, the solver first looks at the problem matrix, figures out what the correct answer must look like, then selects the correct answer from the problem bank. Response elimination, on the other hand, follows a guess-and-check system, where the solver plugs in items from the response bank, eliminating potential responses until he or she reaches the solution.

A third, novel strategy was identified in pilot think-aloud protocols – a hybrid strategy to be called *isolate-and-eliminate*. In this strategy, participants would establish the rules governing one feature of items in the problem matrix, use that rule to eliminate potential responses from the response bank, and then proceed with another feature, until only one item was left. This is reminiscent of the partial rule strategy outlined by Embretson et al. (1986) in verbal analogies and is notably

Table 2
Descriptive statistics for WMC and gF task performance.

	M	SD	Range	Min(Max)	Skew	Kurtosis	Reliability
WMC tasks (factor)	0	1.00	4.78	−2.19(2.59)	0.22	0.18	
Backwards digit span	40.43	9.56	55	11(66)	0.30	1.47	0.71
Operation span	55.57	12.65	55	20(75)	−0.80	0.38	0.82
Running span	47.37	17.47	93	10(103)	0.60	1.16	0.81
Symmetry span	28.98	7.87	32	10(42)	−0.53	−0.11	0.82
gF tasks (factor)	0	1.00	4.03	−2.15(1.87)	0.10	−0.65	
Figural analogies	11.92	4.80	18	4(22)	0.19	−1.08	0.79
Letter sets	19.80	4.74	20	10(30)	−0.09	−0.72	0.72
Series completion	8.10	3.31	13	1(14)	−0.11	−0.72	0.76
RAPM	5.00	2.74	11	0(11)	0.22	−0.56	0.72
Full (N = 85) data							
Symmetry span	29.00	7.78	32	10(42)	−0.50	−0.32	0.80
RAPM	5.33	2.63	11	0(11)	0.06	−0.65	0.70

Note: $N = 60$, except where noted; factor data are presented in italics before individual tasks; reliability calculated using Cronbach's alpha for gF tasks, and parallel forms for WMC tasks.

Table 3
Task performance correlations.

	BDS	OS	RS	SS	FA	LS	SC
1. Backwards digit span							
2. Operation span	0.31*						
3. Running span	0.47*	0.41*					
4. Symmetry span	0.43*	0.31*	0.41*				
5. Figural analogies	0.22	0.28*	0.46*	0.52*			
6. Letter sets	0.38*	0.35*	0.55*	0.54*	0.47*		
7. Series completion	0.23	0.20	0.47*	0.38*	0.49*	0.57*	
8. RAPM	0.37*	0.10	0.29*	0.35*	0.35*	0.40*	0.27*

Note: $^{*}p < .05$; $n = 60$. The correlation of RAPM and symmetry span for the $N = 85$ sample was $r(83) = 0.32$, $p = .003$.

different from a true response elimination strategy. In his work on strategy use during the paper folding task, Snow (1980) describes response elimination as a cyclical process of feature comparison, matching stimulus and response cues and eliminating responses in which features do not match. Little role is given to constructing relationships while using this strategy. Alternatively, those using an isolate-and-eliminate strategy were constructing a partial response, and then applying that partial response to the response bank in order to eliminate one or more answers. Finally, it is worth noting that some participants, on some problems, were unable to verbalize while simultaneously solving problems, even after repeated reminders to think aloud while solving. Thus, problems solved without a coherent verbalization were placed into a *no-verbalization* category. Using the coded verbal protocols, each problem completed by participants was categorized by two raters into one of these four categories, with disagreements decided by a conversation between the two raters to reach a mutual decision. To test for inter-rater reliability an intraclass correlation coefficient was computed, demonstrating moderate agreement, ICC [1, 3] = 0.71, before resolving disagreements.

4.2. Likelihood of strategy use

The likelihood of strategy use by the 80 participants with complete think-aloud data, measured by the proportion of trials on which each strategy was used, is shown in Table 5. The most commonly-used strategy was constructive matching, followed by no-verbalization. The isolate-and-eliminate strategy was used on a small proportion of trials, while a true response elimination strategy was rarely used. The effectiveness of each strategy was determined by looking at the likelihood of correct solution. As seen in Table 5, problems were more likely to be solved correctly when a constructive matching strategy was used, compared to other strategies, $\chi^2(3) = 62.93$, $p < .001$.

Overall, 86% of participants used constructive matching at least once. A total of 63% of participants used the isolate-and-eliminate strategy on at least one trial, and 48% did not verbalize on at least one trial. Only 13% of participants ever used true response elimination on a trial.

4.3. Variations in strategy use due to item characteristics

To test whether strategies change with item difficulty, a multilevel binary logistic regression was used to predict constructive matching use (the most common strategy), with item number (a proxy for difficulty on the RAPM) entered as a fixed effect and participant as a random effect. The model was significant, $F(1, 958) = 22.62$, $p < .001$, and demonstrated that as difficulty increased, likelihood of constructive matching use decreased, $\beta = -0.11$ [−0.16, −0.07]. Converting this coefficient via exponentiation to an odds ratio allows for an assessment of how the probability of using constructive matching changes in response to another variable, such as item number. In this case, doing so gives an odds ratio of 0.90, suggesting that for every unit increase in item number, the odds of using constructive matching decreased by 10%. In contrast, the opposite pattern was observed when predicting a lack of verbalization, $F(1, 958) = 22.47$, $p < .001$, with an increased likelihood of no-verbalization as difficulty increased, $\beta = 0.16$ [0.09, 0.23]. This yields an odds ratio of 1.17, suggesting a 17% increase in the likelihood of a lack of verbalization with each additional item. Item difficulty also predicted the isolate-and-eliminate strategy, $F(1, 958) = 4.27$, $p = .04$. Participants were more likely to use isolate-and-eliminate as difficulty increased, $\beta = 0.06$ [0.003, 0.12], with an odds ratio of 1.06 suggesting a 6% increase in the likelihood of using this strategy for each additional item. Finally, item difficulty did not predict

Table 4
RAPM think-aloud coding for strategy use.

Code	Type	Example phrase
1	Constructive matching	"The first part plus the second part equals the third part which is number 5."
2	Isolate-and-eliminate	"…the third row's all about curves and lines. So I can cross out, um, 1, 3, 7, 8, 6, and..."
3	Response elimination	"So number 1 wouldn't fit because it has the black markings on it. Number 2 is filled out completely…"
4	No-verbalization	After a prompt: "…I'm not thinking, I'm just staring…"

Table 5
Likelihood of use, solution rates, and correlations with individual differences in WMC for each strategy.

	Likelihood of strategy for each participant						Proportion solved when using each strategy			Correlations		
Strategy	All M(SD)	All Min (Max)	Verbal ($n = 71$)	Visual ($n = 71$)	All Skew	All Kurtosis	All	Verbal ($n = 71$)	Visual ($n = 71$)	WMC ($n = 55$)	Sspan	RAPM
Constructive Matching	0.56(0.33)	0(1.00)	0.58	0.53	−0.60	−0.94	0.55	0.60	0.61	0.40**	0.25*	0.52**
Isolate-and-eliminate	0.16(0.20)	0(0.92)	0.15	0.15	1.81	3.47	0.40	0.57	0.33	0.06	−0.02	0.02
Response Elimination	0.03(0.12)	0(1.00)	0.04	0.03	6.65	50.49	0.25	0.36	0.30	−0.33*	−0.34*	−0.21
No-Verbalization	0.24(0.36)	0(1.00)	0.23	0.29	1.27	0.07	0.26	0.26	0.37	−0.27*	−0.10	−0.42**

Note: The *Sspan* column represents correlations with the symmetry span task. Values represent data for all 80 participants with verbal protocols, except where otherwise noted. Exceptions include proportion of problems solved for Verbal and Visual problems (9 participants completed a modified problem set), and correlations with the WMC factor (WMC; 55 participants completed the full battery of WMC tasks).

the likelihood of response elimination use, $F(1, 958) = 1.20$, $p = .27$.

To test whether certain strategies were used more, and were more or less effective on items of a specific type, usage and solution rates were re-examined for subsets of verbal-analytic and visual-holistic items (DeShon et al., 1995).[1] Because it is unclear whether participants approach "both" items with a verbal or a visual strategy, items that could be solved by both approaches were excluded from these analyses. Four multilevel binary logistic regressions were used to examine the likelihood of using each strategy, with the impact of problem type (verbal vs. visual) entered as a fixed effect and participant as a random effect in each analysis. As seen in Table 5, there were no differences in the likelihood of strategy use based on problem type for constructive matching, $F(1, 637) = 2.51$, $p = .11$, isolate-and-eliminate, $F(1, 637) = 0.03$, $p = .87$, or response elimination, $F(1, 637) = 0.79$, $p = .37$. However, when predicting no-verbalization, the model was significant, $F(1, 637) = 6.02$, $p = .01$, demonstrating that the likelihood of no-verbalization increased on visual problems, $\beta = 0.68$ [0.14, 1.23]. Exponentiation yields an odds ratio of 1.98, suggesting an increase in the likelihood of no-verbalization on visual problems of 98%, when compared to verbal items.

Both constructive matching and isolate-and-eliminate strategies were more likely to lead to success on verbal items when compared to other strategies, $\chi^2(3) = 22.58$, $p < .001$, while only constructive matching strategies led to a comparatively higher success rate on visual items, $\chi^2(3) = 22.99$, $p < .001$.

4.4. Variations in strategy use due to WMC

The next question was whether the likelihood of using a strategy depended on individual differences in WMC. As shown in the correlations presented in Table 5, WMC positively predicted the likelihood of using the constructive matching strategy. On the other hand, WMC negatively predicted the likelihood of using a true response elimination strategy; however, it should be noted that very few people ever used response elimination and that these results were driven by a single outlier. The likelihood of using isolate-and-eliminate did not vary significantly with WMC. Finally, though the likelihood of no-verbalization related negatively to a WMC factor in a smaller sample, it did not relate to symmetry span performance in the full sample. Fig. 3 provides scatter plots for each of these relationships.

In order to determine the relationship between WMC and strategy choices, several lag analyses were performed. The simplest of these was a check to see if WMC correlated with the number of strategies that an individual used. This was a possibility, since there was variance in the number of strategies used ($M = 2.19$, $SD = 0.78$). However, no relation was found ($r(78) = -0.03$, $p = .76$) between symmetry span performance and the number of different strategies used. Additional analyses explored when individuals switched strategies. There was no correlation between symmetry span performance and when an individual first changed strategies, $r(78) = 0.02$, $p = .84$, nor was there a correlation between symmetry span performance and how often an individual switched strategies, $r(78) = -0.18$, $p = .11$. Finally, a multilevel binary logistic regression was used to predict constructive matching use with WMC, item number, and their interaction entered as fixed effects, and participant as a random effect. The model was significant, $F(3, 956) = 9.72$, $p < .001$. Although both item number, $\beta = -0.11$ [−0.16, −0.07], and WMC, $\beta = 0.45$ [0.05, 0.85], impacted the likelihood of constructive matching use, the two did not interact, $\beta = -0.04$ [−0.09, 0.01]. Interpreting these findings via odds ratios, item number (odds ratio = 0.90) predicted a 10% decrease in constructive matching use on each additional item, as demonstrated in previous analyses. In contrast, for every 1 *SD* increase in WMC (odds ratio = 1.57), the odds of using constructive matching increased by 57%.

The first strategy that individuals used while solving the RAPM was also explored. Symmetry span performance positively predicted the likelihood of using constructive matching on the first item (Spearman's rho = 0.24, $p = .03$). Thus, high-WMC participants were more likely than low-WMC participants to attempt to solve the first RAPM item using constructive matching.

4.5. Strategy use as a mediator for the WMC-RAPM relationship

To explore the possibility of strategy as a mediator between WMC and the RAPM, a mediational analysis was conducted using Baron and Kenny's (1986) approach, with WMC (symmetry span) predicting RAPM performance, and frequency of constructive matching use as a mediator. As seen in Fig. 4, frequency of use of a constructive matching strategy related to both WMC and RAPM performance, and WMC significantly predicted RAPM performance. Using the PROCESS macro in SPSS, a mediational analysis using Preacher and Hayes's (2008) bootstrapping procedure (5000 samples) was conducted. As seen in Fig. 4, the relationship between WMC and RAPM performance ($r(78) = 0.30$, $p = .008$) was reduced to non-significance after mediation ($r(78) = 0.18$, $p = .08$), suggesting that strategy use may mediate this relationship. However, the inclusion of the direct effect within the confidence interval of the pre-mediation relationship suggests that this was only partial mediation. Additionally, a reverse mediation analysis was conducted, with WMC as a mediator between constructive matching use and RAPM performance. In this case, the strength of the direct effect ($r(78) = 0.49$ [0.29, 0.69], $p < .001$) was almost identical to the pre-mediation relationship ($r(78) = 0.52$ [0.34, 0.73], $p < .001$), suggesting no reverse mediation effect.

Further analyses assessed whether the mediational relationship

[1] This analysis was restricted to the 71 participants who completed the same item set.

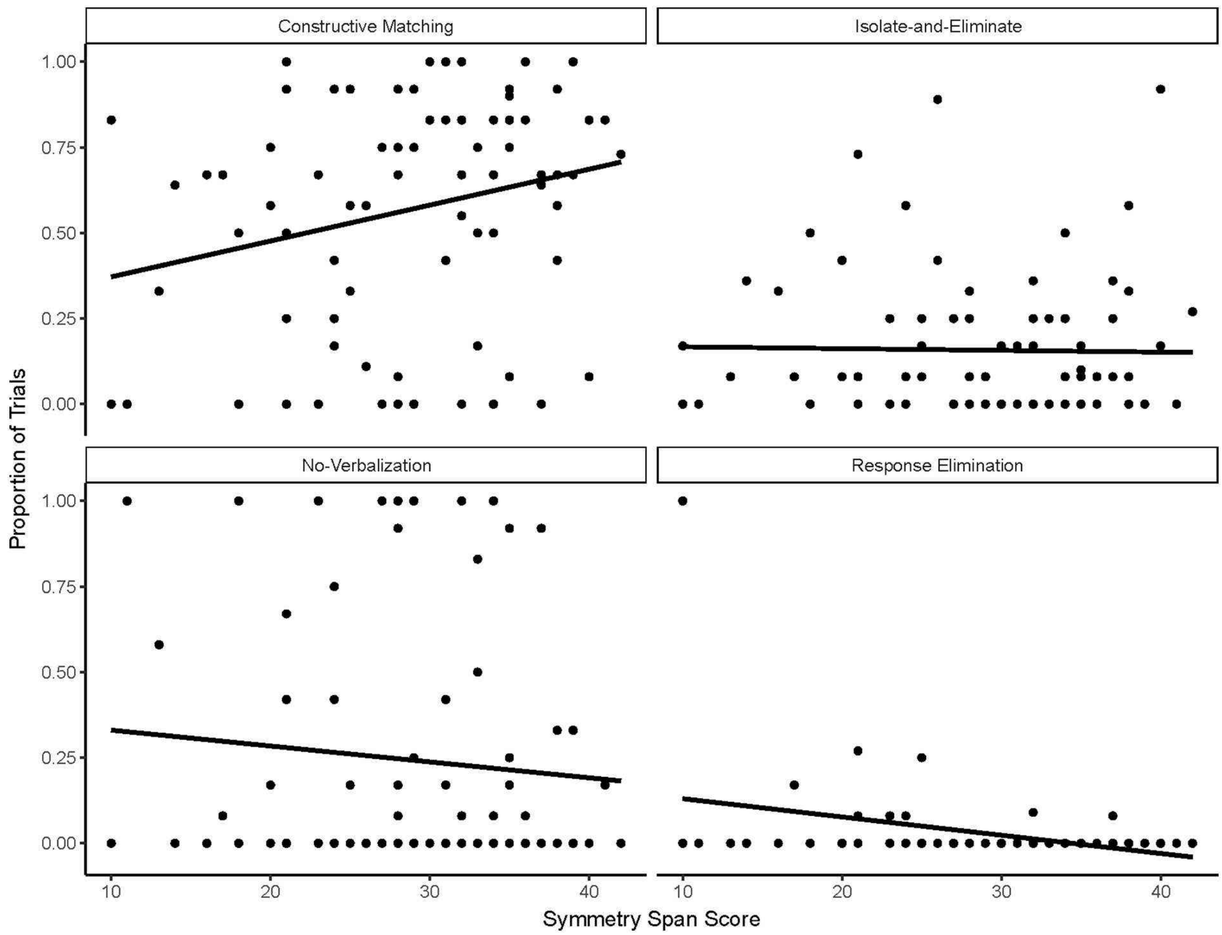

Fig. 3. Proportion of trials using *constructive matching* (top-left), *isolate-and-eliminate* (top-right), *no-verbalization* (bottom-left), or *response elimination* (bottom-right) strategies, predicted by symmetry span performance. Lines on graphs represent regression lines.

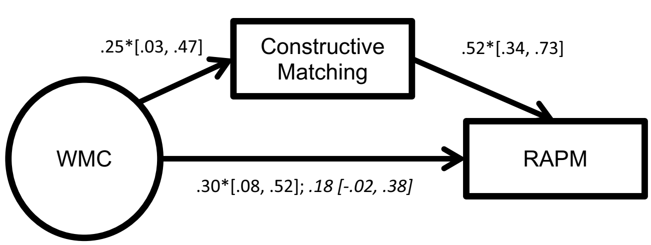

Fig. 4. Constructive matching strategy use at least partially mediates the symmetry span/RAPM relationship. Numbers in brackets represent 95% confidence intervals. Italicized numbers represent the direct effect.

above is unique to the RAPM.[2] WMC correlated with performance on the figural analogies task, $r(53) = 0.46$, $p < .001$, the series completion task, $r(53) = 0.37$, $p = .005$, and the letter sets task, $r(53) = 0.59$, $p < .001$. However, frequency of constructive matching use only correlated with performance on figural analogies, $r(53) = 0.27$, $p = .047$, and letter sets, $r(53) = 0.29$, $p = .03$, and not with series completion performance, $r(53) = 0.23$, $p = .10$. Thus, while mediation is possible for figural analogies and letter sets, the series completion task does not meet the assumptions for such an analysis. Using the same mediation analyses as above and changing only the gF task being predicted, direct effects for both figural analogies ($r(53) = 0.43$ [0.16, 0.70], $p = .003$) and letter sets ($r(53) = 0.57$ [0.33, 0.82], $p < .001$) were not reduced by including constructive matching use as a mediator. Thus, strategy use only mediated performance on the task from which it was measured and was not a mediator between WMC and performance on other gF tasks.

[2] This analysis was only performed on the 55 individuals for whom full factors were available.

5. Discussion

The above study allows one to draw several conclusions in relation to prior work exploring constructive matching and response elimination strategies. First, in agreement with prior studies (Bethell-Fox et al., 1984; Vigneau et al., 2006), there is variability in the strategies used by individuals on a common test of gF, the RAPM. As item difficulty increased, individuals were less likely to try using the constructive matching strategy, and more likely to use the isolate-and-eliminate strategy or to present a lack of verbalization. Additionally, having higher WMC was related to increased use of constructive matching on the RAPM, a strategy previously identified as being more effective than its alternatives (Bethell-Fox et al., 1984; Vigneau et al., 2006). Not only was the constructive matching strategy predicted by WMC, it was also more likely to be used by high-WMC individuals on the first trial, and its use at least partially mediates the WMC-RAPM relationship. The WMC-RAPM correlation in this study ($r = 0.32$) fell between those of previous studies examining strategies on the RAPM (rs of roughly 0.3 and 0.5 for Gonthier & Thomassin, 2015 and Jastrzębski et al., 2018, respectively), and though the confidence intervals surrounding the pre-mediation relationship include the direct effect, the path between WMC and the RAPM was reduced to a non-significant value with mediation. Additionally, the fact that strategy use only showed mediation on the task on which it was measured, while not mediating the effects of WMC on other tasks, suggests that these results are unlikely to be an epiphenomenon or driven by a third variable. Rather, it seems plausible that WMC unlocks the ability to utilize complex, task-specific strategies appropriate for the task at hand that are otherwise unavailable to lower-WMC individuals.

A second major finding from this study was that participants were unable to articulate strategies while solving on about a quarter of the RAPM trials (the no-verbalization category). This was also related to poorer RAPM performance. There are several possible explanations for these results. First, it could be that individuals who were unable to think aloud while solving RAPM items were using a strategy, but it was not an easy one to articulate. That is, it may be that they employed some form of visual or holistic solution strategy (e.g., Hunt, 1974) that could not easily be put into words, or that they relied on iconic, rather than verbal, representations of some problems (Kunda, McGreggor, & Goel, 2013). Either of these possibilities would have resulted in difficulty with verbalization. The increase in no-verbalization on visual items provides some support for this idea, though the absence of verbalization on some verbal items suggests that it is likely not the only factor involved. Alternatively, the lack of verbalization could reflect solution attempts that were devoid of any strategic thought. Another possibility is that thinking aloud put participants into a dual-task situation, with the inability to concurrently verbalize while solving being due to the lack of available WMC resources. Finally, since solvers were being recorded with an experimenter in the room, it is also possible that some individuals were more likely to experience anxiety or fear of being wrong while solving, resulting in them being unwilling or unable to speak; "choking" on reasoning tasks usually impacts high-WMC individuals (Belletier et al., 2015; Gimmig et al., 2006). The mix of these two final propositions may be why the negative relation between symmetry span performance and the incidence of items in this category did not reach significance. Even though no-verbalization responses might have been primarily made by those with low-WMC, some high-WMC participants may also have been affected. Any combination of these reasons could be contributing to the relation that was seen between the inability to produce a verbalization and poor RAPM performance.

Importantly, the think-aloud protocols also revealed that very few individuals used a true response elimination strategy. Prior descriptions of the response elimination strategy tend to refer to it simply as eliminating incorrect responses, either with little regard for how such elimination occurs, or assuming a comparison of features between the target response choice and the matrix (Bethell-Fox et al., 1984; Jarosz & Wiley, 2012; Snow, 1980; Vigneau et al., 2006). Instead, the present study is the first to document an isolate-and-eliminate strategy on RAPM problems, wherein people construct partial responses and eliminate some of the responses, iterating this procedure until left with one answer. Given that response elimination use is generally compared to constructive matching, it is a fair assumption that true response elimination involves little-to-no construction of an ideal response. In contrast, Embretson et al.'s (1986) partial rule strategy suggests that on some problems, solvers may construct part of the answer, eliminate some of the responses, and then guess the answer from the available choices. The present study identified a similar pattern during solution of RAPM items, and the use of this isolate-and-eliminate strategy did not relate to WMC. This discovery resulted from the use of an item-by-item concurrent think-aloud procedure, which is particularly important to note because previous studies using self-report survey responses (Gonthier & Thomassin, 2015; Jastrzębski et al., 2018) or eye tracking (Bethell-Fox et al., 1984; Jarosz & Wiley, 2012; Vigneau et al., 2006) would not have been in a position to discern this strategy from a true response elimination strategy. Indeed, previous eye-tracking studies may very well have combined these two strategies into one group (though Snow briefly mentions that some individuals go through multiple cycles of stimulus analysis and response scanning; 1980). Similarly, studies relying upon retrospective surveys (Gonthier & Thomassin, 2015; Jastrzębski et al., 2018) have presented options relating to only two strategies, constructive matching and response elimination, predetermining the strategies that could be found using that measure. This would make it impossible to discern the use of an isolate-and-eliminate strategy, which would allow for the endorsement of items related to either of those strategies.

The finding that strategies play a part in determining performance on reasoning tasks is an important piece of information though not necessarily novel. What the present work does demonstrate, however, is that past assumptions about strategy use may need to be revised. Strategies do not fall into the simple dichotomy of constructive matching and response elimination, and strategy differences seem to be driven not only by differences between individuals but also by characteristics of the items themselves. Future work assessing strategies on gF tasks should attempt to build upon these findings. It may be that non-think-aloud strategy assessments will require some revision to account for the presence of the isolate-and-eliminate strategy, particularly given the relative scarcity of true response elimination strategies in the present work, and the overlap of isolate-and-eliminate with both the response elimination and constructive matching strategies. Likewise, given the emergence of this new strategy alternative that seems unrelated to WMC, future studies should explore whether prompting an isolate-and-eliminate strategy may be more effective for lower WMC individuals. Finally, another interesting question that needs further exploration is what individuals whose responses fell into the "no-verbalization" category were actually doing. Though this may require drawing upon multiple methodologies, such as concurrent eye tracking and think-aloud protocols, understanding this lack of verbalization is necessary to gain a fuller understanding of how reasoners approach RAPM problems.

Declarations of Competing Interest

None.

Acknowledgements

Portions of this work were completed as part of the dissertation of the first author. The authors thank the dissertation committee for their feedback on this project. They also thank Lauren Harris and Gabriella Valencia for their assistance in coding verbal protocols, and Melissa Meinders, Mimi Nguyen, and Melissa Pasierb for their assistance in data collection.

Appendix A. Supplementary data

Additional analyses related to this article can be found online at https://doi.org/10.1016/j.intell.2019.101387.

References

Ash, I. K., Jee, B. D., & Wiley, J. (2012). Investigating insight as sudden learning. *The Journal of Problem Solving, 4*(2), 1–27.
Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *the psychology of learning and motivation. Vol. 8. the psychology of learning and motivation* (pp. 47–89). New York, NY: Academic Press.
Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition, 36*, 1383–1390.
Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173–1182.
Beilock, S. L., & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 33*, 983–998.
Belletier, C., Davranche, K., Tellier, I. S., Dumas, F., Vidal, F., Hasbroucq, T., & Huguet, P. (2015). Choking under monitoring pressure: Being watched by the experimenter reduces executive attention. *Psychonomic Bulletin & Review, 22*, 1410–1416.
Bethell-Fox, C. E., Lohman, D. F., & Snow, R. E. (1984). Adaptive reasoning:

Componential and eye movement analysis of geometric analogy performance. *Intelligence, 8*, 205–238.
Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods, 42*, 563–570.
Carpenter, P., Just, M., & 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.
Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge, England: Cambridge University Press.
Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1–22.
DeCaro, M. S., Carlson, K. D., Thomas, R. D., & Beilock, S. L. (2009). When and how less is more: Reply to Tharp & Pickering. *Cognition, 111*, 391–403.
DeCaro, M. S., Thomas, R., & Beilock, S. L. (2008). Individual differences in category learning: Sometimes less working memory capacity is better than more. *Cognition, 107*, 284–294.
DeShon, R., Chan, D., & Weissbein, D. (1995). Verbal overshadowing effects on Raven's advanced Progressive Matrices: Evidence for multidimensional performance determinants. *Intelligence, 21*, 135–155.
Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *Quarterly Journal of Experimental Psychology, 60*, 1227–1245.
Ekstrom, R. B., French, J. W., Harman, H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests.* Princeton, NJ: Educational Testing Service.
Embretson, S., Schneider, L. M., & Roth, D. L. (1986). Multiple processing strategies and the construct validity of verbal reasoning tests. *Journal of Educational Measurement, 23*, 13–32.
Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19–23.
Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 972–992.
Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Vol. Ed.), *The psychology of learning and motivation. Vol. 44. The psychology of learning and motivation* (pp. 145–199). NY: Elsevier.
Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*, 215–251.
Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data.* Cambridge, MA, US: The MIT Press.
Fritz, M. S., & MacKinnon, D. P. (2007). Required sample size to detect the mediated effect. *Psychological Science, 18*, 233–239.
Gimmig, D., Huguet, P., Caverni, J. P., & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence (Gf). *Psychonomic Bulletin & Review, 17*, 1005–1010.
Gonthier, C., & Thomassin, N. (2015). Strategy use fully mediates the relationship between working memory capacity and performance on Raven's matrices. *Journal of Experimental Psychology: General, 144*, 916–924.
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*, 93–103.
Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision, 11*, 1–11.
Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence, 48*, 1–14.
Hunt, E. (1974). Quote the raven? Nevermore!. In L. W. Gregg (Ed.). *Knowledge and cognition* (pp. 129–157). Hillsdale, NJ: Erlbaum.
Imbo, I., Duverne, S., & Lemaire, P. (2007). Working memory, strategy execution, and strategy selection in mental arithmetic. *The Quarterly Journal of Experimental Psychology, 60*, 1246–1264.
Jarosz, A. F., & Wiley, J. (2012). Why does working memory capacity predict RAPM performance? A possible role of distraction. *Intelligence, 40*, 427–438.
Jastrzębski, J., Ciechanowska, I., & Chuderski, A. (2018). The strong link between fluid intelligence and working memory cannot be explained away by strategy use. *Intelligence, 66*, 44–53.
Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory, 15*, 634–646.
Kane, M. J., Hambrick, D. Z., & Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 66–71.
Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.
Kotovsky, K., & Simon, H. A. (1973). Empirical tests of a theory of human acquisition of concepts for sequential patterns. *Cognitive Psychology, 4*, 399–424.
Kunda, M., McGreggor, K., & Goel, A. K. (2013). A computational model for solving problems from the Raven's Progressive Matrices intelligence test using iconic visual representations. *Cognitive Systems Research, 22*, 47–66.
Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test: Form 6.* Level H: Riverside Publishing.
Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.). *Metacognition in educational theory and practice* (pp. 117–145). Hillsdale, NJ: Erlbaum.
Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7*, 107–127.
McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition, 29*, 10–17.
Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology, 12*, 252–284.
Oberauer, K., Schulze, R., Wilhelm, O., & Süß, H. M. (2005). Working memory and intelligence–their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61–65.
Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017–1045.
Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40*, 879–891.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices.* San Antonio, TX: Harcourt Assessment.
Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*, 628–654.
Snow, R. E. (1980). Aptitude processes. *Aptitude, Learning, and Instruction, 1*, 27–63.
Thiede, K. W., Griffin, T. D., Wiley, J., & Redford, J. (2009). Metacognitive monitoring during and after reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.). *Handbook of metacognition and self-regulated learning* (pp. 85–106). New York: Routledge.
Turley-Ames, K. J., & Whitfield, M. M. (2003). Strategy training and working memory task performance. *Journal of Memory and Language, 49*, 446–468.
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*, 498–505.
Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory, 17*, 635–654.
Verguts, T., & De Boeck, P. (2002). The induction of solution rules in Raven's Progressive Matrices test. *European Journal of Cognitive Psychology, 14*, 521–547.
Vigneau, F., Caissie, A. F., & Bors, D. A. (2006). Eye-movement analysis demonstrates strategic influence on intelligence. *Intelligence, 34*, 261–272.
Wiley, J., Jarosz, A. F., Cushen, P. J., & Colflesh, G. J. H. (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.