Journal of Experimental Psychology: General
2018, Vol. 147, No. 2, 257–281

0096-3445/18/$12.00 http://dx.doi.org/10.1037/xge0000378

Much Ado About Aha!: Insight Problem Solving Is Strongly Related to Working Memory Capacity and Reasoning Ability

Adam Chuderski and Jan Jastrzębski
Jagiellonian University

A battery comprising 4 fluid reasoning tests as well as 13 working memory (WM) tasks that involved storage, recall, updating, binding, and executive control, was applied to 318 adults in order to evaluate the true relationship of reasoning ability and WM capacity (WMC) to insight problem solving, measured using 40 verbal, spatial, math, matchstick, and remote associates problems (insight problems). WMC predicted 51.8% of variance in insight problem solving and virtually explained its almost isomorphic link to reasoning ability (84.6% of shared variance). The strong link between WMC and insight pertained generally to most WM tasks and insight problems, was identical for problems solved with and without reported insight, was linear throughout the ability levels, and was not mediated by age, motivation, anxiety, psychoticism, and openness to experience. In contrast to popular views on the sudden and holistic nature of insight, the solving of insight problems results primarily from typical operations carried out by the basic WM mechanisms that are responsible for the maintenance, retrieval, transformation, and control of information in the broad range of intellectual tasks (including fluid reasoning). Little above and beyond WM is unique about insight.

Keywords: insight problem solving, reasoning, working memory, personality

Human ability to solve novel abstract problems via knowledge-lean analytical reasoning (called reasoning ability, fluid reasoning, or fluid intelligence), which strongly predicts academic and professional success, socioeconomic status, and even longevity (Deary, 2012), is the key component of general cognitive ability (Holyoak, 2012). Plausible theories of the structure (Carroll, 1993; Cattell, 1971) and processes underlying analytical reasoning and problem solving (Anderson, 2005; Eliasmith, 2013; Hummel & Holyoak, 2003; Johnson-Laird, 2006; Laird, 1929/2012; Newell & Simon, 1972) have been developed. Individual differences in analytical reasoning and its relationships to other psychological constructs are studied by means of valid and reliable tests, large samples, and advanced statistical methodology (see Wilhelm & Engle, 2005). The key finding is that reasoning ability is strongly related (i.e., shares from half to all of the variance; Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Kane, Hambrick, & Conway, 2005) with the storage capacity and control efficiency of the working memory (WM) system (Cowan, 2001).

Simply speaking, WM briefly stores representations crucial for whatever our minds are currently working on, actively switching to new pieces of information as needed to do the work, and at the same time blocking irrelevant and distracting information (Kane & Engle, 2002). The most crucial feature of WM is its limited capacity (WMC): people can simultaneously represent in WM only a few "chunks," such as objects, bindings between objects (or their features), or relational attributes (Cowan, 2001). Although the actual nature of WMC limitation is a matter of debate, a robust finding is that substantial differences in WMC exist within the population (Cowan, 2001; Kane & Engle, 2002; Oberauer, Farrell, Jarrold, & Lewandowsky, 2016). Tasks that impose high loads on WM (e.g., those strongly correlating with WMC and/or strongly depleted by a parallel task) are believed to rely on the systematic, analytical processing that requires complex representations in WM (mental models, mappings, schemas; see Holyoak, 2012; Johnson-Laird, 2006).

In contrast, tasks not loading on WM are believed to rely on nonanalytical processes including automatic and intuitive heuristics and associations (Evans, 2010; Kahneman, 2011). It has been proposed that one potentially important category of nonanalytical problem-solving processes (Ohlsson, 2011; Wiley & Jarosz, 2012), which may lead to new and valuable scientific discoveries, technological inventions, and artistic masterpieces (Finke, 1996; Perkins, 1981; but see Beaty, Nusbaum, & Silvia, 2014), are processes leading to *insight*—a sudden understanding of an idea or a solution resulting from seeing the problem from a completely novel perspective (Koffka, 1935; Köhler, 1947; Maier, 1930; Wertheimer, 1945). One way to analyze such processes is to study the cases of great inventors and artists who came up with breakthrough solutions to important problems of their time (Ohlsson, 2011; but see Weisberg, 2006, 2015). In laboratories, insightful problem solving is studied by means of so-called insight problems (Batchelder & Alexander, 2012; Chu & MacGregor, 2011; Kounios & Beeman, 2014). Like standard analytical problems, insight problems have one correct solution but, unlike the former and regardless of the

This article was published Online First October 23, 2017.

Adam Chuderski, Cognitive Science Department, Institute of Philosophy, Jagiellonian University; Jan Jastrzębski, PhD Programme in Psychology, Institute of Psychology, Jagiellonian University.

This study was supported by National Science Centre of Poland (Project 2013/11/B/HS6/01234). No part of this study has been previously published or presented. All source data are available at osf.io/83jbp.

Correspondence concerning this article should be addressed to Adam Chuderski, Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland. E-mail: adam.chuderski@gmail.com

efforts made, this solution cannot be directly reached (an impasse is encountered). This is so because such problems suggest misleading problem representations, yielding mental constraints that block the correct representations. To solve the problem, these initial obstacles have to be overcome and an effective representation of the problem has to be developed (its *restructuring* has to occur). Once the problem is properly restructured, the solution should be obvious and yield the phenomenological experience of insight (Aha! or Eureka!).

Indeed, some authors suggested that there is a different type of processing during solving analytical problems versus insight problems (e.g., Metcalfe, 1986; Ohlsson, 2011). Moreover, studies reported null or even negative correlations between reasoning/WMC and insight problem performance (see Wiley & Jarosz, 2012). Unfortunately, in contrast to analytical thinking, insight problem solving seems to lack the precise and testable theories (Ash, Cushen, & Wiley, 2009; Batchelder & Alexander, 2012), and only few studies have used advanced methodological tools (e.g., Benedek et al., 2014; Chuderski, 2014; Nusbaum & Silvia, 2011). Many studies on the relationship among insight, reasoning, and WMC included questionable assumptions (e.g., on what given tasks measure), weak methodology (e.g., small samples, unreliable measures), and problematic data analysis (for a critical review see Chuderski & Jastrzębski, 2017). In consequence, the comprehensive study of the differences and similarities between the two paradigmatic types of thinking—analytic and insightful—is still lacking. The present work was aimed at such a comprehensive test of the relationships between insight problem solving and reasoning, with each ability measured by a solid battery of problems. We also accounted for the broad WM construct, measuring five most commonly identified functions of WM, as well as calculating its general capacity. Crucially, we tested if WMC's contribution to insight problem solving is similar to, or different from, its contribution to analytical problem solving.

Insight Problems

Although there is no actual definition of insight problem, Batchelder and Alexander (2012) list eight features shared by most problems used in current insight research. The description of a "good" insight problem (1) is equivocal, (2) suggests initial interpretations that are unlikely to lead to the correct solution, and (3) misleads a solver about the correct interpretation. When (4) such a description is ignored, the finding of the correct solution is facilitated. The solution of a "good" insight problem (5) is quickly found when the problem is represented in the correct manner (when the traps it posed have been overcome), (6) relies on common sense and general knowledge, (7) is obvious to solvers when found, and (8) is obvious to nonsolvers when revealed to them (i.e., it elicits a similar Aha! experience in both cases).

The available catalogue of so-called classical insight problems accumulated during the 20th century as researchers published newly devised problems. Dow and Mayer (2004) collected the hallmark problems, and, on the basis of their content, categorized them into verbal, spatial, and math problems. An example verbal problem is "how an archaeologist could immediately claim that his colleague lied when declaring that he had found a coin dated 21 BC." Solving it requires realizing that things could be dated B.C. no earlier than in the AD times. Probably the best known spatial problem is the nine-dot problem, in which dots arranged in a 3 × 3 grid have to be connected with four straight lines drawn without lifting a pencil from the paper. An aspect of the problem that people are misled by (Kershaw & Ohlsson, 2004) is that the grid suggests that the lines should fit within it (but a maximum of eight dots can be connected in such a way). In fact, there is no such constraint in the description. The task is extremely difficult, as usually only a few participants find the solution without additional hints (see MacGregor, Ormerod, & Chronicle, 2001; Weisberg & Alba, 1981). A simple math problem can ask "The giraffes and ostriches in a zoo altogether have 30 eyes and 44 legs. How many animals are there?" The misleading information suggests that the difference in the number of legs between giraffes and ostriches must be used to tell the exact numbers of giraffes (7) and ostriches (8); this requires some mental arithmetic. In fact, the legs information may be ignored, because the question is not about the numbers of giraffes and ostriches, but about the total number of animals; only the information about the number of eyes is needed to deduce that there are 15 animals in total (each having two eyes). However, most existing insight problems involve more than one modality (e.g., all spatial problems have verbal instructions that can be semantically undetermined), so their univocal categorization is impossible.

Unfortunately, several methodological and practical issues are associated with classical problems. First, they are too complex for laboratory studies: most of them yield long solution times and low accuracy. Second, such problems are highly idiosyncratic (e.g., obstacles they pose belong to disjoint types; Ash et al., 2009), so a few selected problems may not generalize to the whole class of problems. Third, idiosyncrasy may negatively affect the internal consistency of tests composed from highly disparate problems.

In order to attenuate the above weaknesses of classical problems, several other types of insight problems were developed, such as compound remote associates (Bowden & Jung-Beeman, 2003), modeled after Mednick's (1962) remote associates, matchstick equations (Knöblich, Ohlsson, Haider, & Rhenius, 1999), and rebus puzzles (MacGregor & Cunningham, 2008). Remote associates are triples of seemingly unrelated words for which, however, one and only one meaningful semantic associate exists (e.g., "captain", "hotel", "sky" → "stars"). A matchstick problem is an incorrect arithmetic equation including Roman numerals and arithmetic operations composed from matchsticks, which requires a nontrivial decomposition of either the numeral or the operation sign in order to make the equation correct (e.g., "V—I = IX" → "X—I = IX", "III + III = III" → "III = III = III"). Rebus puzzles represent popular phrases or sayings by a combinations of letters, digits, and/or spatial layouts (e.g., "1t345" –> "tea for two"). Substantial collections of such problems were published, with either norms (Bowden & Jung-Beeman, 2003) or clear rules (Knöblich et al., 1999; MacGregor & Cunnigham, 2008) pertaining to their restructuring requirements. Their satisfactory reliability and validity were also supported (e.g., Chuderski, 2014; Cunningham, MacGregor, Gibb, & Haar, 2009).

Theories of Insight Problem Solving

Unlike many robust theories of analytical problem solving and reasoning (e.g., Anderson, 2005; Hummel & Holyoak, 2003; Johnson-Laird, 2006; Laird, 1929/2012), for insight there are only a few theories of limited scope (see Batchelder & Alexander, 2012) whose predictions cannot be easily operationalized (see Ash et al., 2009). The Gestalt psychologists (e.g., Duncker, 1945; Koffka, 1935; Köhler, 1947; Maier, 1930; Wertheimer, 1945) originally viewed insight as a sudden change in the problem representation due to its restructuring in response to an impasse. This line of theorizing has been revived in the eighties as the neo-Gestalt account of insight (Ohlsson, 1984; Metcalfe & Wiebe, 1987), including the *representational change theory* (RC; Knöblich et al., 1999; Ohlsson, 1992; for a related but newer *redistribution theory,* see Ohlsson, 2011), which predicts that initial unsuccessful attempts to solve an insight problem are followed by the relaxation of unnecessary constraints that block access to the correct solution, as well as by a decomposition of the problem chunks into simpler pieces of information to be recomposed more adequately. The most effective insights come from holistic relaxations and decompositions (seeing a problem in a completely new perspective). In sum, theories such as RC, named "special-process view" (Davidson, 1995), claim that insight is a qualitatively different process from analytic problem solving (see Ohlsson, 2011; Wiley & Jarosz, 2012).

The contrasting perspective, developed within modern cognitive psychology (Davidson & Sternberg, 1984; Perkins, 1981; Weisberg & Alba, 1981), attempts to explain insight problem solving in terms of mechanistic chains of specific, well defined (but not necessarily easily identifiable) mental operations, such as attention shifts, memory retrievals, and imagery. This "nothing-special" or "business-as-usual" approach treats insight as involving the same elementary processes that underlie systematic, analytic thinking. The only difference to solving analytic problems is that insight problem solving often consists of an all-or-nothing, instead of a gradual, conscious access to a solution. The best known account in this vein, the *progress monitoring theory* (PM; later renamed the *criterion for satisfactory progress theory*, MacGregor et al., 2001), proposes that people start solving an insight problem using general problem-solving strategies that are often useful in analytic problems (such as hill-climbing or means-ends analysis; Newell & Simon, 1972). According to the PM theory, however, such strategies rarely work in insight problems (e.g., in the nine-dot problem analyzed by MacGregor et al.). Solvers persevere with these unpromising strategies unless, at some point, they represent in their WM either the whole hill-climbing path or the complete subgoals sequence (depending on the strategy adopted) that is available to them (i.e., unless they are able to "look ahead"). When they realize that this path/sequence cannot lead them to the correct solution, they re-represent the problem and change their problem solving strategy to one that helps to solve this specific problem. Thus, the PM theory predicts that the effective solving of insight problems involves substantial loads on WM.

Jones (2003), using a car park game (a nontrivial moving of cars in order to exit a jammed car park), reported eye-tracking data supporting the PM theory: when participants could look three potentially unpromising moves ahead, they reached the correct solution faster than when they looked one or two moves ahead. However, Jones suggested also that his data support the RC theory: moving the car that was key for the unblocking of the car park (interpreted as a nonobvious insight in the game) was directly preceded by an impasse (indicated by prolonged fixations). Thus, to some extent, both theories accounted for the data (although the PM theory accounted for most of the effects).

Fleck and Weisberg (2013; see also Fleck & Weisberg, 2004) reached a similar conclusion after analyzing the verbal protocols of people solving five classical insight problems. Only a minority of participants produced protocols that matched the impasse–restructuring–insight sequence, whereas others solved problems by means of their knowledge and general problem solving heuristics (for similar data see Danek, Wiley, & Öllinger, 2016; Salvi, Bricolo, Kounios, Bowden, & Beeman, 2016; Webb, Little, & Cropper, 2016). Moreover, some people effectively restructured the problem representation with no impasse. In order to explain these results, Fleck and Weisberg (2004, 2013) proposed a hybrid theory of problem solving using four hierarchical stages (for a recapitulation of this theory see Weisberg, 2015). Later stages are elicited only when earlier stages cannot yield a solution. First, people try to recall potential solutions from long-term memory (solutions can match directly or by analogy). Second, if no solution is found after this knowledge search, a solver attempts general problem solving strategies (heuristics), such as hill-climbing. Third, even if neither knowledge search nor heuristics bring a ready-to-go solution, the failure may yield some additional information about the problem (e.g., some kind of feedback) that can help to restructure its representation and enable a new perspective leading to the solution. Only if neither recall, nor heuristics, nor restructuring succeeds in bringing the solution does the classical impasse occur. As a result, at the fourth and final stage the solver may approach the problem description and available data once again, potentially reencoding them as well as relaxing unnecessary constraints, and this time may be able to restructure the problem successfully. Thus, the solution of insight problems is driven by the interaction of various analytical processes. What some scholars take as "insight," may just be a "restart" of this interaction after an impasse.

Overall, relatively little progress has been made so far in developing theory of insight problem solving, as compared with analytical problem solving (see Batchelder & Alexander, 2012). The Fleck and Weisberg (2013) proposal is possibly the most comprehensive, but the empirical tests of all existing theories are scarce (e.g., Danek et al., 2016; Fleck & Weisberg, 2004, 2013; Jones, 2003; Knoblich et al., 1999; Lung & Dominowski, 1985; MacGregor et al., 2001; Salvi et al., 2016; Weisberg & Alba, 1981). Thus, it may be fruitful to compare performance on insight problems to performance on WM and reasoning tasks, because the two latter domains have acquired deep theoretical explanations (e.g., Anderson, 2005; Cowan, 2001; Johnson-Laird, 2006; Oberauer, 2016; Hummel & Holyoak, 2003). Knowledge of relationships between insight, WM and reasoning may allow the transfer of some of these explanations into the insight problem solving domain.

The Psychometric Approach to Insight and Its Relation to Fluid Reasoning and Working Memory

In the psychometric approach to insight, participants take a test involving insight problem solving, and another test capturing some form of analytical thinking. The correlation between these tests reflects the amount of commonality between both types of thinking. The special-process approach expects weak correlations; the nothing-special account predicts substantial positive links. Several studies have correlated a single fluid reasoning test with a composite insight score, and have reported positive correlations ranging from $r = .41$ (Gilhooly & Murphy, 2005) to $r = .65$ (Davidson, 1995). However, psychometricians acknowledge (Wittmann, 1988) that single tasks contaminate wanted variance in a certain ability with unwanted method- and task-specific variance, and can only partly measure this ability. One should apply at least two (ideally three or four) superficially differing tests designed to assess this ability in different ways. Variance shared among all these tests (called a *latent variable*) will filter specific factors out, and will thus more closely represent true ability. Latent variables can be correlated using confirmatory factor analysis (CFA), or regressed in structural equation modeling (SEM; see Kline, 1998). Such a method has been the gold standard of investigating WMC and its correlates for more than 15 years (e.g., Engle et al., 1999).

The CFA/SEM approach must assume that general insight problem solving ability (reflected by the latent variable) really exists; that is, scores on various insight tests correlate more strongly mutually than with other measures. The first attempt (Cinan, Özen, & Hampshire, 2013) to identify such a factor failed, as the CFA model including the insight factor (loading on the remote associate and classical insight problems) and the (noninsight) planning factor showed a worse fit than the one-factor model that assumed insight and planning to be one and the same ability (both variables correlated at $r = 1.0$). However, Chuderski (2014) succeeded in demonstrating an insight factor (on which the verbal and figural classical insight tasks loaded in one study, and additionally the remote associates test loaded on it in another study) that was statistically distinguishable from the reasoning factor (i.e., the two-factor model fit best), even though the factors strongly correlated ($r \approx .8$). Thus, various insight tasks seem to show convergent validity.

However, fluid reasoning tests are themselves complex, and the strong link between performance on them and performance on insight problems might result from the shared complexity and requirements of both types of tasks. In order to avoid possible confounds, several studies examined relation of insight with WM, instead of reasoning, because the WM tasks involve rules of relatively low complexity, and usually involve processing, maintenance, and later recall/recognition of very simple stimuli (single letters, digits, figures, etc.). The extent to which insight problem solving and fluid reasoning differ/agree in their reliance on WM will help us to know not only conceptual but also processual differences/agreements between them.

So far, highly inconsistent results have been reported, from moderately positive (e.g., Chuderski, 2014), through null (Ash & Wiley, 2006), to even negative effects of WMC on insight problem solving (DeCaro, Van Stockum, & Wieth, 2016). Significant positive correlations, varying greatly in strength, were reported between WMC and insight problem solving performance (e.g., Chein, Weisberg, Streeter, & Kwok, 2010; Chuderski, 2014; Chuderski & Jastrzębski, 2017; De-Young, Flanders, & Peterson, 2008; Gilhooly & Murphy, 2005) as well as remote associates performance (e.g., Chein & Weisberg, 2014; De Dreu, Nijstad, Baas, Wolsink, & Roskes, 2012; Kane et al., 2004; Storm & Angello, 2010). In contrast, Wiley and Jarosz (2012) reviewed evidence on null or even negative correlations between WMC and various insight tasks, suggesting that overefficient WM may harm insight problem solving. However, a closer look (Chuderski & Jastrzębski, 2017) at these studies (e.g., Ash & Wiley, 2006; DeCaro et al., 2016; Ricks, Turley-Ames, & Wiley, 2007; Van Stockum & DeCaro, 2014) suggests that most of them examined too small samples, or reported too weak effects, to be able to support the hypothesis of the null/negative WMC-insight link.

Goals of the Present Study

To date no WM-insight study has examined more than one category of insight problems, with the risk that the observed results strongly depended on the peculiarities of the selected problems instead of reflecting broad insight problem solving ability (a probable reason why the results varied so much from study to study). Furthermore, no study has examined substantially more than 100 participants; that is, the sample sizes were greatly below the widely adopted standards for correlational studies (at least 200 people or more, see Kline, 1998). Thus, a comprehensive, large-sample study was necessary for a valid and reliable evaluation of the relationship of insight problem solving performance to both the specific functions of WM and its general capacity (WMC), when controlling for reasoning ability, personality, anxiety, motivation, and age.

Our general goal consisted of carrying out such a comprehensive psychometric test of the relationship between WMC and insight problem solving ability (henceforth, insight ability), with both constructs defined at the latent level. Furthermore, we not only used objective measures of performance on the insight tasks, but also subjective reports on whether or not a person experienced insight during problem solving. The WMC-insight ability relationship was compared with the relationship between WMC and reasoning ability, in order to test whether or not insightful and analytical thinking differed in their WM mechanisms. No study to date has investigated the link between WM and insight problem solving in such a comprehensive way. We attempted to meet our general aim by employing five specific analyses discussed below.

Links Between Specific WM Functions and Insight Problem Solving Ability

The literature on WM suggests that its several specific functions may be distinguished as follows:

a. active maintenance of information, often termed the short-term memory (STM), primary memory, or storage function of WM, usually examined with simple span and change-detection tasks (Cowan, 2001);

b. storage during processing (the need to maintain information in face of a distracting task), measured by the complex span tasks that interleave stimuli encoding with simple decisions (Kane & Engle, 2002);

c. updating (frequent exchange of outdated pieces of information with those that become task relevant), assessed by tasks such as n-back, mental counters, and running memory (Smith & Jonides, 1999);

d. executive control (focusing on a task goal while blocking/inhibiting the interference/distraction/habitual response), expressed as stability of performance (its accuracy or latency) between low- and high-distraction conditions, such as congruent versus incongruent trials in the Stroop task (MacLeod, 1991), prosaccade versus antisaccade trials in the antisaccade task (Hallett, 1978), or go versus no-go/stop trials in the no-go (Eimer, 1993)/stop signal (Logan & Cowan, 1984) task;

e. binding of information (integrating separate pieces of information into more complex, task relevant structures and relations, necessary even when this information is perceptually available), for example measured by means of the relation monitoring tasks (Oberauer, Süß, Wilhelm, & Sander, 2007).

Although several published results pointed to a positive relationship between WMC and insight problem solving ability, it might be possible that such a relationship exists for only some functions of WM, but not others. For example, Wiley and Jarosz (2012), citing Fleck (2008), suggested that insight may be more strongly related to storage than to executive control and to complex span, but did not formally test this difference in the data. Gilhooly and Fioratou (2009), using hierarchical linear regression, indeed found that switching ability (possibly involving executive control) was a significant predictor for performance on noninsight problems, but not on insight ones; however, the difference in their 120-people sample was not significant. Chuderski (2014) found a similar contribution to insight of both executive control and storage, but he noted a weaker link for binding (important for fluid reasoning).

In Analysis 1, individual capacities were measured for each of the five above listed WM functions. Three tasks per function were used (with an exception for updating), and relevant latent variables were calculated. Then, using CFA, we assessed correlations between each WM function and insight ability, and between WM function and reasoning ability. We tested for any differences in correlation strengths between the functions, as well as between insight and reasoning for a particular function.

The True Strength of the Relationship Between General WMC and Insight Problem Solving Ability

Even if all key functions of WM contributed to insight ability, the "true strength" of the relationship between the general WMC factor and insight problem solving is unknown. As most reported effect sizes pertained to single insight tests (often encompassing just a few insight problems), the reliability and validity of such tests must have been (and was) limited, and the links found must have been specific and attenuated, compared with the true strength of the relationship. Simply, the maximum strength of correlation between even two isomorphic measures cannot surpass the square root of the product of their reliabilities. To date, no study has measured insight problem solving with a varied and large enough battery of problems that could reflect insight ability at the latent level. Nor has any study related this ability to WMC defined as a higher-level factor, loading on several WM functions. Only such a method allows for estimation of the true size of the link between insight ability and WMC (unspoiled by low reliability, test peculiarity, and stimulus modality). Thus, the goal of Analysis 2 was to assess the true size of the relationship between insight ability and WMC.

The Relationship Between WMC and Insight Problem Solving Ability for Subjective Measures of Insight

Recent studies (e.g., Danek et al., 2016; Fleck & Weisberg, 2013; Salvi et al., 2016; Webb et al., 2016) that have analyzed participants' reports about their experience of insight during insight problem solving suggest that it may be required for solving only some of the existing problems, whereas other problems do not elicit such an experience. Moreover, differences in solution strategy exist, with only a relatively low proportion of participants reporting insight, while others admit the use of more systematic strategies (for a discussion, see Weisberg, 2015). It is possible that WMC predicts solutions that are not accompanied by subjectively perceived insight, whereas "real" insightful solutions are unrelated to WMC. To test this possibility, for each verbal and spatial insight problem, each remote associate and each matchstick equation (but not for the math problems, for justification see below) we recorded participants' ratings of whether or not "the solution came suddenly and unexpectedly to mind" (henceforth, both correct and incorrect solutions which were rated positively are called insightful). In Analysis 3, we reran Analysis 2, but used two subjective scores for each test: (1) the proportion of correctly solved insightful problems in all problems ("insight accuracy") and (2) the proportion of correctly solved noninsightful problems in all problems ("analysis accuracy"). We tested if insight accuracy yielded either weaker (as expected by the special-process account) or comparable (as predicted by the nothing-special account) correlations with WMC, as compared with analysis accuracy.

Threshold in the WMC-insight Ability Relationship

A well-known prediction pertaining to creativity called the threshold hypothesis (Barron, 1969; Guilford, 1967) assumes that a certain level of cognitive ability is a necessary condition for an individual to produce creative thought (or outcome). Recently, by means of a novel statistical technique, Jauk et al. (2013) suggested that such a threshold does indeed exist. It was rather low (IQ 85) for fluency on a divergent thinking test, and higher for originality (IQ range = 100–120), but there was no threshold for creative achievement (the self-reported real success in creative domains, such as art, literature, and science). Analogously, a significant WMC effect on insight may exist only for people below a given threshold, while participants above this threshold may already have the necessary cognitive resources to solve insight problems (ceiling effect for WMC), so the real link is null. Due to the mathematical properties of correlation, even when only a minority of a sample shows the effect, the correlation can still be substantial (Smolen & Chuderski, 2015). In Analysis 4, in order to test the possibility of threshold in the WMC-insight ability link, we examined invariance of WMC-insight correlations for groups below/above the mean in fluid reasoning.

Mediation of Personality and Motivational Variables

Unlike reasoning ability (Ackerman & Heggestad, 1997; Moutafi, Furnham, & Crump, 2003), creative problem solving has been related to personality and motivational factors (e.g., Amabile, 1985; Fürst, Ghisletta, & Lubart 2016; Prabhu, Sutton, & Sauser, 2008; Feist, 1998). Apart from openness to experience, which notably links to creativity for rather trivial reasons (it reflects novelty seeking as well as is the personality trait most strongly related to intelligence; Kaufman, 2013; Moutafi et al., 2003), the most replicated (Rawlings, 1985) and interpretable link (Eysenck, 1993) exists for psychoticism—a tendency toward egocentric, impulsive, and antisocial behavior. It is suggested that the prolonged self-focus, strong self-assertion, and breaking of the established rules helps to deliver creative outcomes. However, similar factors (e.g., anxiety) were reported to influence WM and fluid reasoning (e.g., Chuderski, 2014; Eysenck, Payne, & Derakshan, 2005; Owens, Stevenson, Hadwin, & Norgate, 2014; Unsworth et al., 2009). Thus, it is at least theoretically possible that individual differences in state anxiety, motivation level, and/or personality traits can mediate to some extent the WMC-insight link. In Analysis 5, the links between residual variances in WM and insight variables were examined after explaining them by age, psychoticism, openness to experience, state anxiety, and motivation.

The Study

In this section we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

Participants

We defined a relatively large sample size, surpassing the samples studied to date in insight research. The sample included 318 volunteers (216 women, 102 men) recruited via advertisements on popular networking websites. Participants were paid the equivalent of 20 euros in Polish zloty. The mean age was 24.5 years (SD = 6.02, range 18 – 46). The sample included 3 secondary school pupils, 17 people who attended high school and another 55 who graduated it, 165 college students, and 78 people holding a BA or MA degree (or equivalent). All participants had normal or corrected-to-normal vision. All participants were informed in a general way that the study investigates human thinking, that their data would be anonymous, and that they could end their participation at will at any moment. The study was carried out in accordance with the ethical principles of the 1964 Declaration of Helsinki as well as the psychological research guidelines suggested by the Polish National Science Center (www.ncn.gov.pl).

Materials

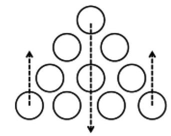
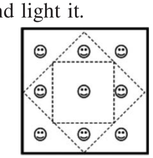
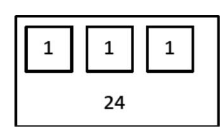
For the insight problem categories that were described above, we chose 19 classical problems (consisting of six verbal, six spatial, and seven math problems) as well as 15 remote associates and 8 matchstick equations. The classical problems were selected on the basis of two criteria: (1) they were used in at least two studies that, for the given task, reported effects typical for insight problems (frequent impasse etc.); and (2) their reported accuracy was around 33%; this is the optimal trade-off between the required difficulty and the avoidance of the floor effect. We tried to select the most frequently used problems. For remote associates, we selected 15 out of 24 items used in our previous study (Chuderski, 2014) which yielded accuracy levels of around 33%. For matchstick equations, we used two chunk decomposition (CD) and two constraint relaxation (CR) items from Knöblich et al. (1999) and designed two CD and two CR items (see below). However, these two new CR items did not yield the expected difficulty and were discarded, so the final set included six matchstick equations. There were 40 insight problems in total—probably the largest set of insight problems ever used in a study. All insight problem solving tests were paper-and-pencil tests. We also used four paper-and-pencil fluid reasoning tests (with 74 items in total), as well as 13 computerized WM tasks. Five questionnaires investigated personality, anxiety, and motivation.

Verbal and spatial problems. In order to decrease transfer and learning effects, we interleaved the six verbal and the six spatial problems on two pieces of paper. See Table 1 for the description of the problems, in the order matching the actual test. The verbal problems included *Woman and policeman*, *Candle*, *The game of checkers*, *Four coins*, *Two people*, and *Two strings*. The spatial problems consisted of *Triangle*, *Pig pen*, *Four dots*, *Sheep pen*, *Figure L*, and *Ten dots*. The task instructions specified flexible and nonstandard thinking, and required the answer for each problem to be written below the problem. Participants were given 20 min to complete the test. The numbers of solutions that satisfied the description of the respective problems were taken as the verbal problems and the spatial problems scores. Each problem also included a box to be checked by a participant if this problem's solution "came to mind suddenly and unexpectedly" (for discussion and justification of this method see Bowden et al., 2005). Participants were reminded that they should report their subjective state when arriving at solutions.

Math problems. The test included seven math problems previously used in insight problem solving studies (e.g., *Lilies, Horse trading*; see Frederick, 2005; Weisberg, 1995). Unlike the verbal and spatial problems, for which initially the solver cannot find any answer, for each math problem there is an intuitive, but incorrect solution that results from a specific misrepresentation of the problem (see Table 2). To solve the problem, one needs to override the initial incorrect solution (or avoid the misrepresentation in the first place) and come up with a proper, more reflective representation of the problem. As some problems have already been popularized in the media, we changed the specific contents of these problems while leaving their internal structures intact (e.g., the spread of water lilies was substituted by the spread of fire in a forest). Descriptions of all math problems were printed on one paper sheet, and participants wrote their solutions directly below each respective problem. The number of correctly solved problems in five minutes represented the score on this test. One easy problem (*Seven sisters*) preceded the seven problems as a run-up, and was not included in the score. Because for this test both incorrect and correct solutions come to mind suddenly, this was the only test for which the subjective reports on solutions were not collected.

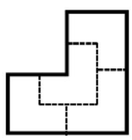
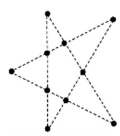
Matchstick equations. The test included 16 matchstick equations printed on two sides of a single sheet of paper (see Table 3 for the equations list, in the order in which they appeared in the test). There were eight standard type (ST) problems that were relatively easy to solve analytically (they acted as controls). Such problems required moving a vertical matchstick from one numeral to another (e.g., "I + V = IV" →

Table 1
Description of Verbal and Spatial Insight Problems Used in the Study

Problem name	Problem description	Sample solution	Accuracy	Insight accuracy	ΔWMC solvers	ΔWMC insight
Triangle	How to make the triangle below point downward by moving only three of the circles.		.43	.28	**.50**	**.46**
Woman and policeman	A woman did not have her driver's license with her. She ignored a stop-sign, then ignored a one-way traffic sign and travelled in the wrong direction down the one-way street. All this was observed by a policeman, who right-fully made no effort to arrest the woman. Why?	The woman was walking.	.40	.29	**.46**	.30
Candle	You have a box of tacks, a box of matches and a candle. How to attach the candle vertically to a door and light it, so that it lights normal?	Pin the box to the door. Then light a match, melt the bottom of the candle, put it into the box and light it.	.53	.15	.04	.22
Pig pen	Nine pigs are closed in a pen. How to separate each pig from all the others using only two additional square fences?		.28	.29	**.56**	.13
Two people	How can one person stand behind a second person when at the same time the second person is standing behind the first one?	Two persons can stand with their backs turned toward each other.	.58	.23	.27	.34
Four dots	Connect all the 4 dots with two straight lines, without lines crossing each other, and without taking the pen off the paper.		.45	.31	**.78**	.34
Checkers game	Two men played five full games of checkers and each won an even number of games, with no ties, draws, or forfeits. How is that possible?	The men did not play with each other	.35	.22	**.56**	.02
Sheep pen	How can you close 27 sheep in 4 different pens, so that inside each pen there is an odd number of animals?		.09	.21	**.95**	−.01
Four coins	A brother with his sister and a husband with his wife found 4 coins. Each of them took 1 coin, but there was still 1 coin left. How is that possible?	The brother and the husband are the same person.	.61	.19	**.54**	.24

(*table continues*)

Table 1 (*continued*)

Problem name	Problem description	Sample solution	Accuracy	Insight accuracy	ΔWMC solvers	ΔWMC insight
Figure L	Divide the figure into four equal parts.		.03	.13	**1.15**	.36
Two strings	Two strings hang from a ceiling but are too far apart to allow a person to hold one and grab the other. In the room there is also a chair and a table, on which there is a hammer and two nails. How could the strings be held together?	Grab one string, get on the table, and nail it to the ceiling. Then, grab both strings while standing on the table.	.32	.05	.01	.17
Ten dots	Arrange the 10 dots below into 5 rows with 4 dots each.		.01		.34	

Note. Differences in WMC with *ps* < .01 are in bold. Only 3 participants solved the *Ten dots* problem, and none of them reported insight.

"I + V = VI"). The eight remaining items were meant to require restructuring in order to discover a nonstandard move of one matchstick that made the equation correct. Two more difficult CD problems called for the decomposing of chunk "V" into "X" or vice versa (Knoblich et al., 1999). Two even more difficult CR problems required relaxation of the constraint that only one equal sign is allowed per equation (Knoblich et al., 1999). Two newly designed items required introducing a negative number by changing the number "I" into a minus sign, and another two such items called for equating two positive/negative numbers by eliminating the number "I" and making a plus/minus sign. Participants were given 20 min to complete all 16 problems and were instructed to think in a nonstandard and creative way. Items from each pair were printed on the opposite sides of the paper sheet, so one could not directly see the analogy between the two items. Unfortunately, problems 2 and 16 appeared too easy (mean accuracy around 90%, matching the ST equations), so the final score on the test was the sum of correctly solved problems out of six equations (Items 4, 5, 7, 9, 12, 13). The check-boxes for subjective ratings (a solution "came to mind suddenly and unexpectedly") were placed beside each equation.

Remote associates. The task included 15 remote associates problems (Mednick, 1962) verified in our previous study (Chuderski, 2014; note that they are untranslatable from Polish to English). Ten minutes were given to complete all remote associates. Also in this test, the subjective ratings were collected, as in the verbal and spatial problems test and the matchstick equations test.

Active maintenance tasks. Active maintenance of information in WM was measured with three variants of the visual arrays task (Cowan, 2001; Luck & Vogel, 1997). Each of the 90 trials in each of the tasks consisted of a virtual, 4 × 4 array filled with five, seven or nine stimuli (i.e., only some cells in the array were filled). The stimuli were 10 Greek symbols (e.g., α, β, χ), 10 figures (e.g., a square, a triangle, a star) or squares in 10 distinctive colors, in the letter, figure and color arrays, respectively. The array was presented for the time equal to the number of stimuli presented multiplied by 0.5 s, and followed by a 1-s black square mask of the same size as the array. In a random half of the trials, the second array was identical to the first, while in the remaining trials both differed by exactly one item at one location. If they differed, the new item was highlighted by a square red border. If they were identical, a random item was highlighted. The task was to press one of two response keys, depending on whether the highlighted item differed or not in the two arrays (see Figure 1 for a sample trial). The second array was shown until a response was given or 8 s elapsed. The task was preceded by six training trials. The score on the task was the estimated number of objects held in the storage capacity buffer (the *k* value; Cowan, 2001), calculated as the difference between the proportion of correct responses for arrays with one item changed and incorrect responses for unchanged arrays, multiplied by the set size. The total scores were the mean *k* in each task.

Storage and processing tasks. Each of the three variants of the complex span task, which was used to assess the storage and processing function of WM, required memorizing four, six, or eight (set size) stimuli, presented for 1.2 s apiece. Each stimulus was followed by a simple decision task to prevent the chunking of stimuli. The participants were instructed to recall as many stimuli

Table 2
Description of Math Problems Applied in the Study

Problem name	Problem description	Correct solution	Implicated incorrect solution	Accuracy	ΔWMC solvers
Bat and ball	A flashlight and a battery cost $11 in total. The flashlight costs $10 more than the battery. How much does the flashlight cost?	10.50	10	.26	**.81**
Machines	If it takes 5 machines 5 minutes to make 5 toys. how long would it take 100 machines to make 100 toys?	5	100	.25	**.76**
Barrels	If John can drink one barrel of water in 6 days. and Mary can drink one barrel of water in 12 days. how long would it take them to drink one barrel of water together?	4	9	.43	**.76**
Fifteenth best/worst	Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class?	29	30	.27	**.74**
Lilies	A forest fire doubles in area every 24 hours. The fire started 24 days ago. Today. the whole forest is on fire. On which day was half of the forest on fire?	23	12	.31	**.93**
Horse trade	A man buys a book for $60. sells it for $70. buys it back for $80. and finally sells it for $90. How much has he made?	20	10 or 30	.28	.06
Socks	If you have black socks and brown socks in your drawer. mixed in the ratio of 4:5. how many socks will you have to take out to be sure of having a pair the same color?	3	4 or 5	.27	.32

Note. Differences in working memory capacity (WMC) with $ps < .01$ are in bold.

as they could (in proper order), and also to provide correct answers to the decision tasks. Five trials for each set size (in increasing order) were presented in each variant. The letter span task required memorizing letters, while indicating with a mouse button whether or not intermittent simple arithmetical equations (e.g., $2 \times 3 - 1 = 5$?) were correct (see Figure 1). The digit span consisted of memorizing digits, and the decision task was to decide whether the five-letter string presented after each stimulus started and ended with a consonant (e.g., "KVTAR), or started and ended with a vowel (e.g., "ENYBA"). In the figure span task, participants memorized simple geometric figures, while judging colors as light (yellow or beige) or dark (brown or navy blue). During response procedure, in each task as many 3×3 matrices as was a particular set size were displayed. Each matrix contained the same set of all nine possible stimuli for a given task. The participants were required to select with the mouse those stimuli that had been presented in a sequence, in the correct order. There was no time limit for responding. The dependent variable for each complex span task was the proportion of correctly selected stimuli of the 90 stimuli presented in the task.

The WM updating task. To assess the updating function of WM, we used just one task—mental counters (Larson, Merritt, &

Table 3
Description of Matchstick Equations Applied in the Study

Number	STandard/ INSight	Description	Correct solution(s)	Accuracy	Insight accuracy	ΔWMC solvers	ΔWMC insight
1	ST	X = VII + V	XI = VI + V	.88	.19	**1.09**	.31
2	skipped	I − VII = VII	+VII = VII/−VII = −VII	.89	.19	**1.17**	.29
3	ST	VII − X = III	VII = X − III	.72	.27	**.85**	.12
4	INS	IV = III − VI	−IV = II − VI	.14	.24	**.75**	.12
5	INS	V = XI − I	V = VI − I	.44	.13	**.46**	.31
6	ST	X + IV = V	X − IV = VI	.91	.17	**1.17**	.13
7	INS	VI + VI = VI	VI = VI = VI	.32	.35	**.67**	.36
8	ST	XI = V − V	X = V + V	.91	.21	**1.17**	.07
9	INS	II − III = II	II = II = II	.32	.24	**.74**	.08
10	ST	V−II = VI	V + I = VI	.82	.18	**.86**	.19
11	ST	V + III = I	V−III = II	.93	.19	**1.27**	.30
12	INS	IV = V + IV	IX = V + IV	.41	.19	**.50**	**.57**
13	INS	I − I = I	−I = −I/+I = I	.34	.35	**.57**	.04
14	ST	V = II − III	V − II = III	.71	.24	**.78**	.14
15	ST	IV + II = VIII	VI + II = VIII	.90	.17	**1.25**	.20
16	skipped	XI = I − XI	XI = +XI/− XI = −XI	.96	.17	**.93**	.30

Note. Differences in working memory capacity (WMC) with $ps < .01$ are in bold.

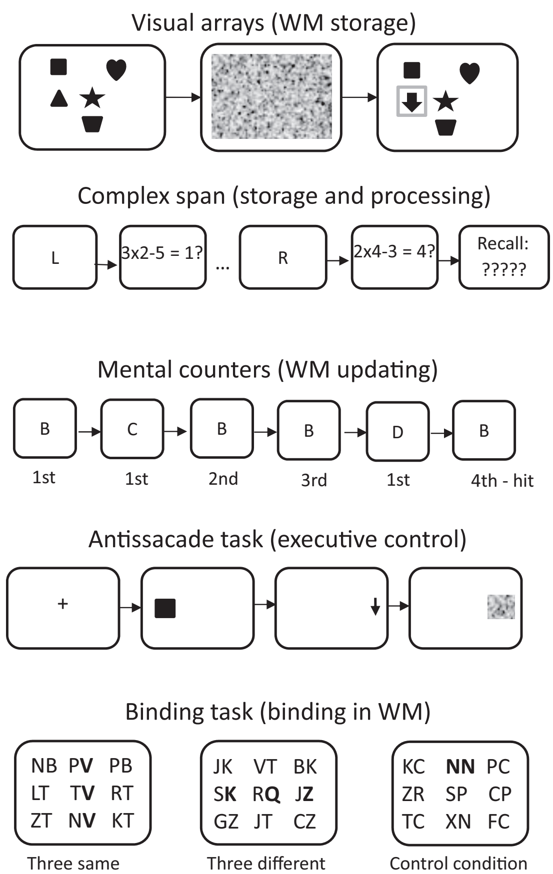


Figure 1. Schematic illustration of sample tasks applied to measure different working memory (WM) functions. From the top to bottom: the storage, storage and processing, updating, executive control, and binding tasks. Letters in the binding task are marked in bold to better present each task conditions (bold was not present in the actual task).

Williams, 1988) – with two conditions loading the updating variable. There were two reasons for using just this task. First, it is the best of several existing updating tasks at conceptually reflecting the updating of information (it requires updating of the number of times a particular stimulus appears in a sequence). Other tasks, such as n-back (Kirchner, 1958) or the running memory task (Pollack, Johnson, & Knaff, 1959), either have been criticized for assessing processes beyond WMC (Chuderski & Necka, 2012; Kane et al., 2007; Roberts & Gibson, 2002) or, on the other hand, have reflected the variance isomorphic to nonupdating WM tasks (e.g., Chuderski, Taraday, Necka, & Smolen, 2012; Ecker, Lewandowsky, Oberauer, & Chee, 2010; Schmiedek, Hildebrandt, Lövdén, Lindenberger, & Wilhelm, 2009). Second, in a previous study of one of us (Chuderski et al., 2012), the mental counters task yielded an internal consistency of $\alpha = .97$ and loaded at $r = .902$ on the updating latent variable (strongly surpassing the n-back and the keep track task; Yntema & Mueser, 1962). Thus, the mental counters task accounts for most of the updating variance. The task consisted of a sequence of 120 single letters (B, C or D) presented in random order in the center of the screen (plus 12 training letters). The goal was to count how many times each letter had been presented, up to four (see Figure 1). Each time a letter was presented, participants provided a response by pressing one of the two response keys to indicate whether they had seen the letter for the fourth time (target) or less than four (nontarget). So, after each fourth occurrence, participants should have reset their mental counter of a particular letter to zero. In order to help resetting the counters, the participants were presented with feedback after false alarms and missed targets. Each letter, approximately 2 × 2 cm in size, was displayed for 2.5 s, followed by a mask displayed for 0.3 s. The final scores were the proportion of correct hits in the target trials and correct rejections in the nontarget trials.

Executive control tasks. Three tasks were used to measure executive control: the antisaccade (Hallett, 1978), the Stroop (see MacLeod, 1991), and the go/no-go (Eimer, 1993). The antisaccade task consisted of five training and 50 test trials. Each test trial consisted of four events (see Figure 1). First, a cue was presented for 1.5 s, reminding participants that a target would be presented in the middle of the side opposite from the flashing square. Next, a fixation point was presented at the center of the screen for between 1 s and 2 s. Then, a rapidly flashing black square (3 cm in size) was shown on the left or right side of the screen, about 16 cm from the fixation point, for 0.2 s. Finally, a small dark gray arrow (0.6 cm long), pointing left, down, or right, was presented in the location opposite the square for 0.15 s and was then replaced by a mask. The visual angle from both the square and the arrow to the fixation point was around 15°. The task was to look away from the flashing square, to detect the direction of the arrow, and to press the arrow key matching the stimulus arrow. The trials were self-paced. The dependent variable in the task was mean accuracy.

The stimuli in the classic color-word variant of the Stroop task were four capital Polish words (approx. 7 cm × 2 cm in size), printed in bold Arial font, naming the colors red, green, blue, and brown. Each word could be displayed in any of the four ink colors. For congruent stimuli (random 200 trials), the meaning of the word and the color of the ink were the same. For incongruent stimuli (another 50 trials), the former differed from the latter. There was a constraint that the direct repetition of the target stimuli was not possible. Trials lasted until a response was given, or for a maximum of 2.2 s, and a 0.5-s mask separated the subsequent trials. The instruction was to avoid reading a word and to press a response key (*Z*, *X*, *N*, or *M*) that was assigned to a respective color. The stimuli were displayed in the center of the screen, and a reminder of the response buttons was displayed at the bottom. The task was preceded by a training session that included 20 neutral stimuli (HHHHHH strings), aimed at strengthening the stimulus-key associations. The dependent variable was the difference between mean accuracy in the incongruent and congruent trials.

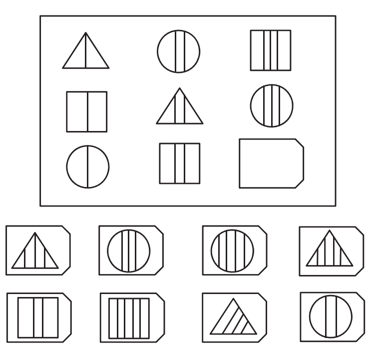
In the go/no-go task, digits (1 or 2) were presented sequentially in the center of the screen. The task was to categorize the presented number by pressing one of the two respective keys when the digit was displayed in a color other than green (red, blue, or yellow; go trials), but to withhold the response if the presented digit was green (no-go trials). Trials lasted until a response was given, or for a maximum of 0.9 s, and each trial was followed by a 0.1-ms mask. There were 325 go and 25 no-go trials. The score on the task was mean accuracy on no-go trials.

Binding task. Three variants of the relation monitoring task, originally introduced by Oberauer et al. (2007), and later modified by Chuderski (2014), were used to measure the capacity to set and maintain bindings between information held in WM. The task consisted of the presentation of a continuous sequence of stimulus patterns (trials). Each trial included a 3 × 3 array (approx. 6 × 6 cm in size) of pairs of stimuli (see Figure 1). The stimuli were Greek letters, digits or figures (e.g., triangle, square) in letter, digit, and figure variants, respectively. On each trial, only one (random) of the three following relations between the stimuli was present: there were three pairs in a row or a column ending with the same stimulus (three-same), three pairs in a row or a column ending with a different stimulus (three-different), or there was a pair of identical stimuli, which was a control condition. The task was to detect the relation in an array, and respond by pressing one of the three respective keys (a reminder of the response keys was displayed at the bottom of the screen). To decrease the amount of information changed from trial to trial, four pairs (at random) in each subsequent array were the same as in the preceding array. Trials lasted until a response was given, to a maximum of 10 s. Each task consisted of 120 trials (and 10 training trials). The total score on each task was the mean accuracy on the three-same and three-different trials. As the task did not require active storage of stimuli in WM (which were present on the screen for the total duration of a trial), but it demanded detecting the relationships between selected elements, the task is assumed to capture the ability to construct bindings in WM (for justification see Chuderski, 2014; Oberauer et al., 2007).

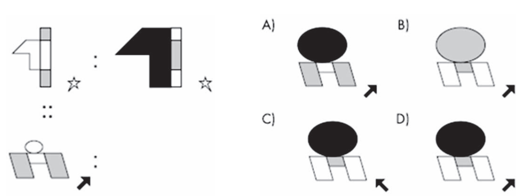
Fluid reasoning tests. Four tests of fluid reasoning were used. Two tests included figural material. The well-known Raven's Advanced Progressive Matrices (Raven, Court, & Raven, 1983) consists of items that include a three-by-three matrix of figural patterns which is missing the bottom-right pattern, and eight response options comprising the patterns that can potentially match the missing one. The goal is to discover the rules that govern the distribution of patterns and to apply them to the response options in order to choose the single correct pattern. The Figural Analogies Test (Chuderski & Necka, 2012) consists of analogies in the form of "A is to B as C is to X," where A, B, and C are relatively simple patterns of figures. A is related to B according to two, three, four, or five rules (e.g., symmetry, rotation, change in size, color, thickness, number of objects), and X is an empty space. The task is to choose one figure from a choice of four which relates to Figure C, as B relates to A. In each test, 18 odd-numbered items out of 36 original items were used, printed on several sheets of paper; the administration time was half of the original (20 min and 15 min, respectively). In the newly designed number series test (Number series), the task is to find the hidden rule according to which a sequence or an array of numbers is constructed, and to complete the sequence or the array with the missing number. Participants were given 18 min to solve 18 number series problems. The fourth reasoning test (logic) required deductive reasoning in order to infer one out of four (one logically valid and three invalid) conclusions from a set of premises in syllogistic, propositional, and predicate calculi. Premises were described in natural language. Some incorrect conclusions were believable (consistent with common knowledge) and this made the task of choosing the logically correct answer even more difficult. The score on the test was the sum of correct conclusions selected in all 20 items within 20 min. There were 74 fluid reasoning test items in total (see Figure 2 for sample items).

Personality questionnaires. We aimed to measure three latent variables reflecting personality traits—extraversion, neuroticism, and psychoticism—assumed by the PEN model of personality (Eysenck, 1990). However, we used more recent scales for measuring the traits; the Polish adaptations of the NEO-Five Factor Inventory (NEO-FFI; Costa & McCrae, 1992), Behavioral Inhibition/Activation System scales (BIS/BAS; Carver & White, 1994), and the brief Dark Triad Scale (Jonason & Webster, 2010), the latter measuring Machiavellianism, narcissism, and psychop-

Raven Matrices



Figural Analogies



Number Series

1	3	5	7
9	12	15	18
21	25	29	33
37	42	47	?

Logic Problems

All flowers need water. All roses need water. Therefore:

a) All flowers are roses
b) All roses are flower
c) Some roses are flowers
d) None of the above conclusions is valid

Figure 2. Sample items of each of four fluid reasoning tasks applied in the study.

athy. The extraversion variable was loaded by the NEO-FFI extraversion subscale and BAS, the neuroticism variable included the neuroticism subscale and BIS, and the psychoticism variable combined the agreeableness subscale (reversed) and the Dark Triad score. We also used the Openness to Experience scale from NEO-FFI.

Anxiety and motivation measures. Anxiety was assessed with 12 items of the Polish adaptation of the X-1 part of the State–Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, & Lushene, 1970). X-1 measures state anxiety and contains questions about the current emotional state of a participant, such as "I am relaxed" (reversed scoring) or "I am nervous". Motivation to deal with the tasks was probed with a set of 10 questions, for example, "It was important to me to perform well in the task". The total score on each motivation and STAI measurement was the sum of answers obtained using a 5-point Likert scale.

Procedure. Participants were tested in a psychological laboratory in groups of six people on average. The session was divided into two 4-hr parts with a 45-min lunch break between them. Each part included a 15-min break and several shorter breaks in order to prevent mental fatigue. Cookies and tea/coffee/water were served throughout to maintain optimal levels of glucose, arousal, and hydration. In order to minimize procedure-dependent variance, all the tasks were applied in the following fixed order. The insight, WM, and reasoning tests were intermixed in order to balance such factors as tiredness, boredom, learning effects, and so forth. The tasks for each WM function were grouped, to minimize the time devoted to understanding the task instructions. The first part of the procedure included: mental counters, math, logic, verbal and spatial problems, letter binding, digit binding, figure binding, letter span, digit span, and figure span. After the lunch break, participants completed: letter arrays, figure arrays, color arrays, Raven matrices, analogies, antisaccade, Stroop, go/no-go, matchstick equations, remote associates, number series, and personality questionnaires (computerized variants). Several other tests serving scientific aims unrelated to the present study were also administered during the session, including a computerized transitive reasoning test, an analogical pattern matching task, a probability assessment task, as well as questionnaires of rationality and religiosity. The motivation and anxiety level of each participant was assessed at the beginning, in the middle, and at the end of data collection. Importantly, regardless of the whole-day procedure, there was no significant drop in motivation level, $F(2, 634) = 0.19$, while anxiety even decreased from its first to the second measurement, $F(1, 317) = 30.99$, $p < .001$.

Data screening and elimination. Regarding the insight, WM, and reasoning scores, there were no missing data. Cook's distance revealed no multivariate outliers (all *D*s < 1.0). For anxiety and motivation, single scores were missing for one participant, and they were substituted with the mean from the two remaining measurements. For personality questionnaires, some scores were missing for 13 participants, and they were not substituted, so Analysis 5 used data from 305 people.

Model fitting. The fit of CFA/SEM models was evaluated by three indices of fit (see Hu & Bentler, 1999; Kline, 1998): the chi-square statistic (its value divided by the number of degrees of freedom should not surpass $\chi^2[2.0]$), the comparative fit index (CFI should exceed .92), and the root mean square error of approximation (RMSEA should be less than .08). Comparing the well-fitting initial model with alternative models (i.e., either more constrained models, with some parameters fixed, or the group models, in which parameters were calculated separately for two groups), we investigated whether a simpler model (with more degrees of freedom) yielded a worse fit, as indicated by a significant increase in the chi-square statistic (if it did not, then this simpler model was preferred).

Results

Table 4 provides descriptive statistics and reliabilities for all 44 measures used in the CFA and SEM analyses. As many as 37 scores fit the widely accepted rule-of-thumb criteria of normality (skew <2.0, kurtosis <4.0). All four insight accuracy measures slightly deviated from normality due to floor effects, which are commonly reported for subjective measures (e.g., Fleck & Weisberg, 2013). Also, the Stroop and go/no-go scores slightly deviated from normality due to ceiling effects, reflecting the common psychometric problems with these measures (see Chuderski, 2014). The same observation applied to the nontarget accuracy in the mental counters task. As none of these deviations from normality was substantial, and the maximal likelihood estimation is relatively robust, we decided not to transform the deviating data in any way. Reliabilities for most of the tasks were acceptable or high (αs $> .65$), with the exception of verbal ($\alpha = .42$) and spatial problems ($\alpha = .48$), as well as the corresponding insight accuracy and analysis accuracy measures. This fact indicated the above mentioned psychometric inadequacy of classical insight problems (probably related to their idiosyncrasy). However, reliability for all the 12 classical problems was higher ($\alpha = .58$).

Accuracy on single insight problems (see Tables 1-3) suggested that most of the problems were sufficiently difficult for participants, with accuracy oscillating around 33%. Problem *Ten dots*, solved only by three participants ($M = .01$), was the most difficult (but was also the last problem on the paper sheet), whereas *Four coins* was the easiest ($M = .61$).

Finally, Table 5 presents the correlation matrix for all variables used in Analyses 1 through 4. Table 6 shows data used in Analysis 5. These data were used to test the consecutive CFA and SEM models presented below.

Relationships of the specific WM functions to insight and reasoning ability (Analysis 1). The results of CFA, which modeled seven factors (the storage, storage & processing, updating, executive control, binding, insight ability, and reasoning ability latent variables) are presented in Figure 3, and the correlations between the WM factors are presented in Table 7. The model had a good fit to the data, $\chi^2(209) = 357.63$, $\chi^2(1.71)$, CFI = .958, RMSEA = .048 [.039–.056]. Inspection of the factor loadings on respective measures indicated that, for the insight ability variable, all tests had satisfactory validity (loadings of at least .565), especially given the low number of items that each test included. The exception was the remote associate test that, despite counting as many as 15 items, showed low validity (.395). For reasoning ability, the two figural tests yielded slightly higher loadings than the two alphanumerical tests, but all loadings were satisfactory (.693 or higher). All WM measures properly loaded the respective variables (.633 or higher), except for the loadings of the Stroop (.459) and go/no-go tasks (.376) on executive control, probably due to their weak psychometric quality. As expected, all WM

Table 4
Descriptive Statistics and Reliabilities for all Measures Used in the Study

Measure	Variable	Mean	*SD*	Min.	Max.	Skew	Kurt.	α
Verbal problems	Insight	.46	.25	.00	1.00	.08	−.65	.42
Spatial problems	Insight	.21	.19	.00	.83	.66	−.2	.48
Math problems	Insight	.30	.26	.00	1.00	.86	−.03	.67
Matchstick equations	Insight	.33	.30	.00	1.00	.63	−.64	.73
Remote associates	Insight	.40	.22	.00	1.00	.13	−.45	.76
Verbal problems—insight acc.	Insight’	.09	.17	.00	.83	2.14	4.64	.67
Spatial problems—insight acc.	Insight’	.06	.12	.00	.50	2.03	3.69	.43
Matchstick equations—insight acc.	Insight’	.08	.15	.00	.83	2.15	4.56	.61
Remote associates—insight acc.	Insight’	.07	.14	.00	.80	2.41	6.16	.83
Verbal problems—analysis acc.	Analysis	.37	.23	.00	1.00	.39	−.4	.37
Spatial problems—analysis acc.	Analysis	.15	.16	.00	.83	.99	.69	.34
Matchstick equations—analysis acc.	Analysis	.25	.24	.00	1.00	.72	−.42	.59
Remote associates—analysis acc.	Analysis	.32	.20	.00	1.00	.41	−.26	.73
Raven	Reasoning	.61	.18	.00	1.00	−.53	.41	.76
Analogies	Reasoning	.67	.18	.06	1.00	−.48	−.26	.74
Number series	Reasoning	.51	.21	.06	1.00	−.08	−.59	.79
Logic	Reasoning	.53	.16	.15	.90	.16	−.42	.72
Letter arrays	Storage	3.32	1.63	−.73	6.60	−.36	−.64	.89
Color arrays	Storage	3.99	1.46	−.38	6.80	−.67	.1	.89
Figure arrays	Storage	3.15	1.61	−.60	6.69	−.34	−.61	.89
Letter complex span	S&P	.71	.17	.02	.98	−1.01	1.19	.86
Digit complex span	S&P	.77	.14	.09	1.00	−1.33	2.94	.84
Figure complex span	S&P	.64	.18	.09	.99	−.6	.26	.89
Mental counters nontargets	Updating	.77	.19	.03	.99	−2.3	5.13	.86
Mental counters targets	Updating	.30	.15	.00	1.00	.48	.95	.70
Antisaccade	Control	.51	.22	.04	.96	−.08	−.92	.91
Stroop	Control	−.18	.22	−.98	.05	−2.16	6.93	.83
Go/no-go	Control	.79	.18	.00	1.00	−2.3	6.43	.91
Letter binding	Binding	.60	.22	.09	.98	−.22	−1.04	.96
Digit binding	Binding	.68	.19	.05	.99	−.63	−.33	.95
Figure Binding	Binding	.70	.18	.01	.99	−.98	.72	.95
NEO–Extraversion	Extraversion	.65	.12	.33	1.00	.03	−.2	.80
NEO–Neuroticism	Neuroticism	.61	.16	.20	.98	.02	−.45	.88
NEO–Agreeableness (reversed)	Psychoticism	.32	.11	.07	.67	.42	.09	.75
NEO–Openness to Experience	Openness	.69	.11	.43	1.00	.13	−.53	.71
BAS	Extraversion	.74	.10	.48	1.00	−.15	−.36	.74
BIS	Neuroticism	.76	.14	.32	1.00	−.63	.53	.73
Dark Triad	Psychoticism	.48	.15	.20	1.00	−.17	−.26	.89
STAI—1. Measurement	Anxiety	.41	.14	.20	.98	1.05	1.45	.91
STAI—2. Measurement	Anxiety	.37	.12	.20	.87	1	1.38	.90
STAI—3. Measurement	Anxiety	.38	.12	.20	.82	.91	.85	.91
Motivation—1. Measurement	Motivation	.79	.12	.40	1.00	−.51	−.02	.81
Motivation—2. Measurement	Motivation	.79	.13	.30	1.00	−.8	.73	.86
Motivation—3. Measurement	Motivation	.79	.13	.20	1.00	−.76	.94	.85
Age	Age	24.45	6.02	18.00	46.00	1.83	3.08	—

Note. $N = 318$ for all the tasks, except for NEO, BAS, BIS and Dark Triad measures for which $N = 305$. All variables, except for the k values in the visual array tasks, were normalized (expressed as a proportion of maximal possible score). Kurt. = kurtosis; acc. = accuracy; BAS = behavioral activation; BIS = behavioral inhibition; S&P = storage and processing.

functions correlated positively and strongly (all $ps < .001$) with the lowest link between storage and binding ($r = .476$), and the strongest between executive control and storage and processing ($r = .809$). However, no link approached unity, so all WM functions were clearly identifiable.

The positive correlations between the consecutive WM functions and insight ability were all significant at $p < .001$ and varied between $r = .491$ (storage) and $r = .759$ (executive control). Analogous links with reasoning were stronger by $\Delta r \approx .15$, with binding being the weakest correlate ($r = .675$), while executive control was the strongest ($r = .920$). As the magnitude of differences in correlations for particular WM functions was limited, the invariance model was tested. This model assumed that the loadings of all five WM functions on the insight ability variable were equal, and that the loadings of these functions on the reasoning ability variable were equal (but different than the loading on insight). Such a model had an additional eight degrees of freedom, but its fit worsened significantly, $\Delta\chi^2(8) = 22.53$, $p = .004$, so it was rejected.

Interestingly, the correlation between the insight and reasoning ability variables was as strong as $r = .920$, suggesting that these variables might not be distinguishable, and thus any further comparisons between the WM contributions to insight versus reasoning might be meaningless. However, due to a negligible standard error, the model which assumed a correlation of $r = 1.0$ between both variables yielded a significant loss of fit, $\Delta\chi^2(1) = 11.02$, $p <$

Table 5
Correlation Matrix of Measures Used in Analysis 1 Through 4

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1. Verbal problems	—																													
2. Spatial problems	.37	—																												
3. Math problems	.39	.46	—																											
4. Matchstick	.38	.40	.54	—																										
5. Remote associates	.32	.20	.26	.26	—																									
6. Verbal problems—insight acc.	.40	.16	.17	.21	.20	—																								
7. Spatial problems—insight acc.	.25	.54	.28	.21	.20	.39	—																							
8. Matchsticks—insight acc.	.28	.23	.39	.59	.30	.33	.33	—																						
9. Remote assoc.—insight acc.	.18	.08	.16	.14	.41	.30	.22	.31	—																					
10. Verbal prob.—analysis acc.	.76	.28	.30	.25	.19	−.28	−.01	.06	−.02	—																				
11. Spatial prob.—analysis acc.	.27	.80	.35	.32	.09	−.09	−.07	.04	−.06	.34	—																			
12. Matchsticks—analysis acc.	.29	.35	.42	.86	.13	.06	.05	.10	−.02	.27	.37	—																		
13. Remote assoc.—analysis acc.	.22	.15	.16	.18	.78	.01	.06	.11	−.25	.22	.14	.15	—																	
14. Raven	.38	.50	.55	.44	.29	.20	.28	.31	.17	.26	.40	.35	.19	—																
15. Analogies	.37	.45	.51	.48	.26	.23	.26	.33	.13	.22	.35	.38	.19	.67	—															
16. Number series	.36	.34	.50	.45	.32	.19	.26	.30	.18	.25	.22	.36	.22	.57	.55	—														
17. Logic	.42	.44	.58	.42	.31	.27	.32	.42	.22	.25	.29	.26	.18	.54	.48	.52	—													
18. Letter arrays	.16	.30	.31	.27	.20	.13	.23	.15	.07	.08	.19	.24	.16	.48	.43	.44	.32	—												
19. Color arrays	.21	.28	.29	.33	.29	.17	.20	.25	.11	.10	.19	.25	.23	.52	.52	.48	.34	.65	—											
20. Figure arrays	.12	.22	.25	.30	.23	.06	.17	.22	.12	.08	.14	.23	.16	.46	.41	.44	.29	.66	.72	—										
21. Letter complex span	.26	.36	.36	.28	.20	.13	.23	.23	.09	.18	.26	.21	.15	.50	.45	.46	.45	.43	.49	.43	—									
22. Digit complex span	.17	.31	.34	.30	.19	.06	.18	.15	.05	.13	.24	.27	.17	.43	.40	.50	.36	.48	.52	.48	.67	—								
23. Figure complex span	.16	.34	.39	.34	.21	.11	.22	.25	.06	.10	.24	.26	.18	.53	.55	.51	.44	.49	.59	.52	.67	.68	—							
24. Men. counters nontargets	.22	.19	.30	.25	.20	.14	.18	.20	.08	.14	.10	.18	.15	.36	.46	.35	.36	.21	.29	.24	.33	.24	.30	—						
25. Men. counters targets	.27	.32	.35	.33	.23	.16	.25	.23	.12	.17	.20	.27	.16	.41	.40	.39	.36	.41	.39	.33	.38	.34	.39	.48	—					
26. Antisaccade	.29	.39	.37	.30	.23	.16	.29	.19	.09	.19	.26	.25	.18	.45	.44	.43	.39	.49	.42	.44	.44	.37	.44	.23	.33	—				
27. Stroop	.24	.29	.25	.18	.11	.14	.14	.12	.07	.16	.24	.15	.07	.40	.38	.30	.34	.22	.27	.20	.29	.27	.33	.20	.23	.28	—			
28. Go/no-go	.05	.12	.17	.17	.14	.01	.09	.07	−.01	.05	.08	.16	.16	.26	.28	.25	.10	.36	.29	.26	.21	.27	.32	.16	.24	.30	.10	—		
29. Letter binding	.23	.33	.33	.19	.23	.18	.28	.23	.18	.11	.20	.09	.12	.45	.37	.33	.42	.26	.33	.29	.36	.30	.38	.22	.34	.32	.28	.05	—	
30. Digit binding	.17	.32	.33	.25	.17	.19	.22	.25	.15	.04	.22	.15	.08	.46	.43	.36	.36	.34	.35	.29	.31	.26	.35	.25	.34	.29	.34	.20	.71	—
31. Figure binding	.21	.36	.38	.29	.24	.20	.25	.28	.12	.08	.24	.18	.17	.51	.51	.44	.44	.41	.43	.35	.42	.37	.44	.30	.39	.39	.38	.22	.63	.77

Note. Nonsignificant ($ps > .05$) rs are underlined. $N = 318$.

Table 6
Correlation Matrix of Measures Used in Analysis 5

Measure	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1. Verbal problems	—																										
2. Spatial problems	.36	—																									
3. Math problems	.39	.45	—																								
4. Matchstick equations	.36	.38	.53	—																							
5. Remote associates	.30	.18	.25	.24	—																						
6. Raven	.36	.49	.54	.42	.28	—																					
7. Analogies	.35	.44	.50	.46	.25	.66	—																				
8. Number series	.34	.32	.49	.43	.30	.55	.52	—																			
9. Logic	.39	.42	.58	.40	.30	.52	.46	.49	—																		
10. WM—Storage	.16	.29	.32	.32	.26	.53	.50	.51	.34	—																	
11. WM—Storage and processing	.20	.36	.39	.33	.21	.53	.51	.54	.45	.64	—																
12. WM—Updating	.27	.27	.37	.31	.23	.42	.47	.40	.40	.40	.39	—															
13. WM—Executive control	.28	.38	.37	.30	.23	.52	.51	.45	.38	.55	.52	.35	—														
14. WM—Binding	.20	.37	.37	.25	.23	.52	.48	.40	.42	.43	.43	.40	.42	—													
15. NEO–Extraversion	.07	.07	.04	.04	.07	.05	.06	.05	.04	.03	.06	.07	−.06	.15	—												
16. NEO–Neuroticism	−.16	−.12	−.09	−.10	−.06	−.08	−.08	−.05	−.12	−.07	−.07	−.18	.00	−.14	−.41	—											
17. NEO–Agreeableness (reversed)	.15	−.05	.06	.04	.03	−.01	−.05	−.02	−.02	.01	−.06	.11	.04	−.11	−.13	−.01	—										
18. NEO–Openness to experience	.11	.15	.07	.14	.07	.14	.14	.00	.22	.02	.03	.04	.10	.18	.09	−.04	.06	—									
19. BAS	.06	−.01	.01	.00	.06	.02	−.04	−.03	.02	−.03	−.02	.00	−.03	.00	.44	−.14	.27	.17	—								
20. BIS	.02	.00	−.01	−.03	.15	.04	.03	.08	.07	.10	.05	−.05	.06	.01	−.27	.62	−.21	−.05	−.10	—							
21. Dark Triad	.15	−.00	.12	.12	.04	.05	−.00	.04	.06	−.05	−.04	.13	−.02	−.00	.04	.03	.57	−.09	.35	−.09	—						
22. STAI—1. Measurement	−.03	−.11	.01	−.05	−.05	−.06	−.03	−.04	.04	−.05	−.06	−.05	−.02	−.15	−.25	.38	.13	.04	−.17	.15	.07	—					
23. STAI—2. Measurement	.03	−.13	.02	−.04	−.03	−.05	−.03	−.02	.05	−.10	−.14	−.03	−.11	−.18	−.21	.33	.17	−.01	−.07	.14	.11	.62	—				
24. STAI—3. Measurement	.05	−.15	−.01	−.05	−.07	−.12	−.12	−.07	.03	−.12	−.13	.00	−.11	−.19	−.26	.30	.19	−.02	−.18	.12	.11	.58	.72	—			
25. Motivation—1. Measurement	.15	.14	.15	.14	.17	.12	.16	.22	.10	.15	.12	.06	.07	.21	.19	−.19	−.16	−.04	.11	.03	−.24	−.32	−.34	−.34	—		
26. Motivation—2. Measurement	.00	.04	.03	.06	.11	.04	.10	.10	−.07	.08	.09	−.01	.07	.24	.04	−.09	−.14	−.08	−.04	.03	−.12	−.21	−.42	−.37	.65	—	
27. Motivation—3. Measurement	.05	.14	.12	.17	.15	.17	.25	.16	.01	.09	.12	.03	.05	.16	.08	−.04	−.11	.04	.06	.06	−.19	−.17	−.24	−.41	.48	.60	—
28. Age	−.08	−.14	−.20	−.14	.10	−.23	−.28	−.19	−.21	−.10	−.22	−.33	−.27	−.23	−.11	−.02	−.06	−.04	−.16	.03	−.10	−.01	.01	.09	.07	.16	.10

Note. Non–significant ($ps > .05$) rs are underlined. $N = 305$. WM = working memory; BAS = behavioural activation; BIS = behavioural inhibition. STAI = anxiety measure.

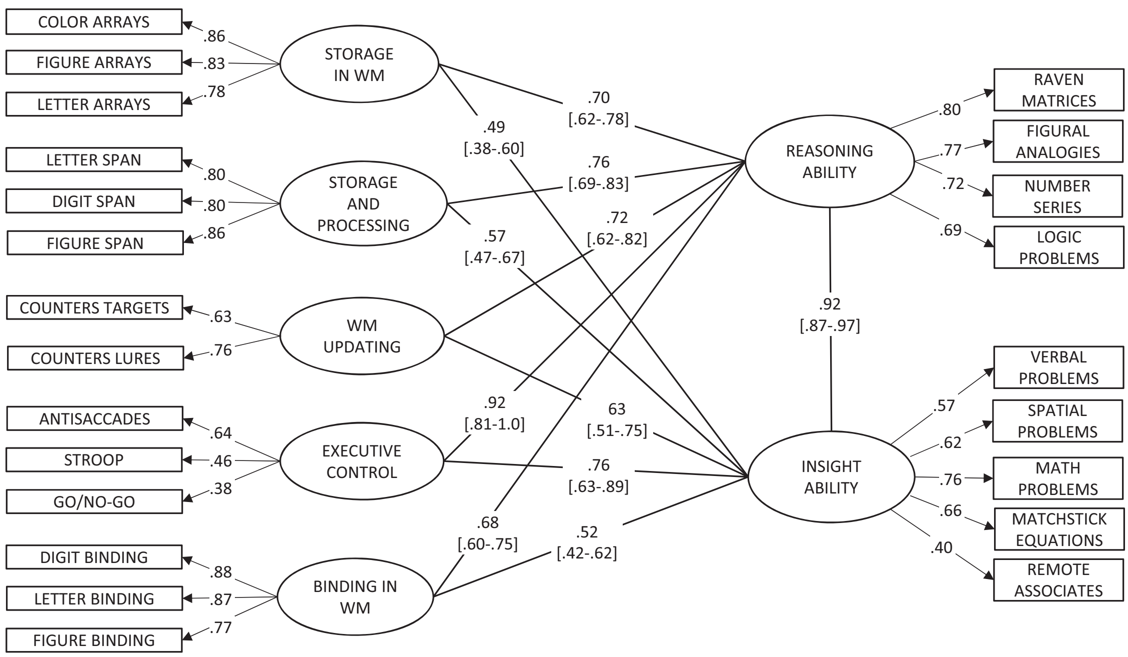


Figure 3. The confirmatory factor analysis (CFA) model from Analysis 1. Boxes represent tasks and ovals represent latent variables. Values between ovals and boxes represent relevant standardized factor loadings (all *ps* $< .001$). Values between ovals represent correlation coefficients among latent variables (all *ps* $< .001$), except correlations between the working memory (WM) functions which are presented in Table 7. Values in brackets indicate 95% confidence intervals.

.001, and so was not tenable. Thus, we assumed that both variables are separate, though extremely strongly intercorrelated, sharing as much as 84.6% of their variance.

The strength of relationship of the high-level WMC factor to insight and reasoning ability (Analysis 2). In order to express in the model the general effectiveness of WM in a wide range of WM tasks, we introduced another latent variable that loaded all five latent variables reflecting the WM functions investigated. The model also included the insight and reasoning ability variables correlated with WMC. The fit of the resulting model (see Figure 4) was good, $\chi^2(222) = 386.26$, $\chi^2(1.74)$, CFI $= .953$, RMSEA $=$ .050 [.042–.058]. The loadings of WM functions on WMC varied between $r = .676$ (binding) and $r = .990$ (executive control). The WMC factor was strongly linked to insight ($r = .720$), and very strongly to reasoning ($r = .933$), though the latter link was significantly below unity, $\Delta\chi^2(1) = 11.59$, $p < .001$. The relationship to WMC of reasoning was significantly stronger than the relationship of WMC with insight, as equating them substantially decreased the fit of the model, $\Delta\chi^2(1) = 30.07$, $p < .001$.

Table 7
Correlations Between Working Memory (WM) Factors in the Confirmatory Factor Analysis Model From Analysis 1

WM factor	1	2	3	4
1. Storage	—			
2. Storage and processing	.74	—		
3. Updating	.56	.58	—	
4. Executive control	.80	.81	.67	—
5. Binding	.49	.51	.53	.66

Note. All *ps* $< .001$.

Next, by means of SEM, we used WMC to predict the insight and reasoning ability variables, in order to investigate the correlation between the residual variances of these variables (i.e., disturbance terms, reflecting for each variable the variance unexplained by WMC). The resulting model had the same structure as the CFA model and therefore the same goodness of fit, as well as the same coefficients for links between insight and WMC, and reasoning and WMC. In this model, the insight and reasoning disturbances correlated weakly ($r = .248$, $p < .001$), meaning that only 6.1% of insight ability variance unaccounted for by WMC was shared with reasoning ability. So, the strong link between insight and reasoning was primarily due to their common WMC mechanisms.

It is possible that various insight problems depend differently on WM (see Batchelder & Alexander, 2012; Fleck, 2008), so the WMC correlation of a given insight test might be driven by a small number of its items, while most items might be unrelated to WMC. To assess the role of WMC for each individual insight problem and matchstick equation, the WMC factor value was compared between solvers and nonsolvers of a given item, and the effect size was expressed as the difference in *SD* (see the respective columns of Tables 1 through 3). For most of these items (19 out of 25), solvers had a higher WMC at $p < .01$ or less. Averaging across all 25 insight problems/equations, the WMC of solvers surpassed that of nonsolvers ($\Delta M = 0.57$, $SD = 0.28$, $p < .001$). The respective average difference for remote associates was $\Delta M = 0.34$, $SD =$ 0.19, $p < .001$. Thus, the substantial WMC correlations of the five

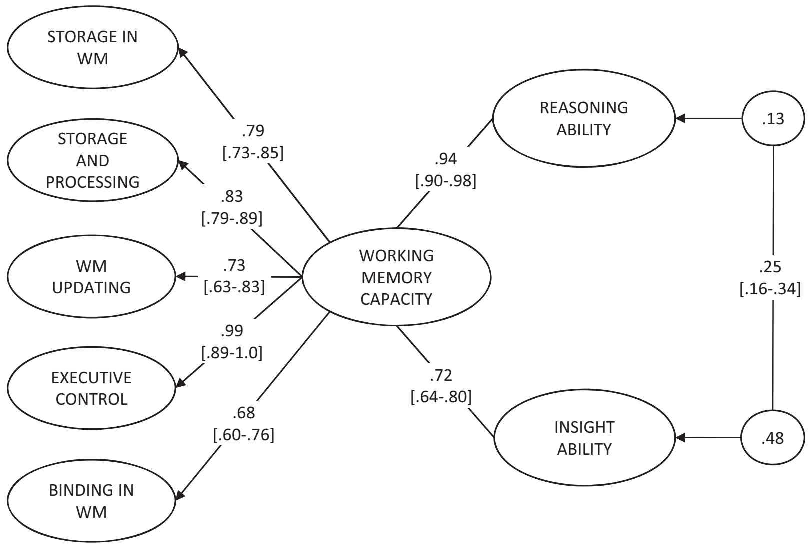


Figure 4. The confirmatory factor analysis (CFA) model from Analysis 2. Boxes represent tasks and ovals represent latent variables. Values between ovals represent correlation coefficients among latent variables (all $ps < .01$). Values in brackets indicate 95% confidence intervals.

insight tests were not driven only by their selected items, but reflected their more general dependence on WM.

Relationships of WMC and reasoning ability to problems solved with insight versus by analysis (Analysis 3). In order to establish whether the solutions of insight problems accompanied by (reported) insight differed in their relationship to WMC and reasoning ability, compared with the solutions without insight (i.e., supposedly reached by some incremental, analytical processes), the insight ability latent variable from the CFA model in Analysis 2 was split into two corresponding variables. The first of these was the insight variable loaded by insight accuracy on the verbal and spatial problems as well as the matchstick equations and remote associates (please remember that no subjective reports for the math problems were collected). The second was the analysis variable, loaded by analysis accuracy on the same four tests. The model (see Figure 5) yielded good fit, $\chi^2(288) = 507.45$, $\chi^2(1.76)$, CFI = .937, RMSEA = .049 [.042–.056]. Naturally, the insight' and analysis variables were not significantly correlated ($r = -.062$), as the more a person reported insightful solutions, the less she or he reported the noninsightful solutions. Although the WMC and reasoning correlation coefficients seemed to be numerically higher for the analysis variable, compared with the insight' variable ($\Delta r \approx .10$), their pairwise equaling between insight' and analysis (to $r = .520$ for WMC and $r = .650$ for reasoning) did not bring about any significant loss of fit, $\Delta\chi^2(2) = 2.07$, $p = .355$.

Analogously to Analysis 2, we also compared the WMC of people who solved each single insight problem/matchstick equation using insight with the WMC of those solving it by means of analysis (see the last columns of Tables 1 and 3). The comparison revealed that significant differences (with $p < .01$ of the two-tailed t test) in WMC for insight versus noninsight solvers were observed in only two items (and those who reported insight showed higher

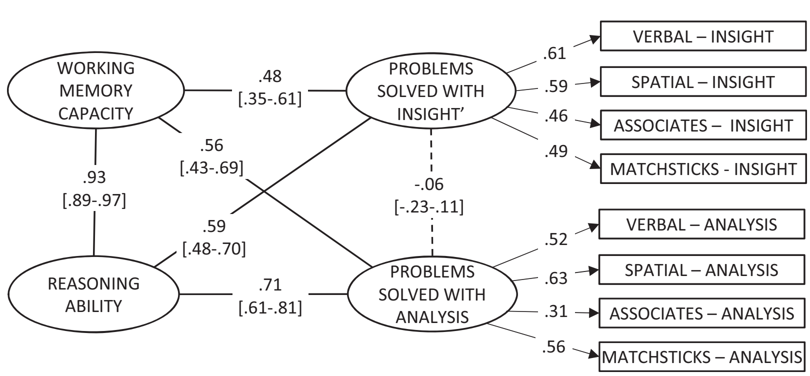


Figure 5. The confirmatory factor analysis (CFA) model from Analysis 3. Boxes represent tasks and ovals represent latent variables. Values between ovals and boxes represent standardized factor loadings (all $ps < .001$), and values between ovals represent correlation coefficients among latent variables. The dashed line indicates a nonsignificant correlation. Values in brackets indicate 95% confidence intervals.

WMC). Averaging across 17 insight problems/equations, the WMC difference in favor of insight solvers was only $\Delta M = 0.24$, $SD = 0.16$ (ns), whereas the average difference for remote associates was even smaller, $\Delta M = 0.12$, $SD = 0.41$ (ns). Thus, whether people declared insight when solving the problems was not apparently related to WMC.

Testing the threshold in the WMC-insight problem-solving relationship (Analysis 4). In Analysis 4, we calculated the CFA model which related the WMC variable only to the insight ability variable, whereas the value of the reasoning ability variable was used as a criterion (breakpoint) for splitting the sample into two groups. Then, although the model assumed invariance in factor loadings, it estimated the WMC-insight ability correlations separately for the groups of people below and above the breakpoint. As existing studies have reported varied breakpoints (see Jauk et al., 2013), we calculated three alternative models, with the breakpoint set at 0.5 standard deviation below the mean in the reasoning ability factor, at that mean, and at 0.5 standard deviation above that mean (i.e., correcting by +5 IQ points for our academic-city, young-adult sample, the thresholds equaled around 97, 105, and 113 IQ, respectively). We tested whether assuming invariance in the WMC-insight ability correlation for the groups would harm the fit in any of the three models (indicating the different relationship of WMC to insight ability between the groups). However, it did not, as the $\Delta\chi^2$ for the invariance models was highly nonsignificant, $\Delta\chi^2(1) = 0.01$, 0.91, and 0.92 (all $ps > .33$), for the 98-, 105-, and 113-IQ model variants, respectively. Thus, it seems that every WMC point counted for insight problem solving ability. For each model variant, we also tested if the invariance of factor loadings of the WMC latent variable on the respective WM functions as well as of the insight variable on the respective insight tests harmed the model fit, as compared with the model variants in which both the WMC-insight ability correlation and the factor loadings could vary freely between the groups. No significant loss of fit was observed in the models assuming invariance, $\Delta\chi^2(11) = 10.31$, 8.24, and 10.60 (all $ps > .47$), respectively.

The WMC contribution to insight ability after accounting for age, personality, anxiety, and motivation (Analysis 5). In order to test if the strong relationship persisted between WMC and insight ability after accounting for age, current levels of anxiety and motivation, and the personality traits of extraversion, neuroticism, psychoticism, and openness to experience, in Analysis 5 we aimed to calculate the SEM model in which the latter seven variables would predict the WMC and insight and reasoning ability variables, and the correlation between the WMC and insight and reasoning disturbance terms would reflect the link after the variance accounted for was extracted. First, we computed the CFA model assuming the seven predictors as intercorrelated latent variables. Unfortunately, such a CFA model yielded unacceptable fit, $\chi^2(61) = 268.06$, $\chi^2(4.39)$, CFI = .852, RMSEA = .103 [.090–.116]. Inspection of the model revealed that eliminating the extraversion and neuroticism variables (note that they were not related to creative problem solving in existing literature) made the fit of this measurement model acceptable, $\chi^2(61) = 268.06$, $\chi^2(3.55)$, CFI = .928, RMSEA = .089 [.070–.110]. See Table 8 for the links among its variables.

In consequence, we calculated the above mentioned SEM model using age, motivation, anxiety, psychoticism, and openness to experience as predictors. Because the hierarchical latent variable of WMC, which was calculated in Analyses 2 through 4, could not be used as an endogenous variable in this Analysis, we submitted each set of three respective WM measures (two in the case of updating) to exploratory factor analysis (all Eigenvalues >1.46), and used the calculated five factors as manifest measures for the WMC factor. The fit of the SEM model (see Figure 6) was also acceptable, $\chi^2(226) = 471.81$, $\chi^2(2.09)$, CFI = .913, RMSEA = .058 [.050–.066]. Unsurprisingly, age negatively predicted performance on WM, insight, and reasoning tasks ($r = -.329$, $-.210$, -336, respectively; note that all rs above $|.17|$ are significant at $p = .05$ level). Also as expected, motivation predicted such performance positively ($r = .223$, .320, .312, respectively). Openness was positively related to both insight problem solving ($r = .358$) and reasoning ($r = .354$) but not to WMC. Psychoticism weakly predicted insight problem solving ($r = .172$) but was unrelated to reasoning ($r = .012$). Anxiety was not significantly related to the three cognitive variables at all ($rs < |.11|$).

Most importantly, after the influence of age, motivation, anxiety, psychoticism, and openness to experience had been partialed out, the residual variances of WMC and insight and reasoning ability still correlated strongly (from $r = .534$ to $r = .728$), losing on average $\Delta r = -.209$ from their respective correlations reported in Analysis 2. This specific observation supports the robustness of relationships among the three cognitive ability variables, which could not be explained away by the most important noncognitive factors.

Table 8
Correlations Between Predictors in the Structural Equation Model From Analysis 5

Predictor	1	2	3	4
1. Openness	—			
2. Psychoticism	.02	—		
3. Motivation	−.11	−.24	—	
4. Anxiety	−.01	.22	−.52	—
5. Age	−.07	−.11	.15	.04

Note. Nonsignificant ($ps > .05$) rs are underlined.

General Discussion

In a study of 318 young and middle aged adults, we administered five several-item tests of insight problem solving, four fluid reasoning tests, and as many as 13 WM tasks assessing the five most important functions of WM. We controlled for age, motivation, anxiety, and the two personality factors that in previous studies were associated with creativity. The insight problems studied showed good convergent validity that allowed for calculation of the insight ability latent variable. One exception was remote associates, which did not share sufficient variance with the other insight ability tests; this finding questions the validity of remote associates as a measure of insightful thinking (for similar concerns see Dailey, 1978; Lee, Huggins, & Therriault, 2014).

Analysis 1 showed that each WM function substantially predicted insight problem solving, although these correlations were weaker than in the case of reasoning. WM functions varied in predictive strength, but the relative advantage of executive control

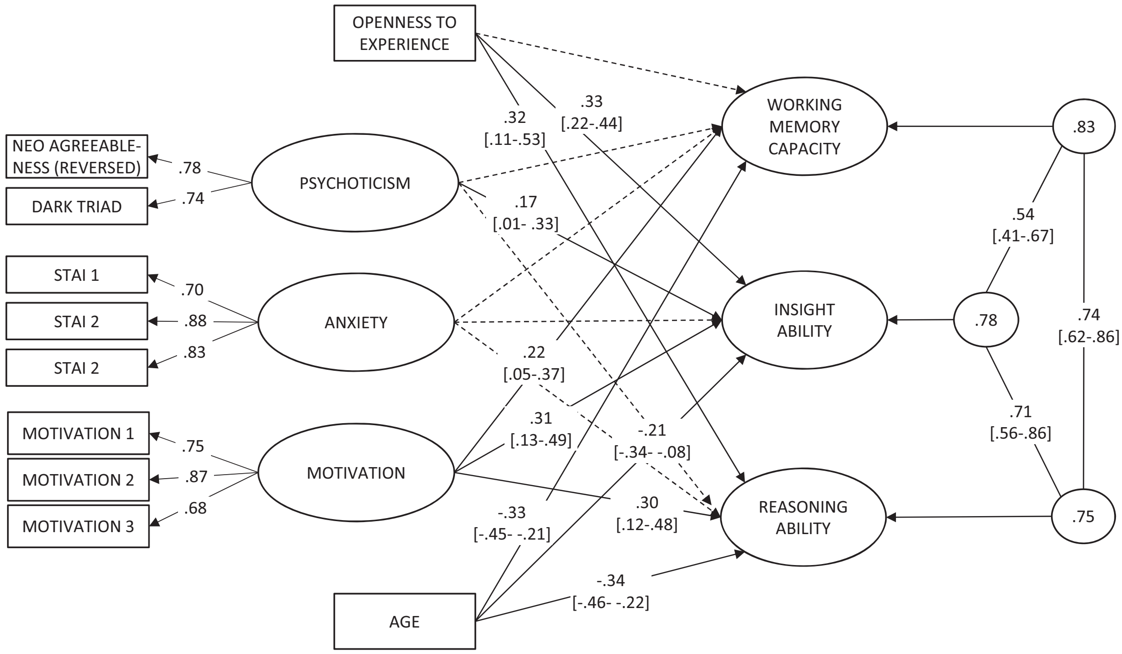


Figure 6. The structural equation model from Analysis 5. Three endogenous latent variables (ovals on the right) reflecting cognitive abilities studied are predicted by three exogenous latent variables (ovals on the left: psychoticism, anxiety, and motivation) as well as openness to experience and age. Boxes represent measures. Values between ovals and boxes represent standardized factor loadings (all *p*s $< .001$). Small circles represent disturbance terms. Numbers in circles indicate the amount of residual variance in each endogenous variable unexplained by all the exogenous variables (predictors). Values between ovals/circles represent standardized path coefficients among variables. Solid lines between ovals indicate *p*s $< .05$; dashed lines –*p*s $> .05$. Values in brackets indicate 95% confidence intervals.

over the other functions was not substantial, and may be related to the fact that executive control was the only WM latent variable with the method-specific variance filtered out (unlike the four remaining WM variables, it was assessed with three conceptually different tasks). We conclude that insight problem solving is related to WM in a general way, and their relationship is not restricted to any particular WM function. Irrespective of which function was used as a predictor, from a quarter (storage) to a half of variance (executive control) in insight problem solving could be explained.

In consequence, Analysis 2 used a general WMC latent variable that loaded each specific WM function, in order to predict insight problem solving versus fluid reasoning. The WMC factor explained 51.8% of insight variance, as well as 87.0% of reasoning variance. Both types of thinking shared above 84.6% of variance; however, the lion's share of this common variance was due to WMC. The strong effects of WMC were revealed not only at the insight ability latent variable level and each insight test level, but were also demonstrated for most single items (insight problems, matchstick equations, remote associates).

Analysis 3 ruled out a possibility that the strong link between WMC and insight problem solving was driven by the items solved by means of some analytic strategy, for which the solvers did not report the sudden and unexpected subjective experience of realizing the solution. Solutions that did or did not yield such an experience were equally strongly related to WMC. Tables 1 and 3 show that the proportion of correct solutions due to (reported) insight, among 17 problems with such data available, was 22.6% (varying between 5% and 35% for particular problems). This outcome matches previous reports (e.g., Fleck & Weisberg, 2013) which suggested that sudden insights in insight problem solving are relatively infrequent. Nevertheless, even if they occur, respective solution rates still strongly depend on WMC.

In Analysis 4, the relationship between WMC and insight problem solving was perfectly linear, with no threshold detected above which this relationship would weaken. Simply speaking, every WMC point counts for insight problem solving ability, exactly as it does for fluid reasoning (Gignac & Weiss, 2015). This result rejects the possibility that the WMC-insight ability link could rely on a minority of outlier participants.

Finally, Analysis 5 demonstrated that the relationship between WMC and insight problem solving remains robust, even after the openness to experience, psychoticism, anxiety, motivation, and age were partialed out. Generally, these variables were weak predictors of insight problem solving and fluid reasoning, in comparison to WMC as a predictor. As expected, openness to experience predicted up to 10% of the variance in both insight and reasoning ability (but not in WMC). A significant but weak link was also replicated between insight problem solving and psychoticism, supporting the renowned theories linking psychoticism and creativity (Eysenck, 1993). That link was found neither for fluid reasoning nor for WMC.

The Validity of Large-Scale Psychometric Methods to the Study of Insight Problem Solving

As, traditionally, most studies on insight have examined, in experimental settings, human performance in just a few insight problems (or even only one such problem), several substantial differences exist between these and the present study. The present study may thus be criticized as measuring insight problem solving in a way not representative of this type of problem. In this section, we discuss several points of such a potential criticism (some of them expressed by reviewers of our already published articles).

For example, one may argue that the CFA/SEM methods applied to many multiitem tests are so powerful that they estimate overly strong correlations between phenomena that might actually be much more weakly correlated (e.g., Ackerman, Beier, & Boyle, 2005). The latent variable calculated, instead of truly representing insight problem solving, might extract some artificial factor out of many problems so idiosyncratic as to lack meaningful commonality. For instance, such a factor might reflect a strategy that a person tended to use in most of the problems, her/his particular way of understanding the instructions (e.g., how easily one was trapped by misleading cues), or her/his task commitment. This possible argument may be rebutted in three ways. First, other noncognitive phenomena included in our models, especially motivation, did not yield substantial correlations regardless of also being latent variables. So, just calculating a latent variable does not suffice to account for the almost perfect correlations reported for the cognitive factors. Second, when the consecutive measures of each ability were just averaged, the mean score of the five insight tests still strongly correlated with both the WMC (r = .562) and reasoning ability (r = .666) mean scores. Third, even when we averaged scores on the first items of each test (i.e., single problems attempted with no preexposure to problems of that type), such a 5-item insight measure still substantially predicted both the WMC (r = .456) and reasoning (r = .550) summary scores. In sum, the factor analysis cannot take the credit for the reported strong correlations between insight problem solving and WMC (and fluid reasoning).

However, although fully convergent with the CFA/SEM analyses, unlike them, the second and third method above underestimated the true relationship between WMC and insight, due to limited reliability of insight tests. For example, the reliability of one 25-item compound (a "metatest") of insight problem solving (including all the verbal, spatial, math, and matchstick items, with remote associates skipped because of their low validity) was imperfect (α = .81). Analogous reliability (α = .89) of the WMC compound (with single tasks serving as "items") was higher but also below unity. So, even if both constructs had been perfectly isomorphic ("real" r = 1.0), the imperfect reliability of their measures would have made their correlation only r = .849 (i.e., the square root for .81 times .89). CFA/SEM helps to resolve this problem.

Another possible line of criticism might argue that the requirement for continuous monitoring of participants' thoughts in order to report whether insight did or did not occur might promote a more analytic way of problem solving (more reflection, more metacognition). This argument cannot stand, as the math problem performance that required no monitoring and had been administered before the other insight tests were administered correlated with WM more strongly (r = .497) than the monitored verbal (r = .375) and spatial problems (r = .465). So, the present results cannot be attributed to the need to report insights.

Many problems of each type were administered; this could also be criticized because of possible learning effects that might have favored high-WMC participants. For example, facing a sequence of problems containing a similar hidden clue (e.g., each math problem required a part of its description to be ignored, enabling a simpler calculation), high-WMC people might recognize the clue more quickly. However, this argument can be rebutted on three grounds. First, the strong correlation between WMC and the initial items of each insight test, as well as the observation that solvers of such items had a significantly larger WMC than nonsolvers (see Tables 1 through 3), suggest that performance on even the initial items before any learning could occur was substantially associated with WMC. Second, having already learned how to deal with problems of a given type, the participants would be unlikely to report insights. So, it might be reasonable to assume that solutions with reported insights did not result from learning; nevertheless, they substantially correlated with WMC. Third, another study by us (Chuderski & Jastrzębski, 2017) divided the insight problems into two separate sessions, each comprising three classical problems and four matchstick equations. If learning had occurred, its effects should have been more pronounced in the second session. However, the WMC correlations of insight scores were perfectly comparable between the sessions (rs were stable between .30 and .35, irrespective of a session and a problem type). So, learning is unlikely to matter for the WM-insight ability link. The latter study also ruled out any influence on the respective correlation strengths of whether the insight problem instructions insisted on the participants thinking in flexible and nonstandard ways (as in the present study) or just required "following test instructions." Finally, the amount of time allowed did not affect these correlation strengths (no difference between speeded and unspeeded sessions).

Given all the above reservations, we claim that the present psychometric approach to insight still captured the most important aspects of insight problem solving, comparably to single-problem studies in experimental settings. At the same time, the use of multiitems tests, latent variables and maximum likelihood estimation helped to correct biases resulting from limited reliability and increased idiosyncrasy of the single-problem studies (e.g., our use of the insight latent variable helped to account for low reliability of our spatial and verbal classical insight problems). We suggest that valid conclusions on the true relationship between WMC and insight problem solving can only be drawn from the CFA/SEM or equivalent methods.

How can WM Contribute to Insight Problem Solving?

Above half of the variance in insight problem solving explained here by WMC constitutes a strong refutation of the recent theoretical claims (e.g., Wiley & Jarosz, 2012) assuming that WM is unrelated, or even negatively related, to insight problem solving. (For a methodological critique of these claims, see Chuderski & Jastrzębski, 2017). Specifically, these claims assume that insights result from weakened executive control, leading to a faster diverting of attention from the initial ineffective problem representation, and a less constrained, broader search for alternatives in long-term memory, and thus high WMC (associated with strong control over attention and memory) hinders insight problem solving. Although such claims are rooted in popular views on creative thinking (Guilford, 1967; Mednick, 1962), the strong positive links between WMC and insight problem solving reported here, especially pronounced in the case of attention control ability captured by the executive control tasks (note: $r = .759$), but also substantial for ordered recall from secondary memory measured by the complex span tasks (note: $r = .572$), make it highly unlikely that the key mechanisms of insight consist of divertible attention and unconstrained associative memory access. In contrast, the present data strongly imply that processing of insight problems relies substantially on multiple components of WM, both in the case of reported insights and solutions provided without such insights.

Moreover, the current study suggests that even if the processing of insight problems can sometimes be facilitated by defocused attention and broad memory search, it is highly unlikely that for such a reason high-WMC people lose their general cognitive advantage over low-WMC ones. Even if the beneficial effects of defocused attention for insight problem solving exist, the latter will still be facilitated by higher WMC. For instance, the dynamic interaction between associative exploration of the state space and its systematic generation, postulated by several models of problem solving (e.g., Finke, 1996; Necka, 2003; Simonton, 2015), allows that, even if high-WMC people permanently focused attention and directed memory search (which might hinder their hitting upon good initial ideas), they could catch up during systematic generation of the final solution. Moreover, even assuming that most WM task scores reflect controlled attention and directed retrievals, people who score high need not display such attention/retrievals all the time. Instead, high WMC may be rooted in the flexible adjustment of WM to processing demands, such as zooming WM in when blocking distractors, but zooming WM out in order to grasp many elements at once (Cowan, 2005). High-WMC people can be better in adjusting their WM than are low-WMC individuals: although they are able to more effectively focus their WM in the WM tasks, they may also better de-focus WM when some stages of insight problem solving require less constrained operation of attention and memory.

What can the knowledge of its strong links with WMC and reasoning tell about insight problem solving? The most plausible explanation suggests that finding solutions for insight problems is supported by the same well-defined incremental processes that also underpin analytical (noninsight) problem solving (Perkins, 1981; Weisberg, 2015; Weisberg & Alba, 1981). An analytical problem can be described as a discrepancy between the starting and goal states within the problem's state space. A solution is the path from the start to the goal. Often, the solution can be just retrieved from memory (the same or an analogous problem has been solved in the past) or potential solutions can be quickly generated and tested (the state space is small). However, in nontrivial problems no memorized solution is available, and the state space is so huge that without intelligent heuristics (e.g., hill-climbing, means-ends analysis; Newell & Simon, 1972), allowing one to explore only the most promising paths, its search results in combinatorial explosion. The primary role played by WM in problem solving was elucidated by its numerous computational models (e.g., Altmann & Trafton, 2002; Anderson, 2005; Laird, 1929/2012), with WM being responsible for storing the subgoals and subproducts during the search/generation of solutions across the problem's state space, as well as for directing that process (via goal representations). WM is also the mental place for binding the crucial pieces of the solution together, either spatially (Lyon, Gunzelmann, & Gluck, 2008; Ragni & Knauff, 2013) or relationally (Chuderski & Andrelczyk, 2015; Hummel & Holyoak, 2003). Existing studies reported substantial positive links between WM and performance on problems that required the search of state space (Gilhooly & Fioratou, 2009; Zook, Davalos, Delosh, & Davis, 2004), complex calculations (Holmes & Adams, 2006), combinatorial operations (Tsaparlis, 2005), and the control over dynamical environment (Bühner, Kröner, & Ziegler, 2008; Schweizer, Wüstenberg, & Greiff, 2013). The same positive relationship with WMC can thus be expected in the case of insight problems if they also depend on a similar search of state space, calculations, combinatorial operations, as well as planning and visualization (MacGregor et al., 2001).

Although the present data evidently support the large commonality of performance on insight and analytical problems, two results seem to question their complete equivalence. First, insight ability was more weakly associated with WMC than was reasoning. WMC explained almost all the variance in reasoning, while it left almost half the variance in insight problem solving unexplained. Second, the residual variance in reasoning, which could be ascribed to the learning of proper problem solving strategies and heuristics from previous items (Chuderski, 2016; Hayes, Petrov, & Sederberg, 2015), only weakly correlated with residual variance in insight problem solving. The relative independence of both residual variances suggests that insight problem solving requires specific strategies and heuristics beyond those involved in reasoning.

Previously, one of us (Chuderski, 2014) proposed that insight problem solving can best be described as "nothing special with special add-ons". In light of the present results as well as the PM theory (MacGregor et al., 2001) and the hybrid account of problem solving (Fleck & Weisberg, 2013), the "nothing special" in this framework refers to general-domain cognitive processes, shared with analytical thinking, that include selective attention, memory access, learning, analogical retrieval and mapping, problem space search, and transformation of mental representations (also using imagery). All these processes are strongly dependent on active maintenance and control of information exerted by the WM system. The special add-ons comprise the set of strategies, heuristics, and abilities specific for insight problem solving. Such add-ons can, for instance, encompass a tendency to frequently check whether a given approach to a problem at hand can effectively lead to its solution (as suggested by the PM theory), an ability to decouple from this approach (if it is unpromising) and to restart the problem solving process (as suggested by the hybrid theory), as well as a general strategy for trying out as many diverse approaches as possible. Most likely, people differ in the extent, to which they employ such add-ons, because of many reasons, including previous experience, training, and personality (psychoticism). Nevertheless, the add-ons contribute to a minor portion of variance in insight problem solving, as compared with general-domain processes underpinned by WM. Moreover, the add-ons themselves constitute analytic processes (high-order ones; Fleck & Weisberg, 2013), and need not imply any role for associative, unconstrained, holistic, and intuitive thinking in insight problem solving ability.

Conclusion

This comprehensive psychometric study included a solid battery of insight problems, fluid reasoning tests, and working memory tasks, and a large sample, as well as adequate methods of data analysis. The findings show that WMC explained above half of the variance in insight problem solving ability as well as almost all of its isomorphism to fluid reasoning ability. Analyses excluded the possibility that the strong link between WMC and insight problem solving might result from some specificity of WM tasks (all five WM functions studied yielded comparable links); from selected insight problems (the relationship held for most of problems investigated); from mixing problems accompanied by reported insight with those solved due to analysis (no difference in WM links for both categories); from specific subsamples of the sample examined (the WM–insight relationship was linear throughout the ability levels); or from mediation of noncognitive factors (age, motivation, anxiety, and personality). These results thus contradict the popular view as well as neo-Gestalt claims on the special, sudden, and holistic nature of insight, and are in line with the nothing-special account. The strong association of insight problem solving with WM implies that this type of thinking primarily results from typical mental operations, important also for analytical thinking, which are carried out by the relatively simple and domain-general mechanisms of WM that are responsible for the maintenance, retrieval, transformation, and control of information in a broad range of thinking tasks. The special-process theories of insight made much ado about the uniqueness of Aha! experience within human cognition; however, it seems that if anything is unique about insight problem solving (i.e., add-ons), it is marginal as compared with the standard machinery of WM. In contrast, the present results speak for a relatively uniform structure of human cognitive abilities, with WM playing the central role in various intellectual tasks.

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. http://dx.doi.org/10.1037/0033-2909.131.1.30

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121,* 219–245. http://dx.doi.org/10.1037/0033-2909.121.2.219

Altmann, E. M., & Trafton, J. G. (2002). Memory for goals: An activation-based model. *Cognitive Science, 26,* 39–83. http://dx.doi.org/10.1207/s15516709cog2601_2

Amabile, T. M. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology, 48,* 393–397. http://dx.doi.org/10.1037/0022-3514.48.2.393

Anderson, J. R. (2005). Human symbol manipulation within an integrated cognitive architecture. *Cognitive Science, 29,* 313–341. http://dx.doi.org/10.1207/s15516709cog0000_22

Ansburg, P. I., & Dominowski, R. L. (2000). Promoting insightful problem solving. *The Journal of Creative Behavior, 34,* 30–60. http://dx.doi.org/10.1002/j.2162-6057.2000.tb01201.x

Asa, I. K., & Wiley, J. (2008). Hindsight bias in insight and mathematical problem solving: Evidence of different reconstruction mechanisms for metacognitive versus situational judgments. *Memory & Cognition, 36,* 822–837. http://dx.doi.org/10.3758/MC.36.4.822

Ash, I. K., Cushen, P. J., & Wiley, J. (2009). Obstacles in Investigating the role of restructuring in insightful problem solving. *The Journal of Problem Solving, 2,* 6–41. http://dx.doi.org/10.7771/1932-6246.1056

Ash, I. K., & Wiley, J. (2006). The nature of restructuring in insight: An individual-differences approach. *Psychonomic Bulletin & Review, 13,* 66–73. http://dx.doi.org/10.3758/BF03193814

Barron, F. (1969). *Creative person and creative process*. New York, NY: Holt, Rinehart & Winston.

Batchelder, W. H., & Alexander, G. E. (2012). Insight problem solving: A critical examination of the possibility of formal theory. *The Journal of Problem Solving, 5,* 6. http://dx.doi.org/10.7771/1932-6246.1143

Beaty, R. E., Nusbaum, E. C., & Silvia, P. J. (2014). Does insight problem solving predict real-world creativity? *Psychology of Aesthetics, Creativity, and the Arts, 8,* 287–292. http://dx.doi.org/10.1037/a0035727

Benedek, M., Jauk, E., Sommer, M., Arendasy, M., & Neubauer, A. C. (2014). Intelligence, creativity, and cognitive control: The common and differential involvement of executive functions in intelligence and creativity. *Intelligence, 46,* 73–83. http://dx.doi.org/10.1016/j.intell.2014.05.007

Bowden, E. M., & Jung-Beeman, M. (2003). Normative data for 144 compound remote associate problems. *Behavior Research Methods, Instruments, & Computers, 35,* 634–639. http://dx.doi.org/10.3758/BF03195543

Bowden, E. M., Jung-Beeman, M., Fleck, J., & Kounios, J. (2005). New approaches to demystifying insight. *Trends in Cognitive Sciences, 9,* 322–328. http://dx.doi.org/10.1016/j.tics.2005.05.012

Bühner, M., Kröner, S., & Ziegler, M. (2008). Working memory, visual-spatial intelligence and their relationship to problem-solving. *Intelligence, 36,* 672–680. http://dx.doi.org/10.1016/j.intell.2008.03.008

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511571312

Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67,* 319–333. http://dx.doi.org/10.1037/0022-3514.67.2.319

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston, MA: Houghton Mifflin.

Chein, J. M., & Weisberg, R. W. (2014). Working memory and insight in verbal problems: Analysis of compound remote associates. *Memory & Cognition, 42,* 67–83. http://dx.doi.org/10.3758/s13421-013-0343-4

Chein, J. M., Weisberg, R. W., Streeter, N. L., & Kwok, S. (2010). Working memory and insight in the nine-dot problem. *Memory & Cognition, 38,* 883–892. http://dx.doi.org/10.3758/MC.38.7.883

Chu, Y., & MacGregor, J. N. (2011). Human performance on insight problem solving: A review. *The Journal of Problem Solving, 3,* 119–150. http://dx.doi.org/10.7771/1932-6246.1094

Chuderski, A. (2014). How well can storage capacity, executive control, and fluid reasoning explain insight problem solving. *Intelligence, 46,* 258–270. http://dx.doi.org/10.1016/j.intell.2014.07.010

Chuderski, A. (2016). Time pressure prevents relational learning. *Learning and Individual Differences, 49,* 361–365. http://dx.doi.org/10.1016/j.lindif.2016.07.006

Chuderski, A., & Andrelczyk, K. (2015). From neural oscillations to complex cognition: Simulating the effect of the theta-to-gamma cycle length ratio on analogical reasoning. *Cognitive Psychology, 76,* 78–102. http://dx.doi.org/10.1016/j.cogpsych.2015.01.001

Chuderski, A., & Jastrzębski, J. (2017). Working memory facilitates insight instead of hindering it: Comment on DeCaro, Van Stockum, and Wieth (2016). *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://dx.doi.org/10.1037/xlm0000409

Chuderski, A., & Jastrzębski, J. (2017). The relationship of insight problem solving to analytical thinking. In F. Vallée-Tourangeau (Ed.), *Insight: On the origin of ideas*. Abingdon, UK: Routledge.

Chuderski, A., & Necka, E. (2012). The contribution of working memory to fluid reasoning: Capacity, control, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1689–1710. http://dx.doi.org/10.1037/a0028465

Chuderski, A., Taraday, M., Necka, E., & Smolen, T. (2012). Storage capacity explains fluid intelligence while executive control does not. *Intelligence, 40,* 278–295. http://dx.doi.org/10.1016/j.intell.2012.02.010

Cinan, S., Özen, G., & Hampshire, A. (2013). Confirmatory factor analysis on separability of planning and insight constructs. *Journal of Cognitive Psychology, 25,* 7–23. http://dx.doi.org/10.1080/20445911.2012.729035

Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs but why? *Intelligence, 36,* 584–606. http://dx.doi.org/10.1016/j.intell.2008.01.002

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and the NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–114. http://dx.doi.org/10.1017/S0140525X01003922

Cowan, N. (2005). Understanding intelligence: A summary and an adjustable-attention hypothesis. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 469–488). London, UK: SAGE. http://dx.doi.org/10.4135/9781452233529.n25

Csikszentmihalyi, M. (1996). *Creativity: Flow and the psychology of discovery and invention*. New York, NY: Harper Perennial.

Cunningham, J. B., MacGregor, J. N., Gibb, J., & Haar, J. (2009). Categories of insight and their correlates: An exploration of relationships among classic-type insight problems, rebus puzzles, remote associates and esoteric analogies. *The Journal of Creative Behavior, 43,* 262–280. http://dx.doi.org/10.1002/j.2162-6057.2009.tb01318.x

Cushen, P. J., & Wiley, J. (2011). Aha! Voila! Eureka! Bilingualism and insightful problem solving. *Learning and Individual Differences, 21,* 458–462. http://dx.doi.org/10.1016/j.lindif.2011.02.007

Dailey, D. P. (1978). An analysis and evaluation of the internal validity of the remote associates test: What does it measure? *Educational and Psychological Measurement, 38,* 1031–1040. http://dx.doi.org/10.1177/001316447803800421

Danek, A. H., Wiley, J., & Öllinger, M. (2016). Solving classical insight problems without aha! experience: 9 dot, 8 coin, and matchstick arithmetic problems. *The Journal of Problem Solving, 9,* 4. http://dx.doi.org/10.7771/1932-6246.1183

Davidson, J. E. (1995). The suddenness of insight. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 125–155). New York, NY: Cambridge University Press.

Davidson, J. E., & Sternberg, R. A. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly, 28,* 58–64. http://dx.doi.org/10.1177/001698628402800203

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63,* 453–482.

DeCaro, M. S., Van Stockum, C. A., Jr., & Wieth, M. B. (2016). When higher working memory capacity hinders insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42,* 39–49. http://dx.doi.org/10.1037/xlm0000152

De Dreu, C. K. W., Nijstad, B. A., Baas, M., Wolsink, I., & Roskes, M. (2012). Working memory benefits creative insight, musical improvisation, and original ideation through maintained task-focused attention. *Personality and Social Psychology Bulletin, 38,* 656–669. http://dx.doi.org/10.1177/0146167211435795

DeYoung, C. G., Flanders, J. L., & Peterson, J. B. (2008). Cognitive abilities involved in insight problem solving: An individual differences model. *Creativity Research Journal, 20,* 278–290. http://dx.doi.org/10.1080/10400410802278719

Dow, G. T., & Mayer, R. E. (2004). Teaching students to solve insight problems. Evidence for domain specificity in training. *Creativity Research Journal, 16,* 389–402. http://dx.doi.org/10.1080/10400410409534550

Duncker, K. (1945). On problem-solving. *Psychological Monographs, 58,* 270. http://dx.doi.org/10.1037/h0093599

Ecker, U. K. H., Lewandowsky, S., Oberauer, K., & Chee, A. E. H. (2010). The components of working memory updating: An experimental decomposition and individual differences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 170–189. http://dx.doi.org/10.1037/a0017891

Eimer, M. (1993). ERPs elicited by go and no go stimuli: Effects of attention and stimulus probability. *Biological Psychology, 35,* 123–138. http://dx.doi.org/10.1016/0301-0511(93)90009-W

Eliasmith, C. (2013). *How to build a brain: A neural architecture for biological cognition*. New York, NY: Oxford University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General, 128,* 309–331. http://dx.doi.org/10.1037/0096-3445.128.3.309

Evans, J. St. B. T. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry, 21,* 313–326. http://dx.doi.org/10.1080/1047840X.2010.521057

Eysenck, H. J. (1990). Biological dimensions of personality. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 244–276). New York, NY: Guilford Press.

Eysenck, H. J. (1993). Creativity and personality: Suggestions for a theory. *Psychological Inquiry, 4,* 147–178. http://dx.doi.org/10.1207/s15327965pli0403_1

Eysenck, M. W., Payne, S., & Derakshan, N. (2005). Trait anxiety, visuospatial processing, and working memory. *Cognition and Emotion, 19,* 1214–1228. http://dx.doi.org/10.1080/02699930500260245

Feist, G. J. (1998). A meta-analysis of personality in scientific and artistic creativity. *Personality and Social Psychology Review, 2,* 290–309. http://dx.doi.org/10.1207/s15327957pspr0204_5

Finke, R. A. (1996). Imagery, creativity, and emergent structure. *Consciousness and Cognition, 5,* 381–393. http://dx.doi.org/10.1006/ccog.1996.0024

Fleck, J. I. (2008). Working memory demands in insight versus analytic problem solving. *Journal of Cognitive Psychology, 20,* 139–176. http://dx.doi.org/10.1080/09541440601016954

Fleck, J. I., & Weisberg, R. W. (2004). The use of verbal protocols as data: An analysis of insight in the candle problem. *Memory & Cognition, 32,* 990–1006. http://dx.doi.org/10.3758/BF03196876

Fleck, J. I., & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology, 25,* 436–463. http://dx.doi.org/10.1080/20445911.2013.779248

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19,* 25–42. http://dx.doi.org/10.1257/089533005775196732

Fürst, G., Ghisletta, P., & Lubart, T. (2016). Toward an integrative model of creativity and personality: Theoretical suggestions and preliminary empirical testing. *The Journal of Creative Behavior, 50,* 87–108. http://dx.doi.org/10.1002/jocb.71

Gignac, G. E., & Weiss, L. G. (2015). Digit Span is (mostly) related linearly to general intelligence: Every extra bit of span counts. *Psychological Assessment, 27,* 1312–1323. http://dx.doi.org/10.1037/pas0000105

Gilhooly, K. J., & Fioratou, E. (2009). Executive functions in insight versus non-insight problem solving. *Thinking & Reasoning, 15,* 355–376. http://dx.doi.org/10.1080/13546780903178615

Gilhooly, K. J., & Murphy, P. (2005). Differentiating insight from non-insight problems. *Thinking & Reasoning, 11,* 279–302. http://dx.doi.org/10.1080/13546780442000187

Guilford, J. P. (1967). *The nature of human intelligence*. New York, NY: McGraw-Hill.

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18,* 1279–1296. http://dx.doi.org/10.1016/0042-6989(78)90218-3

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. http://dx.doi.org/10.1016/j.intell.2014.10.005

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematics curricula. *Educational Psychology: An International Journal of Experimental Educational Psychology, 26,* 339–366. http://dx.doi.org/10.1080/01443410500341056

Holyoak, K. J. (2012). Analogy and relational reasoning. In K. J. Holyoak & R. G. Morrison (Eds.), *The Oxford handbook of thinking and reasoning* (pp. 234–259). New York, NY: Oxford University Press. http://dx.doi.org/10.1093/oxfordhb/9780199734689.001.0001

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55. http://dx.doi.org/10.1080/10705519909540118

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110,* 220–264. http://dx.doi.org/10.1037/0033-295X.110.2.220

Jauk, E., Benedek, M., Dunst, B., & Neubauer, A. C. (2013). The relationship between intelligence and creativity: New support for the threshold hypothesis by means of empirical breakpoint detection. *Intelligence, 41,* 212–221. http://dx.doi.org/10.1016/j.intell.2013.03.003

Johnson-Laird, P. N. (2006). *How we reason*. New York, NY: Oxford University Press.

Jonason, P. K., & Webster, G. D. (2010). The dirty dozen: A concise measure of the dark triad. *Psychological Assessment, 22,* 420–432. http://dx.doi.org/10.1037/a0019265

Jones, G. (2003). Testing two cognitive theories of insight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 1017–1027. http://dx.doi.org/10.1037/0278-7393.29.5.1017

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Strauss, Giroux.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 21–46). New York, NY: Oxford University Press.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review, 9,* 637–671. http://dx.doi.org/10.3758/BF03196323

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131,* 66–71. http://dx.doi.org/10.1037/0033-2909.131.1.66

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. http://dx.doi.org/10.1037/0096-3445.133.2.189

Kaufman, S. B. (2013). Opening up openness to experience: A four-factor model and relations to creative achievement in the arts and science. *The Journal of Creative Behavior, 47,* 233–255. http://dx.doi.org/10.1002/jocb.33

Kershaw, T. C., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: The case of the nine-dot problem. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 3–13. http://dx.doi.org/10.1037/0278-7393.30.1.3

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358. http://dx.doi.org/10.1037/h0043688

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

Knöblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1534–1555. http://dx.doi.org/10.1037/0278-7393.25.6.1534

Koffka, K. (1935). *Principles of Gestalt psychology*. New York, NY: Harcourt, Brace and Company.

Köhler, W. (1947). *Gestalt psychology*. New York, NY: Liveright.

Kounios, J., & Beeman, M. (2014). The cognitive neuroscience of insight. *Annual Review of Psychology, 65,* 71–93. http://dx.doi.org/10.1146/annurev-psych-010213-115154

Laird, J. E. (2012). *The soar cognitive architecture*. Cambridge, MA: MIT Press. (Original work published 1929)

Larson, G. E., Merritt, C. R., & Williams, S. E. (1988). Information processing and intelligence: Some implications of task complexity. *Intelligence, 12,* 131–147. http://dx.doi.org/10.1016/0160-2896(88)90012-8

Lee, C. S., Huggins, A. C., & Therriault, D. J. (2014). A measure of creativity or intelligence? Examining internal and external structure validity evidence of the Remote Associates Test. *Psychology of Aesthetics, Creativity, and the Arts, 8,* 446–460. http://dx.doi.org/10.1037/a0036773

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91,* 295–327. http://dx.doi.org/10.1037/0033-295X.91.3.295

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279–281. http://dx.doi.org/10.1038/36846

Lung, C., & Dominowski, R. L. (1985). Effects of strategy instructions and practice on nine-dot problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 804–811. http://dx.doi.org/10.1037/0278-7393.11.1-4.804

Lyon, D. R., Gunzelmann, G., & Gluck, K. A. (2008). A computational model of spatial visualization capacity. *Cognitive Psychology, 57,* 122–152. http://dx.doi.org/10.1016/j.cogpsych.2007.12.003

MacGregor, J. N., & Cunningham, J. B. (2008). Rebus puzzles as insight problems. *Behavior Research Methods, 40,* 263–268. http://dx.doi.org/10.3758/BRM.40.1.263

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the nine-dot and related problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 176–201. http://dx.doi.org/10.1037/0278-7393.27.1.176

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109,* 163–203. http://dx.doi.org/10.1037/0033-2909.109.2.163

Maier, N. R. F. (1930). Reasoning in humans: I. On direction. *Journal of Comparative Psychology, 10,* 115–143. http://dx.doi.org/10.1037/h0073232

Mednick, S. A. (1962). The associative basis of the creative process. *Psychological Review, 69,* 220–232. http://dx.doi.org/10.1037/h0048850

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 12,* 288–294.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition, 15,* 238–246. http://dx.doi.org/10.3758/BF03197722

Moutafi, J., Furnham, A., & Crump, J. (2003). Demographic and personality predictors of intelligence: A study using the NEO Personality Inventory and the Myers-Briggs Type Indicator. *European Journal of Personality, 17,* 79–94. http://dx.doi.org/10.1002/per.471

Nęcka, E. (2003). Creative interaction: A conceptual schema for the process of producing ideas and judging the outcomes. In M. A. Runco (Ed.), *Critical creative processes* (pp. 115–127). Cresskill, NJ: Hampton Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Nusbaum, E. C., & Silvia, P. J. (2011). Are intelligence and creativity really so different? Fluid intelligence, executive processes, and strategy use in divergent thinking. *Intelligence, 39,* 36–45. http://dx.doi.org/10.1016/j.intell.2010.11.002

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin, 142,* 758–799. http://dx.doi.org/10.1037/bul0000046

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in working memory* (pp. 49–75). New York, NY: Oxford University Press.

Ohlsson, S. (1984). Restructuring revisited: II. An information processing theory of restructuring and insight. *Scandinavian Journal of Psychology, 25,* 117–129. http://dx.doi.org/10.1111/j.1467-9450.1984.tb01005.x

Ohlsson, S. (1992). Information-processing explanations of insight and related phenomena. In M. Keane & K. Gilhooly (Eds.), *Advances in the psychology of thinking* (pp. 1–44). London, England: Harvester-Wheatsheaf.

Ohlsson, S. (2011). *Deep learning. How the mind overrides experience*. New York, NY: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511780295

Owens, M., Stevenson, J., Hadwin, J. A., & Norgate, R. (2014). When does anxiety help or hinder cognitive test performance? The role of working memory capacity. *British Journal of Psychology, 105,* 92–101. http://dx.doi.org/10.1111/bjop.12009

Perkins, D. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press.

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology, 57,* 137–146. http://dx.doi.org/10.1037/h0046137

Prabhu, V., Sutton, C., & Sauser, W. (2008). Creativity and certain personality traits: Understanding the mediating effect of intrinsic motivation. *Creativity Research Journal, 20,* 53–66. http://dx.doi.org/10.1080/10400410701841955

Ragni, M., & Knauff, M. (2013). A theory and a computational model of spatial reasoning with preferred mental models. *Psychological Review, 120,* 561–588. http://dx.doi.org/10.1037/a0032460

Raven, J. C., Court, J. H., & Raven, J. (1983). *Manual for Raven's Progressive Matrices and Vocabulary scales* (Section 4: Advanced progressive matrices). London, UK: H. K. Lewis.

Rawlings, D. (1985). Psychoticism, creativity and dichotic shadowing. *Personality and Individual Differences, 6,* 737–742. http://dx.doi.org/10.1016/0191-8869(85)90084-4

Ricks, T. R., Turley-Ames, K. J., & Wiley, J. (2007). Effects of working memory capacity on mental set due to domain knowledge. *Memory & Cognition, 35,* 1456–1462. http://dx.doi.org/10.3758/BF03193615

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research, 31,* 573–598. http://dx.doi.org/10.1023/A:1021213004302

Salvi, C., Bricolo, E., Kounios, J., Bowden, E., & Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Thinking & Reasoning, 22,* 443–460. http://dx.doi.org/10.1080/13546783.2016.1141798

Schmiedek, F., Hildebrandt, A., Lövdén, M., Lindenberger, U., & Wilhelm, O. (2009). Complex span versus updating tasks of working memory: The gap is not that deep. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 1089–1096. http://dx.doi.org/10.1037/a0015730

Schweizer, F., Wüstenberg, S., & Greiff, S. (2013). Validity of the MicroDYN approach: Complex problem solving predicts school grades beyond working memory capacity. *Learning and Individual Differences, 24,* 42–52. http://dx.doi.org/10.1016/j.lindif.2012.12.011

Simonton, D. K. (2015). On praising convergent thinking: Creativity as blind variation and selective retention. *Creativity Research Journal, 27,* 262–270. http://dx.doi.org/10.1080/10400419.2015.1063877

Sio, U. N., & Ormerod, T. C. (2009). Does incubation enhance problem solving? A meta-analytic review. *Psychological Bulletin, 135,* 94–120. http://dx.doi.org/10.1037/a0014212

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283,* 1657–1661. http://dx.doi.org/10.1126/science.283.5408.1657

Smolen, T., & Chuderski, A. (2015). The quadratic relationship between difficulty of intelligence test items and their correlations with working memory. *Frontiers in Psychology, 6,* 1270. http://dx.doi.org/10.3389/fpsyg.2015.01270

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.

Storm, B. C., & Angello, G. (2010). Overcoming fixation. Creative problem solving and retrieval-induced forgetting. *Psychological Science, 21,* 1263–1265. http://dx.doi.org/10.1177/0956797610379864

Tsaparlis, G. (2005). Non-algorithmic quantitative problem solving in university physical chemistry: A correlation study of the role of selective cognitive factors. *Research in Science & Technological Education, 23,* 125–148. http://dx.doi.org/10.1080/02635140500266369

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. http://dx.doi.org/10.1037/0033-295X.114.1.104

Unsworth, N., Miller, J. D., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., & Goodie, A. S. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences, 30,* 194–200. http://dx.doi.org/10.1027/1614-0001.30.4.194

Van Stockum, C. A., Jr., & DeCaro, M. S. (2014). Enclothed cognition and controlled attention during insight problem solving. *The Journal of Problem Solving, 7,* 73–83. http://dx.doi.org/10.7771/1932-6246.1164

Wallas, G. (1926). *The art of thought*. New York, NY: Harcourt, Erace.

Webb, M. E., Little, D. R., & Cropper, S. J. (2016). Insight is not in the problem: Investigating insight in problem solving across task types. *Frontiers in Psychology, 7,* 1424. http://dx.doi.org/10.3389/fpsyg.2016.01424

Weisberg, R. W. (1995). Prolegomena to theories of insight in problem solving: A taxonomy of problems. In R. J. Sternberg & J. E. Davidson (Eds.), *The nature of insight* (pp. 157–196). New York, NY: Cambridge University Press.

Weisberg, R. W. (2006). *Creativity: Understanding innovation in problem solving, science, invention, and the arts*. Hoboken, NJ: Wiley.

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning, 21,* 5–39. http://dx.doi.org/10.1080/13546783.2014.886625

Weisberg, R. W., & Alba, J. W. (1981). An examination of the alleged role of "fixation" in the solution of several "insight" problems. *Journal of Experimental Psychology: General, 110,* 169–192. http://dx.doi.org/10.1037/0096-3445.110.2.169

Wertheimer, M. (1945). *Productive thinking*. New York, NY: Harper.

Wiley, J., & Jarosz, A. F. (2012). How working memory capacity affects problem solving. *Psychology of Learning and Motivation, 56,* 185–227. http://dx.doi.org/10.1016/B978-0-12-394393-4.00006-6

Wilhelm, O., & Engle, R. W. (2005). *Handbook of measuring and understanding intelligence*. Thousand Oaks, CA: SAGE.

Wittmann, W. W. (1988). Multivariate reliability theory. Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 505–560). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4613-0893-5_16

Yntema, D. B., & Mueser, G. E. (1962). Keeping track of variables that have few or many states. *Journal of Experimental Psychology, 63,* 391–395. http://dx.doi.org/10.1037/h0045706

Zook, N. A., Davalos, D. B., Delosh, E. L., & Davis, H. P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and London tasks. *Brain and Cognition, 56,* 286–292. http://dx.doi.org/10.1016/j.bandc.2004.07.003

Received February 14, 2017
Revision received September 1, 2017
Accepted September 7, 2017 ■