Intelligence 41 (2013) 244–262

Contents lists available at SciVerse ScienceDirect

Intelligence

journal homepage:

When are fluid intelligence and working memory isomorphic and when are they not?☆

Adam Chuderski *

Cognitive Science Department, Institute of Philosophy, Jagiellonian University, Krakow, Poland

ARTICLE INFO

Article history:
Received 16 May 2012
Received in revised form 5 March 2013
Accepted 5 April 2013
Available online xxxx

Keywords:
Fluid intelligence
Working memory capacity
Speeded vs. power tests
Relational learning

ABSTRACT

Study 1 investigated whether the strength of correlation between latent variables representing working memory capacity (WMC) and fluid intelligence (Gf) depends on the time allowed to work on an intelligence test. When the half recommended time was given to fulfill two Gf tests, WMC and Gf were statistically indistinguishable, indicating that working memory and fluid intelligence are fully isomorphic constructs. However, when virtually no time limit was applied, WMC explained only 38% of variance in Gf. Further analyses suggested that only the latter testing conditions allowed low-capacity participants for relational learning during test taking, which allowed them to reduce their distance to high-capacity people. Study 2 corroborated the moderate value of WM–Gf correlation in untimed intelligence testing with a larger number of Gf and WM tasks, as well as showed that the indices of learning in a novel test of relation discovery predict significant amount of Gf variance. In sum, the research suggests that fluid reasoning can be differently related to WMC depending on the time pressure during Gf testing, and it also indicates that learning abstract relational representations may be an important component of unspeeded intelligence, but barely takes place during speeded testing.

1. Introduction

For more than a century (Binet, 1903; Galton, 1883; Spearman, 1904), the nature of general intelligence (*g* factor), the theoretical construct reflecting vast interindividual variability but high intraindividual consistency in coping with diverse cognitive tasks, has been one of the central research problems of psychology and neuroscience. Its importance is highlighted by the fact that *g* has been found to strongly predict educational, professional, and personal success (or lack of it) in everyday life (e.g., Deary & Der, 2005; Gottfredson, 1997; Sternberg, 1996).

A crucial finding in research on the structure of human intellect (e.g., Cattell, 1971; Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Gustaffson, 1984; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002) is that *g* factor seems to rely to a great extent on fluid intelligence (Gf factor, also referred to as fluid ability, reasoning ability, or fluid reasoning). Gf reflects the ability to use abstract relational reasoning in order to solve novel problems, in which prior experience and learned knowledge are of little use. Great efforts have been devoted to the identification of the neuronal and cognitive mechanisms which determine scores on Gf tests, including the well-known Raven's Advanced Progressive Matrices test (Raven, 1962; Raven for short). Apart from fulfilling the scientific goal of explaining the nature of human intelligence, such a finding could also provide researchers with methods for increasing fluid ability (Jaeggi et al., 2010; Klingberg, 2010), which would be especially desirable for the compensation of cognitive deficits in some groups of people, like the mentally deteriorated (Holmes, Gathercole, & Dunning, 2009) or ADHD children (Klingberg, Forssberg, & Westerberg, 2002) as well as healthy aging persons (Schmiedek, Lövdén, & Lindenberger, 2010).

☆ This research was supported by grant N106 417140 funded by the National Science Centre of Poland. The funding source had no role in the design of the study. Thanks to Maciej Taraday, Anna Anzulewicz, Krzysztof Cipora, Dominika Czajak, and Jolanta Wojcik for their help in conducting the study. I am also grateful to Anna Chuderska for valuable comments and Mike Timberlake for correcting the text.

* Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland.
E-mail address: adam.chuderski@gmail.com.

0160-2896/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.intell.2013.04.003

So far, the most important conclusion drawn from the research on the neurocognitive basis of fluid intelligence shows that the capacity of working memory (WM) is its strongest predictor. WM denotes processes and mechanisms responsible for the active maintenance and transformation of information crucial for the current goal/task/operation, occurring within a time scale of several seconds (Baddeley, 2007). WM is usually assessed with tasks requiring the encoding, storage, and recall or recognition of stimuli. WM capacity (WMC) has been operationalized as the direct number of items that a person can reproduce (Engle & Kane, 2004), or the indirectly estimated number of items that one is presumed to keep in WM (Rouder, Morey, Morey, & Cowan, 2011). The most surprising finding concerning WMC is the fact that, most probably, human WM is able to reliably store in parallel only a few items at best. Average capacity has been estimated to be four items (Luck & Vogel, 1997), and it seems to vary in people from two to six items (Cowan, 2001). Early studies assumed that proper estimation of WMC has to require some form of concurrent processing (i.e., complex span tasks; Engle & Kane, 2004), but recently it has been suggested that tasks without any processing component (i.e., simple span or short-term memory tasks; STM tasks) are also excellent measures of WMC as well as good predictors of Gf (e.g., Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Cowan, Fristoe, Elliott, Brunner, & Saults, 2006; Oberauer, 2005; Unsworth & Engle, 2007). Due to the possibility of significant measurement errors and task-specific variance reflected by WM scores, WMC is usually not derived from a single WM task, but instead is assessed on the basis of scores from several WM tasks. The strength of correlation between Gf and WMC is usually estimated on the latent variable level, with the use of confirmatory factor analysis (CFA) and/or structural equation modeling (SEM).

1.1. The strength of the Gf–WM relation and its explanations

Many studies have investigated the precise strength of the relation between Gf and WMC. The results of most of these studies indicate that both constructs are at least moderately correlated, with *r*s usually falling in the .30–.80 range (e.g., Ackerman, Beier, & Boyle, 2005; Buehner, Krumm, Ziegler, & Pluecken, 2006; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Friedman et al., 2006; Kane et al., 2004; Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Shelton, Elliott, Matthews, Hill, & Gouvier, 2010; Unsworth, 2010; Unsworth, Brewer, & Spillers, 2009; Unsworth, Spillers, & Brewer, 2010; Unsworth, Miller et al., 2009).

The lower-level mechanisms underlying the variance shared by the Gf and WMC latent variables have been hotly debated. One of the proposals suggested that both variables can be explained by the differences in mental speed, for example assessed with simple perceptual-motor tasks involving stimuli comparison (for reviews see Jensen, 2006; Sheppard & Vernon, 2008). Such an explanation seems to be valid with regard to Gf tests administered under severe time constraints, in which the speed of mental operations may determine if a participant is able to attempt all Gf test items or not (Wilhelm & Schultze, 2002), and so the strong intercorrelation of Gf and mental speed can simply be attributed to the shared method variance. However, mental speed indices also seem to correlate with scores on unspeeded (i.e., power) tests (Jensen, 2006). Another explanation for the WM–Gf link pertains to the sheer storage capacity of the active and highly accessible memory buffer (called primary memory or the focus of attention). Individual Gf level has been related to the number of elements (Colom et al., 2008; Cowan et al., 2006), the number of temporary bindings among elements (Oberauer, Süß, Wilhelm, & Sander, 2007), or the number of variables within a relation describing elements (Halford, Cowan, & Andrews, 2007) that such a buffer can simultaneously maintain and/or process. Evidence for the contribution to the WMC and Gf of both mental speed and storage capacity is vast, but their estimates are often intercorrelated (e.g., Ackerman, Beier, & Boyle, 2002; Conway et al., 2002; Süß et al., 2002).

Consequently, it is currently disputed which of these factors is a genuine predictor of Gf and which is not. Some theorists (e.g., Jensen, 1998; Salthouse, 1996) proposed that mental speed determines storage capacity, because representations in WM quickly decay, and the faster these representations can be processed before falling below a retrieval threshold, the more of them can be recalled, bound, or related. However, such theories assume that decay in memory really exists, while many studies question the role of decay in forgetting (e.g., Lewandowsky & Oberauer, 2009; Saito & Miyake, 2004). Moreover, the variables reflecting capacity usually correlated with Gf more strongly than the variables reflecting speed (e.g., Colom et al., 2008; Conway et al., 2002; Kaufman et al., 2009; Martínez et al., 2011). So, Wilhelm and Oberauer (2006) argued that storage capacity determines psychometric speed, because tasks that measure speed require active storage of stimulus–response (S–R) bindings, and as storage capacity is very limited, the low-capacity persons often lose the required S–R bindings from their buffer, therefore requiring additional time to restore these bindings, and so leading to prolonged response latencies, especially when bindings are arbitrary.

The debate is far from being settled, and the relations between speed and capacity may be even more complex (see Rypma & Prabhakaran, 2009) than in the views presented above. Moreover, there could be another factor that determines both speed and capacity. For example, one proposal pertains to attention control (Engle & Kane, 2004; Vogel, McCollough, & Machizawa, 2005), which consists of focusing attention on task-relevant information while blocking distraction and interference. Effective control may be crucial not only for storage capacity, as only relevant elements/bindings/relations are maintained in the active buffer (so available capacity is used optimally), but also for processing speed, as irrelevant elements do not capture attention (so no time needs to be wasted for overriding the capture).

Furthermore, if the strength of correlation between WMC and Gf really amounts to a value that falls between $r = .30$ and $r = .80$, then the underlying factor, regardless of what it really is, will explain only part of the variance shared by both these constructs (probably half of it; see metaanalysis done by Kane, Hambrick, & Conway, 2005). Thus, an interesting question concerns the other part of this variance, which is unexplained by WMC. What factor could be related to fluid intelligence above and beyond WMC? One alternative is that more intelligent people possess more efficient learning abilities. Indeed, recent research has found that associative learning contributes to Gf independently from WMC (Kaufman et al., 2009; Tamez, Myerson, & Hale, 2008; Williams & Pearlberg, 2006), though the amount of variance accounted for by learning is not so impressive (e.g., 10% in Kaufman et al.'s study).

Another possibility worth considering is that fluid intelligence tests require not only reasoning, but also divergent problem solving, including phenomena like insight, and involving proper strategic control. For example, in two studies (Davidson, 1995; Paulewicz, Chuderski, & Nęcka, 2007), tests including insight problems (i.e., ones that have trivial solutions, though not available to participants at first glance) yielded $r =$ .65 correlation with Raven. Specifically, the link between problem solving and reasoning may be supported by selective attention, because the need to effectively select a subset of elements from the whole scene (e.g., a figural matrix, a description of a problem) seems to be a key factor influencing the difficulty of both items in reasoning tests (Primi, 2001) and insight problems (Davidson, 1995). So, successful elimination of irrelevant information, leading to better abstraction, may explain a substantial part of variance in fluid intelligence (Garlick & Sejnowski, 2006), which cannot be accounted for by WM. Also Carpenter, Just, and Shell (1990) argued that strategic control over processing, responsible for the setting and management of processing goals as well as backtracking from wrong ones, which can be only partially substituted for by increased storage capacity, is the crucial factor determining high scores on Gf tests.

Finally, also factors beyond the cognitive domain may be responsible for some variance in fluid intelligence. For example, though correlations between intelligence and nonpathological personality traits are usually insignificant, some data indicate that a weak link between fluid ability and neuroticism does exist (e.g., Unsworth, Brewer, & Spillers, 2009; Unsworth, Miller, et al., 2009), and that state anxiety has a negative impact on performance in reasoning tests (e.g., Marjorie & Revelle, 1985).

1.2. Isomorphism between Gf and WM?

However, what if we do not really need to seek any factors underlying fluid intelligence above and beyond WM, because Gf and WM are isomorphic (Kyllonen, 2002)? Some results regarding the Gf–WMC link presented independently by Colom and his collaborators (Colom, Abad, Rebollo, & Shih, 2005, Colom et al., 2004; Martínez et al., 2011) and by Oberauer, Süß, Wilhelm, and Wittman (2008) suggest that this may be the case. In two studies (Colom et al., 2004; Martínez et al., 2011), using the standard STM tasks, Colom's team reported estimates of WMC–Gf correlation equaling one; the remaining study (Colom et al., 2005) showed a slightly lower but still extremely high estimate ($r = .89$). Oberauer et al. (2008), using very simple WM tasks that required the detection of certain relations among either items maintained in memory or perceptually available stimuli (e.g., identifying whether three rhyming words appeared in a row, column, or diagonal line of the three-by-three matrix of words), also found a correlation close to one ($r = .94$). A metaanalysis of a larger sample of studies on WM and Gf led Oberauer et al. (2007) to conclude that at least 75% of Gf variance is shared with WMC. So, little room would be left for any Gf contributor other than WM.

The aforementioned research on Gf–WM link yields a very interesting question: why, in some studies, did Gf and WMC seem to be (almost) isomorphic constructs, while in other studies these constructs were only moderately related? Unfortunately, providing an exhaustive answer to this question by reviewing all the aforementioned studies on the strength of Gf–WM relation does not seem possible, as these studies substantially differed in applied WM tasks, administered Gf tests, the presence or not of some other measures (e.g., mental speed, attention, long-term memory tests, etc.), as well as in sizes and characteristics of examined samples. Moreover, those studies differed in the structure of links between WM and Gf latent variables. In some cases, the WM variable was directly correlated with Gf (e.g., Conway et al., Engle et al., 1999; Martínez et al., 2011; Oberauer et al., 2008), while in other cases the former was linked with the general intelligence, which was loaded not only by Gf but also by other ability constructs like verbal and crystallized intelligence (Colom et al., 2005), or even WM and Gf loaded the same "g" variable (Colom et al., 2004). Nevertheless, when looking for variables responsible for the amount of variance shared between Gf and WMC, it can be noticed that all studies which showed Gf–WM isomorphism used some time constraints. Results obtained in Colom's (personal communication) lab were based on standard administration times (e.g., 40 min for Raven), which usually constitute moderate time constraints. Oberauer et al. (2008) used reasoning tests from the Berlin Intelligence Structure (BIS) model, which can be treated as a severely timed test, including 45 tasks in total (15 — in the reasoning parcel), with the average limit of a few minutes per task (see Wilhelm & Schultze, 2002). So, it seems that a certain level of time pressure during intelligence testing can positively moderate the strength of the WMC–Gf link.

However, while at least two studies that applied power testing (Conway et al., 2002; Engle et al., 1999), and another one (Kaufman et al., 2009) which involved a weak time constraint (45 min in Raven), yielded relatively low estimates of Gf–WMC link (i.e., *r*s in range .34 to .60), several other studies applied severe time constraints, but did not yield strong correlations between WM and Gf latent variables (e.g., Kane et al., 2004; Unsworth, 2010; Unsworth, Brewer, & Spillers, 2009; Unsworth, Miller et al., 2009; Unsworth, Spillers, & Brewer, 2010). Because of differences among existing studies with regard to both WM and Gf test loading respective latent variables, as well as due to the aforementioned incompatibility of models' structure from study to study, comparing correlations between latent variables in the function of time allowed for Gf testing does not seem possible, unfortunately, and a more constrained comparison is necessary. In order to evaluate the influence of time on the estimated relation between WM and Gf, while minimizing the impact of all other variables, I decided to compare only those studies that applied Raven (probably the most widely used Gf test) and STM/complex span tasks, and I focused solely on psychometric studies using healthy adult samples, excluding studies which had a different goal than relating Gf (or *g*) to WM (e.g., focused on ability training), or which examined children or elderly or patient samples. Finally, I compared only studies published not earlier than in year 1999 (i.e., starting from the seminal study by Engle et al., 1999), because at that time common methods of WM testing seem to become properly developed. I also tried to restrict myself only to studies that used at least three different measures of WM, but the surprisingly scarce number of studies which applied untimed Raven (i.e., four in total) made me to include two studies that used only one measure of WMC (however, a highly reliable one). For each study, I estimated the mean zero-order correlation between Raven and WM scores used in that study.

The final choice included 26 studies, which seems to constitute quite a representative sample. All these studies were divided into three sets: "highly speeded" Raven administration (30 min or less allowed for all 36 Raven items, or a proportional time when only a subset of items were applied; the total of 55 correlations and $N = 2253$), "moderately speeded" administration (40 min allowed, that is, as in the test's manual; 41 correlations, $N = 1847$), and "unspeeded" administration (at least 45 min allowed; 11 correlations, $N = 553$). Table 1 presents sample sizes, administration times, types of WM tasks used (i.e., either STM or WMC tasks or both), and mean correlation coefficients between Raven and WM tasks for each study. The (grand) mean correlation in highly speeded studies was $r = .395$, in moderately speeded studies it equaled $r = .316$, while in unspeeded set it was $r = .263$. Though the analyzed studies did not report standard errors for observed correlations, assuming that the error value estimated in the present Study 1 ($N = 890$, $SE = .031$) well approximates an error of the correlation in each data set, then the WM–Gf correlation in the highly speeded set was significantly stronger in comparison to the moderately speeded set, $t(4098) = 2.54$, $p = .005$, and also the correlation in the latter set was significantly stronger than in the unspeeded set, $t(2398) = 1.71$, $p = .044$. So, as I expected, increasing the time pressure of Raven's administration seems to increase the zero-order correlations between Gf and WM scores. However, the analysis of studies being so diverse may constitute only an imperfect cue about the importance of time allowed for intelligence testing, and the comparison of zero-order correlations can say nothing about an isomorphism between Gf and WM. So, more data, coming from fully comparable studies, and allowing for calculating correlations between the WM and Gf latent variables, was very needed.

1.3. Goals of the present study

The main goal of the present study was thus to demonstrate that WM and Gf (a) could be isomorphic constructs when speeded Gf tests were applied, while both these constructs (b) would share relatively little variance if the same but unspeeded intelligence tests were used. More specifically, I aimed to treat the strength of correlation between latent variables representing Gf and WM as a dependent variable, and to test if an increase in the amount of time allowed for solving intelligence tests could decrease that correlation from the expected perfect link in the case of speeded testing, thus replicating the close-to-one correlations observed by Colom et al. (2004, 2005), Martínez et al. (2011) and Oberauer et al. (2007, 2008), to only a moderate link when power testing will be applied. In order to achieve this goal, I examined three relatively large samples of participants (each N was around 300), with two Gf tests applied in one of three

Table 1
Studies included in the metaanalysis of the strength of the WM–Raven correlation.

Study	Sample size	Minutes allowed	No. of tested items	Type(s) of tasks	No. of tasks	Mean r value
Highly speeded testing						
Broadway & Engle, 2010 (dep. cond.)	151	10	18	STM, WM	4	.54
Broadway & Engle, 2010 (ind. cond.)	143	10	18	STM, WM	4	.49
Fukuda, Vogel, Mayr, & Awh, 2010	79	30	36	STM	3	.44
Kane et al., 2004	236	10	18	STM, WM	12	.37
Mackintosh & Bennett, 2003	138	30	35	STM	3	.36
Shelton, Elliott, Hill, Calamia, & Gouvier, 2009	174	15	36	WM	3	.33
Shipstead, Redick, Hicks, & Engle, 2012 (set A)	505	10	18	STM, WM	7	.45
Shipstead et al., 2012 (set B)	170	10	18	STM, WM	7	.47
Unsworth, Brewer, & Spillers, 2009	173	10	18	WM	3	.26
Unsworth, Redick, Lakey, & Young, 2010	138	5	12	WM	3	.51
Unsworth, 2010	165	10	18	WM	3	.32
Unsworth, Spillers, & Brewer, 2010	181	10	18	WM	3	.20
Moderately speeded testing						
Ackerman et al., 2002	135	45	48	STM, WM	7	.32
Burgess, Gray, Conway, & Braver, 2011	102	40	36	WM	4	.41
Colom et al., 2008	289	20	18	STM, WM	4	.33
Dang, Braeken, Ferrer, & Liu, 2012	368	40	36	WM	6	.16
Martens & Johnson, 2009	97	40	36	WM	4	.34
Martínez & Colom, 2009	265	20	18	WM	3	.28
Martínez et al., 2011	185	20	18	STM, WM	6	.25
Schweizer & Moosbrugger, 2004	120	40	36	WM	2	.33
Shelton et al., 2010	172	40	36	WM	2	.21
Wilhelm & Schultze, 2002	114	20	18	WM	3	.42
Unspeeded testing						
Conway et al., 2002	120	unlimit.	60	WM	3	.24
Engle et al., 1999	133	unlimit.	60	STM, WM	6	.27
Kaufman et al., 2009	169	45	36	OSPAN	1	.30
Mrazek et al., 2012	131	20	12	OSPAN	1	.24

Note. WM = working memory tasks, STM = short-term memory tasks, OSPAN = operation span task, unlim. = unlimited time, dep./ind. cond. = dependent/independent condition.

possible time constraint conditions: (a) half the test's recommended administration times, (b) their recommended administration times, and (c) virtually no time constraint (i.e., 1 h allowed for each test). I assessed WMC with two WM tasks applied in the same manner in all tested conditions.

Previous research comparing speeded and unspeeded administrations of intelligence tests suggested which results may be expected in the cases of speeded versus unspeeded testing. Firstly, the data indicates that participants solve most of the test items they can solve at all in the first 20 min of testing, and mean scores on speeded tests are close to scores on unspeeded tests (Hamel & Schmittmann, 2006; Heron & Chown, 1967; Salthouse, 1993). Secondly, the former scores nicely predict the latter, being similarly reliable (Hamel & Schmittmann, 2006; Wilhelm & Schultze, 2002). In contrast to these observations, speeded versus unspeeded tests yield different patterns of errors, and the constructs of "speeded intelligence" and "unspeeded intelligence" can be easily differentiated (Partchev & De Boeck, 2012). This result suggests that though timed and untimed tests can show similar reliability, they may have different construct validity, and so they may yield different relations to other cognitive abilities, including WMC.

2. Study 1

2.1. Method

The most important decision regarding the design of the present study dealt with WM measurement. For practical reasons, due to the large sample to be examined, I had to choose a small number of tasks from a broad pool of existing computerized WM tests. I decided to use one visual recognition memory (STM) task, which required participants to maintain a pattern of items in memory and to compare one of the items with a probe, and to apply one relation monitoring task, which did not require any memorization but involved the integration of relations among perceptually available stimuli. Each task involved letter stimuli, in order to reject any possibility that the relation between WMC and Gf tests could arise due to the use of the similar material as figural stimuli included in Gf tests. Both WM tasks yielded significant correlations with Gf as well as they seemed to be especially simple and to have relatively clear theoretical interpretations pertaining to the sheer capacity of the active buffer of WM (e.g., Chuderski, Taraday, Nęcka, & Smoleń, 2012; Cowan et al., 2006; Oberauer et al., 2008; Rouder et al., 2011). Indeed, with computational simulations, it was demonstrated (Chuderski, Andrelczyk, & Smoleń, 2013) that the number of bindings maintained by the model of each task allows for the exact replication of a distribution of scores which had been observed in human participants. Finally, those STM tests and monitoring tasks yielded almost perfect WM–Gf links in previous studies (Colom et al., 2005, 2004; Martínez et al., 2011; Oberauer et al., 2008).

No controversy regarded the choice of fluid intelligence tests. The first test applied was Raven, which is a hallmark test of Gf. The second tool that I used, a Figural Analogies Test (Analogies; Orzechowski & Chuderski, 2007), was especially designed in order to precisely match scores on Raven: it also contains 36 figural items and yields a mean score and *SD* comparable to those observed in Raven. Both tests were applied in paper-and-pencil versions in order to rule out any interpretation of the WM–Gf link pertaining to the shared (i.e., computerized) method of testing.

2.1.1. Participants and procedure

A total of 1377 people participated in the fluid intelligence tests (855 women, *M* age = 22.9 years, *SD* age = 4.2 years, range 15–46). Most of them were recruited via publicly accessible social networking websites, in order to obtain a wide range of intellectual abilities (i.e., it was not a student sample). The testing took place in the professional laboratory of The Institute of Psychology, Jagiellonian University, Krakow, Poland, except for an examination of 170 participants, which took part in another laboratory, in Lodz, Poland. For participation, each person received the equivalent of 5 to 10 EUR (depending on the experiment, see below) in Polish zloty. Each person filled a written consent to participation and was informed that she or he could stop and leave the laboratory at will. Soft drinks and sweets were available to participants for the whole duration of the study. During WM testing, participants were provided with headphones in order to block any distraction.

Participants were assigned to three groups, depending on the time allowed. The highly speeded group, which had 20 min to solve Raven and 15 min to work on Analogies, contained 410 participants. The moderately speeded group, which had 40 min to complete Raven and 30 min for Analogies, contained 494 participants. The unspeeded group, which was allowed 1 h to work on each test, contained the remaining 473 participants. These samples were used to analyze the reliability of the Gf tests and for the manipulation check. A subset of 890 participants were tested also for WMC (298 in the highly speeded group, 289 in the moderate speed group, and 303 in the unspeeded group). These latter samples were used for the calculation of CFA models.

Because of the large number of participants examined and the practical problems related to that number (e.g., recruitment, financing), the measurements reported below were administered during several experiments that were related to other research projects, involved additional tasks, and had distinct research aims. These experiments have been reported elsewhere (Chuderski & Nęcka, 2010, 2012; Chuderski & Stettner, 2013; Chuderski et al., 2012). However, no data presented in this paper has been previously published. Moreover, in all experiments which included the WM measurement, a standardized procedure was used to apply the two WM tasks and two Gf tests. In each experiment, a few (from three to five, depending on the experiment) computerized tasks unrelated to the present study were applied first, then the STM and relation monitoring tasks were applied (in this order). The whole computerized session lasted from 1 to 2 h, depending on the number of tasks involved. After a short break, Raven was administered, followed by Analogies.

2.1.2. Administered tests

2.1.2.1. Visual recognition memory task. I used a modified so-called change detection paradigm (Luck & Vogel, 1997; Rouder et al., 2011). Each of the 90 trials of the task consisted of a virtual, four by four array filled with a few stimuli (i.e., only some cells in the array were filled). The stimuli were ten Greek symbols (e.g., α, β, χ, and so on), each approximately

2×2 cm in size. The number of stimuli within the array varied from five to seven items. The array was presented for the time equal to the number of its items multiplied by 400 ms, and then followed by a black square mask of the same size as the array, presented for 1.2 s. In a random 50% of trials, the second array was identical to the first one, while in the remaining trials both arrays differed by exactly one item at one position. If they differed, then 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 two arrays. The second array was shown until a response was given or 4 s elapsed. The trials were self-paced.

The score on this task is the estimated sheer capacity of the active buffer of WM (Cowan, 2001; Rouder et al., 2011), based on the proportion of hits (*H*, correct responses for arrays with one item changed) and the proportion of false alarms (*FA*, incorrect responses for unchanged arrays). The capacity of the buffer is estimated to be *k* items (out of *N* items of a memory load), on the assumption that a participant produces a correct hit or avoids a false alarm only if a cued item is transferred to his or her buffer (with the k/N chance). If a non-transferred item is cued, then a participant is assumed to be guessing the answer. In consequence, the following formula evaluates the score on the task for each N: $k = N \times (H - FA)$. The total score on this task was the mean from the values of *k* in the three *N* conditions (i.e., 5–7).

2.1.2.2. Relation monitoring task. I used a slightly modified, no-memory version of the monitoring task introduced by Oberauer et al. (2008). The task consisted of the presentation of 80 patterns. Each pattern consisted of a three by three array of two-letter syllables (each approx. 3.0×2.5 cm in size). Each syllable was composed of a capital consonant and one of four vowels (A, E, O, or U). Two subsequent arrays differed by exactly one syllable. A participant's goal was to detect if three syllables ending with the same vowel were located in one row, column, or diagonal line (i.e., this was a simpler version of a "rhyming" task). As many as 20 patterns matched this simple relation. Responses given when stimuli did not form such a relation were interpreted as false alarm errors. In order to minimize the influence of either processing speed or visual search efficiency, I allowed 250% more time (5 s) for each trial than the time allowed in Oberauer et al.'s (2008) original study (2 s). The score on this task was the number of correct responses for the specified relation minus one third of the false alarm errors (as three times more non-target patterns occurred than target patterns).

2.1.2.3. Raven's Advanced Progressive Matrices. The test (Raven, Court, & Raven, 1983) consists of 36 items that include a three-by-three matrix of figural patterns which is missing the bottom-right pattern, and eight response options which are the patterns that can potentially match a missing one. The participant's task is to discover the rules that govern the distribution of patterns (see Carpenter et al., 1990) and to apply them to response options in order to choose the one and only right pattern. Responses were recorded on an answer form. The total number of correctly answered items was the score on this test.

2.1.2.4. Figural Analogy Test. This test includes 36 figural analogies in the form 'A is to B as C is to X', where A, B, and C are the types of relatively simple patterns of figures, A is related to B according to two, three, four, or five latent rules (e.g., symmetry, rotation, change in size, color, thickness, number of objects, etc.), 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. Responses were recorded on an answer form. As in Raven, the total number of correct answers was taken as the score.

2.1.3. Calculation and evaluation of confirmatory factor analysis (CFA) models

For CFA computations, I used Statistica software (version 9) with maximum-likelihood estimation. For each group, I calculated a simple model which correlated the Gf latent variable, loaded by the scores on two intelligence tests, with the WMC latent variable, loaded by the scores on two WM tasks. The goodness of fit of CFA models was evaluated with two measures: Bentler's comparative fit index (CFI), and the root mean square of approximation (RMSEA). I adopted the following criteria of a good fit of models: CFI should be higher than .95 and RMSEA should not surpass the value of .08. Because large samples were used (i.e., larger than 200), in estimating an absolute fit of a model, I did not consider χ^2 statistic, which underestimates fits in such a case. Nevertheless, the increase in χ^2 was used for comparing the relative fits of nested models. I expected that all three models would fit well. The main aim was to compare correlation coefficients between the latent variables in the three models.

2.2. Results

2.2.1. Descriptive statistics

Table 2 presents descriptive statistics and reliabilities for the WM tasks, and for the Gf tests in each group. All variables approximated the normal distribution. Reliabilities of all scores were high.

2.2.2. Manipulation check

In the total sample, the effect of time allowed for fulfilling the Gf tests (i.e., the effect of the time pressure) on Gf scores was significant in both Raven, $F(2, 1374) = 30.24$, $p < .001$, $\eta^2 = .04$, and Analogies, $F(2, 1374) = 99.04$, $p < .001$, $\eta^2 = .13$. Planned comparisons revealed that the moderately speeded group scored substantially higher than the highly speeded group, in both Raven, $\Delta = 3.38$, $F(1, 1374) = 60.47$, $p < .001$, and Analogies, $\Delta = 5.32$, $F(1, 1374) = 168.23$, $p < .001$. However, the unspeeded group scored on Raven slightly but significantly lower than the moderately speeded group, $\Delta = -1.47$, $F(1, 1374) = 12.96$, $p < .001$ (no significant difference in the Analogies score pertained to both these groups, $\Delta = -0.51$, $p = .197$). As an additional time should allow participants to attempt more test items, this result was surprising. A closer inspection of data indicated that a subsample of 170 participants in the unspeeded group, who were tested in another city (all of them were not tested with WM tasks), was solely responsible for such an unexpected decrease in Raven scores, for unknown reasons (but possibly due to a lower socioeconomical status of the city from which those people came). Therefore, I re-ran ANOVA

Table 2
Descriptive statistics and reliabilities for all measures in Study 1.

Measure	*M*	*SD*	Range	Skew	Kurtosis	Reliability
Short-term memory task	3.20	1.18	0.00–5.89	−0.40	−0.37	.86
Relation monitoring task	0.81	.15	0.00–1.00	−1.43	2.96	.86
Raven 20-min version	18.95	5.63	2.00–32.00	−0.57	0.06	.84
Raven 40-min version	22.23	6.05	5.00–36.00	−0.46	−0.01	.85
Raven 60-min version	20.76	7.12	2.00–36.00	−0.35	−0.31	.89
Analogies 15-min version	18.38	5.26	3.00–32.00	−0.09	−0.14	.77
Analogies 30-min version	23.70	6.16	2.00–36.00	−0.52	−0.04	.83
Analogies 60-min version	23.19	6.81	3.00–36.00	−0.54	−0.25	.86

Note. $N = 890$ for the short-term memory and relation monitoring tasks. In the cases of Raven and Analogies, $N = 410$ for the 20/15 min versions, $N = 494$ for the 40/30 min versions, and $N = 473$ for the 60 min versions. Reliability = Cronbach's alpha.

using the data set that excluded those participants from the unspeeded group. In result, the unspeeded group did not significantly differ in Raven in comparison to the moderately speeded group ($M = 22.78$ vs. $M = 22.23$, respectively), $p = .207$, while in Analogies the unspeeded group scored significantly higher than the moderately speeded group ($M = 25.03$ vs. $M = 23.70$, respectively), $F(1, 1204) = 10.14$, $p = .001$.

Next, I looked for test items which mostly suffered from imposing the time pressure. I aggregated scores on particular test items into four bins, representing the first, second, third, and last set of nine items, separately for each test. The 3 (group) × 4 (bin) ANOVA yielded a significant interactive effect in Raven, $F(6, 4122) = 25.92$, $p < .001$, $\eta^2 = .04$, as well as in Analogies, $F(6, 4122) = 61.35$, $p < .001$, $\eta^2 = .08$. The accuracy of solving consecutive items of each test in relation to allowed time is presented in Fig. 1. When comparing the speeded group with the two other groups, this figure clearly indicates that, under time pressure, accuracy decreased mostly in the late, most difficult items, while accuracy in the early, easiest items suffered less or even did not change at all. In the first halves of both Gf tests (averaged), the mean difference in accuracy between the highly and moderately speeded groups and the unspeeded group was only $\Delta = 0.66$ (3.7%), while in the second halves of the tests, the same difference equaled as much as $\Delta = 3.17$ (17.6%).

2.2.3. Correlations and CFA models

Table 3 presents correlations between all measures for each group. Statistical control over age (with the use of partial correlations) barely changed this pattern of correlations: each correlation coefficient varied by no more than $\Delta r = .04$. As all groups fulfilled the WM tasks under the same conditions, obviously no significant difference in the WM tasks' correlation between groups was observed (all $ps > .06$, two-tailed t tests). Correlations between the Gf tests in all groups were also comparable and no significant difference was found between them (all $ps > .10$, two-tailed t tests). The two latter observations indicate that in each group both WM and Gf latent variables should represent similar amounts of variance shared by the WM tasks and the Gf tests, respectively. Mean correlations between WM task scores and Gf test scores were $r = .38$, $r = 40$, and $r = .29$, for the highly speeded, moderately speeded, and unspeeded groups, respectively. These data suggest that the correlation between both latent variables may decrease in the case of the latter group. That observation was tested by CFA analysis.

In each calculated model (one for each group), all presented in Fig. 2, the loadings of observed variables on latent variables (all highly significant) were comparable. All models fitted the data very well: $N = 298$, $df = 1$, CFI = 1.0, RMSEA = .00, in the highly speeded group, $N = 289$, $df = 1$, CFI = .994, RMSEA = .076, in the moderate speed group,

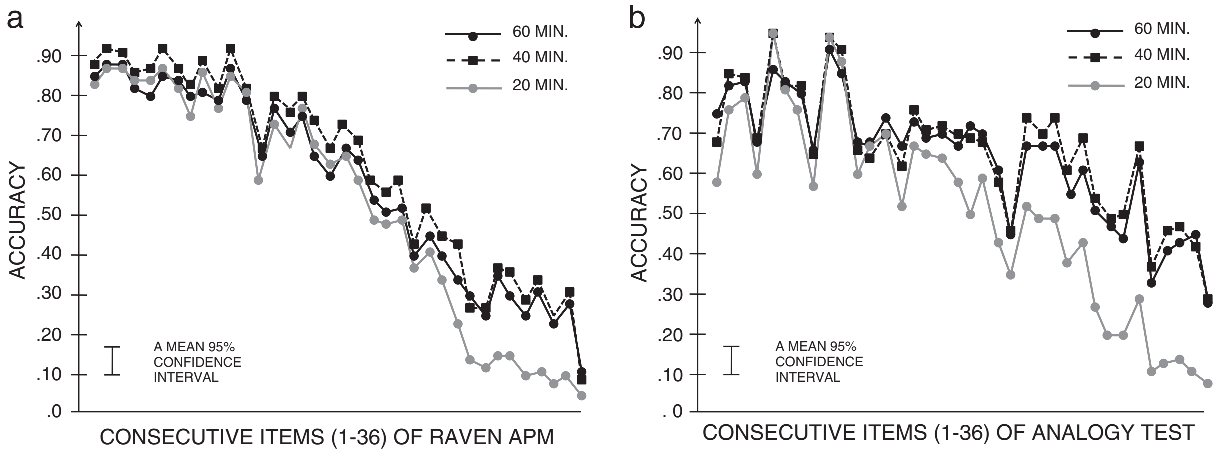

Fig. 1. The accuracy for each problem in Set II of the Raven Advanced Progressive Matrices test (Panel a) and for each problem in the Figural Analogy Test (Panel b), in the speeded (gray solid line), the medium speed (black dashed line), and the unspeeded (black solid line) groups.

Table 3
Correlation matrix for all scores in Study 1.

Measure	1	2	3	5	7
1. Short-term memory	–				
2. Relation monitoring	.32	–			
3. Raven 20 min	.42	.39	–		
4. Analogies 20 min	.38	.32	.59		
5. Raven 40 min	.38	.41		–	
6. Analogies 30 min	.43	.37		.67	
7. Raven 60 min	.31	.34			–
8. Analogies 60 min	.23	.29			.61

Note. $N = 890$ for the correlation of the short-term memory and relation monitoring tasks. In the cases of correlations between WM tasks and Gf tests and between Gf tests, which were used in CFA computations, $N = 298$ for the speeded group, $N = 289$ for the medium group, and $N = 303$ for the unspeeded group. All correlations were significant at least at $p < .05$.

and $N = 303$, $df = 1$, CFI = 1.0, RMSEA = .00, in the unspeeded group. Most importantly for the aims of the present study, the CFAs indicated that the Gf–WMC correlation, as estimated on the latent variable level, was significantly stronger in the highly speeded group ($r = 1.0$) than in the moderate speed group ($r = .83$), $\Delta = .17$, $t(590) = 2.40$, $p = .017$ (a two-tailed test), and it was significantly stronger in the moderate speed group than in the unspeeded group ($r = .62$), $\Delta = .21$, $t(585) = 2.88$, $p = .004$ (a two-tailed test). In the case of the highly speeded group, WMC explained all the variance in fluid intelligence. On the contrary, in the case of the unspeeded group, WMC accounted for only 38% of Gf variance. Such a difference was not dependent on the use of latent variables and the maximum-likelihood estimation, because when I tested Pearson correlations (corrected for attenuation) between mean *z* scores on two Gf tests and mean *z* scores on two WM tasks, respectively, the difference between the highly speeded ($r = .65$) and unspeeded ($r = .46$) groups was still highly significant, $\Delta = .19$, $t(599) = 3.76$, $p < .001$ (a two-tailed test). Of course, those

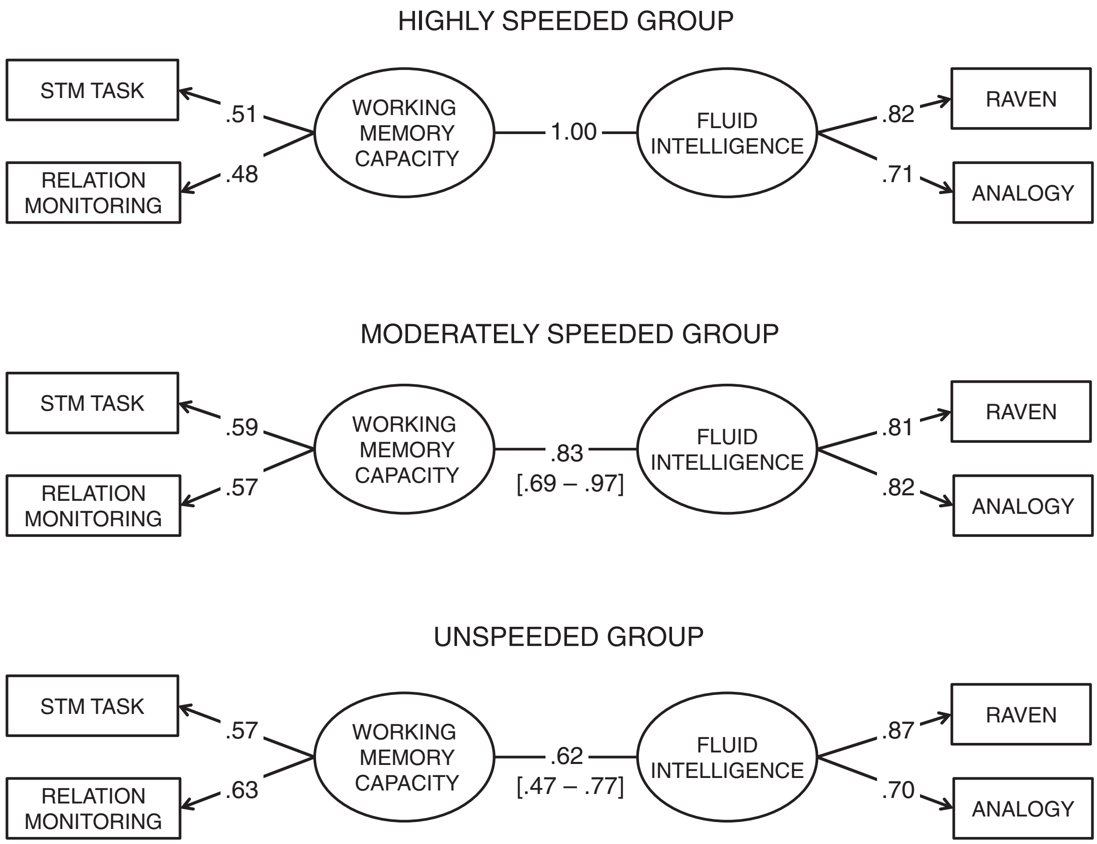

Fig. 2. Confirmatory factor analysis models relating the working memory capacity and fluid intelligence latent variables (represented by ovals), for the speeded (top panel), medium speed (middle panel), and unspeeded (bottom panel) groups. Boxes represent manifest variables, while values between ovals and boxes represent relevant standardized factor loadings (all *p*s < .001). Values between ovals represent correlation coefficients between latent variables. Respective 95% confidence intervals are shown in brackets.

latter correlations were much weaker than the respective links derived from CFA, because of the commonly-known differences between both estimation methods.

As the WM–Gf correlation coefficient was relatively high in the moderately speeded group, I also tested whether fixing it to one will yield a comparable fit of the resulting model, in comparison to the original model. The new model fitted acceptably, $N = 289$, $df = 2$, CFI = .982, RMSEA = .093, but the results of comparison between the models suggested that it fitted marginally worse than the original model, $\Delta\chi^2 = 4.66$, $\Delta df = 1$, $p = .031$. So, it seems that the most plausible estimate of the "true" strength of the WM–Gf correlation in moderately speeded Gf testing is definitely very high, though probably it does not reach unity.

2.2.4. Additional analyses

One possible explanation for the observed difference in the strength of the WMC–Gf correlation may be due to the fact that the drop in accuracy in the highly speeded group was mostly seen in the second parts (i.e., 18 late items) of both Raven and Analogies. As the second half of each test was more difficult than the first, they might have required other types of processes or resources. For instance, the easier items might have relied on the effective manipulation of matrix elements in WM, while the more difficult items could not be solved by a direct recomposition of these elements, but additionally they might have involved some longer lasting processes above and beyond WM. Because in the highly speeded group, the participants rarely succeeded to solve the late items, their total scores reflected scores on easier items to a greater extent than the total scores in the unspeeded group, and due to that fact the respective correlation with WMC might have been stronger in the former case. In the latter group, the total score might have reflected a compound of scores related to WM and scores related to some other faculty measured by the late items, which might have decreased the respective correlation. I tested this possibility by calculating CFA models including two latent variables reflecting Gf, one loaded by scores from 18 early items in each test, and the other loaded by scores from 18 late items. The three models (all CFIs > .959) are presented in Fig. 3. No difference between the coefficients of correlations between WMC and the variables representing the first and second halves of Gf tests in respective groups exceeds $\Delta r = .10$, and none was significant. Thus, there is little support for a hypothesis predicting that either taking or leaving the late items of intelligence tests can result in either weaker or stronger links between WMC and Gf. The correlation between WMC and the scores on early items was influenced by the time

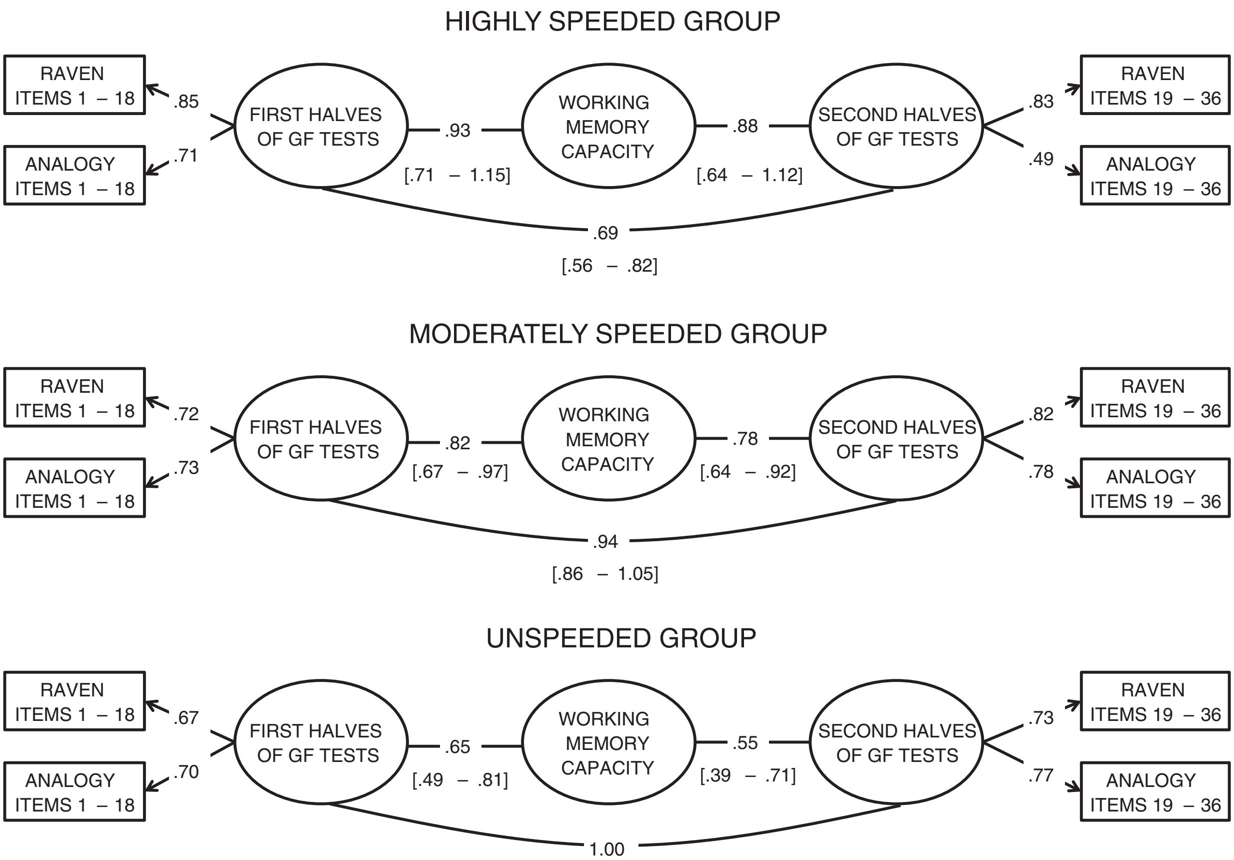

Fig. 3. Confirmatory factor analysis models relating the working memory capacity latent variable and the variables reflecting the early versus late items of the Gf tests, for the speeded (top panel), medium speed (middle panel), and unspeeded (bottom panel) groups. The latent variables are represented by ovals. Boxes represent manifest variables, while values between ovals and boxes represent relevant standardized factor loadings (all *ps* < .001). Values between ovals represent correlation coefficients between latent variables. Respective 95% confidence intervals are shown in brackets.

pressure in a comparable way as was the correlation between WMC and the scores on late items.

So, what other factor could be responsible for such a huge increase in WM–Gf correlation strength, relating to both halves of each Gf test, in the effect of time pressure? One plausible hypothesis can be derived from the aforementioned research suggesting that effective learning is an important component of fluid intelligence. Participants who are able to use their experience of coping with previous test items to enhance their discovery of relations in subsequent items will be judged more psychometrically intelligent than those who cannot learn. However, how does time pressure influence such learning during intelligence testing?

One study, which could provide important insights related to the above question was done by Lerch, Gonzalez, and Lebiere (1999). They reported data which indicate that time pressure present during a real-time dynamic decision making task (requiring the control of a virtual water distribution system) impaired learning of the task. It was shown that people scoring low in Raven learned the task effectively only if time pressure was eliminated, while high ability people could learn the task regardless of whether time pressure was applied or not, though under no time pressure conditions the former learned more than did the latter. According to Lerch et al., learning requires spare cognitive resources. Under time pressure, low-resource participants have to devote all their resources to coping with the task, and this blocks their learning capability. High-resource participants are able to perform the speeded task and simultaneously learn it. However, when no pressure is present, low-resource persons can interleave processing and learning, and as they start from a lower level of performance than the high-resource ones, they may benefit more from learning than the latter people, who have already approached ceiling. The reported study included only a small sample of 33 people and did not allow for decisive conclusions. However, because Lerch et al.'s decision making task was a spatial one and it required the induction of abstract rules from events on the computer screen, it seems that the predictions from this task might be also valid for my study.

Because in the present study I always applied Raven before Analogies, and the latter was designed to provide a score close to Raven's score, I was able to analyze the difference in both scores in relation to the tests' administration time. Such a difference was treated as the index of learning,[1] indicating how much participants had benefited from coping with Raven when fulfilling Analogies. If the resource-dependent learning hypothesis is right, then under time pressure, low-WMC people should be unable to significantly increase their scores on Analogies (compared to Raven). However, when time pressure is reduced, learning ability may become unrelated to WMC (as both low- and high-WMC people will have spare resources to learn), and because of ceiling effects, high-WMC participants might now not be able to learn as much as low-WMC participants. The latter, due to discovered knowledge, may override their WM limitations and catch up with more capacious people. Thus, when time pressure is high, Gf scores will be mostly determined by the amount of WM resources that can be devoted to the processing of relations. On the contrary, under low pressure conditions, WM may determine only part of Gf variance, while the WM-unrelated learning effects would account for some other part of this variance.

I tested this prediction by comparing participants belonging to the 33 and 66 percentiles of a mean from z scores on both WM tasks (referred to as WMC-low and WMC-high groups; for examples of using a similar method see Kane & Engle, 2003; Unsworth & Engle, 2007). Due to the large sample, around 100 participants could be included in each group × each time condition. I submitted both these factors (i.e., the two groups and three conditions) to ANOVA in the index of learning. In the unspeeded condition, the mean gain in score equaled $\Delta = 2.18$ ($SD = 0.37$). In the moderately speeded condition, the learning effect was also significant ($\Delta = 1.69$, $SD = .36$), while in the highly speeded condition, there was no learning effect ($\Delta = -.16$, $SD = .39$). The interaction of group and time pressure factors was significant, $F(2, 595) = 3.63$, $p = .027$, $\eta^2 = .01$ (see Fig. 4), and indicated that the WM groups significantly differed in the index of learning, $F(1, 595) = 13.62$, $p < .001$, only in the unspeeded condition. In this case, WMC-low participants increased their score by $\Delta = 3.52$ on average ($SD = 0.51$), while WMC-high did it by only $\Delta = 0.83$ ($SD = 0.52$). The differences in the index of learning between the WM groups in the highly and moderately speeded conditions were not significant, $p = .182$ and $p = .958$, respectively. So, contrary to Lerch et al.'s (1999) results, under time pressure exerted in this study, highly capacious people could not learn from test to test similarly as could not low capacious ones.

2.3. Discussion

The most important result of the Study 1 consists of the observation that in one group of participants, who were allowed only half the recommended administration time for each Gf test, WM and Gf were found to be isomorphic, namely

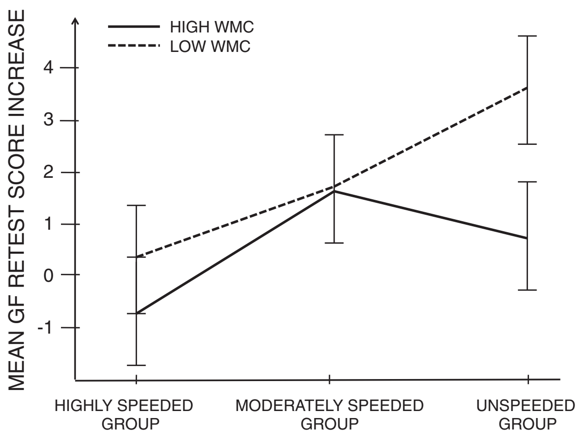

Fig. 4. The index of learning, computed as the mean difference in scores between the Raven APM and Figural Analogy Tests, for participants from both 33 (low WMC people) and 66 (high WMC people) percentiles of WMC distribution, in the speeded, medium speed, and unspeeded groups, respectively. Vertical bars represent 95% confidence intervals.

[1] I fully acknowledge that it is a simplification, as most probably scores on some late items of Raven's test could already reflect some amount of learning from the early items.

there was no amount of variance in Gf above and beyond the variance explained by WMC. So, WMC was the sole determinant of scores on highly speeded tests. The former factor was also able to explain the major part of variance in scores on moderately speeded tests, that is, tests applied in their recommended time. Importantly, both those results were obtained even though I measured Gf with paper-and-pencil figural tests, while WMC was assessed with computerized letter tasks. In contrast, in the group of participants for whom Gf administration time was barely limited, WMC accounted for only 38% of variance in Gf, and so the Gf–WMC relation was far from being isomorphic. Interestingly, the reliability of Gf tests was only slightly influenced by the induced time pressure, compared to the reliability estimates in the unspeeded condition.

Although my time limit manipulations primarily influenced scores on the more difficult Gf test items, this manipulation altered the strength of correlation between WMC and both the easy and difficult items in a similar way. Thus, the substantial difference in the amount of variance shared by WMC and Gf, related to time pressure, cannot be explained away in terms of the bare number of test items that had been attempted/solved by participants. Most probably, under no time pressure, the scores on a Gf test are determined not only by WMC, but also by other factors not directly pertaining to WM.

However, one limitation of the present study regards the fact that due to the large sample, only two Gf tests were used. One could argue that Gf operationalized in such a way may have low criterion validity, and when measured with a larger battery of tests it could yield a different pattern of relations with WM. However, I used Raven, which is probably the most central test to fluid intelligence (see Snow, Kyllonen, & Marshalek, 1984), and has been most widely used in diverse types of studies on intelligence. Another test that was used showed internal reliability comparable to Raven and strongly correlated with it (r equaled .69 in the whole sample). So, it seems that the obtained results can easily be generalized on a more broadly defined construct of fluid intelligence.

A relatively more serious limitation may be connected with my use of only two WM tasks. Especially, it may be argued that if more WM tasks were used, then WMC would be tapped more comprehensively, and the correlation in the unspeeded condition could also approach one. However, in previous studies both tasks which I used appeared especially effective in the measurement of individual differences in WMC, highly correlated with Gf, and had thorough theoretical justification (e.g., Cowan et al., 2006; Oberauer et al., 2008). Moreover, similar results regarding speeded (i.e., the Raven and single WM task correlations reaching about $r = .4$) versus unspeeded (the Raven and single WM task correlations falling below $r = .3$) intelligence testing were found in the presented metaanalysis of existing studies, which relied on greater numbers of WM tasks (though no study directly compared the timed and untimed conditions). Finally, the lower Gf–WM correlation was observed solely in the unspeeded condition, while the WM latent variable based on only two tasks was able to explain the whole/major part of Gf variance in the highly/moderately speeded conditions. So, it seems that the WM latent variable based on only two WM tasks nevertheless validly captured most of the WM construct.

The second important observation provided by Study 1 consists of the identification of a factor which seems to contribute to Gf above and beyond the contribution of WM: it is the ability to learn from testing experience and to transfer learned skills/knowledge to consecutive test items. In the unspeeded condition, when Raven's score was taken as a baseline, participants having low WMC increased their mean score on Analogies by as much as 3.5 correctly solved items, while high-WMC participants only added to Raven's score less than 1 item on average. The former result – an increase of almost 18% from Raven to Analogies – constitutes a considerable performance improvement of less capacious participants, as compared to the highly speeded condition which yielded no significant difference between scores on Raven and Analogies (i.e., little learning). The fact that, surprisingly, low-WMC persons were able to learn better than high-WMC ones might have resulted from that less capacious participants started from lower levels of performance and due to the steep rise of the learning curve at this point they quickly gained a few additional items on the basis of their experience of coping with Raven, while more capacious persons, who had already approached their ceilings, could not increase their scores due to learning by that much.

However, the hypothesis assuming the important contribution of learning to Gf surely needs more data, as in Study 1 the level of processing on matrix items and the ability to learn relational knowledge were confounded (they both were estimated on the basis of the scores from Gf tests). Ideally, the latter factor should be measured by dedicated tests of relational learning, other than Gf tests, and then used to predict Gf.

The next study was designed with the goal to overcome the above mentioned limitations of Study 1, with regard to unspeeded intelligence testing. Firstly, three figural tests of intelligence were used, instead of two. Secondly, each WM task (i.e., the monitoring task and the STM task) was used in two versions, resulting in four tasks in total. Finally, I introduced a novel test of relation (i.e., relational concept) discovery, which consisted of two structurally isomorphic parts (but differing in the involved stimuli and the imposed difficulty), applied in the fixed order, so as an index of relational learning could be computed (i.e., the difference in performance in the part applied as the second one in comparison to the part applied as the first).

3. Study 2

This study was a part of a larger project, which consisted of seventeen computerized tasks, including four complex span tasks, two Stroop tasks, three antisaccade task, three n-back tasks, and the stop signal task, all related to other research goals and thus not reported in this paper, as well as the STM tasks and relation monitoring tests described below. The computerized session lasted about 4 h (including proper breaks), and the two reported monitoring tasks were presented as a third and a fourth task in a row, while the STM tasks were test numbers eleven and fifteen in a row. Four ability tests, which are reported below, were administered in a separate session on the same day, which also lasted 4 h. The order of tests was the following: relational discovery test, computerized analogy test, Raven, and paper-and-pencil Analogies.

3.1. Method

3.1.1. Participants and procedure

A total of 243 people participated (142 women, *M* age = 24.3 years, *SD* age = 5.0, range 18–45 years). All of them were recruited via publicly accessible social networking websites, and tested in the laboratory of The Institute of Psychology, Jagiellonian University, Krakow, Poland. For participation, each person received the equivalent of 15 EUR in Polish zloty. The same testing conditions applied as in Study 1. In the final sample, data from six people were discarded because of their failure to provide even one elaborate description in the relation discovery test. Also, another 79 participants (50 women, *M* age = 24.7 years, *SD* age = 5.4, range 19–41 years) were tested only with the second part of the relation discovery test (it was preceded by five computerized tasks, neither reported here). These participants constituted a control group used in the analysis of whether the transfer of relational knowledge from the first to the second part of the relation discovery test indeed took place.

3.1.2. Administered tests

3.1.2.1. Visual recognition memory task. I used the same task as in Study 1, with three modifications. Firstly, apart from the letter version of the task, the digit version was used, which included digits 0–9 as stimuli, each approximately 2 × 2 cm in size. Secondly, the number of stimuli within the array in each version of the task could count five, seven, or nine items. Finally, the array was presented for the time equal to the number of its items multiplied by 300 ms, instead of 400 ms applied in Study 1. All other details of the recognition memory task were the same as in Study 1.

3.1.2.2. Relation monitoring task. The same task was used as in Study 1, with three exceptions. One was that the critical relation (i.e., to find three strings ending with identical symbols) now could appear only in either rows or columns (i.e., diagonal lines were excluded). Secondly, I used also the number version of the task, with three-digit numbers as stimuli. Thirdly, each task consisted of the presentation of 40 patterns, 20 containing the relation, and another 20 lacking it. Thus, the score on each task was the number of correct responses for the specified relation minus a total of false alarm errors. All other details of the relation monitoring task were the same as in Study 1.

3.1.2.3. Raven's Advanced Progressive Matrices. The same version of the test was used as in Study 1. One hour was allowed for fulfilling the test.

3.1.2.4. Figural Analogy Test. The same version of the test was used as in Study 1. Forty five minutes were allowed for fulfilling the test, as previous studies indicated that it appeared to be the sufficient time for the majority of participants.

3.1.2.5. Computerized Figural Analogy Test. This test is a computerized and substantially modified version of the paper-and-pencil Figural Analogy Test. The test includes 48 figural analogies in the form 'A is to B as C is to X', where A, B, and C are the types of relatively complex patterns of figures, each including either 5 or 8 figures (depending on the test item). In each item, A is related to B according to two to eight latent rules (rotation, change in location, color, thickness, filling, etc.), and X has to be selected by clicking with a mouse on one out of seven alternative answer patterns. The one and only pattern should be chosen which relates to pattern C, as B relates to A. After two training items, the participants were allowed up to 4 min for solving each test item. The total number of correct answers was taken as the score.

3.1.2.6. The test of discovery of relations. The DREL (Discovery of RELations) paper-and-pencil test consists of two, letter and digit, parts. Each part includes 15 items. Each item consists of six four-symbol strings, which are governed by a to-be-discovered relation, and another three strings, which form counterexamples for that relation, that is, the discovered relation must exclude all three counterexamples. A participant is required to write down a concise and abstract description of a relation that matches six positive exemplar strings. The counterexamples are introduced in order to prevent describing too general relations (e.g., all strings consist of four symbols). In each part of the test, there are five binary, five ternary, and five quaternary relations, and item positions for each complexity level with regard to the beginning of the test are balanced. Each part of the test is preceded by a detailed instruction, which explains the way in which relations should describe example strings, while excluding counterexamples, and which presents two sample unitary relations.

In the first part of the test, symbols in each string are two different letters, and a relation governs the place of each letter relative to some number of remaining letters in a string. An example of a binary-relation item requires the discovery of a relation *the same two letters in the middle are different from the same two letters on the extremes*:

OEEO LSSL BVVB

ZKKZ NUUN YAAY

~~RRVV AKAK PPLL~~

Binary relations require pairwise comparisons of two symbols. There is only one mental model corresponding to binary relations (in the case of this example: *abba*).

An instance of ternary relation is *one and only letter different from three other identical letters is always placed in the middle* (it requires the simultaneous relating of three symbols, and the corresponding models are: *aaba* and *abaa*):

ZEZZ LLUL NRNN

ASAA JJWJ PBPP

~~OLLL KKKN VVVB~~

In the most difficult, quaternary-relation items, all four symbols have to be related in one step. An example relation is *the first letter is different from the second one or the third one or*

Table 4
Descriptive statistics and reliabilities for all measures in Study 2.

Measure	*M*	*SD*	Range	Skew	Kurtosis	Reliability
Letter STM task	2.48	1.41	−1.33–6.40	−0.04	−0.14	.73
Digit STM task	4.65	1.43	−0.63–7.00	−1.10	1.11	.83
Letter monitoring task	0.76	0.19	−0.10–1.00	−1.64	3.57	.75
Number monitoring task	0.75	0.20	−0.05–1.00	−1.40	2.23	.74
Learning index — comp. 2	1.73	5.49	−9.00–13.00	0.10	−0.95	.87
Learning index — comp. 3 & 4	0.73	1.90	−4.00–5.50	−0.21	−0.31	.88
Raven	22.02	6.56	4.00–35.00	−0.45	−0.13	.88
Paper-and-pencil analogies	22.32	6.48	6.00–35.00	−0.25	−0.75	.86
Computerized analogies	22.06	11.73	1.00–48.00	0.42	−0.86	.93

Note. $N = 237$. STM = short-term memory, comp. = complexity. Reliability = Cronbach's alpha.

both, *and the third letter is different from the fourth one* (three corresponding models: *aaba*, *abab*, and *abba*):

GGRG NHNH FDDF

BEEB OOXO ACAC

~~FFFF NNNP JJSS~~

The only difference between the first and the second part of the test is that symbols are digits, and relations pertain to their evenness or oddness. However, the abstract structure of the relations of corresponding items in both tests is identical. For example, the digit version of the aforementioned binary relation would be: *two digits in the middle are both odd or both even*, *and in the former case two extreme digits are even*, *while in the latter case two extreme digits are odd*. This part is more difficult, as the crucial feature (evenness/oddness) is not linked to the appearance of a symbol, while the crucial feature of the letter part (identity/difference) is.

The scoring on the test depended on the abstractness of given descriptions. One point was scored if a described relation was correct and properly abstract (as in the examples), no matter what exact formulation was used by participants. Half point was scored if a description was correct, but it was not abstract enough, instead it was composed of particular subcategories of strings (usually corresponding to possible models). No score was given for incorrect descriptions, no matter if they excluded valid instances of strings or included counterexamples. The dependent variables were the differences in the total scores (in range 0 to 5) for each level of relational complexity, between the second and first part of the test (i.e., indices of how well people could apply the relational structures discovered in the letter strings to the more abstract number items). A half an hour was allowed for each part of the DREL test.

3.2. Results

3.2.1. Descriptive statistics

Table 4 presents descriptive statistics and reliabilities for the WM tasks, the Gf tests, and the indices derived from the DREL test. Scores in this latter test significantly decreased with an increasing relational complexity ($M_2 = 3.75$, $M_3 = 1.74$, and $M_4 = 1.20$), $F(2, 472) = 1523.80$, $p < .001$, $\eta^2 = .87$. Because the results on quaternary relations approached floor, in order to increase the psychometric parameters of the DREL learning indices, the indices from ternary and quaternary items were aggregated into one measure. As the check of whether relational learning really took place in the DREL, I compared the mean score on the number subtest with the respective mean score in a control group, who could not learn from the letter version. The control group scored $M = 1.34$ per condition, comparing to $M = 1.77$ in the present data, that is, as expected, there was a highly significant learning effect resulting from the previous experience with the letter part of DREL, $t(314) = 3.46$, $p < .001$.

3.2.2. Correlations and SEM models

Table 5 presents correlations between all measures. Having obtained two indices of relational learning, I could now compute the latent variable reflecting learning. I estimated a path model (SEM), which included two exogenous latent variables, one representing WMC, loaded by scores on four WM tasks, and the

Table 5
Correlation matrix for all scores in Study 2.

Measure	1	2	3	4	5	6	7	8
1. Letter STM task	–							
2. Digit STM task	.41	–						
3. Letter monitoring task	.38	.42	–					
4. Number monitoring task	.27	.33	.52	–				
5. Learning index — comp. 2	.17	.23	.33	.33	–			
6. Learning index — comp. 3 & 4	.13	.20	.26	.22	.64	–		
7. Raven	.28	.39	.46	.41	.45	.25	–	
8. Paper-and-pencil analogies	.21	.23	.36	.35	.36	.20	.65	–
9. Computerized analogies	.19	.15	.30	.23	.40	.26	.49	.49

Note. $N = 237$. STM = short-term memory, comp. = complexity. All correlations were significant at least at $p < .05$.

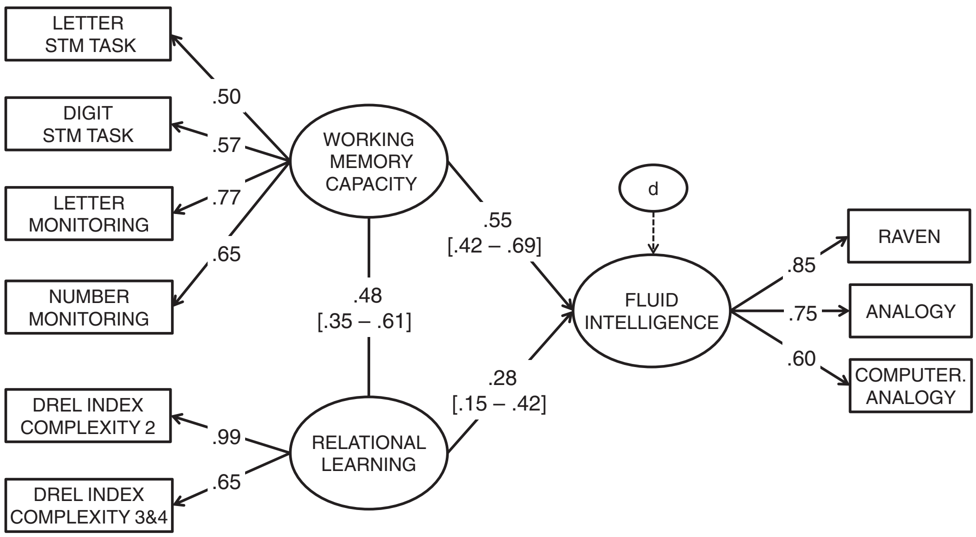

Fig. 5. The path model relating the intercorrelated working memory capacity and relational learning latent variables to the variable reflecting fluid intelligence. The latent variables are represented by ovals; small oval marked "d" stands for a disturbance term. Boxes represent manifest variables, while values between ovals and boxes represent relevant standardized factor loadings (all $ps < .001$). Values between ovals represent path coefficients between latent variables. Respective 95% confidence intervals are shown in brackets.

other reflecting the effectiveness of relational learning in the DREL test. Both these variables predicted the Gf endogenous variable that was loaded by scores on three intelligence tests. The model had a very good fit, $N = 237$, $df = 24$, CFI = .985, RMSEA = .04, and is presented in Fig. 5. The two observations from Study 1 were fully replicated with this improved method. Firstly, the WM variable was not a perfect predictor of Gf, yielding a correlation of (only) $r = .55$, even though four different measures of WMC were used. Secondly, the relational learning variable was a significant predictor of Gf, explaining about eight percent of variance in Gf.

The relational learning variable moderately and positively correlated with WM variable ($r = .48$). However, even when the former variable was calculated as an endogenous one (i.e., WM predicted both Gf and relational learning), the amounts of variance unexplained by WM (i.e., the disturbance terms) correlated significantly, $r = .23$, suggesting that relational learning predicts independently from WM at least 5.3% of Gf variance. In such a model, presented in Fig. 6, the WM–Gf path coefficient equaled $r = .67$.

In the final analysis, I tested whether calculating the model shown in Fig. 6 with the use of only pairs of tasks that were also applied in Study 1 (i.e., the letter WM tasks as well as Raven and Analogies) would in any way change the strengths of links between latent variables. The resulting, much simpler model had the perfect fit, $N = 237$, $df = 6$, CFI = 1.0, RMSEA = .00, and its WM–Gf ($\Delta r = -.03$) and WM-learning ($\Delta r = -.02$) path coefficients were virtually the same as in the original model, as was the link between Gf and DREL disturbance terms ($\Delta r = 0$). Moreover, in order to confirm that the moderate value of WM–Gf correlation does not depend on a particular choice of WM tasks, in the model shown in Fig. 6 , I substituted the original WM measures with four scores on, unreported in this paper, complex span tasks (Conway et al., 2005), which required the encoding of letter, number, spatial, and figural stimuli, respectively, intermixed with some simple processing (decisional) task, for later recall. Each task had reliability of no less than $\alpha = .85$. I observed virtually the same WM–Gf path coefficient as in the original model ($r = .68$; $\Delta r = .01$).

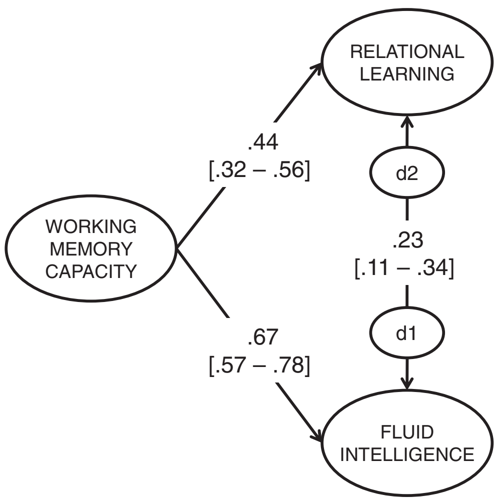

Fig. 6. The path model relating the working memory capacity latent variable to the relational learning and fluid intelligence variables, whose disturbance terms are intercorrelated. The latent variables are represented by ovals; small ovals marked "d1" and "d2" stand for respective disturbance terms. Values between ovals represent path coefficients between latent variables. Respective 95% confidence intervals are shown in brackets.

3.3. Discussion

Study 2 successfully replicated the results found in the unspeeded group of Study 1. Again, under no time pressure during intelligence testing, WM appeared to be only a moderate predictor of fluid intelligence, explaining about one third of its variance. Also, relational learning, operationalized as a relative increase in the efficiency of processing of number relations due to the earlier coping with the letter relations (note: not as an absolute score on that test, which of course would be another measure of Gf), independently contributed to the additional amount of Gf variance. Importantly, as the DREL test used in Study 2 relied solely on the alpha-numeric stimuli, the variance shared between relational learning and Gf cannot be attributed to the use of the same material in both learning and Gf tests.

Methodological concerns pertaining to whether the WM–Gf correlation coefficients, estimated in the moderately speeded and the unspeeded conditions of the present research, could be negatively affected by the use of only pairs of (Gf and WM) tasks can be fully overruled, as neither the increase in the number of tasks nor using more complex tasks were able to affect the observed strength of the WM–Gf link in the unspeeded condition. Moreover, although it must be noted that the WM tasks, which were used in Studies 1 and 2, involved spatial organization of (letter and number) stimuli, it is very unlikely that such their feature contributed to the perfect correlation between Gf and WMC, because (a) spatial relations need not be directly remembered, and (b) the WM–Gf correlation in the unspeeded group was, anyway, far from perfect. On the other hand, a moderate correlation in untimed testing cannot be explained in terms of unreliable measurements of Gf or WMC, because (a) reliability of all applied tasks was high, and (b) the induction of the perfect correlation in the speeded condition was successful.

4. General discussion

4.1. Summary of results

This research examined whether the WMC and Gf latent variables, as estimated by the CFA method, (a) can be perfectly correlated when Gf is assessed with the highly speeded tests, and (b) are only moderately related when the unspeeded Gf tests are used. Both these expectations have been fully confirmed, shedding some new light on the issue of the "true" relationship between WM and fluid intelligence, and suggesting that fluid reasoning can be differently related to WMC depending on time pressure during Gf testing. Furthermore, as untimed intelligence testing resulted in a large amount of variance that could not be explained by WM, I identified another source of several percents of variance in fluid intelligence: the ability to learn relational representations.

4.2. Implications for fluid intelligence research

The main implication of the presented research pertains to the fact that although time pressure barely disrupts the internal and external reliability of the fluid intelligence measures, it substantially impacts their construct validity. Simply, fluid intelligence tests administered under severe time pressure versus those applied under no pressure seem to measure not the same things (for a similar conclusion see also Partchev & De Boeck, 2012). During strictly timed intelligence testing, participants seem to be forced to represent and transform in WM the representations of complete relations reflected by test items, which is a process which relies heavily on the cognitive resources available "here and now", namely on the available capacity. In such cases, there is probably no time to implement more complex and long-lasting processes. On the contrary, in untimed intelligence testing, the process of reasoning can be more iterative (see Kubose, Holyoak, & Hummel, 2002), meaning that it can be divided into a larger number of steps that operate on parts of the eventual relational representation. In such cases, mental faculties beyond WM can be employed in order to supplement the cognitive processing fulfilled by WM. Such processes may involve learning (as advocated in the present paper) as well as other mechanisms leading to better abstraction and more effective problem-solving strategies (Carpenter et al., 1990; Davidson, 1995; Garlick & Sejnowski, 2006). Most importantly, due to these faculties, low-WMC participants may be able to compensate for their capacity limitations. In consequence, in unspeeded conditions, only a moderate correlation between Gf and WMC can be observed.

Two further conclusions follow from the above. Firstly, many recent studies which aimed to estimate the strength of relationship between intelligence and the effectiveness of various elementary cognitive functions supported by WM (e.g., recognition or recall, relational integration, executive functions, etc.), applied strictly timed intelligence testing. For instance, several such studies applied half Raven's items in 10 min (e.g., Kane et al., 2004; Unsworth, Brewer, & Spillers, 2009; Unsworth, Redick, Lakey, & Young, 2010; Unsworth & Spillers, 2010), or even required solving twelve its items in 5 min (Unsworth, Miller et al., 2009), used the time-constrained Berlin Intelligence Scale (e.g., Oberauer et al., 2008; Schmiedek, Hildebrandt, Lövdén, Wilhelm, & Lindenberger, 2009; Schmiedek, Oberauer, Wilhelm, Süß, & Wittmann, 2007; Süß et al., 2002), or included other Gf tests administered under extreme time pressure (Buehner et al., 2006; Unsworth, 2010; Unsworth, Spillers, & Brewer, 2010). My results suggest that these studies might have overestimated the strength of the relationships examined, in comparison to the strength which would be observed when standard administration of intelligence measures was applied. Gf tests administered under strict time constraints might have primarily captured the capacity of WM-based processing, so such tests – as well as elementary cognitive tasks – might have tapped processes so mutually similar that substantial correlations might have arisen just by definition.

Secondly, an analogous problem concerns some famed research on fluid intelligence training. Jaeggi and collaborators (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi et al., 2010) trained participants on versions of an n-back task, for a few weeks, until participants' scores on that task were greatly enhanced. Crucially, participants also demonstrated increased scores on Gf tests, as compared to pre-testing. On the contrary, no transfer from the n-back related to the complex span task (Jaeggi et al., 2010). So, the authors concluded that their training method selectively increased Gf. The most serious problem related to that research (as noticed by Moody, 2009) is the fact that the Gf tests were applied under extreme time constraints (10–11 min for 18 Raven's items, and 10–16 min for 29 BOMAT items, depending on the study). Also Schmiedek et al. (2010) observed a generalized transfer from WM tasks to a battery of ten fluid reasoning tests, but as in the Jaeggi studies, the latter tests were highly speeded. In the light of my results, Gf tests administered in such a way should be treated as not much more than WM tests. So, it is no surprise that training on one WM test helped participants to solve another such test. Any conclusions assuming that a Gf itself was increased by these training methods do not seem to be warranted.

In line with the above interpretation, Colom et al. (2010) tested participants who were trained on either WM or speed/attention tasks, with Gf tests administered in the standard way. The authors found no selective effect of WM training on intelligence. Gf test scores were higher due to both WM and speed/attention training for unknown reasons, but probably at least in part due to learning how to deal with the battery of Gf measures during the pre-test. Moreover, in Chein and Morrison (2010), virtually no effect of training on the performance on unspeeded Raven was found, though strong training effects pertained to reading comprehension and cognitive control. It seems that firm conclusions about the possibility of increasing Gf with WM training can only be made if the selective effect of WM training (i.e., one compared to null effect in a control group trained on some other, low WM-demanding tasks) is observed in untimed Gf tests.

The important effect of administration time on the WM–Gf link seems at first glance to be inconsistent with studies which showed that mental (processing) speed, closely related to neuronal efficiency (Jensen, 1998), is a substantially worse predictor of intelligence than are WM tasks imposing loads on capacity and/or control mechanisms (e.g., Colom et al., 2008; Conway et al., 2002; Kaufman et al., 2009; Martínez et al., 2011). A straightforward interpretation of this effect might assume that WM scores depend on speed (see Salthouse, 1996), and that in speeded intelligence testing these scores so strongly predicted Gf because speed was also responsible for the number of attempted Gf test items, while in unspeeded testing participants were able to attempt all items, and their scores depended on other factors than speed. Of course, such an interpretation cannot be ruled out by the present study (as processing speed was not measured), but it seems unlikely. Firstly, recent neurophysiological research strongly indicates that the pattern of predictive power of neuronal efficiency (measured with various imaging techniques) with regard to higher-order cognition is very complex, and in contributing to cognitive abilities processing speed interacts with various factors, like available capacity, task complexity, and adopted strategies (for reviews see Toffanin, Johnson, de Jong, & Martens, 2007; Rypma & Prabhakaran, 2009). Secondly, it was also shown that those abilities are most strongly predicted not by the mean latency of processing, but by the latency of longest trials ("the worst performance rule"; Coyle, 2003), which is the fact interpreted in terms of control (but not speed) of processing (e.g., in terms of time required to recover from errors or interference). Finally, no influential model of human reasoning assumes that the accuracy of that process primarily depends on parameters reflecting speed (e.g., Carpenter et al., 1990; Goodwin & Johnson-Laird, 2011; Halford, Wilson, & Phillips, 2010; Hummel & Holyoak, 2003; Kemp & Tenenbaum, 2009; Rasmussen & Eliasmith, 2011). So, the view that even in highly speeded intelligence testing participants who solved more items did so because they simply were faster does not seem to be sufficiently supported by existing literature.

The results regarding the unspeeded condition imply that even though, for the last twenty years, the search for the cognitive basis of fluid intelligence has been (maybe, too much) focused on underlying WM processes, the story is not that simple (see also Burgess, Braver, & Gray, 2006; Kane et al., 2005; Kaufman et al., 2009). WM, which indeed fully determined Gf scores on my speeded tests, cannot be the only explanation of individual differences in unspeeded fluid reasoning. My results indicate that fluid intelligence seems to be a more complex phenomenon than – although very important for fluid processing – the operation of WM, and other factors may account for a substantial amount of variance in Gf. Fruitful directions in the identification of such factors should cover psychometric studies of the relations between Gf tests applied in less strict time conditions and measures of candidate abilities, as well as formal analyses of the nature of fluid reasoning process and its crucial components, like problem-solving strategies (Carpenter et al., 1990), adaptive coding (Duncan, 2001; Garlick, 2002), selective attention and abstraction (Davidson, 1995; Garlick & Sejnowski, 2006; Primi, 2001), and mutual dynamical interactions among diverse abilities (van der Maas et al., 2006).

4.3. The possible role of relational learning in fluid intelligence

Taking into account the fact that reasoning in Gf test consists of processing abstract relations, my results suggest that a particular kind of learning which is above and beyond associative learning, namely relational learning, may be an especially plausible explanation of some part of Gf variance that is unexplained by WMC. According to influential approaches to relational learning, including the LISA theory of analogy making (Doumas, Hummel, & Sandhofer, 2008; Hummel & Holyoak, 2003), and the structured statistical (Bayesian) approaches to inductive reasoning (Kemp & Jern, 2009; Kemp & Tenenbaum, 2009), such learning consists of schema induction. A schema is created by transforming specific examples of relations (e.g., in Raven, "in each row, triangles from the top and middle row are juxtaposed to produce respective triangles in the bottom row") into a general structural representation (e.g., "any figures in a bottom row can be juxtapositions of any respective figures from top and middle rows"), defining relational roles (e.g., either "being a result of juxtaposition" or "being a part of juxtaposition") that can be fulfilled by any objects. For example, both two squares producing a rectangle and two rectangles producing a square will constitute the same relation of juxtaposition, though in one case the same figure is being juxtaposed, while in the other case it constitutes a result of juxtaposition.

Most probably, people who can better learn how to induce relational representations governing matrices or analogies when working on a Gf test, and can better discover how to use them in order to enhance detection, transformation, and application of rules governing future test items, will score higher on that and similar tests and thus will be judged more psychometrically intelligent than people who cannot construct and use relational knowledge of such a kind (Halford et al., 2010). As people most probably start to learn from the first test items they attempt (with a possible exception in the case of highly speeded intelligence testing), relational learning seems to be an inherent component of the measurement of fluid intelligence.

4.4. Final remarks

This study has drawn psychology nearer to the resolution of the hotly debated problem of whether fluid intelligence and working memory are (almost) isomorphic constructs or, alternatively, whether the latter is only a moderate predictor of the former, while other substantial cognitive predictors of Gf can be identified. The most original result of the present study shows that either case is true, depending on the methodological decision regarding how to estimate Gf. Imposing extreme time pressure during intelligence testing makes participants rely mostly on the capacity of their WM, while allowing them more time for coping with Gf tests greatly reduces the contribution of WM to Gf, and makes room for other factors to explain the mechanisms underlying Gf. Thus, my study may yield important implications for future research on the nature of fluid intelligence.

In particular, the study suggests that each researcher has to face a dilemma: whether intelligence should be measured with speeded tests, or with power tests. The former testing method will measure the ability to cope with complexity in a dynamic environment, thus having a high real-world validity, as the technological and informational pressure of the world increases rapidly, but it may underestimate people who regardless of their limited capacity would work out good solutions in less dynamic environments. The latter method will give a more comprehensive account of reasoning ability, including the contribution of intellectual faculties that lay beyond WM, and seem to be complementary to it, but it could also include a lot of noise (e.g., learned task-dependent strategies) negatively influencing the evaluation of future effectiveness of an individual in demanding, timed, and completely novel tasks. The former method will surely be more and more predictive in the cases in which new informational technologies are being coped with, but the latter method seems to provide a richer understanding of what intelligent behavior in various situations really is.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*, 567–589.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*, 30–60.

Baddeley, A. D. (2007). *Working memory, thought and action.* Oxford: Oxford University Press.

Binet, A. (1903). *Etude expérimentale de l'intelligence.* Paris: Schleicher.

Broadway, J. M., & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, *42*, 563–570.

Buehner, M., Krumm, S., Ziegler, M., & Pluecken, T. (2006). Cognitive abilities and their interplay: Reasoning, crystallized intelligence, working memory components, and sustained attention. *Journal of Individual Differences*, *27*, 57–72.

Burgess, G. C., Braver, T. S., & Gray, J. R. (2006). Exactly how are fluid intelligence, working memory, and executive control related? Cognitive neuroscience approaches to investigating the mechanisms of fluid cognition. *The Behavioral and Brain Sciences*, *29*, 128–129.

Burgess, G. C., Gray, J. R., Conway, A. R. A., & Braver, T. S. (2011). Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology: General*, *140*, 674–692.

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

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

Chein, J., & Morrison, A. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin and Review*, *17*, 193–199.

Chuderski, A., Andrelczyk, K., & Smoleń, T. (2013). An oscillatory model of individual differences in working memory capacity and relational integration. *Cognitive Systems Research*, *24*, 87–95.

Chuderski, A., & Nęcka, E. (2010). Intelligence and cognitive control. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of individual differences in cognition: Attention, memory, and executive control* (pp. 263–281). New York: Springer Verlag.

Chuderski, A., & Nęcka, E. (2012). The contribution of working memory to fluid intelligence: Capacity, control, or both? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 1689–1710.

Chuderski, A., & Stettner, Z. (2013). Testing the dual-component account of working memory with a serial recognition task. *Review of Psychology Frontier*, *2*, 10–16.

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence while executive control does not. *Intelligence*, *40*, 278–295.

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.

Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent variable approach. *Intelligence*, *33*, 623–642.

Colom, R., Quiroga, M. A., Shih, P. C., Martínez, K., Burgaleta, M., Martínez-Molina, A., et al. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence*, *38*, 497–505.

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by *g*. *Intelligence*, *32*, 277–296.

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

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review*, *12*, 769–786.

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

Cowan, N., Fristoe, N. M., Elliott, E. M., Brunner, R. P., & Saults, J. S. (2006). Scope of attention, control of attention, and intelligence in children and adults. *Memory & Cognition*, *34*, 1754–1768.

Coyle, T. R. (2003). A review of the worst performance rule: Evidence, theory, and alternative hypotheses. *Intelligence*, *31*, 567–587.

Dang, C. -P., Braeken, J., Ferrer, E., & Liu, C. (2012). Unitary or non-unitary of working memory? Evidence from its relation to general fluid intelligence and crystallized intelligence. *Intelligence*. http://dx.doi.org/10.1016/j.intell.2012.05.002.

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

Deary, I., & Der, G. (2005). Reaction time explains IQ's association with death. *Psychological Science*, *16*, 64–69.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review*, *115*, 1–43.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews Neuroscience*, *2*, 820–829.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation*, *Vol. 44*. (pp. 145–199) New York, NJ: Elsevier.

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.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*, 172–179.

Fukuda, K., Vogel, E. K., Mayr, U., & Awh, E. (2010). Quantity not quality: The relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin and Review*, *17*, 673–679.

Galton, F. (1883). *Inquiries into human faculty*. London: Macmillan.

Garlick, D. (2002). Understanding the nature of the general factor of intelligence: The role of individual differences in neural plasticity as an explanatory mechanism. *Psychological Review*, *109*, 116–136.

Garlick, D., & Sejnowski, T. (2006). There is more to fluid intelligence than working memory capacity and executive function. *The Behavioral and Brain Sciences*, *29*, 134–135.

Goodwin, G. P., & Johnson-Laird, P. N. (2011). Mental models of Boolean concepts. *Cognitive Psychology*, *63*, 34–59.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79–132.

Gustaffson, J. -E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Sciences*, *11*, 236–241.

Halford, G. S., Wilson, W. H., & Phillips, S. (2010). Relational knowledge: The foundation of higher cognition. *Trends in Cognitive Sciences*, *14*, 497–505.

Hamel, R., & Schmittmann, V. D. (2006). The 20-minute version as a predictor of the Raven Advanced Progressive Matrices test. *Educational and Psychological Measurement*, *66*, 1039–1046.

Heron, A., & Chown, S. M. (1967). *Age and function.* Boston: Little-Brown.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*, F9–F15.

Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, *110*, 220–264.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *PNAS*, *105*, 6829–6833.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y. -F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — Implications for training and transfer. *Intelligence*, *38*, 625–635.

Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.

Jensen, A. R. (2006). *Clocking the mind: Mental chronometry and individual differences.* Oxford: Elsevier.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: The contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*, 47–70.

Kane, M. J., 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.

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 visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189–217.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Brown, J., & Mackintosh, N. (2009). Associative learning predicts intelligence above and beyond working memory and processing speed. *Intelligence*, *37*, 374–382.

Kemp, C., & Jern, A. (2009). Abstraction and relational learning. *Advances in Neural Information Processing Systems*, *22*, 1–9.

Kemp, C., & Tenenbaum, J. B. (2009). Structured statistical models of inductive reasoning. *Psychological Review*, *116*, 20–58.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*, 317–324.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, *24*, 781–791.

Kubose, T. T., Holyoak, K. J., & Hummel, J. E. (2002). The role of textual coherence in incremental analogical mapping. *Journal of Memory and Language*, *47*, 407–435.

Kyllonen, P. C. (2002). g: Knowledge, speed, strategies, or working-memory capacity? A systems perspective. In R. J. Sternberg, & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lerch, F. J., Gonzalez, C., & Lebiere, C. (1999). Learning under high cognitive workload. *Proceedings of the 21st Conference of the Cognitive Science Society* (pp. 302–307). Mahwah, NJ: Erlbaum.

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1545–1551.

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

Mackintosh, N. J., & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence*, *31*, 519–531.

Marjorie, L. R., & Revelle, W. (1985). Effects of anxiety on analogical reasoning: A test of three theoretical models. *Journal of Personality and Social Psychology*, *49*, 1302–1315.

Martens, S., & Johnson, A. (2009). Working memory capacity, intelligence, and the magnitude of the attentional blink revisited. *Experimental Brain Research*, *192*, 43–52.

Martínez, K., Burgaleta, M., Román, F. J., Escorial, S., Shih, P. C., Quiroga, Mª. A., et al. (2011). Can fluid intelligence be reduced to 'simple' short-term storage? *Intelligence*, *39*, 473–480.

Martínez, K., & Colom, R. (2009). Working memory capacity and processing efficiency predict fluid but not crystallized and spatial intelligence: Evidence supporting the neural noise hypothesis. *Personality and Individual Differences*, *46*, 281–286.

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, *37*, 327–328.

Mrazek, M. D., Smallwood, J., Franklin, M. S., Chin, J. M., Baird, B., & Schooler, J. W. (2012). The role of mind-wandering in measurements of general aptitude. *Journal of Experimental Psychology: General*, *144*, 788–798.

Oberauer, K. (2005). Binding and inhibition in working memory: Individual and age differences in short-term recognition. *Journal of Experimental Psychology: General*, *134*, 368–387.

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). Oxford: Oxford University Press.

Oberauer, K., Süß, H. -M., Wilhelm, O., & Wittman, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, *36*, 641–652.

Orzechowski, J., & Chuderski, A. (2007). Test analogii obrazkowych [A figural analogies test]. Unpublished manuscript, Jagiellonian University, Krakow, Poland.

Partchev, I., & De Boeck, P. (2012). Can fast and slow intelligence be differentiated? *Intelligence*, *40*, 23–32.

Paulewicz, B., Chuderski, A., & Nęcka, E. (2007). Insight problem solving, fluid intelligence, and executive control: A SEM approach. In S. Vosniadou, D. Kayser, & A. Protopapas (Eds.), *Proceedings of the 2nd European Cognitive Science Conference* (pp. 586–591). Hove: Erlbaum.

Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, *30*, 41–70.

Rasmussen, D., & Eliasmith, C. (2011). A neural model of rule generation in inductive reasoning. *Topics in Cognitive Science*, *3*, 140–153.

Raven, J. C. (1962). *Advanced Progressive Matrices, Set II.* London: H. K. Lewis.

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

Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin and Review*, *18*, 324–330.

Rypma, B., & Prabhakaran, V. (2009). When less is more and when more is more: The mediating roles of capacity and speed in brain–behavior efficiency. *Intelligence*, *37*, 207–222.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing–storage relationship in reading span performance. *Journal of Memory and Language*, *50*, 425–443.

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

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review*, *103*, 403–428.

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

Schmiedek, F., Lövdén, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, *2*, 1–10.

Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, *136*, 414–429.

Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M. R., & Gouvier, W. D. (2009). A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence. *Intelligence*, *37*, 283–293.

Shelton, J. T., Elliott, E. M., Matthews, R. A., Hill, B. D., & Gouvier, W. D. (2010). The relationships of working memory, secondary memory, and general fluid intelligence: Working memory is special. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 813–820.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences*, *44*, 535–551.

Shipstead, Z., Redick, T. S., Hicks, K. L., & Engle, R. W. (2012). The scope and control of attention as separate aspects of working memory. *Memory*, *20*, 608–628.

Schweizer, K., & Moosbrugger, H. (2004). Attention and Working Memory as Predictors of Intelligence. *Intelligence*, *32*, 329–347.

Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*, *Vol. 2*. (pp. 47–103)Hillsdale, NJ: Lawrence Erlbaum Associates.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *The American Journal of Psychology*, *15*, 201–293.

Sternberg, R. J. (1996). *Successful intelligence.* New York, NY: Simon & Schuster.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability — And a little bit more. *Intelligence*, *30*, 261–288.

Tamez, E., Myerson, J., & Hale, S. (2008). Learning, working memory, and intelligence revisited. *Behavioural Processes*, *78*, 240–245.

Toffanin, P., Johnson, A., de Jong, R., & Martens, S. (2007). Rethinking neural efficiency: Effects of controlling for strategy use. *Behavioral Neuroscience*, *121*, 854–870.

Unsworth, N. (2010). Interference control, working memory capacity, and cognitive abilities: A latent variable analysis. *Intelligence*, *38*, 255–267.
Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory–fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin and Review*, *16*, 931–937.
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.
Unsworth, N., Miller, J., Lakey, C. E., Young, D. L., Meeks, J. T., Campbell, W. K., et al. (2009). Exploring the relations among executive functions, fluid intelligence, and personality. *Journal of Individual Differences*, *30*, 194–200.
Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: An individual differences investigation. *Intelligence*, *38*, 111–122.
Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention, memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*, 392–406.
Unsworth, N., Spillers, G. J., & Brewer, G. A. (2010). The contributions of primary and secondary memory to working memory capacity: An individual differences analysis of immediate free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 240–247.
van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842–861.
Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *423*, 500–503.
Wilhelm, O., & Oberauer, K. (2006). Why are reasoning ability and working memory capacity related to mental speed? An investigation of stimulus–response compatibility in choice reaction time tasks. *European Journal of Cognitive Psychology*, *18*, 18–50.
Wilhelm, O., & Schultze, R. (2002). The relation between speeded and unspeeded reasoning with mental speed. *Intelligence*, *30*, 537–554.
Williams, B. A., & Pearlberg, S. L. (2006). Learning of three-term contingencies correlates with Raven scores, but not with measures of cognitive processing. *Intelligence*, *34*, 177–191.