Personality and Individual Differences 85 (2015) 98–104

Contents lists available at ScienceDirect

Personality and Individual Differences

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

The broad factor of working memory is virtually isomorphic to fluid intelligence tested under time pressure

Adam Chuderski *

Jagiellonian University in Krakow, Poland

ARTICLE INFO

Article history:
Received 19 January 2015
Received in revised form 3 April 2015
Accepted 27 April 2015

Keywords:
Fluid intelligence
Working memory
Confirmatory factor analysis

ABSTRACT

How much overall variance in fluid intelligence (Gf) can be predicted by four working memory (WM) functions: storage capacity, attention control, relational integration, and updating was investigated under time pressured Gf testing. Confirmatory factor analysis indicated that the broad WM factor, which was subsumed by these four WM functions, shared 83.4% of variance with Gf tested under pressure, whereas a reanalysis of previous data with the same model showed that only 58.2% variance was shared with virtually untimed Gf tests. Moreover, in timed Gf tests, only the easiest, early items contributed to the WM-Gf correlation, whereas in untimed tests also the hardest, late items were linked with Gf. These results suggest that the measurement of "fast" intelligence primarily taps the functions of WM, whereas "slow(er)" intelligence depends also on some other cognitive processes beyond WM.

1. Introduction

Fluid intelligence (fluid reasoning, reasoning ability; Gf), which consists of using reasoning to solve novel abstract problems that cannot be solved on the basis of existing knowledge, is an important component of human general intelligence. Because of strong predictive power of Gf for such psychological variables as socio-economical status (academic, professional, financial, etc.), one of the most important themes in psychology consists of identification of mechanism underlying Gf.

Probably the strongest known predictor of fluid intelligence is the capacity (WMC) of working memory (WM) – the neurocognitive mechanism responsible for the active maintenance and transformation of the limited amount of information in service of the current task. There are several theories on what in WMC makes it so strongly correlate with Gf. One theory (Kane & Engle, 2002) assumes that individual performance in both WM tasks and Gf tests depends on *attention control* exerted over cognitive processes, which includes goal-driven directing attention and filtering out distraction. Alternatively, it was shown that performance on sheer *storage* tasks, which require little attention control, was also a very good predictor of Gf (Cowan, 2001), probably because an individual needs to keep the subproducts of reasoning in the most accessible part of WM. WM may also play an important role in Gf because it affects what relations can be constructed among WM items. Notably, Oberauer, Süß, Wilhelm, and Wittman (2008) proposed that crucial for Gf is *relational integration*, which consists of the construction of flexible, temporary bindings between a number of chunks held in WM, in order to develop a more complex, relational structure. The tasks that require participants to detect simple relations have been shown to be excellent predictors of WMC and Gf (Chuderski, 2014; Oberauer et al., 2008). Finally, also proper *updating* of WM contents, that is, their substitution and transformation in line with the demands of the current task, has been indicated as a key WM function (Jonides & Smith, 1997).

Although each listed WM function probably contributes to Gf to some extent (see Conway, Getz, Macnamara, & Engel de Abreu, 2011), the question of precisely *how much variance in Gf can be explained by the broad WM construct* (e.g., including all four above mentioned WM functions) have not received a univocal answer. Metaanalyses demonstrated that WMC usually explains between half (Kane, Hambrick, & Conway, 2005) and three quarters (Oberauer, Schultze, Wilhelm, & Süß, 2005) of Gf variance. At the same time, some studies reported Gf-WM correlations below $r = .30$ (e.g., Kaufman, DeYoung, Gray, Brown, & Mackintosh, 2009; Unsworth, Spillers, & Brewer, 2010). Finally, some studies reported WMC to be isomorphic to Gf (e.g., Colom, Rebollo, Palacios, Juan-Espinosa, & Kyllonen, 2004; Martínez et al., 2011; Oberauer et al., 2008). Thus, the question whether Gf reduces or not to the effectiveness of WM functioning remains open.

Attempting to answer this question, Chuderski (2013) has suggested that how much Gf variance is determined by WMC depends

* Address: Institute of Philosophy, Jagiellonian University, Grodzka 52, 31-044 Krakow, Poland.

E-mail address: adam.chuderski@gmail.com

http://dx.doi.org/10.1016/j.paid.2015.04.046
0191-8869/

on time pressure during intelligence testing. When Gf tests were administered in strict time limit (e.g., 20 min for Raven APM), WMC and Gf were statistically indistinguishable (i.e., isomorphic), whereas when ample time was given (e.g., one hour for APM), WMC predicted only one third of Gf variance. These results suggest that time pressure somehow blocks longer lasting and more complex cognitive processes contributing to reasoning, and under such pressure people have to rely on simultaneous, online manipulation of stimuli, primarily using their WM resource. However, this study had one major drawback related to the fact that it relied on the reanalysis of already existing datasets. In result, the WM-Gf isomorphism under pressure has been shown with regard to only two WM tasks: a letter variant of storage task, and a number variant of relation integration task (no non-verbal task was used). In result, the WM-Gf isomorphism could be objected, as the strong correlations found might have resulted from unknown peculiarities of the WM tasks used that made these tasks in some way highly similar to the timed Gf tests. Although, Experiment 2 replicated a moderate WM-Gf link with more WM tasks, in this study only virtually untimed Gf was examined, but not timed. These results are important because, although most of Gf tests' instructions recommend a reasonable time allowed (e.g., 40 min in APM) that yields a good compromise between power and speed, in the substantial number of studies on the WM-Gf relation (almost 50%; see Chuderski, 2013, Table 1) this time was strongly reduced. At the same time, several such studies (see ibidem) that were highly influential used unlimited administration time. Thus, it seems important to know the psychometric and theoretical consequences of using non-standard administration times in intelligence testing for the research on WM and Gf.

The goal of the present study was to replicate and extend the Chuderski results by evaluating the strength of link between Gf, tested under time pressure, and the broad WM construct, measured by as much as eight WM tasks, reflecting four above presented WM functions, and both verbal and non-verbal types of stimuli. In line with the previous results, it was expected that the link between timed Gf and WMC is close to unity, reflecting the possibility that solving timed Gf problems requires primarily simultaneous maintenance and online manipulation of information – the same processes that are also required for the WM performance. Moreover, as the same WM and Gf tests were used here and in Chuderski (2014), where virtually untimed Gf tests were applied, we were able to reanalyze the latter data using the model fitted in the present study (reflecting the broad WM latent variable loading the four WM functions), and in result we could test more reliably whether the link between the broad WM latent variable and timed Gf is indeed significantly stronger than for untimed Gf.

Table 1
Descriptive statistics for working memory and fluid intelligence measures ($N = 264$).

Task	M	SD	Range	Skew	Kurt.
Color arrays	3.61	1.27	−1.97 to 6.30	−0.94	1.90
Letter arrays	3.28	1.44	−2.05 to 6.12	−0.47	−0.03
Arrow antisaccade	0.78	0.21	0.02–1.00	−1.52	1.97
Letter antisaccade	0.83	0.20	0.02–1.00	−1.95	3.66
Same relation	0.69	0.18	−0.15 to 0.95	−1.36	2.46
Different relation	0.30	0.23	−0.23 to 0.78	−0.39	−0.77
Figural 2-back	0.69	0.18	0.00–0.95	−1.19	1.99
Number 2-back	0.81	0.17	0.00–1.00	−1.61	3.93
Raven APM	19.07	5.38	1–31	−0.62	0.41
Paper analogies	18.54	5.65	3–34	0.05	−0.16
Computerized analogies	0.20	0.14	0.00–0.83	0.78	1.03

Note: The arrays task score = the mean number of items held in WM. The antisaccade, *n*-back, relation monitoring (same/different relation), and computerized analogies scores = proportion correct. The Raven and paper analogies scores = the number of correctly solved items.

2. Method

Three highly timed intelligence tests were administered. All participants attempted also the battery of eight WM tasks, two (one verbal and one non-verbal, except for the relational integration tasks, in which verbal and non-verbal aspects were combined) per each above mentioned function of WM (storage, updating, control, and integration), thus measuring the broad WM construct to a large extent. The four types of tasks used are schematically presented in Fig. 1, and described below.

2.1. Participants

A total of 264 volunteer participants (166 women) were recruited via publicly accessible social networking websites. Each participant gave informed consent, was told that he or she can leave the experiment at will and at any moment, and was paid the equivalent of 15 euro in Polish zloty. The mean age of participants was 23.2 years ($SD = 4.4$, range 18–46). Another four participants were excluded from analysis due to missing some tasks.

2.2. Measures of fluid intelligence

Two paper-and-pencil tests of reasoning were applied, the widely used Raven Advanced Progressive Matrices (Raven, Court, & Raven, 1983), and a figural analogy test (for description of the test see Chuderski, 2014). Half of the standard administration time was allowed for each test (20 and 16 min, respectively). Also, a computerized figural analogy test including 16 items was applied (for description of the test see Chuderski, 2015a), with 2 min allowed per item (in comparison to 4 min used in Chuderski, 2014). The total number of correct answers in each test was taken as a respective score.

2.3. Storage tasks

Two variants of an array-comparison task were used. Each variant consisted of 90 trials. On each trial a virtual 4×4 array was filled with five to nine stimuli, picked from a set of ten Greek symbols (e.g., α, β, χ, and so on), or colored squares (i.e., the letter and color variants of the task, respectively), then followed by a black square mask of the same size as the array, presented for 1.2 s, and then another array was shown. In a random 50% of trials, the second array was identical to the first; in the remaining trials the second array differed from the first by exactly one item in one position, which was always a new item (not a duplicate of an item from another position). The task was to press one of two response keys to indicate whether the highlighted item was the same or different in the two arrays. The tasks were self-paced.

2.4. Attention control tasks

Two variants of the antisaccade task were used, measuring the attention control ability, each consisting of 40 self-paced trials. Each test trial consisted of four events. First, a cue was presented for 1.5 s to prompt subjects to look at the side opposite to a rapidly flashing black square. Next, a fixation point was presented in the center of the screen for 1–2 s. Then, the flashing square was shown in the middle of the left or right side of the screen, about 16 cm from the fixation point, for 0.15 s. Finally, a small dark gray arrow or a string was presented in the middle of the opposite side of the screen to the square for only 0.2 s before being replaced by a mask. The task was to look away from the flashing square in order to observe the direction of the arrow or the identity of the string and press the associated key.

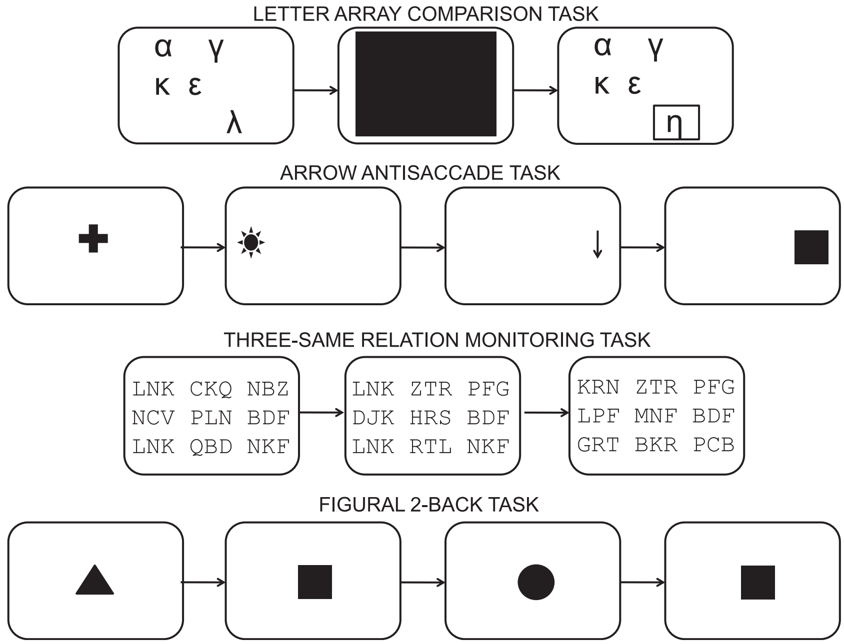

Fig. 1. The schematic illustration of stimuli in the array comparison (measuring storage in WM), antisaccade (tapping attention control), relation monitoring (capturing relational integration), and 2-back tasks (reflecting updating).

2.5. Relation integration tasks

No-memory version of the alphanumeric monitoring task, originally devised by Oberauer et al. (2008), were used. The stimulus for each trial on the task consisted of a 3×3 array of three-symbol strings. In the letter version of the task, the strings contained three letters from a set of ten consonants; in the number version they were three-digit numbers. Depending on the task variant, participants were asked to detect whether any of the rows or columns consisted of three strings ending with the same (the three-same variant) or different digit or letter (the three-different variant; see Chuderski, 2014). On half the trials, the array included one of the specified configurations; on these trials participants were required to press the space key to indicate that they had detected this configuration. On the rest of the trials, the array did not contain any of the specified configurations. Trials lasted 5.5 s and were followed by a 0.1 s blink separating subsequent arrays. In each version of each task variant there were forty test trials.

2.6. Updating tasks

Stimuli in two 2-back tasks, adapted from Chuderski and Nęcka (2012), were two-digit numbers or figures, each approximately 2.5×2.5 cm in size, presented for 1.2 s plus a 0.6 s mask. A total of 80 stimuli were presented serially to participants in each session. Two sessions were used in each task, preceded by the 40-stimuli training. Each session included sixteen 2-back target repetitions of stimuli. No other stimuli could be repeated in a time window of ten stimuli. Participants were instructed to respond to repetitions and to suppress responses to all other items. The dependent variable was mean accuracy of repetition detection.

2.7. Procedure

Participants were tested in a cognitive psychology lab, in groups from six to twelve people. The WM testing session that lasted about two hours preceded the Gf testing session that lasted around one hour. The order of tasks within a session was fixed. Both sessions were separated by a sufficient break, in which sweets and drinks were provided to participants.

3. Results

Table 1 presents descriptive statistics for measures used in the study. All measures approximated the normal distribution. As timed Gf testing results in generally decreased test scores, for the sake of comparability with the Chuderski study, in correlational analyses the Gf test scores were converted to standardized scores. Table 2 shows the correlation matrix and reliabilities. All measures yielded satisfactory reliability except for computerized analogies, which were strongly affected be the decrease in administration time. Anyway, this measure was retained for comparability.

The contribution each investigated WM function made to Gf was tested using confirmatory factor analysis (CFA). First, three models were compared in order to establish the best fitting structure of WM. One model assumed that all eight WM measures load on one latent variable. The second model included four inter-correlated latent variables, each representing one WM function that was loaded by two respective measures. The last model included four uncorrelated latent variables reflecting WM functions, which this time loaded one higher-level latent variable (the broad WM construct). The fits of the models were $\chi^2(20) = 189.83$, CFI = .738, RMSEA = .175, SRMR = .092, $\chi^2(14) = 12.78$, CFI = 1.0, RMSEA = .000, SRMR = .021, and $\chi^2(16) = 12.97$, CFI = 1.0, RMSEA = .000, SRMR = .021, respectively. The first model was unacceptable, whereas the two latter models showed an excellent fit. However, as the higher-level model yielded a slightly lower χ^2/df ratio than the other model (0.81 vs. 0.91, respectively), as well as it allowed for representing the broad WM construct (which could be used for calculating one overall correlation between general WMC and Gf), this very model was chosen. Next, the final CFA model was calculated by supplementing the higher-level model with the latent variable reflecting Gf that was loaded by three intelligence test scores, and correlated with the broad WM factor. The fit of this model, shown in Fig. 2, was very good, $\chi^2(39) = 52.89$, CFI = .986, RMSEA = .058,

Table 2
Correlation matrix for working memory and fluid intelligence measures ($N = 264$).

Task	1	2	3	4	5	6	7	8	9	10	11
1. Color arrays	**.79**										
2. Letter arrays	.59	**.83**									
3. Arrow antisaccade	.39	.34	**.93**								
4. Letter antisaccade	.37	.33	.69	**.93**							
5. Same relation	.34	.42	.25	.31	**.88**						
6. Different relation	.28	.34	.23	.25	.51	**.86**					
7. Figural 2-back	.36	.37	.26	.24	.21	.26	**.77**				
8. Number 2-back	.37	.41	.30	.29	.30	.25	.55	**.80**			
9. Raven APM	.40	.39	.33	.40	.51	.43	.41	.44	**.84**		
10. Paper analogies	.38	.42	.36	.43	.40	.34	.32	.39	.62	**.82**	
11. Computerized analogies	.22	.25	.20	.23	.20	.15	.27	.26	.27	.20	**.37**

Note: For all correlations $p < .05$. Reliabilities (Cronbach alpha) are presented in bold on the diagonal.

Fig. 2

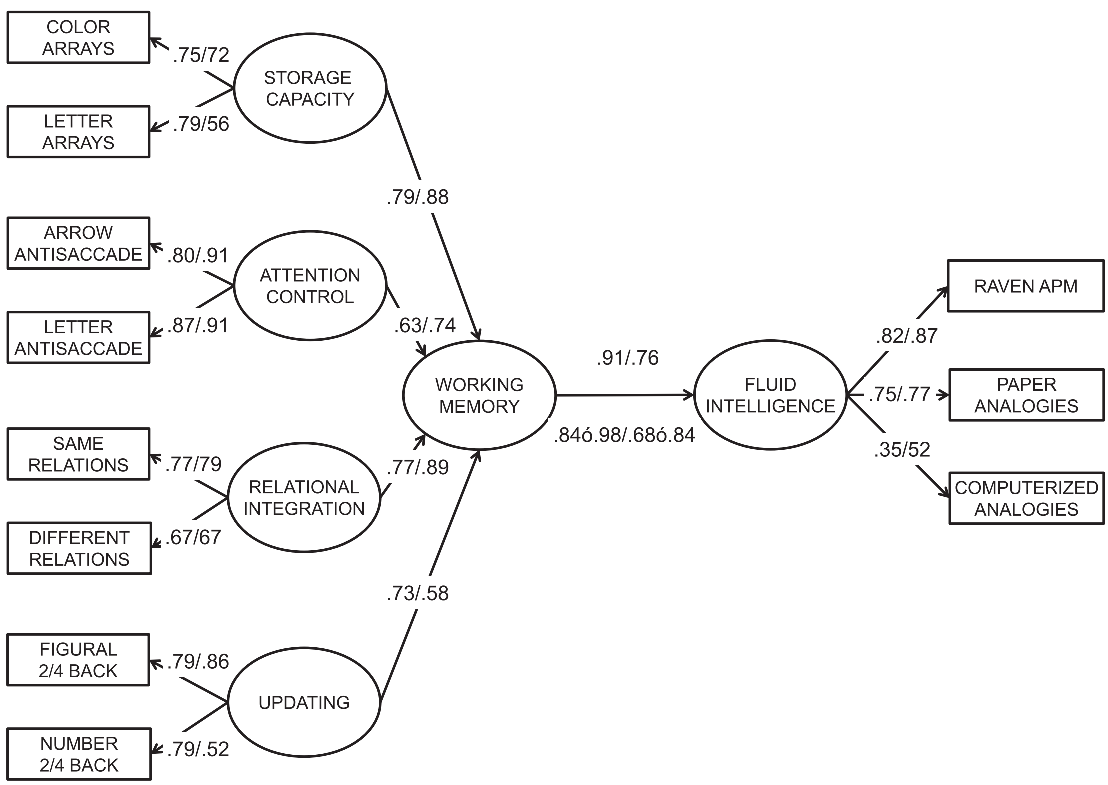

Fig. 2. The confirmatory factor analysis model, including four latent variables (ovals) representing storage capacity, attention control, relational integration, and updating, loaded by respective working memory (WM) tasks. These variables subsume the higher-level latent variable that represents the broad WM construct, and correlates with the fluid intelligence (Gf) latent variable, loaded by scores on the Raven and two analogy tests. Boxes represent manifest variables. Values that accompany directed arrows between ovals and boxes or ovals represent relevant standardized factor loadings. The line between WM and Gf represents correlation. The values on the left represents coefficients for the current dataset (timed Gf testing), whereas the values on the right reflect coefficients for the data from Chuderski (2014), who applied untimed testing.

SRMR = .039. Most importantly, the WMC-Gf correlation was exceptionally strong ($r = .913$). Although fixing this correlation at unity significantly decreased the fit of the model, $\chi^2(1) = 4.94$, $p = .026$, such a strong correlation indicated that the broad WM construct and the Gf construct are almost isomorphic, sharing 83.4% of variance.

Because in Chuderski (2014), in which the same but virtually untimed Gf tests were administered, the battery of 17 WM tasks (for descriptive statistics and correlation matrix see ibidem) included the eight types of tasks also used in the current study (with a negligible difference that here 2-back tasks were used instead of 4-back tasks), the CFA model whose structure was identical to the model presented in Fig. 2 could be calculated ($N = 243$), $\chi^2(39) = 31.94$, CFI = 1.0, RMSEA = .000, SRMR = .030. The coefficients for this model are also presented in Fig. 2. Crucially, in this model the WMC-Gf correlation equaled $r = .763$, and both latent variables shared only 58.2% of variance. As the confidence intervals of WMC-Gf correlations did not overlap for models of timed [.844–.982] versus untimed Gf [.683–.843], it can be reliably concluded that the former yielded significantly stronger correlation with the broad construct of WM. At the same time, the loadings of measures on respective latent variables, as well as the loadings of these variables on the broad WM construct, did not display any substantial difference between the models.

The broad WM factor and the Gf factors from the CFA models are also depicted in Fig. 3 as scatter plots, separately for the groups with timed and virtually untimed Gf testing, in order to show what exactly stands for a decreased WMC-Gf correlation under virtually

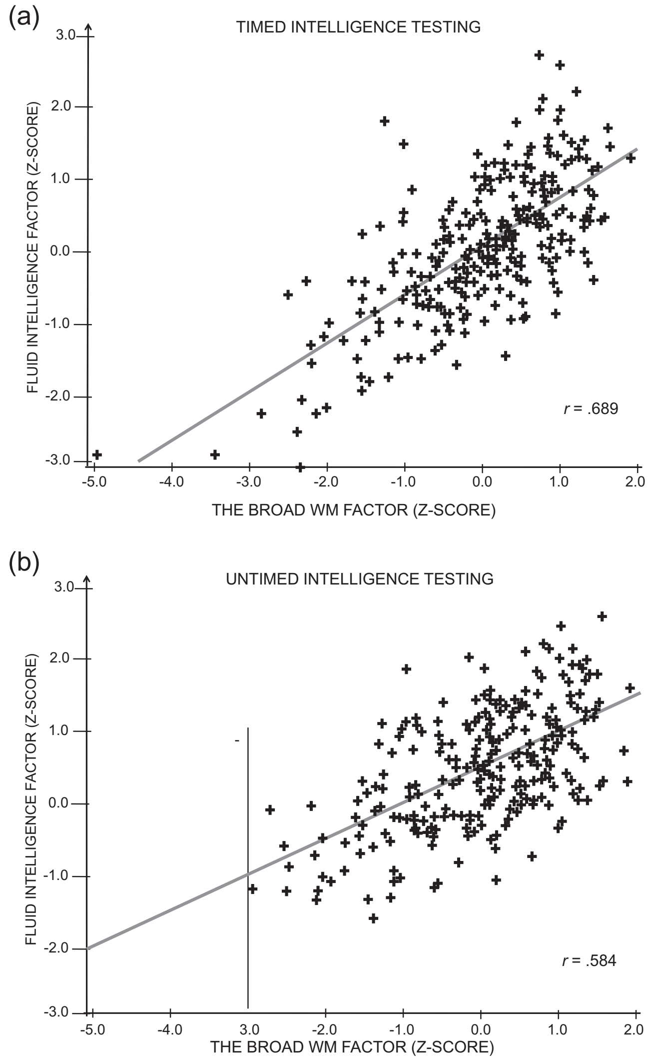

Fig. 3. Scatter plots of the relationship between the broad WM factor (*X* axis) and the Gf factor (*Y* axis), separately for the timed (the current study) and untimed (Chuderski, 2014) Gf test administration.

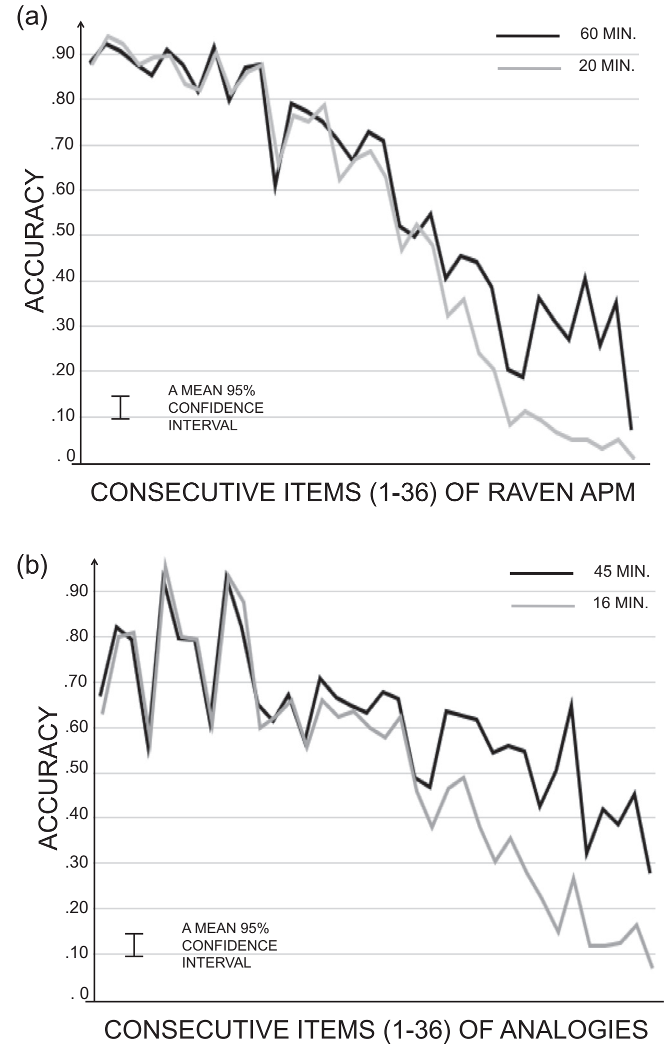

Fig. 4. Proportion correct for the consecutive items of the Raven and paper-and-pencil analogy tests as a function of either the timed (the current study) or untimed (Chuderski, 2014) test administration.

untimed Gf testing. As can easily be seen, the timed testing yields Gf scores that are less dispersed around the best fitting regression line, yielding a stronger WMC-Gf correlation ($r = .689$) than more dispersed untimed Gf scores ($r = .584$), $\Delta r = .105$, $p = .024$, one-tailed test (note that the absolute WMC-Gf correlation coefficients are lower in regression than in the CFA model, because only the latter uses maximum likelihood estimation and reliability corrections, yielding overall higher correlation estimates).

We also analyzed the way in which accuracy for consecutive items of the Raven and paper-and-pencil analogies varied between the current and Chuderski (2014) study. As shown in Fig. 4, time pressure primarily hurt performance on later items of each test, most likely because most participants did not have enough time to attempt those items (or attempted them just by quickly guessing an answer). In consequence, it is likely that the WMC-Gf correlation under pressure vs. no pressure may be driven by different parts of Gf tests.

This ad hoc hypothesis was confirmed by calculating the correlation between our broad WM factor and scores on six bins of the Raven test items (items 1–6, 7–12, 13–18, 19–24, 25–30, 31–36, respectively), separately for each investigated group of participants. The results shown in Fig. 5 indicate that in the case of timed testing, the WMC-Raven correlation was primarily driven by the first half of Raven items (i.e., the easiest ones). In contrast, under virtually untimed testing, both easier and harder Raven items equally contributed to the WMC-Raven correlation.

4. Discussion

The study replicated the observation from Chuderski (2013) that fluid intelligence measured by timed Gf tests is more strongly predicted by WM than when measured by virtually untimed tests. The present result is important, because by extending the original measurement of WM (i.e., one verbal storage task and one relational integration task) with two additional types of tasks (antisaccade and *n*-back), and the figural variants of respective tasks, the conclusion drawn can be generalized onto the broad (i.e., general) factor of WM. Although, unlike in Chuderski (2013), this time WM was not isomorphic to timed Gf, the link between the two variables was anyway so strong that both studies can be validly interpreted as highly compatible.

It has to be noted that our prediction that timed Gf tests yield stronger correlations with WM than do virtually untimed tests has recently been criticized by Colom et al. (2015), who have reported that the correlation between WM and the latent variable calculated using three Gf tests barely changed when one of those

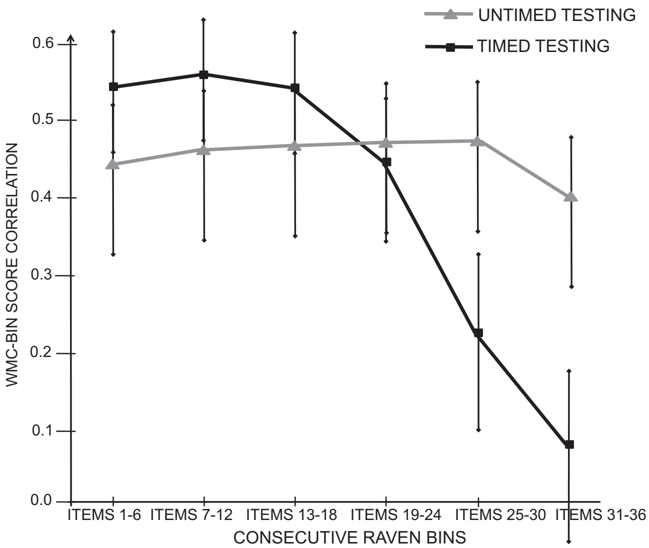

Fig. 5. The strength of correlation between the broad WM factor and scores on six consecutive bins of the Raven test as a function of either the timed (the current study) or untimed (Chuderski, 2014) Raven test administration. Bars represent 95% confidence intervals.

tests (Raven APM) was administered either in the pressured (around hundred people tested in 20 min) or standard (another hundred tested in 40 min) way. However, the present and the Colom et al. results do not necessarily stay in contrast, and the latter unlikely refutes the former. First, our claim that loosening time pressure substantially decreases the Gf-WM correlation pertains primarily to virtually untimed tests (e.g., Raven APM applied in at least 60 min), whereas the standard administration of APM used by Colom et al. can still be considered a relatively pressured test, and we agree with Colom et al. that it will be strongly connected with WM, especially as in our own study (Chuderski, 2013) it also yielded a strong (and comparable to the one found by Colom et al.) correlation ($r = .83$). More importantly, in Colom et al.'s study time pressure was applied to only one Gf test, whereas the two remaining Gf tests, which also strongly loaded Gf variable, were applied within the same time in both groups examined. Thus, in both groups the Gf latent variable might not represent the variance specifically related to time pressure or lack of it; instead, both calculated variables likely reflected the same source of variance.

Overall, the fact that the effectiveness of reasoning processes, which last minutes and operate on quite complex stimuli, is almost isomorphic to accuracy on tasks that require storage, control, integration, or updating of relatively simple stimuli in the course of seconds or even milliseconds, indicates that reasoning measured by the hugely or even moderately timed Gf tests can be reduced to WMC to a large extent. Thus, why timed Gf depends so strongly on available WMC? Two probable hypotheses can be considered.

First, under time pressure the late, most difficult Gf test items, which most probably require also the most complex processing, add less to the overall Gf score than under no pressure (see Fig. 4). In consequence only the early and middle items of a timed Gf test drive its correlation with WMC, but not the late items (see Fig. 5). In light of the Carpenter, Just, and Shell (1990) prominent model of performance on the Raven, which predicted that easier Raven items can be solved solely by properly transforming item rules and rule tokens in WM (as evidenced by their FAIRRAVEN model), whereas more difficult items require both WM operations and more complex processes, like strategic control, learning, abstraction, and rule discovery (as shown by their BETTERRAVEN model), it can be concluded that timed Gf tests so strongly correlate with WMC because their scores are primarily loaded by (early) items that are solved due to using WM. At the same time, participants do not have ample time to run the above mentioned complex processes necessary to solve late items, which might affect Gf scores. In contrast, in untimed Gf tests both the items whose solutions rely on WM and those that need more complex processes contribute to overall Gf score, and drive its correlation with WMC, which due that fact is attenuated in comparison to timed Gf test administration.

Another potential factor that may contribute to the especially strong link between WMC and timed Gf testing is anxiety. A recent study (Chuderski, 2015b) has demonstrated that WM of less intelligent people is especially deteriorated by high levels of state anxiety, whereas WM of more intelligent participants was barely affected by increased anxiety. Thus, as severe time pressure during Gf testing most likely induces increased anxiety levels at least in some participants (e.g., in neurotics), it may additionally decrease WM efficiency in the least intelligent ones (who already suffer from low WMC), and in result it may visibly amplify the effect of WMC on Gf.

In conclusion, the present study substantially extends the Chuderski (2013) study by showing that the fact that the introduction of substantial time pressure during Gf testing results in the visible increase of WMC-Gf correlation, in comparison to virtually untimed testing, is a reliable effect which pertains also to the broad WM factor, encompassing four different functions of WM as well as its both verbal and non-verbal modality. Together, both the Chuderski and the current study might provide important insights on what in fact we measure when we measure fluid intelligence under non-standardized conditions.

References

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.

Chuderski, A. (2013). When are fluid intelligence and working memory isomorphic and when are they not? *Intelligence, 41*, 244–262.

Chuderski, A. (2014). Relational integration task explains fluid reasoning above and beyond other working memory tasks. *Memory & Cognition, 42*, 448–463.

Chuderski, A. (2015a). Why people fail on the fluid intelligence tests. *Journal of Individual Differences*. http://dx.doi.org/10.1027/1614-0001/a000164.

Chuderski, A. (2015b). High intelligence prevents the negative impact of anxiety on working memory. *Cognition & Emotion*. http://dx.doi.org/10.1080/02699931.2014.969683.

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

Colom, R., Privado, J., Garcia, L. F., Estrada, E., Cuevas, L., & Shih, P.-Ch. (2015). Fluid intelligence and working memory capacity: Is the time for working on intelligence problems relevant for explaining their large relationship? *Personality and Individual Differences, 79*, 75–80.

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., Getz, S. J., Macnamara, B., & Engel de Abreu, P. M. J. (2011). Working memory and fluid intelligence: A multi-mechanism view. In R. J. Sternberg & S. B. Kaufman (Eds.), *The Cambridge handbook of intelligence*. Cambridge, UK: Cambridge University Press.

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

Jonides, J., & Smith, E. E. (1997). The architecture of working memory. In M. D. Rugg (Ed.), *Cognitive neuroscience* (pp. 243–276). Cambridge, MA: MIT Press.

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

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.

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.

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

Oberauer, K., Schultze, R., Wilhelm, O., & Süß, H.-M. (2005). Working memory and intelligence – Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61–65.

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

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.

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.