Which working memory functions predict intelligence?

Klaus Oberauer [a],[*], Heinz-Martin Süß [b], Oliver Wilhelm [c], Werner W. Wittmann [d]

[a] *Department of Experimental Psychology, University of Bristol, 12A Priory Road Bristol, BS8 1TU United Kingdom*
[b] *University of Magdeburg, Germany*
[c] *Humboldt University Berlin, Germany*
[d] *University of Mannheim, Germany*

Abstract

Investigates the relationship between three factors of working memory (storage and processing, relational integration, and supervision) and four factors of intelligence (reasoning, speed, memory, and creativity) using structural equation models. Relational integration predicted reasoning ability at least as well as the storage-and-processing construct. Supervision, measured as specific switch costs, was not related to intelligence, but general switch costs were moderately correlated to the reasoning factor. The results question the view of working memory as a device for storage and processing, and the executive-attention account of working memory. They are better explained by theories describing working memory as a system for building relational representations through temporary bindings between component representations.
© 2008 Elsevier Inc. All rights reserved.

Measures of working memory capacity (WMC) have been shown repeatedly to be excellent predictors of intelligence, in particular reasoning ability (for reviews see Ackerman, Beier, & Boyle, 2005; Conway, Kane, & Engle, 2003; Kane, Hambrick, & Conway, 2005; Oberauer, Schulze, Wilhelm, & Süß, 2005). We have proposed a model of the factorial structure of WMC that distinguishes three cognitive functions: concurrent storage and processing, relational integration (previously called coordination), and supervision (Oberauer, Süß, Wilhelm, & Wittmann, 2003). The first function, *concurrent storage and processing*, captures the commonly accepted definition of WMC. It is usually assessed with

complex span tasks in which participants must remember a number of items over a brief period, and perform a processing task in between or after encoding the memory items. *Relational integration* refers to the ability of building new relations between elements and thereby creating structural representations (Waltz et al., 1999). The elements can be held in memory, but can also be given perceptually. Examples are the construction of a mental model of a spatial array from a description (Byrne & Johnson-Laird, 1989), grasping an interaction from a statistical graph (Halford, Baker, McCredden, & Bain, 2004), or "seeing" a constellation in a collection of stars. *Supervision* refers to the control of cognitive processes by goal representations; it includes the prevention of distraction, the setting of response criteria, and the shifting of task sets. These control processes are usually subsumed under the concept of executive functions.

* Corresponding author.
  E-mail address: k.oberauer@bristol.ac.uk (K. Oberauer).

In a comprehensive factor-analytic study (Oberauer et al., 2003) we found the storage and processing factor and the relational-integration factor (called "coordination capacity" in that paper) to be highly correlated but distinguishable. Supervision, as measured by a task-set switching paradigm (Rogers & Monsell, 1995), was only weakly correlated with the other two factors. This result led us to suggest that WMC should be conceptualized as consisting of two closely related aspects, concurrent storage and processing and relational integration; supervision is better regarded as a separate construct.

The purpose of this article is to investigate how the two functional factors of WMC and the supervision factor relate to factors of intelligence. Previous work on the WMC-intelligence relation has nearly exclusively used storage-and-processing tasks to measure WMC (for an exception see Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). When defined in this way, WMC can account for approximately half the variance in tests of reasoning or fluid intelligence (Kane et al., 2004). We believe that measuring WMC only through storage-and-processing tasks is an unfortunate narrowing of the empirical representation of the construct. One goal of this article is to redress this unwarranted restriction of scope.

We predict that the relational-integration factor contributes to the prediction of reasoning ability over and above conventional storage-and-processing tasks. This prediction is justified by two converging theoretical arguments. One line of argument starts from an analysis of demands posed by typical working memory tasks (Oberauer, 2005b). These tasks require short-term maintenance of a small set of elements and of relations between elements. For example, remembering a list of words in order requires remembering the words, and each word's relation to a list position. Other tasks require memory for digits and their relations to spatial locations (e.g., the "memory updating" tasks used by Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). Short-term memory for new, arbitrary relations requires a mechanism for quickly establishing and updating temporary bindings. Based on experimental work (Wheeler & Treisman, 2002) and simulations of temporary bindings in neural networks (Raffone & Wolters, 2001) we argue that the brain has a limited capacity for upholding multiple bindings simultaneously. We believe that this limiting factor is reflected in working memory capacity (Oberauer & Kliegl, 2006).

Our second line or argument derives from an analysis of demands posed by reasoning tasks. We found that reasoning tasks have in common the requirement to build new relational representations (Oberauer, Süß, Wilhelm, & Sander, 2007). For example, inductive reasoning tasks such as series completion require that people construct a representation of the relations between elements of the series, and transfer that representation to a later segment of the series to generate the next element. Deductive reasoning tasks require that people construct mental models of the combined premises to derive a valid conclusion (Johnson-Laird, 1999). Planning requires the construction of hierarchical structures of goals and sequential structures of actions. To construct new relational representations, elements must be bound to each other, or to argument roles in relations (Halford, Wilson, & Phillips, 1998). Therefore, a limit on the number bindings that can be upheld simultaneously posits a limit on the complexity of new relational representations, and thereby limits our reasoning ability. We believe that that limit underlies the common variance of WMC tasks and reasoning tests. In the present study we test the hypothesis that tasks that measure relational integration directly predict reasoning ability at least as well as conventional storage-and-processing tasks.

Our hypothesis about why WMC is closely related to reasoning can be contrasted with two other popular views. One is that WMC tasks and reasoning tasks have in common the requirement of simultaneous storage and processing of information (e.g., Case, 1985; Daneman & Carpenter, 1980). To test this view, we compare the predictive power of relational-integration tasks with a storage component (thus matching the description of "simultaneous storage and processing") with that of parallel tasks without a storage component. The storage-and-processing view predicts that only the tasks with a storage component should be good predictors of reasoning, whereas we predict that the storage demand makes little difference for the predictive power of relational-integration tasks for reasoning tests.

The other alternative is the hypothesis that WMC and fluid intelligence both reflect the efficiency of executive attention, that is, the ability to maintain goals and goal-relevant information in the face of distraction (Kane & Engle, 2002). The task-switching paradigm represented by our supervision factor reflects one aspect of executive attention. It does not, in contrast, require the construction of complex new relations. Thus, whereas the executive-function view of WMC predicts that the supervision factor should contribute substantially to the prediction of reasoning ability, our view does not engender that prediction.

Friedman et al. (2006) have recently found that only one of three factors reflecting executive functions, the

one called *updating*, predicts measures of fluid intelligence. The updating factor is formed by tasks that are virtually identical to tasks that emerged as markers of WMC in a psychometric study (Oberauer et al., 2000), and therefore we regard the updating factor as a WMC factor. Neither the task-set switching factor nor the inhibition factor predicted fluid intelligence in the Friedman et al. study. We investigate whether this negative result can be confirmed for task-set switching as a predictor, using a differentiated set of intelligence factors as criteria.

1. Method

The present data come from the study first reported in Oberauer et al. (2003). That report focused exclusively on the structure of WMC.

1.1. Participants

Participants were 135 students from the University of Mannheim. Their mean age was 25.8 years (SD=3.8), 44% were female. Four participants were excluded due to missing data.

1.2. Materials and procedure

1.2.1. WMC and supervision tasks

There were four tasks for assessing concurrent storage and processing (SP), eight tasks for measuring relational integration (RI), and four tasks for measuring supervision (SUP), described in detail in Oberauer et al. (2003). All tasks were computer based. The SP tasks all followed the same schema: after sequential presentation of a variable number of items to be remembered in correct order, participants worked on an unrelated choice reaction time (CRT) task for 5 s, trying to complete as many trials as possible. After that, the memory items were recalled on an answer sheet. There was one task using words, one using numbers, and two tasks using different kinds of visual–spatial material (arrows and partially filled matrices). The CRT tasks always used material from the same content domain as the memory material. Performance was scored by counting the number of items recalled in their correct positions.

There were four pairs of RI tasks, one using words, one using numbers, and two using visuo-spatial materials. Each task was realized in a no-memory version, in which the elements to be integrated were continuously visible on the screen, and a memory version in which some or all of the elements were presented only briefly and then had to be remembered. In the no-memory version of the verbal task, called *monitoring-verbal*, participants saw one word in each cell of a 3×3 grid. Every 2 s a randomly chosen word was replaced by a new word. Participants had to press the space bar if three words in a row, a column, or a diagonal rhymed with each other. In the memory version, the grid was reduced to a cross with 5 cells,

and each word was removed after being presented, so that participants had to remember the last word presented in each cell, and decide whether these words formed a rhyming row or column. The numerical tasks, *monitoring-numerical*, followed the same procedure. Here, three-digit numbers were presented instead of the words, and participants had to press the space bar when they detected a row, column, or diagonal in which the numbers shared the last digit. Performance in the monitoring tasks was scored as the number of hits minus the number of false alarms.

The first spatial task, flight control, involved monitoring the trajectories of five to nine triangles (representing airplanes) moving in different directions across the screen. Whenever one airplane was about to crash into another plane or a mountain (represented by brown patches) participants were to stop the video and redirect one airplane. Each stop came at a small cost, but each lost airplane incurred a large cost, so that a good score could be obtained by intervening if and only if necessary to prevent crashes. In the memory version, the mountains were displayed only briefly before onset of the movements in each trial and had to be remembered; in the no-memory version they remained visible. Performance was scored by counting the number of plane crashes (reverse-coded so that higher scores reflect better performance). For the second spatial task, *finding squares*, the screen was covered by a 10×10 grid. In the no-memory version, between eight and twelve dots were placed in randomly selected cells. Every 1.5 s two of the dots moved to randomly selected new locations. Participants had to press the space bar whenever four dots formed a square. In the memory version, participants saw between 6 and 10 dots in a sequence and then had to decide whether four of them formed a square if they were visible simultaneously. Performance in the no-memory version was scored as hits minus false alarms, and performance in the memory version was the number of correct decisions across 20 trials.

The RI tasks were all constructed to tap the ability of mentally building and integrating multiple relations between given elements (i.e., the rhyming relations or final-digit identity relation in the *monitoring* tasks, the spatial relations between airplanes and mountains in *flight control*, the relations between dots in *finding squares*). These tasks not only require the detection of pair-wise relations between given elements, but the integration of several relations into the representation of a new configuration. The manipulation of memory demand served to investigate whether a demand on the storage function of working memory was a necessary condition for capturing the source of variance that WMC tasks have in common with each other and with intelligence measures.

Supervision was assessed by indicators of task-set switching costs computed from four versions of the paradigm of Rogers and Monsell (1995). Stimuli were presented in one quadrant of a 2×2 grid, with successive stimuli moving clockwise through the grid. Participants performed one of two speeded two-choice tasks on each stimulus; one task had to be used for stimuli in the upper two quadrants and the other for stimuli in the lower two quadrants. Thereby, every second reaction involved a task switch. In addition to these mixed-task blocks, the two tasks

were also administered in two separate pure blocks. Specific switch costs were computed as difference in log-transformed reaction times (RTs) between switch and no-switch trials in the mixed blocks. General switch costs were computed as the difference between RTs in no-switch trials and mean RTs in the two corresponding pure blocks (again after log-transformation of RTs). There was one mixed block and two pure blocks for each of four types of materials, words, numbers, arrows, and partially filled matrices.

1.2.2. Intelligence test

We used the test for the Berlin Intelligence Structure (BIS) model (Jäger, Süß, & Beauducel, 1997; for an English description see Süß & Beauducel, 2005). The test consists of a large number of different task types, classified according to two dimensions. On the content dimension, verbal, numerical, and visual–spatial contents are distinguished. On the functional dimension there are four factors, reasoning, creativity, memory, and speed. Each of the 12 cells of the model was tested by three task types. The task types were administered in an order that maximizes changes between contents and functions; each task type consisted of one or two pages in a booklet and was time limited.

1.2.3. Data treatment

Before computing general and specific switch costs, RTs from correct responses were logarithmically transformed, so that switch costs reflect proportional increases in RTs. Switch costs were then z-transformed and reverse-coded so that high values reflect better performance. All scores from the WMC and the BIS tasks were z-transformed. BIS task scores were combined into three composite scores for each functional category (i.e., reasoning, creativity, memory, and speed). Each composite consisted of one task from each content domain so that content-related variance was suppressed through aggregation (Wittmann, 1988). These composites were used as manifest variables in the structural equation models. For relational integration we computed four composites by averaging the z-scores from the memory and the no-memory version of each task type. Reliability estimates and descriptive statistics of the WMC data are given in Table 2 of Oberauer et al. (2003).

2. Results

We analyzed the data through a series of structural equation models (the full correlation matrix is given in the Appendix A). First, separate measurement models for intelligence and for the WMC and supervision variables were established. Intelligence was modeled by four correlated factors representing the functional factors of the BIS (Fig. 1, left). The measurement model for WMC and supervision (Fig. 1, right) was the one presented in Oberauer et al. (2003). We explored two ways of specifying the supervision factor, using either the four general switch cost scores or the four specific switch costs as indicators. Moreover, we explored three versions of specifying
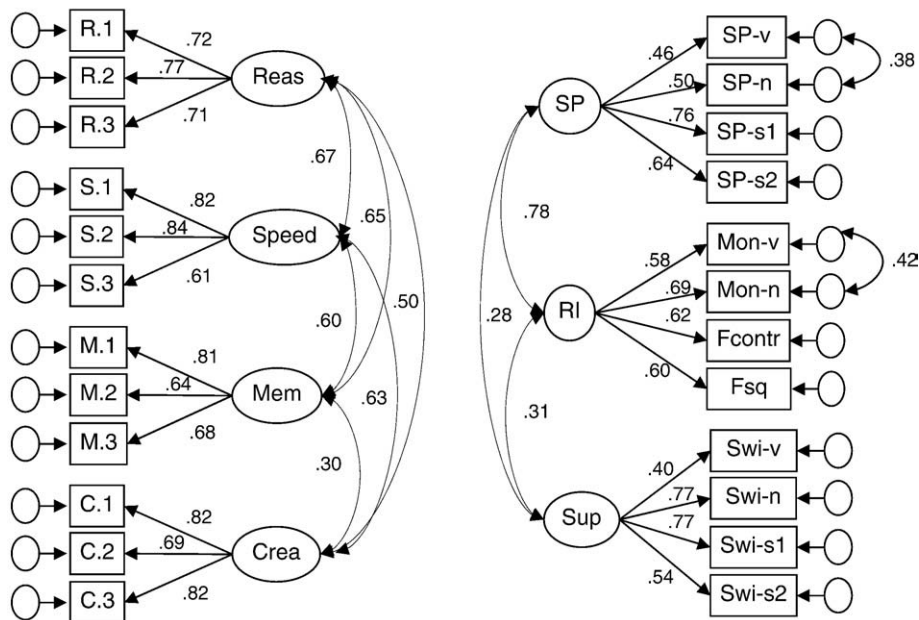
Fig. 1. Left side: Structure of SEM model for intelligence. Large ovals represent latent factors, rectangles represent manifest variables, and small empty ovals represent error terms. Right side: Structure of SEM model for working memory. The factors are storage and processing (SP), relational integration, and supervision (Sup). The manifest variables are Mon = monitoring task, Fcontr = flight control, Fsq = finding squares, Swi = switch cost; suffixes: v = verbal, n = numeric, s1 and s2 refer to the two kinds of spatial material. The parameter estimates are for the model with specific switch costs as manifest variables for supervision, and with composites of memory and no-memory task versions as manifest variables for relational integration (second column in Table 2).

Table 1
Fit statistics for structural equation models

Model	Chi^2 (df)	CFI	RMSEA	SRMR
Intelligence	77 (48)	.954	.068	.063
WMC (RI-full, SSwi)	55 (49)	.984	.031	.053
WMC (RI-full, GSwi)	53 (49)	.988	.025	.057
WMC (RI-no mem, SSwi)	73 (49)	.931	.062	.060
WMC (RI-no mem, GSwi)	60 (49)	.960	.042	.059
WMC (RI-mem, SSwi)	46 (49)	1.0	0	.049
WMC (RI-mem, GSwi)	54 (49)	.984	.027	.060
Intelligence+WMC (RI-full, SSwi)	318 (229)	.925	.055	.069
Intelligence+WMC (RI-full, GSwi)	321 (229)	.919	.056	.068
Intelligence+WMC (RI-no mem, SSwi)	344 (229)	.899	.062	.073
Intelligence+WMC (RI-no mem, GSwi)	331 (229)	.905	.059	.070
Intelligence+WMC (RI-mem, SSwi)	298 (229)	.938	.048	.068
Intelligence+WMC (RI-mem, GSwi)	319 (229)	.917	.055	.068

Note: WMC (SSwi) = working memory capacity model including supervision factor defined by specific switch costs; WMC (Gswi) = version with general switch costs. RI-full = relational-integration factor defined by full set of tasks; RI-no mem = relational-integration factor defined by no-memory task versions only, RI-mem = relational-integration factor defined by memory task versions only. Corresponding correlation and regression models had identical fits. CFI = comparative fit index, recommended to be >.95; RMSEA = root mean square error of approximation, recommended to be <.06; SRMR = standardized root mean residual, recommended to be <.08 (Hu & Bentler, 1999).

the RI factor, one using the composites of memory and no-memory versions of each task as indicators, one using only the no-memory versions, and one using only the memory versions. All six measurement models had good fits to the data

Table 3
Correlations between latent factors of intelligence and working memory, and regression weights for working memory factors predicting intelligence factors

	Reasoning	Speed	Memory	Creativity
Correlations				
Storage and processing	.81	.40	.66	.28
Relational integration (full)	.94	.64	.58	.47
RI-no memory	.88	.63	.59	.48
RI-memory	.92	.51	.52	.40
Supervision (SSwi)	(.21)	(.13)	(.18)	(−.17)
Supervision (GSwi)	.34	(.15)	.34	(.07)
Regression weights WMC → Intelligence				
Storage and processing	(.28)	(−.18)	(.59)	(−.17)
Relational integration (full)	.71	.80	(.06)	.64
Supervision (SSwi)	(−.10)	(−.05)	(−.03)	−.33
Supervision (GSwi)	(.08)	(−.08)	(.25)	(−.11)

Note: The supervision factor was implemented with either general switch costs (GSwi) or specific switch costs (SSwi) as indicators. "RI-no memory" refers to the relational-integration factor using only the no-memory task versions as indicators; RI-memory is the RI factor using only the memory task versions. Correlations in parentheses were not significant at .05 as determined through a bootstrap procedure with 2000 samples (bias-corrected percentile method).

(see Table 1). The loadings of the six models are presented in Table 2.

We combined the measurement models for intelligence and for WMC to investigate the correlations between the latent factors of both sides; these correlations are presented in Table 3. The results of these analyses can be summarized as follows:

Storage and processing (SP) was correlated mostly with the reasoning factor, replicating earlier results (Süß et al., 2002), but also quite substantially with the memory factor — probably because the SP factor reflects performance in memory tasks. Relational integration (RI) was correlated even higher

Table 2
Loadings of tasks on working memory factors in different model versions

Tasks	RI-full GSwi	RI-full SSwi	RI-no mem, GSwi	RI-no mem, SSwi	RI-mem, GSwi	RI-mem, SSwi
SP-v	.52	.46	.52	.52	.52	.51
SP-n	.45	.50	.44	.44	.44	.44
SP-s1	.73	.76	.72	.73	.75	.75
SP-s2	.65	.64	.66	.66	.64	.64
Mon-v	.69	.58	.62	.62	.71	.69
Mon-n	.74	.69	.61	.61	.74	.74
Fcontr	.53	.62	.51	.52	.50	.51
Fsq	.59	.60	.63	.63	.29	.29
Swi-v	.54	.40	.54	.39	.56	.39
Swi-n	.75	.77	.74	.76	.74	.77
Swi-s1	.56	.77	.57	.78	.55	.77
Swi-s2	.08	.54	.08	.54	.08	.54

Note: Column headers represent different model versions, specifying the tasks used to define the relational integration and the supervision factor, respectively; RI = relational integration (full = all tasks, no mem = no-memory tasks, mem = memory tasks), SSwi = specific switch costs, GSwi = general switch costs. For task labels see Fig. 1.
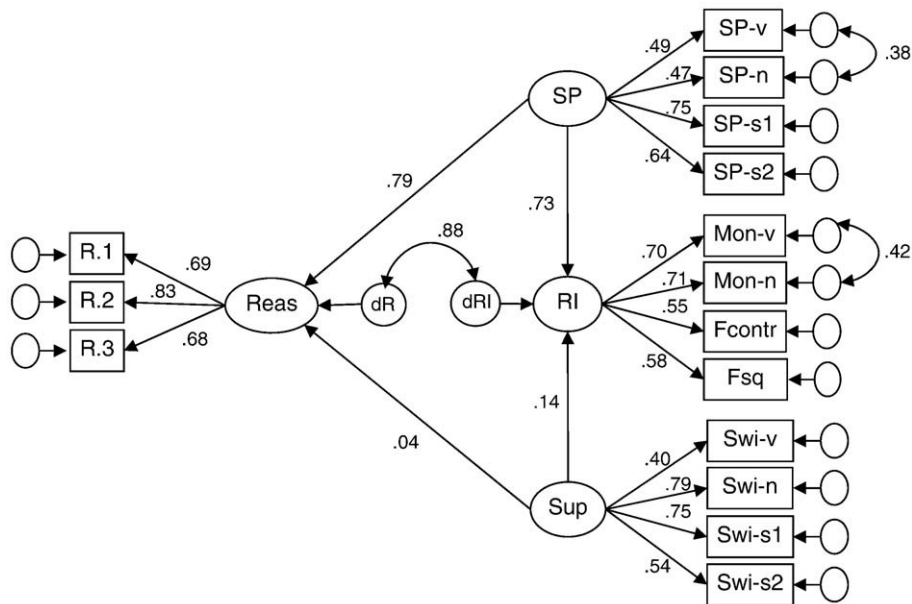
Fig. 2. Model correlating residuals (disturbances) of RI and reasoning (dRI = residual of RI; dR = residual of reasoning). The parameter estimates are for the model with specific switch costs as manifest variables for supervision, and with composites of memory and no-memory task versions as manifest variables for relational integration.

than SP with reasoning, and substantially also with the other three intelligence factors. Constraining the model so that the correlations of SP and RI with reasoning were equal did not lead to a significant loss of fit (ΔChi^2=2.2, Δdf=1). Therefore, the correlations of SP and of RI with reasoning were statistically indistinguishable. They were both in the upper range of correlations between measures of WMC and reasoning observed in previous studies (Kane et al., 2005). Supervision, defined as specific switch costs, was not correlated significantly with any intelligence factor. Supervision, defined as general switch costs, showed a small correlation with reasoning and with memory.

SP and RI were strongly correlated, and therefore we investigated whether they contributed independently to the intelligence factors by running regression models, in which the correlations were replaced by directed paths from working memory factors to intelligence factors. The regression weights are shown in the bottom half of Table 3. The only significant paths, determined by bootstrap analysis, were from RI to reasoning, speed, and creativity, and from supervision (specific switch costs) to creativity (with a negative coefficient, reflecting a suppressor effect). The paths from SP and from supervision (specific switch costs) to reasoning could be fixed to zero without significant loss of fit (ΔChi^2=2.3, Δdf=2).

We also investigated whether the correlations between the RI factor and the other factors in the model were diminished when RI was defined solely through the no-memory versions of the RI tasks. This RI factor's correlations with the intelligence factors were hardly diminished relative to the full RI factor, or relative to the RI factor defined through only the memory task versions (see Table 3, rows with "RI-no-memory" and "RI-

memory", respectively). The correlations of the no-memory RI factor with SP was somewhat smaller (r=.65, compared to .78 for the full RI factor), and that with supervision, defined through general switch costs, was somewhat larger (r=.40 compared to .31 for the full RI factor); the correlation with supervision (specific switch costs) remained unchanged (.28). The RI-memory factor correlated somewhat higher than the full RI factor with SP (r=.85), and moderately with specific switch costs (r=.26) and general switch costs (r=.20).

Another way of testing whether the RI factor accounted for variance in reasoning over and above that accounted for by SP and Supervision is to compute an RI residual and relate it to the reasoning factor.[1] Fig. 2 represents the model structure: RI is predicted by SP and Supervision, and the disturbance of the RI factor represents the residual variance of RI that is not shared with SP or with Supervision. Reasoning is predicted by SP and Supervision. The residual variance of reasoning is correlated with the residual of RI. This correlation was large and significant in all four model versions tested (see Table 4).

We were concerned that the high correlation of RI with reasoning might in part be mediated by shared variance of processing speed, because the RI tasks all involved some degree of time pressure. We used an extension of the model structure of Fig. 2 to test this hypothesis: A factor representing processing speed was added as a further predictor of the reasoning factor and of the RI factor, thereby removing the speed-related variance from the residuals of both constructs. In one variant, we used the processing speed factor of the BIS in this function, and in another variant, we introduced a speed factor defined by the eight pure

---

[1] We thank Roberto Colom for suggesting this analysis.

Table 4
Test of models with RI-residual as predictor of reasoning

Model version	Chi^2 (df)	CFI	RMSEA	SRMR	Correlation (SE)	ΔChi^2
RI-full, Sswi	109.8 (83)	.956	.050	.0767	.88 (.19)	12.6
RI-full, GSwi	106.5 (83)	.958	.047	.0587	.85 (.28)	9.0
RI-no-memory, SSwi	125.7 (83)	.924	.063	.0794	.79 (.36)	11.5
RI-no-memory, GSwi	105.1 (83)	.955	.045	.0604	.73 (.33)	6.9

Legend: RI: relational integration; SSwi: supervision factor defined by specific switch costs; GSwi: supervision factor defined by general switch costs; RI-full: RI factor defined by all RI tasks; RI-no memory: RI factor defined by no-memory tasks only; Correlation = correlation coefficient between residual of RI and residual of reasoning (with bootstrap standard error); ΔChi^2: increase of Chi^2 when correlation between residuals is removed ($\Delta df = 1$). All ΔChi^2 are significant at $p < .01$.

blocks of choice RT tasks (i.e., the no-switching baseline measures of the task-switching tests). Both variants were applied to the four models in Table 4. Across the resulting eight models, the correlation between the reasoning residual and the RI residual ranged from .62 to .86. We conclude that RI and reasoning share a substantial portion of variance that is neither related to SP nor to processing speed.

3. Discussion

Our results provide strong evidence for the hypothesis that the common variance of WMC and reasoning centrally includes the ability to form new structural representations (Oberauer et al., 2007). The RI tasks were explicitly constructed to capture this construct, and they have proven to be at least as good as dual-task combinations of storage and concurrent processing in predicting reasoning. The predictive power of RI was retained even with tasks that did not require any storage in the traditional sense of maintaining information that is no longer perceptually available. SP tasks have so far dominated psychometric research on WMC. Therefore, the success of RI tasks is considerable progress in refining the construct WMC, and in developing a more profound understanding of individual differences in reasoning ability. Two further studies following up on the present one, using the no-memory versions of our RI tasks, replicated their strong correlation with tests of reasoning ability (Buehner, Krumm, & Pick, 2005; Buehner, Krumm, Ziegler, & Pluecken, 2006). In contrast, supervision, measured by general or specific switch costs, contributed very little to the explanation of intelligence. This finding replicates Friedman et al. (2006) and strengthens our contention that WMC should not be equated with executive functions.

These findings have far-reaching implications for our view of working memory and intelligence. The traditional interpretation of the relationship between working memory and fluid intelligence or reasoning is that working memory provides resources for simultaneous storage and processing, that is, the ability to remember

information not currently present in the environment, and to manipulate this or other information at the same time. Both abilities are arguably required in many complex tasks — for example, remembering intermediate results while carrying out further operations in multi-step mental arithmetic tasks (Hitch, 1978). On this account, however, it is hard to understand why relational-integration tasks without any demand on storage should predict reasoning so well. The main difference between specifying the RI factor through memory task versions and specifying it through no-memory task versions was that in the former case, the factor correlated more with SP, confirming that the variation of memory in the RI tasks was effective. This variation had little effect, however, on the RI factor's correlation with reasoning or the other intelligence factors. We conclude that a demand on short-term storage is not a necessary feature of a good measure of WMC. Other research (Colom, Rebollo, Abad, & Shih, 2006; Oberauer et al., 2000) has already shown that a processing component is no necessary feature either. Thus, "simultaneous storage and processing" is a good description for one effective and very popular class of tasks used to measure WMC, but it should not be used to define WMC as a construct.

This is not to say that our RI tasks capture all there is to the construct WMC, and that SP tasks are redundant. Rather, we argue that the construct WMC should be conceptualized in a broader way than before, and operationalized by a broader set of tasks. The present RI tasks were intentionally designed to be different from conventional SP tasks, with the goal to establish a separate factor of RI besides SP, and to test the hypothesis that despite their dissimilarity with SP tasks, RI tasks predict reasoning ability. The finding that both our RI tasks and the SP tasks, despite their superficial dissimilarity, share a large amount of variance and account for large amounts of variance in reasoning, raises the need for a conceptualization of WMC that covers both kinds of tasks.

One might object — as suggested by one anonymous reviewer — that our finding of a high correlation between RI and reasoning is trivial because we predict the criterion

by another instance of the criterion. The argument raised by the reviewer is twofold: (1) the RI factor correlated more with reasoning than with SP, and therefore it should be regarded as part of the reasoning construct rather than as part of an extended WMC construct, and (2) the RI tasks involved cognitive components that have been argued to be involved in reasoning tasks, too, in particular the detection of relations between given elements, as captured by Spearman's (1927) notion of "education of relations" and by the "inference" component in Sternberg's (1985) componential theory. We believe that both arguments are unconvincing.

In response to (1) we argue that a variable does not become part of the reasoning construct simply because it is highly correlated with a reasoning factor. Constructs such as reasoning ability and WMC are not just vectors in factor space, they also have a conceptual meaning. Reasoning, for instance, is well defined in philosophy and cognitive psychology as deriving a conclusion from given information by a rationally justifiable line of argument; the argument forms are often classified as deductive, inductive, and abductive (Wilhelm, 2005). If a variable correlates highly with reasoning ability but does not match the definition of the construct, it cannot be argued to represent reasoning. Our monitoring tasks clearly don't involve inductive, deductive, or abductive inferences, they don't even involve deriving a conclusion, not to speak of a rational argument. Subsuming them under the reasoning construct simply because of their high correlation with reasoning ability would mean to blame the predictor for its success. In contrast, it is conceptually fully justified to regard the RI factor as part of the WMC construct because the RI tasks were constructed to capture a theoretically justified extension of the concept of WMC.

In response to (2) we fully acknowledge that others before us have noticed the central role of relations in reasoning.[2] This does not imply that every task involving the representation and processing of relations is by definition a reasoning task — as argued above, the concept of reasoning involves more than processing of relations. Representing and processing relations is one of several characteristics of reasoning tasks — as shown most clearly by the fact that "inference" is one of several components Sternberg (1985) assumes to underlie reasoning performance. Individual differences in all these characteristics or components could be responsible

for individual differences in reasoning ability, and it is far from trivial to show that the ability to represent relations is an important predictor. Moreover, whereas previous authors highlighted the role of relations, our hypothesis goes beyond these ideas by emphasizing the *integration* of relations. Tasks that only require processing of individual relations don't qualify as RI tasks and therefore should not be as good predictors of reasoning as comparable tasks that require relational integration. For instance, we predict that tasks such as deciding which of two numbers is larger, or whether two given words rhyme, are not as good predictors of reasoning ability as our RI tasks. In fact, many tasks measuring the Speed component of the BIS are of this kind, and Speed correlates less with reasoning than RI does.[3]

Our findings support a view of working memory that goes beyond its characterization of a system for temporary

---

[2] It is worth mentioning in this context that "eduction of relations" and the "inference" component in Sternberg's theory refer to the discovery of relations, a process that lies at the heart of inductive reasoning but seems to be less involved in deductive reasoning. In Sternberg's componential theory, "inference" is not among the components assumed for deductive reasoning.

[3] In a further round of reviews, the anonymous reviewer raised the following points: (1) The strong relationship between WMC, defined as SP, and reasoning offers a parsimonious reductionist explanation of reasoning in terms of a lower-order processing mechanism (i.e., storage and processing) that does not include relational processing. Including RI into the concept of WMC renders the WMC construct more complex and blurs the boundary between WMC and reasoning, because reasoning also involves relational processing. (2) It can be argued that the relational processing involved in our RI tasks is very similar to the relational processing necessary in analogical reasoning tasks such as those used by Sternberg (1985) to measure the "inference" component (i.e., discovering the relationship between the first 2 terms of an analogy). (3) The RI factor may have unique components of variance beyond relational processing, but that claim would have to be demonstrated more unequivocally. In response to (1) we point out that recent experimental work is just beginning to unravel the processes involved in typical SP tasks such as the complex span task (Barrouillet, Bernardin, & Camos, 2004; Hudjetz & Oberauer, 2007; Unsworth & Engle, 2007), and the emerging picture is far from simple. One insight emerging clearly from a careful analysis of errors on the complex span task is that it involves setting up temporary bindings between memory list items and representations of their serial positions (Unsworth & Engle, 2006). Thus, the complex span task already involves relational processing, among other things, and we argue that this is why they correlate highly with reasoning and with our RI tasks. What is new about our RI tasks is that they were intentionally constructed to measure relational integration, and therefore they are arguably purer measures of what we think is the common source of variance of WMC and reasoning. In response to (2) we fully agree that discovering the relation between two terms in an analogy is the same kind of relational processing as is measured in our RI tasks. Yet we insist that WMC is not a limit on processing individual relations but on *integrating* relations, and this is what is measured by the RI tasks. Analogy tasks also require integrating relations, but this is necessary only in the next step, where the relation between the first two terms is applied to the third term to complete the analogy — Sternbergs components "mapping" and "application". This is why, in our view, analogy tasks correlate with WMC. In response to (3) we readily concede that the evidence for the construct validity of the RI tasks as measures of relational *integration*, as opposed to processing of individual relations, is only preliminary, and it is crucial to follow up our study with experimental and correlational research testing this claim.

---

storage and processing. We assume that working memory primarily serves to integrate information — whether perceptually given or memorial — and construct new relational representations from them (Halford et al., 2004; Oberauer et al., 2007; Waltz et al., 1999). Relational representations also underlie performance in complex tasks such as text comprehension and reasoning. Building new relational representations requires a mechanism for temporary binding of the elements that are integrated, either by binding them to positions in a common cognitive coordinate system or by binding them directly to each other. The common denominator of working memory and other complex tasks such as reasoning could be the capacity to build and maintain such bindings (Oberauer, 2005a; Wilhelm & Oberauer, 2006).

This *binding hypothesis* can also explain why SP tasks are successful measures of WMC. Typical SP tasks can be regarded as one instance of relational integration: Participants usually must report back a memory list in order, and representing an ordered list is a special case of a relational representation. Many contemporary models of serial-order recall represent order through bindings between each item and its position on a temporal or ordinal context representation (e.g., Burgess & Hitch, 1999). Bindings between memory items and their contexts become particularly important when other representations not belonging to the memory set are also highly activated. For example, in the complex span procedure (Conway et al., 2005) encoding of memory items alternates with processing of other material that should not be recalled. Items to be recalled must be distinguished from those involved in the processing task by binding the former but not the latter to a context representation that serves as a cue for recall. Likewise, in situations with high proactive interference, bindings between the currently relevant memory items to a context representing the present list is necessary to distinguish these items from previous, no longer relevant lists. This explains why immediate-memory tasks correlate more with measures of WMC when proactive interference is strong (Bunting, 2006; Kane & Engle, 2000).

An alternative, widely endorsed view is that WMC reflects the efficiency of executive attention (Kane & Engle, 2002). This view could cover the RI tasks by arguing that relational integration requires the control of attention such that the elements to be integrated are attended to simultaneously. The executive-attention account, however, has difficulties in explaining the lack of a strong relationship between task-set switching and reasoning, because task-set switching is a prototypical executive-control task. Kane and Engle (2002, p. 638) define executive attention as the "capability whereby memory representations are maintained in a highly active state in the presence of interference, and these representations may reflect action plans, goal states, or task relevant stimuli in the environment". Task-set switching requires maintaining one goal state and its associated action plan highly active in the presence of interference from the other, currently not relevant goal state and action plan. General switch costs reflect the amount of continuous interference from the currently irrelevant task set, whereas specific switch costs reflect the efficiency of the executive system in changing relative levels of activation between the two goal states and action plans, that is, the efficiency of controlling attention on a moment-to-moment basis. Thus, task switching matches perfectly the definition of executive attention, yet its efficiency correlated only weakly with other measures of WMC and with reasoning in the present study.

A modified version of the executive attention view of Kane and Engle (2002), however, could be reconciled with the binding hypothesis by making the following two assumptions: (1) Building and maintaining bindings of multiple elements requires simultaneous attention to these elements, and (2) WMC reflects the ability to direct attention to multiple elements at the same time.

Acknowledgement

Appendix A: Correlation table as used for the models

Variable	S.1	S.2	S.3	M.1	M.2	M.3	C.1	C.2	C.3	R.1	R.2	R.3	SP-v	SP-n	SP-s1	SP-s2	Mon-v	Mon-n	Fcontr	Fsq
S.2	.71																			
S.3	.47	.48																		
M.1	.35	.40	.39																	
M.2	.17	.30	.32	.54																
M.3	.36	.39	.42	.51	.49															
C.1	.49	.44	.29	.23	.15	.18														
C.2	.35	.43	.23	.28	.21	.32	.53													

(continued on next page)

Appendix A (*continued*)

Variable	S.1	S.2	S.3	M.1	M.2	M.3	C.1	C.2	C.3	R.1	R.2	R.3	SP-v	SP-n	SP-s1	SP-s2	Mon-v	Mon-n	Fcontr	Fsq
C.3	.40	.39	.23	.08	.11	.14	.68	.59												
R.1	.46	.41	.45	.46	.17	.28	.28	.20	.25											
R.2	.43	.36	.47	.44	.22	.32	.37	.33	.26	.56										
R.3	.32	.36	.35	.39	.26	.38	.29	.35	.30	.50	.56									
SP-v	.24	.22	.44	.32	.35	.25	.02	.15	.05	.32	.37	.41								
SP-n	.19	.14	.27	.13	.10	.18	.20	.14	.06	.18	.27	.22	.52							
SP-s1	.27	.22	.36	.34	.30	.35	.27	.15	.20	.44	.51	.40	.34	.31						
SP-s2	.12	.08	.22	.38	.30	.31	.11	.04	.08	.36	.38	.35	.28	.43	.49					
Mon-v	.33	.32	.42	.37	.23	.34	.31	.34	.17	.47	.62	.43	.27	.21	.35	.22				
Mon-n	.43	.43	.38	.38	.21	.39	.33	.27	.25	.49	.52	.48	.29	.31	.41	.31	.65			
Fcontr	.14	.12	.28	.17	.09	.21	.17	.14	.18	.31	.45	.27	.16	.15	.46	.28	.40	.39		
Fsq	.42	.31	.28	.24	.16	.25	.27	.20	.19	.34	.48	.37	.21	.29	.34	.26	.34	.46	.35	
Mon-v-nm	.34	.40	.38	.33	.23	.32	.29	.35	.21	.39	.52	.37	.24	.11	.18	.10	.89	.52	.34	.29
Mon-n-nm	.40	.43	.35	.32	.16	.37	.28	.22	.22	.39	.35	.35	.14	.28	.24	.22	.54	.87	.28	.40
Fcontr-nm	.10	.07	.22	.12	.04	.20	.13	.13	.14	.33	.41	.29	.13	.12	.42	.27	.40	.36	.93	.32
Fsq-nm	.34	.27	.19	.28	.26	.29	.27	.25	.23	.36	.48	.35	.23	.27	.38	.31	.34	.47	.40	.77
Mon-v-m	.23	.17	.35	.31	.17	.28	.26	.24	.09	.45	.58	.40	.24	.26	.44	.30	.87	.63	.37	.31
Mon-n-m	.35	.31	.32	.34	.20	.30	.30	.24	.22	.47	.54	.48	.36	.26	.48	.31	.60	.87	.40	.39
Fcontr-m	.15	.15	.30	.18	.13	.19	.19	.14	.19	.25	.42	.21	.18	.15	.44	.25	.35	.37	.93	.33
Fsq-m	.30	.21	.24	.09	−.01	.09	.15	.06	.06	.15	.26	.22	.09	.17	.14	.08	.18	.23	.14	.77
SSwi-v	−.11	−.17	−.03	−.05	.01	−.05	−.09	−.11	−.14	−.14	−.03	−.11	.03	.01	.14	.09	−.12	−.03	.11	.12
SSwi-n	.11	.11	.25	.18	.15	−.03	−.04	−.06	−.12	.07	.15	−.01	.16	.10	.14	.06	.10	.11	.11	.07
SSwi-s1	.05	.06	.19	.14	.12	.05	−.07	−.06	−.16	.16	.18	.15	.25	.14	.18	.15	.11	.24	.22	.18
SSwi-s2	.02	.04	.14	.05	.14	.09	−.04	−.10	−.16	.13	.12	.10	.03	.01	.15	−.01	.18	.19	.11	.07
GSwi-v	.07	.10	.03	.17	.01	.15	.04	−.04	.08	.14	.21	.13	−.04	−.09	−.02	.10	.03	.16	.02	.02
GSwi-n	.05	.05	.12	.19	.06	.18	.05	.12	.01	.19	.14	.13	.02	−.02	.02	.06	.09	.16	.13	.18
GSwi-s1	.06	.09	.16	.22	.14	.24	.03	.00	−.01	.24	.13	.23	.07	.04	.14	.20	.11	.21	.22	.18
GSwi-s2	.16	.24	.07	.18	.03	.07	.02	.11	−.05	−.02	.01	−.01	−.02	−.11	.02	−.11	.02	.12	−.03	−.03

Variable	Mon-v-nm	Mon-n-nm	Fcontr-nm	Fsq-nm	Mon-v-m	Mon-n-m	Fcontr-m	Fsq-m	SSwi-v	SSwi-n	SSwi-s1	SSwi-s2	GSwi-v	Gswi-n	Gswi-s1
Mon-n-nm	.47														
Fcontr-nm	.35	.26													
Fsq-nm	.32	.42	.39												
Mon-v-m	.55	.47	.36	.28											
Mon-n-m	.43	.51	.36	.39	.62										
Fcontr-m	.28	.25	.73	.34	.34	.39									
Fsq-m	.12	.19	.09	.17	.19	.21	.17								
SSwi-v	−.18	−.03	.10	.10	−.02	−.03	.09	.08							
SSwi-n	.08	.13	.10	.09	.10	.07	.11	.01	.36						
SSwi-s1	.07	.26	.24	.15	.12	.16	.16	.12	.26	.59					
SSwi-s2	.13	.13	.15	.01	.19	.20	.06	.09	.21	.41	.43				
GSwi-v	.10	.17	.04	.11	−.05	.10	−.01	−.07	−.12	.12	.08	.15			
GSwi-n	.17	.16	.14	.15	−.03	.12	.10	.12	−.14	−.13	.08	.15	.42		
GSwi-s1	.16	.20	.18	.24	.03	.17	.23	.03	.00	−.05	.00	−.03	.25	.43	
GSwi-s2	.02	.14	−.08	−.02	.00	.08	.03	−.02	.02	.07	.09	−.03	.12	−.02	.13

Legend: S = BIS-speed; M = BIS-memory; C = BIS-creativity; R = BIS-reasoning; SP = storage and processing (v = verbal, n = numerical, s = spatial); Mon = monitoring (v = verbal, n =numerical); Fcontr= flight control, Fsq = finding squares (no suffix: combined memory and no-memory score, nm = no-memory, m = memory); SSwi = specific switch costs, GSwi =general switch costs (v = verbal, n = numerical, s = spatial).

References

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

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83−100.

Buehner, M., Krumm, S., & Pick, M. (2005). Reasoning=working memory≠attention. *Intelligence*, *33*, 251−272.

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.

Bunting, M. F. (2006). Proactive interference and item similarity in working memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *32*, 183−196.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106*, 551−581.

Byrne, R. M. J., & Johnson-Laird, P. N. (1989). Spatial reasoning. *Journal of Memory and Language, 28*, 564−575.

Case, R. (1985). *Intellectual development: Birth to adulthood.* Orlando: American Press.

Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory & Cognition, 34*, 158−171.

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 & Review, 12*, 769−786.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7*, 547−552.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450−466.

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.

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2004). How many variables can humans process? *Psychological Science, 16*, 70−76.

Halford, G. S., Wilson, W. H., & Phillips, S. (1998). Processing capacity defined by relational complexity: Implications for comparative, developmental, and cognitive psychology. *Behavioral and Brain Sciences, 21*, 803−864.

Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology, 10*, 302−323.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1−55.

Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition, 35*, 1675−1684.

Jäger, A. O., Süβ, H. -M., & Beauducel, A. (1997). *Test für das Berliner Intelligenzstrukturmodell (BIS).* Göttingen: Hogrefe.

Johnson-Laird, P. N. (1999). Deductive reasoning. *Annual Review of Psychology, 50*, 109−135.

Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 336−358.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin and 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 (2004). *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.

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. (2005). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 393−408). Thousand Oaks: Sage.

Oberauer, K., & Kliegl, R. (2006). A formal model of capacity limits in working memory. *Journal of Memory and Language, 55*, 601−626.

Oberauer, K., Schulze, R., Wilhelm, O., & Süβ, H. -M. (2005). Working memory and intelligence — Their correlation and their relation: A comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 61−65.

Oberauer, K., Süβ, H. -M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity — Facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017−1045.

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

Oberauer, K., Süβ, H. -M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory — Storage, processing, supervision, and coordination. *Intelligence, 31*, 167−193.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience, 13*, 766−785.

Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*, 207−231.

Spearman, C. (1927). *The nature of "intelligence" and the principles of cognition.* London: MacMillan.

Sternberg, R. J. (1985). *Human abilities.* An Information-Processing Approach New York: Freeman.

Süβ, H. -M., & Beauducel, A. (2005). Faceted models of intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 313−332). Thousand Oaks: Sage.

Süβ, 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.

Unsworth, N., & Engle, R. E. (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., & Engle, R. W. (2006). A temporal-contextual retrieval account of complex span: An analysis of errors. *Journal of Memory and Language, 54*, 346−362.

Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., de Menezes Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychological Science, 10*, 119−125.

Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General, 131*, 48−64.

Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 373−392). Thousand Oaks: Sage.

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.

Wittmann, W. W. (1988). Multivariate reliability theory. Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp. 505−560). New York: Plenum.