ELSEVIER

Intelligence 30 (2002) 261–288

INTELLIGENCE

Working-memory capacity explains reasoning ability—and a little bit more

Heinz-Martin Süß*, Klaus Oberauer, Werner W. Wittmann,
Oliver Wilhelm, Ralf Schulze

Lehrstuhl Psychologie II, University of Mannheim, Schloss, EO, D-68131 Mannheim, Germany

Received 1 December 1996; received in revised form 9 February 2000; accepted 21 February 2001

Abstract

Working-memory capacity is conceptually differentiated according to functions and contents. The resulting two-faceted structure parallels the structure of intellectual abilities in the Berlin Intelligence Structure Model (BIS) [Diagnostica 28 (1982) 195.]. A battery of 17 working-memory tasks, chosen to represent the proposed facet structure of working memory, was administered together with a test for the BIS to 128 young adults. General working-memory capacity was highly related to general intelligence. The prediction of intellectual abilities by working-memory capacity was also tested by differentiating predictor and criterion according to the functional and to the content facet. Moreover, the paths from working memory to intelligence factors appear to be highly specific. This suggests that specific working-memory resources, as opposed to a general capacity, are the limiting factors for their corresponding counterparts in the structure of mental abilities.

1. Introduction

The interplay between experimental research and research on individual differences has led to considerable progress in the field of intelligence research. The point of convergence for both approaches is the question: "Which factors limit human cognitive abilities?" Two main paths have been followed in the attempt to use the theoretical concepts of cognitive psychology to explain individual differences in intelligence.

---

* Corresponding author. Tel.: +49-621-181-2131; fax: +49-621-181-2129.
E-mail address: hms@rumms.uni-mannheim.de (H.-M. Süß).

PII: S0160-2896(01)00100-3

The first approach aims to decompose the cognitive processes of task solution into single components (e.g., Sternberg, 1985). Persons differ in the efficiency with which they perform these single components. The ability to solve an intelligence test task is explained by the composition of the efficiency parameters of the components involved. The positive intercorrelation of different tasks can be explained by overlap in the sets of components needed to solve the tasks. The second approach postulates factors that limit intellectual abilities to be found in the general features of the cognitive architecture (i.e., a set of structures and processes that characterize human information processing). The most promising candidates at present are general mental speed (for reviews, see Juhel, 1991; Kail & Salthouse, 1994) and working-memory capacity (Kyllonen & Christal, 1990). These factors are assumed to limit performance, regardless of the specific processes that lead to a task solution. The present work follows the latter line of reasoning, focusing in particular on the role of working-memory capacity in cognitive abilities.

Working memory is assumed to have a capacity limit that is relevant especially for complex tasks. It is therefore related to reasoning ability, a central subconstruct within the structure of intelligence. A strong relationship between working-memory capacity and reasoning has been found in several studies (Engle, Tuholski, Laughlin, & Conway, 1999; Fry & Hale, 1996; Kyllonen, 1994a; Kyllonen & Christal, 1990; Salthouse, 1992; Salthouse, Babcock, Mitchell, Palmon, & Skovronek, 1990). At the same time, these studies suffer from one or more of the following shortcomings:

1. Working memory was often measured using a small, unsystematic set of tasks. The resulting representation of the construct is therefore contaminated by task-specific variance and does not necessarily reflect all relevant aspects of working memory.
2. In most studies on individual differences, working-memory capacity was operationalized as an undifferentiated construct. This contrasts sharply with numerous experimental findings that suggest a multicomponential view of working memory (e.g., Baddeley, 1986). Different components of working memory, i.e., distinguishable cognitive resources, could be expected to contribute to different extents to different intellectual abilities.
3. Similarly, studies relating working memory to intelligence constructs usually focused on a single mental ability, such as reasoning or reading ability, as the criterion. This does not provide a clear picture of how working memory relates to the structure of intelligence, in other words, which abilities depend to what degree on working memory.
4. The tasks used to measure working memory were very similar to and sometimes indistinguishable from common reasoning tasks. This problem is apparent, e.g., in the work of Kyllonen and Christal (1990), where the same type of task served as a working-memory task in one study and as a reasoning task in another. If tasks used to measure working-memory capacity have many common features with those used to measure reasoning, it is not clear which features are responsible for the high correlation obtained between them.

The present study addressed these issues by using working-memory tasks carefully selected from previous research to represent all aspects of working-memory capacity, but at the same time to minimize overlap with intelligence tasks. These tasks were then related to a well-validated test for the Berlin Intelligence Structure Model (BIS) (see next section).

1.1. Facets of intelligence: the BIS

Several alternate structure models of cognitive abilities have been developed in the long tradition of intelligence research. The present research is based on the BIS developed by Jäger (1982, 1984). The original construction of the BIS was based on a representative sample of all intelligence test tasks documented in the psychometric intelligence literature, which makes the BIS a model with a high degree of generality. The structure of the model has been confirmed repeatedly, using different methods and samples from different populations (for a summary, see Jäger, Süß, & Beauducel, 1997; for a recent replication study, see Bucik & Neubauer, 1996). The BIS structure has also been successfully replicated using different samples of tasks, such as a German version of French, Ekstrom, and Price's (1963) *Kit of Reference Tests* (Jäger & Tesch-Römer, 1988). The BIS is a hierarchical system of intellectual abilities structured on several levels of generality (Fig. 1). On the most general level, *g* is assumed to be an integral part of all ability components. On the second level of generality, several ability components are postulated. These belong to two facets: the *operational* facet (the type of cognitive function demanded by the task) and the *content* facet (the different classes of task content). The model includes four operational components: *processing capacity* (K) corresponds to reasoning factors in other models; *creativity* (E) refers to the

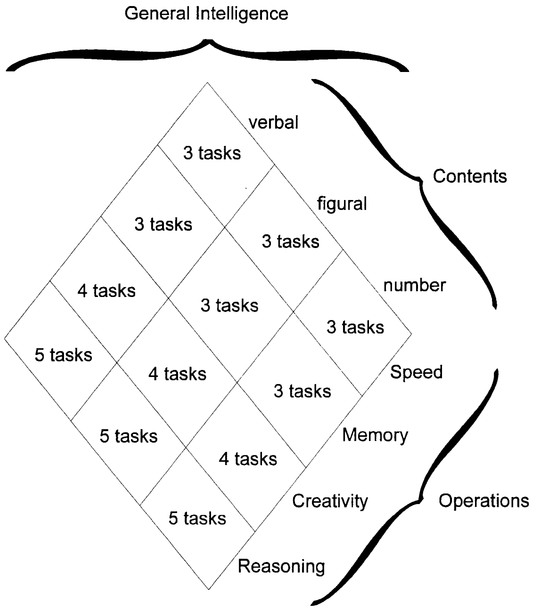

Fig. 1. The BIS. Four functional abilities are cross-classified with three content abilities. General intelligence represents the integral of all operational and content-related abilities, respectively.

ability to fluently produce many different ideas; *memory* (M) represents the ability to recall lists and configurations of items a few minutes after learning them; *speed* (B) refers to the ability to perform quite simple tasks quickly and accurately. The three content components postulated are: verbal (V), numerical (N), and spatial–figural (F) abilities. The cross-classification of the four operational and three content-related components yields 12 cells on a third, more specific level (Jäger, 1982, 1984).

In the following, the BIS is compared to other, more familiar, intelligence models. The four operational and the three content factors of the BIS are comparable to the second stratum factors in Carroll's (1993) Three-Stratum Theory. For example, reasoning ability in the BIS closely matches Carroll's fluid intelligence; memory ability corresponds to general memory and learning; creativity in the BIS corresponds closely to broad retrieval ability; and the BIS speed factor roughly covers the factors broad cognitive speediness and processing speed. Carroll's crystallized intelligence factor can best be mapped onto verbal and numerical abilities in the BIS, and broad visual perception onto figural ability. Broad auditory perception is the only one of Carroll's second stratum factors without a counterpart in the BIS.

The facet structure of the BIS shares some similarities with Guilford's Structure-of-Intellect model (Guilford, 1956, 1967), Guttman's radex model (Marshalek, Lohman, & Snow, 1983), and the CAM framework (Kyllonen, 1994b). Guilford's model has a three-faceted structure combining to produce a total of 120 factors and, in a later version, 180 factors, whereas the BIS differentiates only two facets with three and four components in each, resulting in 12 cells. Moreover, Guilford proposed ability factors in the cells with different levels of generality, whereas Jäger (1982) assumed only performances at the cell level. Jäger postulated that each performance is the result of at least two abilities, one operational and one content ability. The BIS is much more parsimonious than Guilford's model, and it is therefore plausible to assume that the constructs are also much more general. This latter point could be the reason that the BIS has been empirically well validated in many different studies.

The radex model of intelligence specifies a complexity dimension ranging from the center to the periphery, with *g* at the center. This complexity dimension corresponds to the general-to-specific dimension in a hierarchical model like the BIS. Like the BIS, the content facet of the radex model also differentiates the three content classes. The location of reasoning, speed, memory, and creativity tasks in the radex depend on their content and on their *g* loading.

The CAM framework is a more general approach to human abilities that includes intelligence, working memory, and knowledge constructs, and employs a four-facet structure. The first facet, cognitive processes, closely resembles the operational facet in the BIS, but is much more comprehensive. The second facet, knowledge domains, contains the three content classes.

In all these models, the reasoning factor ("processing capacity" in the BIS), which is of particular interest in this study, is measured by tasks that can be further differentiated into four categories: inductive reasoning, deductive reasoning, construction, and planning tasks. Inductive reasoning tasks are those in which a rule must be abstracted from individual exemplars. Typical inductive tasks are analogies, series completion, and matrix tasks. *Deductive reasoning* refers to inferences that are necessarily true if the premises are true.

Typical tasks are syllogistic inferences, mathematical text problems, and surface development tasks. *Construction tasks* such as anagrams, for example, require the integration of given elements into a configuration with certain required features. *Planning tasks* require step-by-step action to produce an appropriate sequence to reach a given goal. Tower of Hanoi, for example, belongs to this category. The four reasoning task categories proposed here are not intended to correspond to different mental abilities. They serve only to systematize the broad pool of reasoning tasks and give a more precise description of the concept's intended scope of application.

1.2. Facets of working memory

Similar to intelligence, working-memory capacity can be structured along two facets (Fig. 2): the content facet and the function facet (see Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). As in the BIS and other intelligence structure models, the same three-content domains can be distinguished on the *content facet*: verbal, numerical, and spatial–figural working memory. From Baddeley's (1986) working-memory model, a difference between a subsystem for verbal and numerical content (the phonological loop) and a spatial–figural one (the visuospatial sketch pad) can also be derived.

On the *functional facet,* three functions that have been ascribed to working memory in previous research can be differentiated. The most common function is the *simultaneous storage and processing of information* (Baddeley, 1986; Daneman & Carpenter, 1980; Salthouse, 1990). This can be defined as the dual requirements of retaining a set of information elements in memory for a brief period and, at the same time, manipulating these or some other elements of information.

A second set of functions associated with working memory can be summarized under the heading of "executive functions." These functions were first ascribed to working memory by Baddeley (1986), who specified the "central executive" in terms of the "supervisory attentional system" (SAS) of Norman and Shallice (1980). In the meantime, a broad set of cognitive functions has been subsumed under the label of "executive" (see Smith & Jonides, 1999). In this work, we focused on the more narrow construct originally circumscribed by the SAS model. That is, monitoring of mental operations, controlling their efficiency, inhibiting inappropriate schemata, and activating more appropriate ones. Therefore, we call this the *supervision function* of working memory. A recent study by Miyake et al. (2000) investigated three executive functions ("task set shifting," "updating," and "inhibition") and showed that they had some variance in common. Only one of these functions, "updating," had a significant relationship to a classical working-memory task, Operation Span (Turner & Engle, 1989). In fact, the tasks used to measure "updating" were very similar to the "storage and processing" tasks in our battery, so that their "updating" factor probably corresponds closely to the "storage and processing" construct as defined here. The tasks we used to measure supervision, on the other hand, correspond mainly to the "task set shifting" factor in Miyake et al. (2000).

The third function connected with working memory is *coordination*. Coordination refers to the integration of information from different content domains (Yee, Hunt, & Pellegrino,

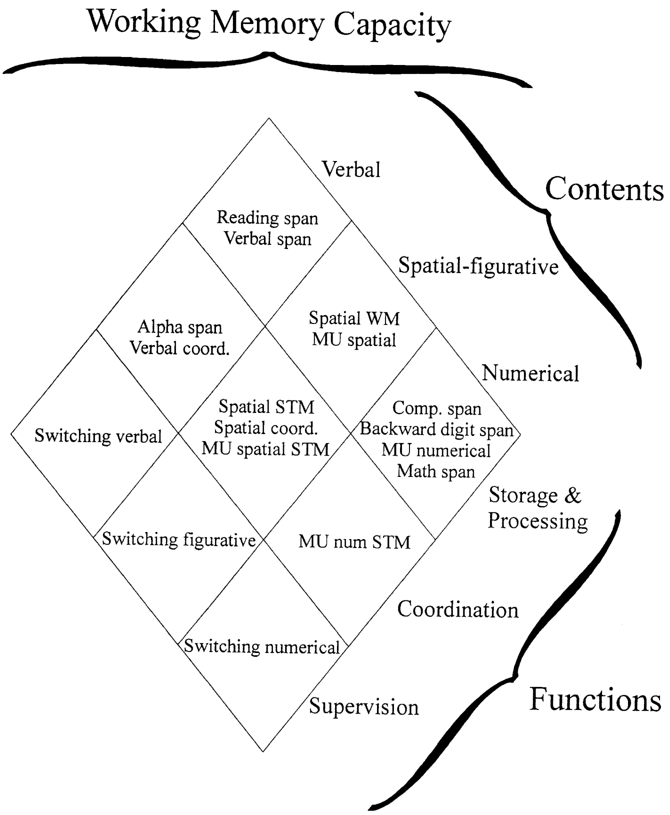

Fig. 2. Hypothetical two-facet structure of working-memory capacity. Three functional categories are cross-classified with three content categories. For abbreviations, see Table 1.

1991), the sequencing of mental operations (Hagendorf & Sá, 1995), and the integration of elements into new coherent structures (Oberauer, 1993). In this study, we focus on the last kind of coordination. By this, we mean the ability to integrate isolated pieces of information into new structures, e.g., mentally constructing a pattern from dots presented individually at different places, or constructing a word from letters presented in a random order. The integration of elements into new structures is required for the construction of mental models, which in turn are assumed to play an important role in different kinds of reasoning tasks (Gentner & Stevens, 1983; Johnson-Laird, 1983, 1994). Coordination requires simultaneous access to several distinct information elements in order to use them as elements in new relationships. We found the notion of coordination useful as a common denominator to describe certain tasks that did not require any obvious information manipulation but were successfully used to operationalize working-memory capacity (see descriptions in the Method section).

Exploratory and confirmatory factor analyses (CFA) performed on the working-memory data from this study showed that the differentiation between and along both facets was necessary to account for the covariance pattern. However, not all distinctions offered by the facet matrix were needed to fit the data. The distinctions between simultaneous storage and processing of information and coordination on the functional facet, as well as the distinction between numerical and verbal working memory on the content facet, did not seem to be necessary. For a description of these analyses, see Oberauer et al. (2000).

1.3. What distinguishes working memory from intelligence test tasks?

One major shortcoming in previous studies relating working memory and intelligence was that the tasks used to operationalize the two constructs were nearly indistinguishable. In the present study, we carefully selected working-memory tasks for which the overlap with intelligence test tasks of the same category, in particular with reasoning tasks, was minimized. This goal could only be approximated because there is no clear-cut, agreed-upon definition of the class of potential intelligence test tasks. In our view, "reasoning" refers to a relatively complex cognitive process consisting of an organized sequence of steps that must be appropriately scheduled in order to reach a goal. Process models for typical reasoning tasks such as analogies and syllogisms reflect this (cf. Sternberg, 1985). Working-memory tasks, in contrast, require only simple cognitive operations. The sequence of operations is highly restricted by the instructions and the task display; only a minimum amount of freedom remains for participants to decide by themselves what to do next. Thus, the processing requirements of working-memory tasks are not more demanding than those in typical speed tasks or in "elementary cognitive tasks." The relative difficulty of working-memory tasks arises from the additional load on some aspect of the cognitive architecture: short-term storage (in storage and processing tasks), simultaneous access to several distinct representations (in coordination tasks), or efficient inhibition of irrelevant action schemata (in task set switching paradigms). This additional requirement distinguishes working-memory tasks from speed tasks. Working-memory tasks differ from the memory test tasks in the BIS and other intelligence batteries in that the latter use mainly supra-span lists that are learned and remembered over time intervals falling mostly in the range of long-term memory.

One might argue that not every task fitting the above characterization is a good working-memory task because the concept of working-memory capacity itself is still developing. We view it as promising strategy in such a situation to choose a sample of well-established working-memory tasks from previous research to ensure, on the one hand, that the construct is represented in all its breadth, and on the other, that the tasks are valid indicators for the construct. Having shown that the proposed two-facet framework is indeed a good representation of the construct (Oberauer et al., 2000), we proceeded one step further in the present study by exploring the relationship between working memory and intelligence. More specifically, we propose working-memory capacity to be a very good predictor of *g* on the most general level and to explain most of the variances in reasoning ability on the "second stratum level" of intellectual abilities. On the other hand, working-memory capacity is not claimed to be the sole predictor of intelligence; therefore, a certain amount of reliable variance on the side of intelligence will remain unexplained by working memory.

2. Method

2.1. Participants

Participants in this study were 113 students and 15 university staff members, ranging from 18 to 46 years of age (mean = 26.2 years, S.D. = 5.0); 49 participants were female (38%). Ten persons indicated that German was not their mother tongue; they had been exposed to German for a range of 2–21 years, with a mean of 10 years. The participants were paid DM 80 (about US$40) for a total of 9 h of testing on two separate days. Eight additional individuals who could not be tested on the second day were excluded from the analysis.

2.2. Materials

2.2.1. Working-memory tasks

Working-memory tasks were chosen from previous research in the field and adapted as psychometric tools for group administration. The selection of tasks was guided by the following criteria:

1. The tasks should be distributed equally among all the postulated facets of working memory and, when possible, tap one content domain and one function exclusively.
2. Ideally, the only relationship between the working-memory tasks and the reasoning tasks should be their common demands on working-memory capacity. Any other relationships should be minimized. Reasoning tasks were classified into four categories based on the cognitive processes they required: inductive reasoning, deductive reasoning, construction, and planning. (The reasoning tasks included in the test for the BIS fell into the first three categories.) Care was taken that no working-memory tasks contained demand characteristics of any of these categories.
3. Furthermore, our working-memory tasks were based on very simple elements and cognitive operations, with limited degrees of freedom for the person taking the test to minimize the need for strategies and special knowledge as much as possible.

In the following, we briefly describe the working-memory tasks used in the study, sorted according to the function category; the character in parenthesis (V for verbal, N for numerical, S for spatial–figurative) indicates the content category. More detailed descriptions, including reasons for the selection and classification of each task, can be found in Oberauer et al. (2000). Unless otherwise stated, two practice trials preceded the test trials.

2.2.1.1. Simultaneous storage and processing tasks

Backward digit span (N). Four to eight digits were presented successively on the screen for 1 s each. Participants had to repeat the series in reversed order, immediately following presentation, by typing their responses into free blanks that appeared on the screen such that only responses in the reversed order were possible; the number of blanks matched the number of digits that had been presented. Fifteen test items were presented; each digit span (ranging from four to eight digits) was presented three times, increasing in length. The number of correctly recalled digits in the correct position was the score for each item; the scale score was the sum of the scores for the 15 items.

Reading span (V) and computation span (N) (Daneman & Carpenter, 1980; Turner & Engle, 1989). Reading span is the task most frequently used for psychometric assessment of working-memory capacity. Because we tried to minimize all demands not related to working memory, our version differed in some details from those used by other research groups.

In the reading span task, sentences were presented successively on the screen for 3 s each, followed by a 1-s pause until the next sentence appeared. Within the 4-s interval, participants had to rate the sentence as "true" or "false" by pressing a key labeled accordingly. Meanwhile, the last word of each sentence had to be memorized. After a series of sentences, these words had to be written down in their order of presentation. The sentences were between four and seven words long, and they were all trivially true or false, to avoid individual differences in knowledge contributing to performance. The last words were all familiar nouns with one to three syllables. Within an item, the last words of all the sentences had no apparent semantic relation to each other.

Computation span is a parallel task with numerical content. Instead of sentences, participants had to verify simple addition and subtraction equations with one- and two-digit numbers. The results of the equations in each sequence had to be remembered and written down in the correct order after the end of the sequence. Presentation time was the same as for reading span.

Each sequence length was presented three times; the length of the lists increased from three to seven sentences for reading span, and from four to eight numbers for computation span. Two practice trials and 15 test trials were presented for each task. As in Daneman and Carpenter (1980) and previous studies using these tasks, scoring was based exclusively on memory performance.

Spatial working memory (S). This task was constructed as a spatial equivalent to reading span. Eight items with two patterns and six items with three patterns were presented. At the beginning of each item, the instruction to rotate the pattern 90° to the left or to the right appeared on the screen. A series of 3×3 matrix patterns was then presented sequentially for 3 s each, followed by a 200-ms interstimulus interval. Participants had to mentally rotate the patterns according to the instruction and remember the resulting patterns. After two or three consecutive patterns, they drew the results on a row of blank 3×3 matrices in the correct order. Scoring was analogous to reading span.

Verbal span (V) and math span (N). To represent the "dual-task" constellation frequently used in experimental studies, two parallel tasks were constructed. Two unrelated tasks — one requiring storage, the other processing of information — were combined. A list of words (verbal span) or digits (math span) that were presented simultaneously on the screen had to be recalled. Presentation time was 1000 ms times the number of elements in the list. Following presentation of the list, one, two, or three successive decision tasks had to be performed quickly. In the verbal task, four words were presented, one in each corner of the screen, and a fifth word in the center. Participants had to select the word that was a subordinate concept of the word in the center (e.g., "table"–"furniture"). In the numerical task, two-digit numbers were displayed in the corners and a one-digit number in the center. The target was the number that could be added to the number in the center so that the result was divisible by five. Participants chose from the four possibilities by typing the number (1–4) shown in front of the correct answer. After completing the decision task, participants had to repeat the initial list in correct order. Word lists were written in blanks on the answer sheet, whereas numbers were typed into blanks on the computer screen. Two scores were derived from these tasks: the number of elements recalled correctly, regardless of decision accuracy (memory), and reaction times for correct decisions (RT).

Memory updating (N and S) (adapted from Salthouse, Babcock, & Shaw, 1991). Both the numerical and the spatial versions of this task were used. The screen was divided by a 3 × 3 matrix into nine cells. The number of "active" cells in the matrix varied for each item; the rest of the nine cells were shaded gray. In the numerical version, numbers were displayed one at a time in each of the active cells. Instructions to perform computations like "+5" or "−3" then appeared successively in selected cells. Each computation has to be performed on the current value of the cell, and the result has to be remembered as the new value for this cell. After a number of computations, a question mark appeared in one cell selected at random. The current value of this cell had to be typed as the answer. Two or three cells (depending on the number of "active" cells) were tested for each item.

In the spatial version, dots appeared in the "active" cells at one of nine possible positions within a cell. Participants had to memorize the positions of the dots and then move the dots mentally from one position to another within the cell, according to the arrows appearing in the cell. They then had to memorize the new position. After a number of successive arrows, question marks appeared in two or three cells to test the dot position. The final dot position reached in the cell was indicated by pressing the appropriate key of the number pad on the right of the keyboard, to which the nine possible dot positions were affixed.

Each version of the task had 15 test items. Stimulus duration for the initial elements and the operation instructions was 1300 ms each.

2.2.1.2. Coordination tasks

Memory updating (short-term memory version) (N and S). Both the numerical and the spatial versions of memory updating were presented a second time without the required operations. This reduced the task demand to one of storage without processing. Since the memory elements are arranged in a two-dimensional array, and must be retrieved from their position in this array, we believe that, in addition to storage, the task requires coordination in order to organize the working-memory contents. If we accept Daneman and Carpenter's (1980) argument that a good working-memory indicator needs a processing component, we would expect this version of memory updating to show a considerably weaker correlation with reasoning ability than the full version would. We hypothesized instead that the demand on coordination ability involved in memory updating would suffice to make the short-term memory version an equally good predictor of reasoning ability.

Fifteen spatial items (5 with three active cells and 10 with five active cells) and 15 numerical items (5 with six active cells and 10 with nine active cells) were presented. Directly following the presentation of the last elements within the "active" cells, three cells selected at random were probed.

Alpha span (V) (Craik, 1986). Three to seven words were presented successively for 1 s each. Following the appearance of the last word in the sequence, the first letter of every word had to be repeated in alphabetical order. Responses were typed into free blanks on the screen. Two parallel blocks of nine items were constructed. The blocks consisted of one item with three words and two items with four, five, six, and seven words each. Scores were computed in the same way as for backward digit span.

Spatial short-term memory (S) (Oberauer, 1993). The screen was divided by a 10 × 10 matrix into cells. Dots appeared sequentially for 1 s in different cells. Participants had to remember the relative positions of the dots, and reproduce the dot pattern by placing crosses into the cells of an empty matrix on the answer sheet. Participants were told that the relative positions of the crosses, and not the absolute positions, would be scored. This was to emphasize that the task requirement was to coordinate the dots into patterns rather than to remember each dot position individually. The 15 items had from two to six dots, with the number of dots increasing by one every third item.

Spatial coordination (S) (Oberauer, 1993). This task is an expansion of spatial short-term memory. Before placing their crosses on the empty matrix, participants had to decide whether the pattern of dots that would emerge, if all the dots were to appear simultaneously, would be symmetrical along a vertical axis or not. Participants indicated their decision by giving their response in a matrix on the left or right side of the answer sheet. The decision step was introduced to further enforce the coordination of dots into a single structure. The scoring procedure was the same as for spatial short-term memory, irrespective of the correct decision.

Verbal coordination (V) (Oberauer, 1993). A number of empty square cells appeared in a row on the screen, and letters appeared one by one in the cells in random order until there was a letter in each cell. Presentation time was 1 s per letter. The letters in all cells had to be remembered and combined mentally into a sequence according to their position in the row. The letters had to be written into one of three different empty rows in the answer sheet, depending on whether the sequence formed a word forwards, backwards, or no word at all. The number of correct letters at the correct location in the row was the score for each item, irrespective of whether the word decision was correct.

2.2.1.3. Supervision tasks

Switching (N, V, and S). Three versions of this task were used, one numerical, one verbal, and one figurative (Allport, Styles, & Hsieh, 1994; Zimmermann & Fimm, 1993). The numerical version was designed according to Allport et al. (1994). Successive displays on the computer screen had to be responded to quickly. Each display contained a varying number of digits arranged at random. Participants had to alternate between reading the digits and counting the number of digits in the display. A short reminder about which kind of response was to follow was displayed on the top part of the screen during each trial. Responses were typed on the number pad of the keyboard. The correct answer varied between one and nine, and the value of the digits was never equal to the number of digits. The time between a response and the next stimulus was 400 ms. After 10 practice trials, 30 test trials were conducted.

The figurative and verbal versions were based on Zimmermann and Fimm (1993). Again, the required task was to alternate between two types of responses to a sequence of displays. In the figural version, two geometrical figures appeared in each display, one left and one right. One figure was angular, the other was rounded. Participants had to respond "left" or "right" to indicate on which side the angular or the rounded figure appeared. The figures were selected at random from a pool of nine angular and nine rounded figures, and their placement was randomly determined. In the verbal version, two words belonging to two prespecified categories appeared instead of the figures. The participants had to indicate on which side the word belonging to the given category was. In both tasks, a one-word reminder indicating the next decision criterion was displayed on the top of the screen. The interval between response and the next stimulus was 300 ms. After 10 practice trials, 40 test trials were presented for each version.

2.2.1.4. Tasks excluded from the data analysis. Seven additional working-memory tasks were used in this study, but excluded from the data analysis in response to a reviewer's critique of an earlier version of this paper. These tasks could be interpreted as being similar to reasoning test tasks and, in one case, to verbal fluency tasks. The complete test battery, including these seven tasks, was analyzed for a second paper testing different theoretical models of working memory (Oberauer et al., 2000).

The working-memory tasks excluded were *Spatial Integration* (S) (Salthouse & Mitchell, 1989), *Multi-Element Visual Tracking* (S) (Pylyshyn & Storm, 1988), *Pattern Transformation* (S) (Mayr & Kliegl, 1993), *Random Generation* (S) (Baddeley, 1986), *Category Generation* (V, switching version) (Baddeley, 1986), *Star-Counting Test* (N) (Das-Smaal, de Jong, & Koopmans, 1993), and the *Gauss Task* (N) (Schmuck, 1996).

2.2.2. Intelligence test

A comprehensive test battery for the BIS consisting of three tasks for each of the speed and memory cells of the model, four tasks representing each creativity cell, and five tasks for each of the reasoning cells was used (Jäger et al., 1997). Each task assigned to a cell in the model is used to measure one operational as well as one content component (see Fig. 1). Each of the four operational components is measured by a scale consisting of 9–15 tasks from the three cells corresponding to the content classes. Analogously, each of the three content components is measured by tasks from the four cells corresponding to the four operational classes. The scores for the operational components and those for the content components are not statistically independent. The score for "general intelligence" can be computed by aggregating all items.

In the following, we give a short description of a representative task for each cell of the BIS, and a complete description of the reasoning tasks, which are of special interest in the present work. The task descriptions are sorted according to the operational facet, and the content facet is indicated by the character in parenthesis (V for verbal, F for figural, and N for numerical).

2.2.2.1. Reasoning (N). Number sequences and letter sequences: The rule governing a given sequence of numbers or letters had to be detected, and the next two elements in the sequence produced. *Reading tables:* Questions had to be answered based on computations on data in a frequency table. *Estimation:* The correct solutions for equations involving large numbers had to be selected from five alternatives on the basis of computational laws (e.g., that the sum of two even numbers has to be even). *Computational reasoning:* Mathematical text problems had to be solved.

2.2.2.2. Reasoning (F). *Figural analogies:* Analogies of the form A:B = C:?, in which the elements were geometric patterns, had to be completed. *Charkow:* Series of abstract drawings had to be completed according to the rule underlying the sequence. *Bongard:* The distinguishing features of two sets of six geometrical patterns each had to be detected. Three new patterns had to be classified into the two sets. *Figure assembly:* Out of five pictures, the geometrical picture that could be formed by rearranging a given set of pieces had to be selected. *Surface development:* The two-dimensional surface of a three-dimensional object was presented, together with five objects shown in a three-dimensional perspective. The object that could result from folding the surface had to be selected.

2.2.2.3. Reasoning (V). *Verbal analogies:* Word analogies had to be completed. *Fact–opinion:* Sentences had to be judged as to whether they stated a fact or an opinion. *Senseless inferences:* Syllogistic inferences from absurd premises had to be evaluated. *Syllogisms:* The logical validity of deductions from statements about everyday matters had to be evaluated, irrespective of their truth or plausibility. *Word knowledge:* The one word not belonging to a set of four words had to be crossed out. The words of successive items were increasingly unfamiliar ones.

2.2.2.4. Creativity. *Masselon:* As many different sentences as possible using three given words had to be written down (V). *Drawing objects:* As many different objects as possible using four given geometrical elements had to be drawn (F). *Telephone numbers:* As many different six-digit telephone numbers had to be written down. The numbers had to follow different principles that could help in remembering them (N). For all creativity tasks, scoring was based on the number of different categories of ideas produced, or on the number of ideas produced.

2.2.2.5. Speed. *Part–whole:* In a list of words, every word immediately following a word to which it stood in a part–whole relation had to be crossed out (e.g., tree–leaf) (V). *Old English:* All letters written in the type Old English had to be crossed out (F). *X-larger:* All numbers in a row that were larger than their predecessors by a given amount had to be crossed out (N).

2.2.2.6. Memory. *Paired associates:* A list of pairs of three-digit numbers had to be learned within 2 min. At retrieval, the first number of each pair was given in a different order, and the second number had to be recalled and written down (N). *City map:* Buildings marked black on a city map had to be memorized and then crossed out on an unmarked map (F). *Story:* As much information as possible from a short story had to be memorized. At retrieval, questions about details had to be answered (V).

2.2.3. Control measures

To statistically control for possible personality factors and computer skills that might affect performance scores on either working memory or intelligence task test, we included the following scales:

(a) The subscales for achievement motivation and susceptibility to stress from the Freiburg Personality Inventory (FPI-R; Fahrenberg, Hampel, & Selg, 1989);
(b) the conscientiousness subscale from the German version of the NEO-FFI (Borkenau & Ostendorf, 1994);
(c) a newly constructed questionnaire in the form of a semantic differential, containing six items assessing subjective stress and fatigue;
(d) a questionnaire from a prior study about complex problem solving to assess experience and familiarity with computers (Süß, Kersting, & Oberauer, 1993);
(e) a test of typing speed, in which participants had to type 75 characters or numbers appearing successively on the screen as fast as possible.

2.3. Procedure

The tests were administered to groups of 3–17 participants on 2 days, in sessions lasting $4\frac{1}{2}$ h. The intelligence and working-memory tasks were grouped into sets requiring between 20 and 40 min. One intelligence test set was combined with one working-memory set to build a task block. Four task blocks were separated by short 10-min breaks on each day. The other tests and questionnaires were interspersed among the task blocks. The stress questionnaire was presented five times on the first day (at the beginning of testing and after each of the four task blocks) and twice on the second day (at the beginning and at the end of testing). The experimenter read the instructions for each task aloud. After the experimenter explained the examples and answered questions, participants worked on their own. There were time limits for the intelligence tasks but not for the working-memory tasks (except as otherwise indicated in the task description in the Materials section). Participants who completed a task before the rest of the group did were asked to remain silent until the others had finished. All the working-memory tasks were presented on computer; the responses for some of the tasks were written on an answer sheet. The intelligence test was a pure paper-and-pencil test.

3. Results

In the first analysis, scale scores from the BIS tasks were formed according to the instructions given in the test manual. Scale scores for the operational factors were built by summing up the z standardized task scores within one operational category over all three content categories. Content scales were built analogously. The seven resulting scales and the general intelligence scale were correlated with the 17 working-memory tasks (see Table 1). Table 1 also includes the reliability estimates for the working-memory tasks (Cronbach's α).

Table 1
Intercorrelations of intelligence scale scores and correlations of working-memory tasks with intelligence

	S	C	M	R	V	F	N	BIS *g*	Reliability
Speed									.88
Creativity	.48								.88
Memory	.49	.22							.87
Reasoning	.61	.41	.51						.90
Verbal	.70	.57	.57	.66					.90
Figural	.70	.56	.64	.63	.54				.88
Number	.70	.53	.61	.75	.49	.55			.90
General intelligence	.85	.66	.73	.82	.81	.83	.84		.95
Switching numerical	−.59	−.30	−.38	−.50	−.45	−.47	−.52	−.58	.78
Switching verbal	−.49	−.26	−.28	−.33	−.41	−.45	−.27	−.45	.94
Switching figurative	−.36	−.26	−.20	−.19	−.28	−.37	−.18	−.33	.94
Backward digit span	.39	.39	.29	.51	.31	.40	.55	.51	.81
Reading span	.48	.30	.48	.56	.53	.41	.52	.59	.84
Computation span	.32	.29	.34	.53	.35	.32	.51	.48	.85
Spatial working memory	.26	.23	.38	.57	.21	.49	.46	.47	.86
Verbal span	.35	.34	.41	.45	.44	.40	.41	.51	.82
Math span	.32	.23	.35	.40	.27	.32	.45	.43	.79
MU numerical STM	.28	.26	.31	.40	.23	.32	.44	.40	.66
MU numerical	.50	.35	.41	.67	.46	.47	.61	.63	.81
MU spatial STM	.31	.26	.44	.58	.28	.51	.49	.51	.83
MU spatial	.36	.35	.44	.58	.26	.54	.58	.56	.85
α span	.41	.31	.39	.57	.41	.31	.61	.55	.81
Spatial STM	.31	.28	.45	.59	.29	.47	.54	.53	.75
Spatial coord.	.38	.28	.48	.63	.31	.53	.59	.58	.75
Verbal coord.	.42	.28	.52	.51	.46	.42	.51	.57	.84

Reliabilities are presented for the BIS scales from the standardization sample in Jäger et al. (1997), and Cronbach's α for the working memory tasks. S = speed; C = creativity; M = memory; R = reasoning; V = verbal; F = figural; N = numerical. Comp. = Computation; WM = working memory; MU = memory updating; STM = short-term memory; Spat. = spatial; Coord. = coordination.

The data in Table 1 show that the reliability of the working-memory tasks is satisfactory. The working-memory scales correlate strongly with the intelligence scales, particularly reasoning and *g*. Particularly interesting is that those tasks that do not require manipulation of information (spatial short-term memory, spatial coordination, and the two short-term memory versions of memory updating) are no less related to reasoning than the classical "storage and processing" tasks (e.g., reading span, math span, or the memory-updating tasks with updating operations). This shows that a good predictor of complex cognitive performance need not necessarily be a combination of storage and processing demand.

This evidence is contrary to the hypothesis that a combination of storage and manipulation demands is necessary to predict complex cognitive performance.

We further investigated the relationship between working memory and intelligence by using confirmatory tools of analysis (Amos [computer program]; Arbuckle, 1997). First, we performed CFA for intelligence and working memory separately. Then, using structural equation models, we predicted the intelligence factors from the working-memory factors. The direction of prediction in the models reflects our hypothesis that working memory is a causal factor for intelligence and not the other way around (although a correlational study like this one cannot decide between these alternatives).

Because we measured both working memory and intelligence with a broad selection of different tasks, we decided to aggregate our data. We did this for two reasons. Our general strategy was to aggregate tasks classified into the same category of the facet taxonomy, in other words, according to a theory-based classification that had been supported empirically in previous factor-analytic studies (summarized in Jäger et al., 1997, for the BIS, and in Oberauer et al., 2000, for the working-memory tasks). Theory-guided aggregation emphasizes wanted variance and suppresses unwanted variance, especially task-specific variance (Wittmann, 1988). Second, in consideration of our rather small sample, we aggregated our data to keep the number of estimated parameters within a reasonable range.

The data were aggregated in three ways: (a) aggregation across contents within one functional category (“operational” category in the BIS), which led to parcels that suppressed content variance as part of the unwanted variance, leaving mainly functional (or operational) variance; (b) aggregation across functions within one content category, which resulted in parcels that brought content variance to the fore, but let functional variance count as unwanted; and (c) aggregations within the cells of the two-dimensional matrix used to structure working memory and intelligence, respectively; here both content and functional variance were wanted variance, and only task-specific variance was suppressed. Aggregated variables were the means of z-transformed task scores.

Fig. 2 shows how the functional and the content parcels were built from the pool of working-memory tasks. Because the numbers of variables in the nine cells were unequal, in some cases, one parcel includes the mean of two variables, and two parcels contain no variable representing numerical coordination. These difficulties did not occur with the BIS test data because the test contained sufficient variables for each cell of the matrix.

3.1. Confirmatory factor analysis

Two CFA were performed on the intelligence test variables, one with the operational and one with the content parcels. The factors were the four operational and three content factors postulated in the BIS (see Fig. 1). General intelligence was included as a second-order factor. As can be seen in Table 2, both models had satisfactory fits. The models are presented[1] in Figs. 3 and 4.

Similarly, two CFA were computed with the parcels built from the working-memory tasks. The model with the function parcels had one factor for coordination together with simultaneous storage and processing, and a second factor for supervision. In all models, the three items classified as measuring supervision were not aggregated because this would have resulted in only one single parcel. To ensure identification of the supervision factor, we

---

[1] In all figures presented here, numbers placed at a corner of an indicator or an oval representing a factor stand for variance explained (R^2). All estimated path coefficients and covariances are standardized.

Table 2
Summary of fit statistics for CFA

Model	χ^2	*df*	*P*	B–S *P*	CFI	AGFI	RMSEA	LO90	HI90
(1) BIS (OP)	105.45	86	.076	.16	.98	.86	.04	.00	.07
(2) BIS (CO)	46.33	24	.004	.009	.96	.85	.09	.05	.12
(3) WMC (OP)	15.60	12	.210	.270	.99	.93	.05	.00	.11
(4) WMC (CO)	17.98	8	.021	.044	.98	.88	.10	.04	.16

OP = operations; CO = contents; *P* = *P* value of the χ^2 statistic; B–S *P* = Bollen–Stine bootstrapped *P* based on 1000 resamples; CFI = comparative fit index; AGFI = adjusted goodness of fit index; RMSEA = root mean square error of approximation, LO90 = lower bound of the 90% confidence interval for RMSEA; HI90 = upper bound of the 90% confidence interval for the RMSEA.

used the item scores—at the risk of having to model unwanted item-specific variances. Only the verbal and numerical switching items were of the same type so we estimated the covariance between the error terms of these two items to account for item-specific covariances not explained by the supervision factor. We did not separate the two categories, storage and processing and coordination, because comprehensive factor analyses (Oberauer et al., 2000) revealed that these categories were indistinguishable in the present data. Analogously, the verbal and numerical parcels loaded on the same factor in the content model of working memory because our previous analyses showed that these two categories did not form distinct factors. Because there were only two primary factors in the two working-memory models, no second-order factor could be established. Instead, we let the two factors correlate freely to account for the common variance among all working-memory

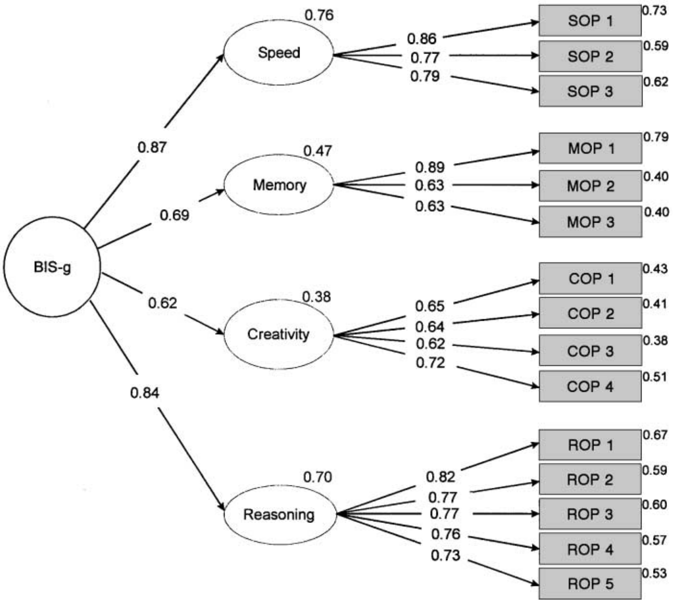

Fig. 3. CFA of the BIS (operational facet). SOP = speed parcel; MOP = memory parcel; COP = creativity parcel; ROP = reasoning parcel; d = disturbance term.

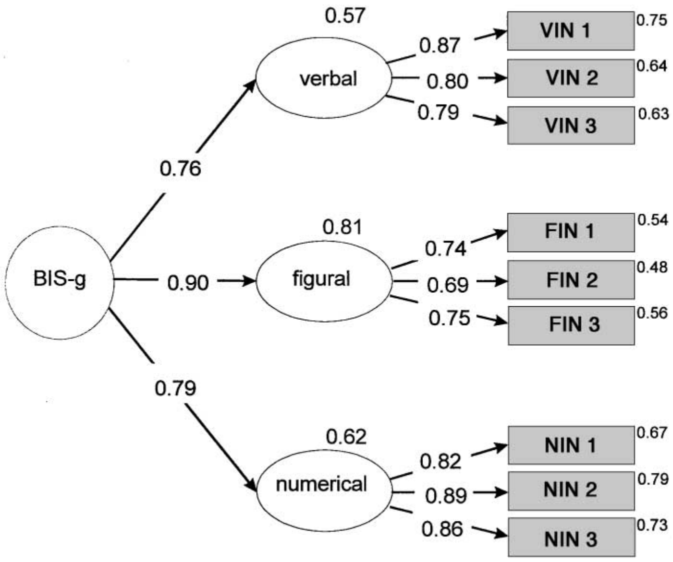

Fig. 4. CFA of the BIS (content facet). VIN = verbal parcel; FIN = figural parcel; NIN = numerical parcel; d = disturbance term.

variables. The two factor models, presented in Figs. 5 and 6, had satisfactory fits, as summarized in Table 2.

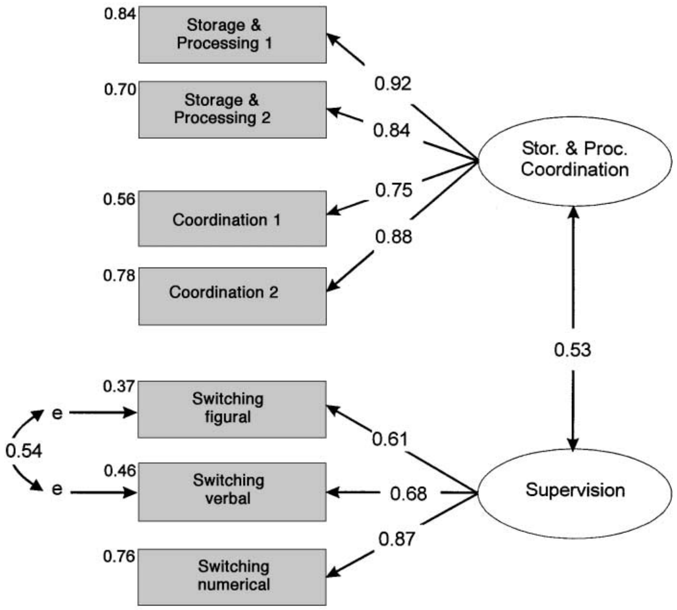

Fig. 5. CFA of working-memory structure (operational facet). e = error term.

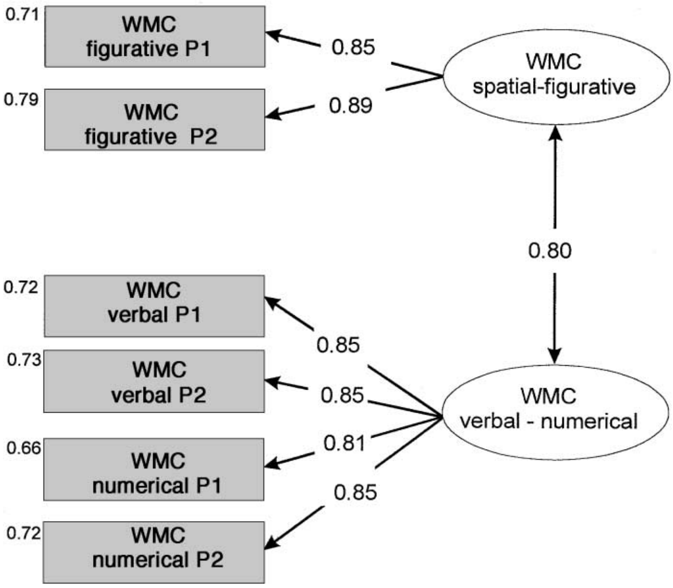

Fig. 6. CFA of working-memory structure (content facet). WMC = working-memory capacity; P = parcel.

3.2. Structural equation models

In the next step, we investigated confirmatory models for the relationship between working memory and intelligence. Models 1 and 2 relate the two constructs at the highest level of aggregation. Thus, the two working-memory factors predict intelligence exclusively through the second-order factor of general intelligence. The function-specific parcels were used for Model 1 (Fig. 7), and the content-specific ones for Model 2 (Fig. 8).

Both models have acceptable, but far from optimal, fits. Summaries of the goodness-of-fit statistics are presented in Table 3. A strong relationship between working memory and general intelligence is apparent from the high path coefficients relating the latent factors.

The next two models focus on a lower level of aggregation. The second-order factor of intelligence was replaced, and the two working-memory factors were related to each of the primary factors of intelligence. Model 3 differentiates working memory and intelligence along the functional dimension; Model 4 differentiates them along the content dimension. The correlations between the disturbance terms in both Models 3 and 4 are freely estimated because we do not propose that all variance common to the primary factors of intelligence can be explained by working-memory capacity.

Model 3 (see Fig. 9), which relates working memory to intelligence on a more specific level, shows a slightly better overall model fit than Model 1, as evidenced by the fit statistics reported in Table 3 (e.g., CFI = .97 as compared to CFI = .95). Similarly, Model 4 (see Fig. 10), which relates the content factors on a specific level, does show an increase in overall fit of the model as well, though all models relating content factors show a somewhat inferior fit compared to models relating operation factors. Tentatively, one can conclude that Models 3

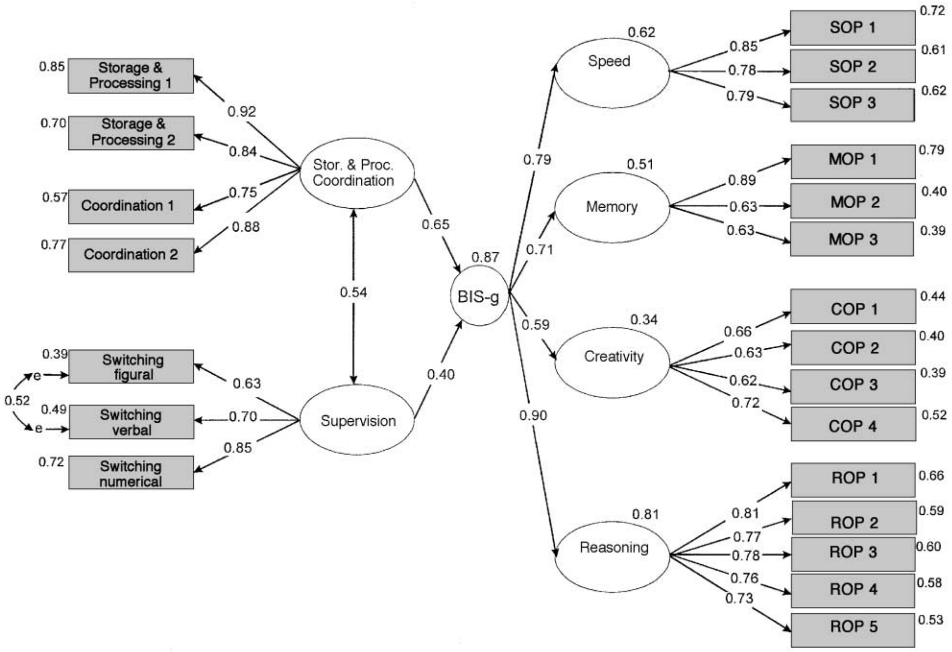

Fig. 7. Structural equation model relating working memory and intelligence with *g* at the apex of the intelligence hierarchy (operations). For abbreviations, see previous figures.

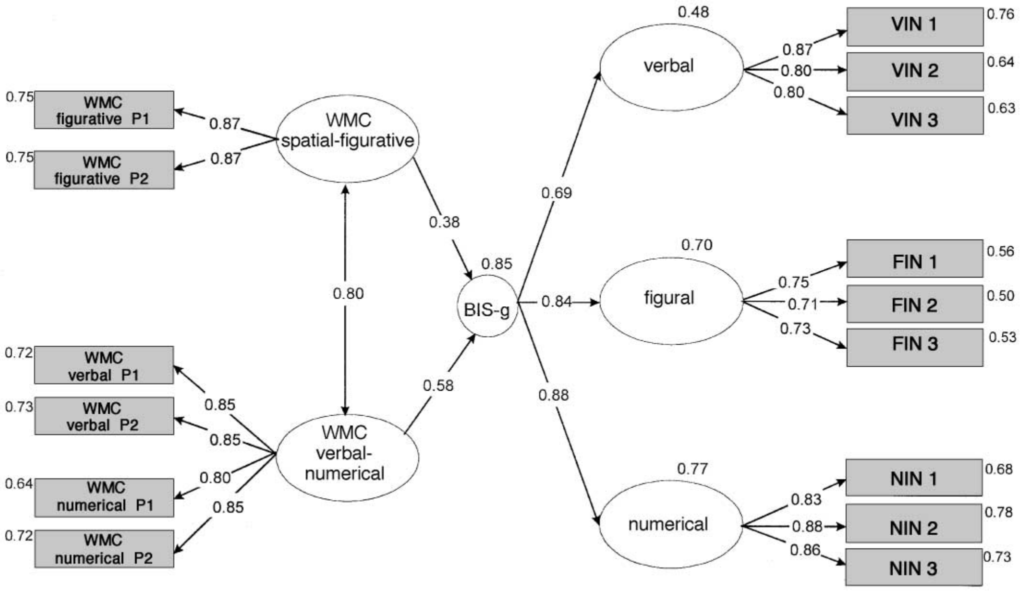

Fig. 8. Structural equation model relating working memory and intelligence with *g* at the apex of the intelligence hierarchy (contents). For abbreviations, see previous figures.

Table 3
Summary of fit statistics for structural equation models

Model	χ^2	*df*	*P*	B–S *P*	CFI	AGFI	RMSEA	LO90	HI90
(1) WMC — intelligence *g* (OP)	265.02	201	.002	.098	.96	.81	.05	.03	.07
(2) WMC — intelligence *g* (CO)	163.39	84	.000	.001	.94	.78	.09	.07	.11
(3) WMC — intelligence group factors (OP)	233.53	193	.025	.232	.97	.82	.04	.02	.06
(4) WMC — intelligence group factors (CO)	134.61	80	.000	.012	.96	.81	.07	.05	.09

OP = operations; CO = contents; *P* = *P* value of the χ^2 statistic; B–S *P* = Bollen–Stine bootstrapped *P* based on 1000 resamples; CFI = comparative fit index; AGFI = adjusted goodness of fit index; RMSEA = root mean square error of approximation, LO90 = lower bound of the 90% confidence interval for RMSEA; HI90 = upper bound of the 90% confidence interval for the RMSEA.

and 4 show moderately improved fit and provide insight into how operational and content factors of working memory relate to parallel factors of more complex intellectual abilities and do, therefore, merit special consideration.

The path coefficients of the resulting models are in accordance with theoretical expectations and previous research. The reasoning factor of the BIS was strongly dependent on the

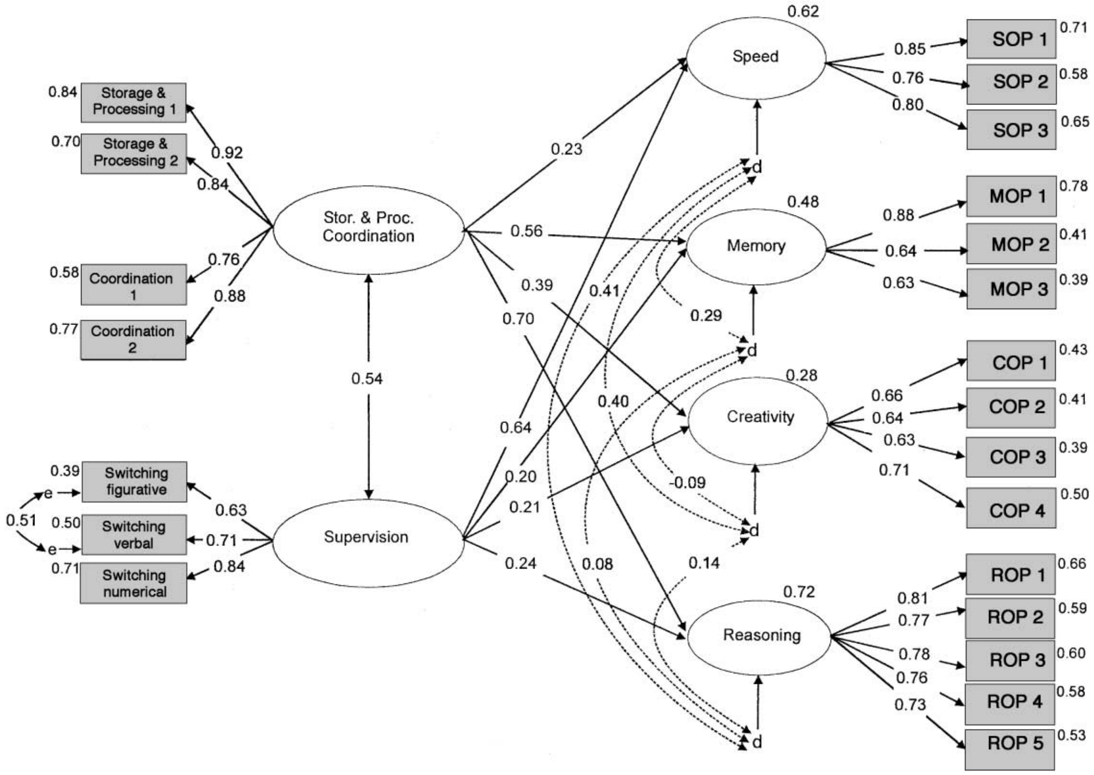

Fig. 9. Structural equation model relating working memory and intelligence group factors (operations). For abbreviations, see previous figures.

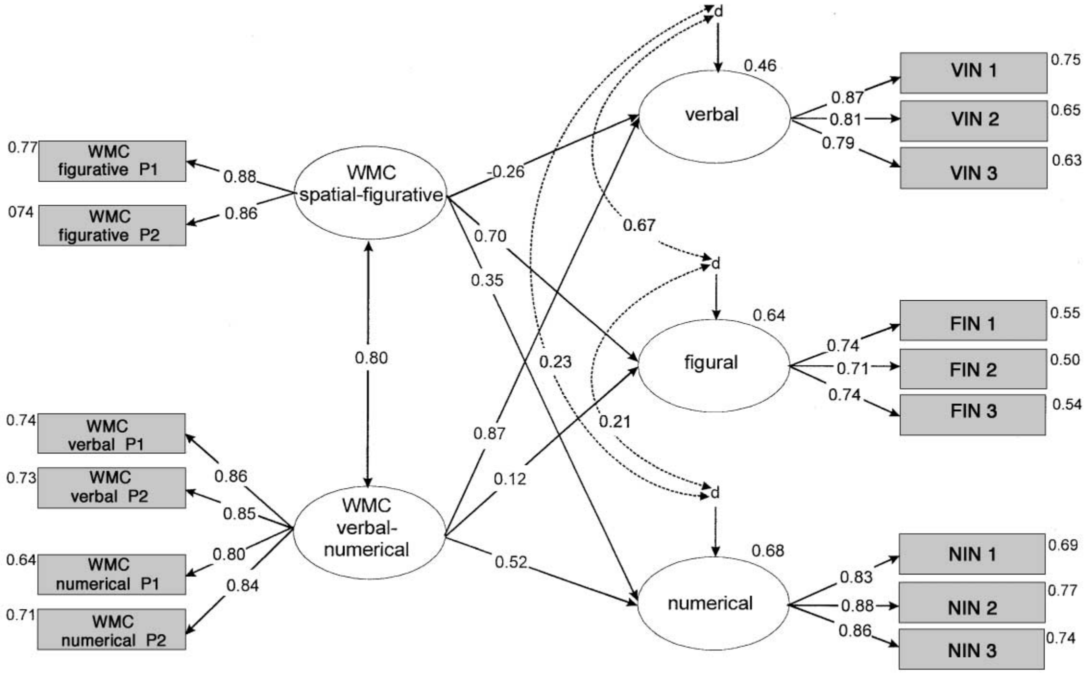

Fig. 10. Structural equation model relating working-memory operations and intelligence group factors (contents). For abbreviations, see previous figures.

storage and processing plus coordination factor in Model 3. This replicates the findings from Kyllonen (1994a) and Kyllonen and Christal (1990). The supervision factor also contributed to the variance of reasoning ability, but much less so than did storage and processing plus coordination. The reverse was the case for the speed factor of the BIS, which was predicted mainly by the supervision factor. Storage and processing plus coordination were also strongly related to the memory factor and the creativity factor of the BIS. In sum, Model 3 confirmed that working-memory capacity is an excellent predictor of both reasoning and speed, as the four-sources model (Kyllonen, 1994a) would predict. On the other hand, there were nonnegligible relations to other intelligence factors as well.

It should be noted that, from the six-factor intercorrelations in the intelligence model, three remained significant after working memory was introduced as predictor. This implies that the common variance of the intelligence factors cannot be accounted for completely by working memory. The three significant factor intercorrelations all involved the speed factor of the BIS. This might be because all the intelligence test tasks were speeded, whereas most working-memory tasks were not. Alternatively, it could indicate that the common variance of all intelligence factors, expressed in *g* or the second-order factor integrating all the primary factors, consists of a composite of variance associated with working-memory capacity and speed. After removing the common variance associated with working memory, the remaining factor intercorrelations reflect the common variance associated with speed.

In Model 4, the highest path coefficients appeared for corresponding content factors. Spatial–figural working memory was an excellent predictor for the figural factor of the BIS, and not related at all to verbal intelligence. Verbal–numerical working memory, on the other hand, was strongly related to verbal abilities, as measured in the BIS, and not at all to the figural factor. The numerical intelligence factor seems to depend on a mixture of verbal–numerical and spatial working memory. The results of a bootstrap analysis for Models 3 and 4 can be found in Table 5 in Appendix.

3.3. Control variables

Finally, we checked whether the relationship between working memory and intelligence could have been mediated by personality factors, subjective stress, or computer skills. The correlations of the control variables with the functional dimensions of working memory and the operational intelligence scales are presented in Table 4.

Personality factors were not significantly related to any of the ability scales employed in this study. A consistent pattern of correlations was obtained between the subjective stress at the end of the first day and the intelligence test scales. This pattern was not replicated on the second day. It should be noted that the intelligence scales consist of about equal numbers of tasks from the first and the second day. Subjective stress was not systematically related to working memory, so this factor cannot have inflated the relationship between the two constructs. The two indicators of computer skills, typing speed and the computer experience questionnaire, were moderately correlated with some intelligence and working-memory scales. The correlation between an aggregated score for general intelligence and an

Table 4
Correlations of intelligence and working-memory operations with control variables

	Intelligence				Working memory			
Control measures	C	M	S	R	S&P	Coord.	Sup.	Rel.
NEO-FFI conscientiousness	.00	−.02	.04	−.10	.00	−.01	.01	.81[a]
FPI achievement	.18*	.15	.17	.13	.13	.08	.15	.65[a]
FPI stress	−.11	−.06	−.14	−.06	.05	.03	−.08	.67[a]
Stress beginning, Day 1	−.04	−.14	−.15	−.11	−.11	−.04	−.13	.87[b]
Stress end, Day 1	−.19*	−.18*	−.21*	−.23*	−.16	−.12	−.10	.87[b]
Stress beginning, Day 2	.03	.00	−.07	.00	−.07	−.03	.07	.92[b]
Stress end, Day 2	−.06	−.13	−.20*	−.14	−.13	−.03	.00	.87[b]
Typing speed	−.12	−.15	−.34**	−.34**	−.17	−.21*	−.36**	.90[b]
Computer experience	.26**	.15	.18*	.31**	.15	−.18*	.11	.91[c]
Reliability	.88[a]	.87[a]	.88[a]	.90[a]	.89[b]	.83[b]	.82[b]	

C = creativity; M = memory; S = speed; R = reasoning; S&P = storage and processing; Coord. = coordination; Sup. = supervision; Rel. = reliability of measure.

[a] Standardized parcel α, taken from the respective test manual.

[b] Standardized item α.

[c] Retest–reliability reported in Süß (1996).

* $P < .05$.

** $P < .01$.

aggregated score for general working memory fell from .83 to .82 after partialling out typing speed and computer experience. These two variables therefore do not mediate the strong relationship between intelligence and working memory.

4. Discussion

The present study yielded three main results. (1) Working-memory capacity is highly related to intelligence. The strongest relationship found was to reasoning ability, thereby replicating results found by Kyllonen (1994a) and Kyllonen and Christal (1990). The working-memory tasks in our test pool were selected so that overlap in cognitive processes and strategies with the reasoning tasks was minimal. The common variance between the working memory and the reasoning tests therefore can hardly be attributed to something other than working-memory capacity. (2) The strong relationship between working memory and reasoning (and, more broadly, general intelligence) holds over a large range of different tasks, including dual-task combinations of unrelated memory and processing demands (e.g., verbal span, math span), tasks where the memory content itself must be manipulated (e.g., spatial working memory, memory-updating with operations), and tasks where the memory content is not manipulated at all, but must be coordinated into a new structure (e.g., spatial short-term memory). The relationship of working-memory tasks to complex intellectual abilities, therefore, does not seem to depend critically on any specific feature of a task. (3) The fit of structural equation models could be improved by relating working memory and intelligence on the level of specific primary factors. Moreover, the paths from working memory to intelligence factors appear to be highly specific. This suggests that specific working-memory resources, as opposed to a general capacity, are the limiting factors for their corresponding counterparts in the structure of mental abilities. The two-facet structure postulated here seems to be a fruitful, general framework for classifying cognitive abilities. A similar proposal was developed by Kyllonen (1994a,b).

At present, working-memory capacity is the best predictor for intelligence that has yet been derived from theories and research on human cognition. Components of intelligence process models cannot account for comparable amounts of variance in general mental abilities (cf. Lohman, 1994). Indicators of mental speed, like RT and inspection time, also failed to predict intelligence test scores with the same success (for reviews, see, e.g., Juhel, 1991; Kranzler & Jensen, 1989; Neubauer, 1997).

The strong relationship between working memory and intelligence paves the way for a better understanding of psychometric ability constructs through theories of cognition. Establishing this general association, however, is only the first step. Working memory itself is not a precisely defined construct. It is widely accepted that working-memory capacity is an important limited resource for complex cognition; however, which functions of working memory affect which part of the cognitive process in a given reasoning task is not well understood. What is critical for successful performance in reasoning tasks? Is it storage capacity, processing capacity, the combination of both, or something beyond storage and

Table 5
Bootstrap results for standardized parameter estimates in Models 3 and 4 in Table 3

Model 3			Model 4		
Parameter estimate	Mean	S.E.	Parameter estimate	Mean	S.E.
Speed — S&P/Coord.	.225	.106	Verbal — WMCVN	.879	.162
Memory — S&P/Coord.	.561	.098	Numerical — WMCVN	.526	.144
Creativity — S&P/Coord.	.382	.129	Figural — WMCVN	.129	.159
Reasoning — S&P/Coord.	.703	.071	Verbal — WMCF	− .268	.179
Speed — supervision	.644	.095	Numerical — WMCF	.338	.157
Memory — supervision	.201	.104	Figural — WMCF	.694	.148
Creativity — supervision	.214	.143	VIN1 — Verbal	.863	.036
Reasoning — supervision	.229	.101	VIN2 — verbal	.808	.039
SOP1 — speed	.844	.032	VIN3 — verbal	.794	.034
SOP2 — speed	.760	.043	FIN1 — figural	.737	.057
SOP3 — speed	.802	.050	FIN2 — figural	.704	.066
MOP1 — memory	.880	.046	FIN3 — figural	.732	.064
MOP2 — memory	.639	.067	NIN1 — numerical	.827	.035
MOP3 — memory	.626	.067	NIN2 — numerical	.879	.032
COP1 — creativity	.653	.069	NIN3 — numerical	.859	.033
COP2 — creativity	.636	.077	WMCF-P1 — WMCF	.872	.034
COP3 — creativity	.621	.075	WMCF-P2 — WMCF	.859	.037
COP4 — creativity	.707	.070	WMCV-P1 — WMCVN	.860	.028
ROP1 — reasoning	.805	.044	WMCV-P2 — WMCVN	.855	.033
ROP2 — reasoning	.766	.043	WMCN-P1 — WMCVN	.795	.033
ROP3 — reasoning	.774	.041	WMCN-P2 — WMCVN	.842	.036
ROP4 — reasoning	.759	.040	WMCF — WMCVN	.794	.050
ROP5-reasoning	.728	.048	dFigural–dNumerical	.220	.196
S&P1 — S&P/Coord.	.915	.023	dVerbal–dNumerical	.219	.150
S&P2 — S&P/Coord.	.835	.029	dVerbal–dFigural	.692	.146
Coord1 — S&P/Coord.	.757	.044			
Coord2 — S&P/Coord.	.875	.026			
Switch-F — supervision	.628	.075			
Switch-V — supervision	.713	.058			
Switch-N — supervision	.841	.070			
S&P/Coord. — supervision	.537	.101			
dCreative–dReasoning	.146	.152			
dMemory–dReasoning	.070	.169			
dSpeed–dReasoning	.412	.178			
dMemory–dCreativity	− .091	.136			
dSpeed–dCreativity	.411	.151			
dSpeed–dMemory	.283	.146			
eSwitch-F–eSwitch-V	.505	.092			

The number of resamples for both models was 1000. Parameter estimates between combinations of factors correspond to path coefficients between factors as indicated in Figs. 9 and 10. Combinations of parcels and factors correspond to loadings, whereas combinations of variables preceded by a "d" or an "e" correspond to correlations between disturbances and error terms, respectively. Indicators and factors are labeled as in Figs. 9 and 10, with S&P/Coord. = storage, processing, and coordination; WMCVN = verbal/numerical working memory; WMCF = spatial–figurative working memory.

processing, like coordination of mental contents or supervision of mental processes? What crucial steps in the solution of a reasoning problem are limited by these functions? Now that the relationship between working memory and intelligence has been established on a molar level, further research with more fine-grained analyses needs to be done.

Acknowledgments

We thank Günter Trendler and Raphael Vogel for their help in collecting the data and Markus Werz and Heinz Wöbken-Blachnik for programming part of the working-memory task battery. This research was supported by Grant Wi 1390/1-1 of the German Research Foundation (Deutsche Forschungs Gemeinschaft).

Appendix

Bootstrap analyses were run for Models 3 and 4 (Table 3) to estimate the stability of the path coefficients and loadings. Table 5 presents the results for Model 3 in the first two columns and those for Model 4 in the last two columns. Only means and standard deviations for the parameters are shown because goodness-of-fit statistics are known to be severely biased in bootstrap analyses (Bollen & Stine, 1993).

References

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. In C. Umiltá, & M. Moscovitch (Eds.), *Attention and performance, vol. XV* (pp. 421–452). Cambridge, MA: MIT Press.

Arbuckle, J. L. (1997). *Amos* (version 3.61) [Computer software]. Chicago, IL: Small Waters.

Baddeley, A. D. (1986). *Working-memory.* Oxford: Clarendon Press.

Bollen, K. A., & Stine, R. A. (1993). Bootstrapping goodness-of-fit measures in structural equation models. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 111–135). Newbury Park, CA: Sage.

Borkenau, P., & Ostendorf, F. (1994). *Neo-Fünf-Faktoren Inventar (NEO-FFI) nach Costa and McCrae, Handanweisung* (Neo-Five-Factor Inventory following Costa and McCrae Handbook). Göttingen: Hogrefe.

Bucik, V., & Neubauer, A. (1996). Bi-modality in the Berlin Model of Intelligence Structure (BIS): a replication study. *Personality and Individual Differences*, *21*, 987–1005.

Carroll, J. B. (1993). *Human cognitive abilities: a survey of factor-analytic studies*. New York: Cambridge University Press.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix, & H. Hagendorf (Eds.), *Human memory and cognitive capabilities* (pp. 409–422). North Holland: Elsevier.

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

Das-Smaal, E. A., de Jong, P. F., & Koopmans, J. R. (1993). Working-memory, attentional regulation and the star-counting-test. *Personality and Individual Differences*, *14*, 815–824.

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.

Fahrenberg, J., Hampel, R., & Selg, H. (1989). *Das Freiburger Persönlichkeits-Inventar FPI. Revidierte Fassung FPI-R und teilweise geänderte Fassung FPI-A1. Handanweisung.5, ergänzte Auflage* (The Freiburg Personality Inventory FPI. Revised version FPI-R and partially modified version FPI-AI. Handbook, 5th augmented ed.). Göttingen: Hogrefe.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors*. Princeton, NJ: Educational Testing Service.

Fry, A. F., & Hale, S. (1996). Processing speed, working-memory, and fluid intelligence: evidence for a developmental cascade. *Psychological Science*, *7*, 237–241.

Gentner, D., & Stevens, A. L. (1983). *Mental models*. Hillsdale, NJ: LEA.

Guilford, J. P. (1956). The structure of intellect. *Psychological Bulletin*, *53*, 267–293.

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

Hagendorf, H., & Sá, B. (1995). Koordinierungsleistungen im visuellen Arbeitsgedächtnis (Coordination in visual working-memory). *Zeitschrift für Psychologie*, *203*, 53–72.

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells (Multimodal classification of intelligence performance. Experimentally controlled development of a descriptive intelligence structure model). *Diagnostica*, *28*, 195–226.

Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven (Intelligence structure research: competing models, new developments, perspectives). *Psychologische Rundschau*, *35*, 21–35.

Jäger, A. O., Süß, H.-M., & Beauducel, A. (1997). *Test für das Berliner Intelligenzstrukturmodell (BIS)* (Test for the Berlin Intelligence Structure Model (BIS)). Göttingen: Hogrefe.

Jäger, A. O., & Tesch-Römer, C. (1988). Replikation des Berliner Intelligenzstrukturmodells (BIS) in den "Kit of Reference Tests for Cognitive Factors" nach French, Ekstrom and Price (1963). Eine Reanalyse der Daten von Scholl (1976) (Replication study of the Berlin Intelligence Structure Model (BIS) using the Kit of Reference Tests for Cognitive Factors (French, Ekstrom, and Price, 1963): a reanalysis of Scholl's (1976) data). *Zeitschrift für Differentielle und Diagnostische Psychologie*, *9*, 77–96.

Johnson-Laird, P. N. (1983). *Mental models*. Cambridge, MA: Cambridge University Press.

Johnson-Laird, P. N. (1994). Mental models and probabilistic thinking. *Cognition*, *50*, 189–209.

Juhel, J. (1991). Relationship between psychometric intelligence and information-processing speed indexes. *Cahiers de Psychologie Cognitive*, *11*, 73–105.

Kail, R., & Salthouse, T. A. (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*, 199–225.

Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: a meta-analysis. *Intelligence*, *13*, 329–348.

Kyllonen, P. C. (1994a). Aptitude testing inspired by information processing: a test of the four-sources model. *Journal of General Psychology*, *120*, 375–405.

Kyllonen, P. C. (1994b). CAM: a theoretical framework for cognitive abilities measurement. In D. K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence* (pp. 307–359). Norwood, NJ: Ablex.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, *14*, 389–433.

Lohman, D. F. (1994). Component scores as residual variation (or why the intercept correlates best). *Intelligence*, *19*, 1–11.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, *7*, 107–127.

Mayr, U., & Kliegl, R. (1993). Sequential and coordinative complexity: age-based processing limitations in figural transformations. *Journal of Experimental Psychology, Learning Memory and Cognition*, *19*, 1297–1320.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*, 49–100.

Neubauer, A. C. (1997). The mental speed approach to assessment of intelligence. *Advances in Cognition and Educational Practice*, *4*, 149–173.

Norman, D. A., & Shallice, T. (1980). Attention to action: willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation: advances in research and theory, vol. 4* (pp. 1–18). New York: Plenum.

Oberauer, K. (1993). Die Koordination kognitiver Operationen. Eine Studie zum Zusammenhang von "working-memory" und Intelligenz (The coordination of cognitive operations: a study on the relationship between working-memory and intelligence). *Zeitschrift für Psychologie*, *201*, 57–84.

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

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*, 179–197.

Salthouse, T. A. (1990). Working-memory as a processing resource in cognitive aging. *Developmental Review*, *10*, 101–124.

Salthouse, T. A. (1992). Working-memory mediation of adult age differences in integrative reasoning. *Memory and Cognition*, *20*, 413–423.

Salthouse, T. A., Babcock, R. L., Mitchell, D. R. D., Palmon, R., & Skovronek, E. (1990). Sources of individual differences in spatial visualization ability. *Intelligence*, *14*, 187–230.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working-memory. *Psychology and Aging*, *6*, 118–127.

Salthouse, T. A., & Mitchell, D. R. D. (1989). Structural and operational capacities in integrative spatial ability. *Psychology and Aging*, *4*, 18–25.

Schmuck, P. (1996). *Flexibilität menschlichen Verhaltens. Differentialdiagnose mit objektiven Tests* (The flexibility of human behavior. Differential diagnosis with objective tests). Frankfurt: Lang.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science*, *283*, 1657–1661.

Sternberg, R. J. (1985). *Beyond IQ*. Cambridge, MA: Cambridge University Press.

Süß, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen* (Intelligence, knowledge, and problem solving. Cognitive prerequisites for successful action with computer-simulated problems). Göttingen: Hogrefe.

Süß, H.-M., Kersting, M., & Oberauer, K. (1993). Zur Vorhersage von Steuerungsleistungen an computersimulierten Systemen durch Wissen und Intelligenz (The predictability of performance in the control of computer-simulated systems by knowledge and intelligence). *Zeitschrift für Differentielle und Diagnostische Psychologie*, *14*, 189–203.

Turner, M. L., & Engle, R. W. (1989). Is working-memory capacity task-dependent? *Journal of Memory and Language*, *28*, 127–154.

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

Yee, P. L., Hunt, E., & Pellegrino, J. W. (1991). Coordinating cognitive information: task effects and individual differences in integrating information from several sources. *Cognitive Psychology*, *23*, 615–680.

Zimmermann, P., & Fimm, B. (1993). *Testbatterie zur Aufmerksamkeitsprüfung (TAP), Version 1.02. Handbuch* (Testing battery for attention, version 1.02. Handbook). Würselen: Vera Fimm/Psychologische Testsysteme.