VU Research Portal

Attention and Intelligence: The Validity of the Star Counting Test.

de Jong, P.F.; Das-Smaal, E.A.

published in

The Journal of Educational Psychology
1995

DOI (link to publisher)

[10.1037/0022-0663.87.1.80](https://doi.org/10.1037/0022-0663.87.1.80)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

de Jong, P. F., & Das-Smaal, E. A. (1995). Attention and Intelligence: The Validity of the Star Counting Test. *The Journal of Educational Psychology*, 87, 80-92. <https://doi.org/10.1037/0022-0663.87.1.80>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Attention and Intelligence: The Validity of the Star Counting Test

P. F. de Jong and E. A. Das-Smaal
Vrije Universiteit

The mechanisms underlying performance on the Star Counting Test (SCT) and its nomothetic span were investigated along with the relationships between working memory capacity, fluid intelligence (Gf), speed, and school achievement. The SCT is an attention test for children that requires the alternation of forward and backward counting. The test is based on A. D. Baddeley and G. J. Hitch's (1974) model of working memory in conjunction with D. A. Norman and T. Shallice's (1986) theory of central executive functioning. Tests were administered to 1,122 boys and 1,100 girls in 4th grade from 111 Dutch schools. The SCT required flexible alternation, counting speed, and sustained effort. Factor analysis showed that the SCT forms one factor with other indicators of working memory capacity. There was also a strong association between working memory capacity and Gf. The two clearly differ, however, in their relation to speed.

Attention and intelligence are important determinants of school achievement (Fraser, Walberg, Welch, & Hattie, 1987; Horn & Packard, 1985; Rowe, 1991). During this century, correlational studies have provided considerable insight into the structure of intelligence (e.g., Carroll, 1993). More recently, experimental researchers have identified some of the mechanisms underlying performance on specific intelligence tests (Butterfield, Nielsen, Tangen, & Richardson, 1985; Carpenter, Just, & Shell, 1990; Geary & Widaman, 1992; Sternberg, 1985). In contrast, in the field of attention there is still a gap between what Cronbach (1957) called the "two disciplines of psychology" (p. 671), the experimental and the correlational approach.

In the present study, both approaches were used to investigate the construct validity of the Star Counting Test (SCT), an attention test for children (de Jong & Das-Smaal, 1990). First, the construct representation of the test, including its dimensionality and the mechanisms that are involved in test performance, was determined. Next, we examined the nomothetic span of the SCT by considering its relationships with working memory, intelligence, speed, and school achievement. In what follows, we first describe the theoretical background of the test, and subsequently, we consider its construct representation and nomothetic span.

Theoretical Background of the SCT

In recent theories of attention researchers have emphasized its central role in the regulation of processes in the human information processing system (Navon, 1989a,

1989b; Neumann, 1987; Norman & Shallice, 1986). The SCT was designed to measure the regulation of processes in working memory. Working memory is a system for the temporary storage and processing of information and is used in a broad range of everyday cognitive tasks (e.g., Baddeley, 1986; Gilhooly, Logie, Wetherick, & Wynn, 1993; Hitch, 1978).

The SCT is based on Baddeley and Hitch's working memory model (Baddeley, 1986; Baddeley & Hitch, 1974). The model incorporates a central executive system as a control center and has two specialized slave systems for the temporary storage of information. The central executive is a limited capacity system that researchers assume initiates and modulates the various mental processes associated with working memory (Morris & Baddeley, 1988). The SCT was designed to tap the functioning of the central executive. It should be noted that several researchers do not consider the modality-specific storage systems to be part of working memory, and they reserve the term *working memory* to indicate the limited capacity system only (Just & Carpenter, 1992; Kyllonen & Christal, 1990; Salthouse & Babcock, 1991; Swanson, 1992). For sake of clarity, we use the term *central executive* throughout this article to denote the control system proper, and we use the term *working memory capacity* to refer to individual differences in the functioning of the central executive. We use the term *attention* to indicate the broader range of attentional capabilities, including working memory capacity.

Several models have been presented to specify the functioning of the central executive (Baddeley, 1986; Just & Carpenter, 1992; Salthouse & Babcock, 1991). Baddeley (1986) proposed Norman and Shallice's (1986; Shallice, 1988) conceptualization that emphasizes the top-down influence of the central executive (or "supervisory attentional system" as Norman & Shallice, 1986, p. 6, called it) on lower order processes. According to Norman and Shallice, a sequence of (mental) activities can be represented by a series of schemas that are run off successively and, in many situations, automatically. The central executive, however, is

P. F. de Jong and E. A. Das-Smaal, Department of Special Education, Vrije Universiteit, Amsterdam, The Netherlands.

We gratefully acknowledge the constructive comments of Jeffery Braden on a draft of this article.

Correspondence concerning this article should be addressed to Peter F. de Jong, Vrije Universiteit, Department of Special Education, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Electronic mail may be sent via Internet to Peter_de_Jong@psy.vu.nl.

able to control processing by the deliberate activation and inhibition of relevant and irrelevant schemas respectively.

The SCT is directly aimed at measuring a person's ability to activate and inhibit processes in working memory. The test requires the control of a very simple process, namely counting, which is known to involve working memory (Healy & Nairne, 1985; Logie & Baddeley, 1987; Nairne & Healy, 1983). More specifically, the SCT asks for the alternation of forward and backward counting. Each item in the test consists of a pattern of stars with signs in between (see Figure 1). The open spaces have no significance and are only there to prevent the person from counting by fives. The examinee has to start counting the stars from a given number. The plus and minus signs indicate the direction (forward or backward, respectively) in which the following stars should be counted. Although occasionally a sign may be repeated, most times the plus and minus signs are given in alternation. By counting and regularly alternating forward and backward counting, the last star will be reached, which is the answer to the item. On the basis of this principle, items can be constructed that differ systematically in the counting list that is necessary for the production of a correct solution.

Construct Representation of the SCT

Construct representation refers to the dimensionality of the test and the mechanisms that underlie test performance. We hypothesized that the test items could be scaled on one dimension and that the same mechanisms would underlie the performance on all items.

To study the mechanisms involved, we examined the relationship between an item's components and its difficulty (for a justification of this approach, see Embretson, 1983). We hypothesized that three components affect the demands

on the central executive and, hence, influence item difficulty. The first component is the number of alternations in counting direction. Each alternation is supposed to require the inhibition of an ongoing process and the activation of a new one, thus making extra demands on the central executive. As the number of alternations increases, the central executive is more often involved and the probability of an error on the item concerned increases. The second component is the duration of a process before it must be stopped. If an ongoing process is running automatically, inhibition needs to be stronger to interrupt the process (cf. Norman & Shallice, 1986). We assumed that forward counting is an overlearned skill and that the counting process runs automatically after a number of counts have been made. Accordingly, we hypothesized that the interruption of forward counting requires less inhibition if only a few counts were made than would be the case if the process were already running for some time. The third item component concerns the meaning of the signs denoting a change in the direction of counting. Normally, a plus sign denotes forward counting and a minus sign implies backward counting. When the meaning of the signs is reversed, however, a well-learned response has to be suppressed. The reversal of meaning is, therefore, expected to pose extra demands on the central executive and to increase the probability of an error.

The difficulty of an item might also be influenced by item components that cannot be related directly to the central executive part of working memory. First, counting lists of the SCT items will inevitably contain repeated-digit numbers (22, 33, etc.). These numbers can elicit counting errors (Healy & Nairne, 1985; Nairne & Healy, 1983). Therefore, we hypothesized that the number of repeated-digit numbers in the counting list would influence the item difficulty. Second, the items of the SCT differ in the size of the numbers in the counting list. As counting with higher numbers is probably less familiar than counting with lower ones, we assumed that more errors would be made with higher numbers in the counting list. Finally, items vary according to their position in the test. At the beginning of a task, attention is usually at its peak, gradually decreasing to an asymptote (e.g., Sanders, 1983). As the SCT is aimed at measuring attention, the items at the end of the test were expected to elicit more errors than items at the beginning, although the influence of practice could counteract this effect.

Nomothetic Span of the SCT

We determined the nomothetic span of the SCT by assessing the place of the SCT within the nomological network that is supposed to surround the test. We examined the relationship of the SCT with attention, especially working memory capacity, fluid intelligence (Gf), speed, and school achievement. Because the theoretical relationships of working memory capacity with Gf, speed, and school achievement are not fully established, the structure of interrelations between the constructs is, in itself, also theoretically important.

Figure 1. An example of an item from the Star Counting Test. The answer to this item is 37. Reprinted from *Personality and Individual Differences*, 11, P. F. de Jong and E. A. Das-Smaal, The Star Counting Test: An attention test for children, p. 599. Copyright (1990), with permission from Elsevier Science, Ltd., Pergamon Imprint, The Boulevard, Langford Lane, Kidlington OX5 1GB, United Kingdom.

Stankov (1988) reports a factor analysis of a large battery of ability tests in which several attention tests, originally devised by Wittenborn (1943), form a separate factor at a primary level. In a second-order analysis on the factor intercorrelations, this factor loaded heavily on Gf. In a recent study, Crawford (1991) also used several of Wittenborn's tests. Crawford found no separate attention factor. Wittenborn's tests had their major loading on Gf. Thus, the results of studies by Stankov (1988) and Crawford (1991) indicate a close relationship between attention and Gf. Although the Wittenborn tests were not based on Baddeley's (1986) model of working memory, they primarily seem to reflect working memory capacity (e.g., Stankov, 1988). Consequently, the findings of Stankov (1988) and Crawford (1991) suggest that working memory capacity and Gf are very much related.

Recent experimental (Carpenter et al., 1990; Gilhooly et al., 1993) and correlational (Kyllonen & Christal, 1990) research provides some direct evidence for a link between working memory capacity and Gf. Carpenter et al. (1990) showed that an important determinant of performance on the Raven Advanced Progressive Matrices Test (Raven, 1965), a common marker for Gf, is the ability to simultaneously generate, evaluate and maintain goals in working memory. This ability obviously depends on working memory capacity. In several independent studies, Kyllonen and Christal (1990) obtained correlations ranging from .80 to .90 between working memory capacity and Gf. Results of recent research therefore indicate a significant relationship between working memory capacity and Gf.

With respect to the relationship between working memory capacity and speed, a distinction should be made between general and task-specific processing speed (e.g., Salthouse & Babcock, 1991). *General processing speed* refers to the efficiency of the carrying out of elementary processing operations, as, for example, in perceptual comparison speed. As argued by Salthouse and Babcock (1991), however, most of the task-specific processes, such as reading (Daneman & Carpenter, 1980) or simple arithmetic problems (Salthouse & Babcock, 1991), are of moderate complexity in working memory capacity tasks and probably do not reflect general processing speed only. *Task-specific processing speed* refers to the speed of execution of the specific operations required in a particular task designed to measure working memory capacity. If one assumes that there is a trade-off between the storage and the processing component of working memory (e.g., Case, Kurland, & Goldberg, 1982; Just & Carpenter, 1992), then an increase in task-specific speed should result in an increase in storage capacity, which is usually the measure of interest in working memory capacity tasks. Because general speed will be related to task-specific processing speed, a relationship between these simple speed measures and working memory capacity is likely. Indeed, both task-specific processing speed (Case et al., 1982; Crammond, 1992; Hitch & McAuley, 1991; Salthouse & Babcock, 1991) and general processing speed (Kyllonen & Christal, 1990; Salthouse, 1992; Salthouse & Babcock, 1991) have been shown to be related to working memory capacity, albeit not strongly, in most cases.

Finally, evidence suggests that there is a relationship between working memory capacity and school achievement. Results from experimental studies show that working memory is involved in tasks that are indicative of school achievement, such as reading comprehension (Baddeley, 1986; Just & Carpenter, 1992) and arithmetic (Hitch, 1978). Consistent correlations have also been found between these tasks and working memory capacity (Daneman & Carpenter, 1980; Swanson, 1992; Turner & Engle, 1989).

In short, results of previous research indicate that working memory capacity is associated with Gf, with school achievement, and to a lesser extent, with speed. It should be noted, however, that most of this research, especially concerning the relationship of working memory capacity with Gf and speed, has been performed with adults. The present study involved elementary school children. In addition to the SCT, a battery of tests was administered. The tests were selected to reflect working memory capacity, Gf, speed, and school achievement. We used measures for working memory capacity that required the simultaneous retention and manipulation of information and that covered various content (Baddeley, 1986; Kyllonen & Christal, 1990). Gf was indicated by a test requiring verbal and figural reasoning. Speed measures reflected the examinee's speed of executing simple mental operations. Finally, we used common tests of reading comprehension and arithmetic to measure school achievement.

In summary, in the present study we sought to examine the construct validity of the SCT. In the first part, we investigated the dimensionality of the test and the mechanisms contributing to SCT performance. In the second part of the study we focused on the nomothetic span of the test. First, we examined the relationship between the SCT and other tests of working memory capacity. Next, we explored the interrelations between working memory capacity, reasoning ability, speed, and school achievement.

Method

Participants

Participants included 2,222 Dutch fourth-grade elementary school children aged 9 years, 10 months ($SD = 5.0$ months) who participated in the Dutch National Assessment Study of Attentional Deficit Disorders (see further description of this sample in de Jong, 1991). In short, a national sample of 111 elementary schools in The Netherlands participated in the study. One randomly selected fourth-grade class was selected per school. Five children, aged 9 years at a prespecified date, were also randomly selected. This latter sample of 552 children is denoted as the subsample. Note that, because of the restriction of the age range, the subsample is not completely representative of the total sample.

Of the 2,588 children in the (total) sample (including the subsample), 321 were omitted because these children had at least one parent born outside The Netherlands. In addition, 45 third-grade children were removed. After omission of these children, 2,222 children (1,122 boys and 1,100 girls) remained in the sample, 443 (208 boys and 235 girls) of whom were left in the subsample.

Star Counting Test (SCT)

Each item of the test (see Figure 1) consisted of a pattern of stars with plus and minus signs between them and of a number in the left top corner. The children were instructed to "count the stars from left to right and from top to bottom, like you would if you were reading." However, the children were instructed "not to count from one, but from the number in front of the item." Each item had a different starting number. The signs denoted the direction (forward or backward) in which subsequent stars were to be counted. Thus, the test required alternating forward and backward counting until the last star was reached. An item always started with forward counting. The number of the last star was the answer to "write on the dotted line below."

An item pool of 44 items was constructed. For theoretical reasons, items differed in the number of changes in counting direction (two, four, or six changes), the number of stars preceding a change from forward to backward counting (small versus large)¹, and the meaning of the signs (normal versus reversed). In addition, the starting number of the items varied from 14 to 78, and items were distinguished according to the number of double digits that occurred during counting (three versus six). The 44 items in the pool formed 22 pairs. The items of a pair were similar in every respect except for a small difference in the starting number, and the items were meant to be parallel. For 12 pairs of items, the meaning of the signs was normal (i.e., after a plus counting was to go forward, but counting was to go backwards after a minus). The meaning of the signs for 10 pairs was reversed, implying backward counting after a plus and forward counting after a minus sign.

The 44 items in the pool were used to construct three versions of the test. Each version of the test consisted of two parts, which were administered separately. The first part of each version contained 12 items, and the second part contained 10 items. The items of the pool were distributed as follows: One item of an item pair was randomly assigned to a first version, and the other was randomly assigned to the third. A second version was formed by the odd and the even items of the first and the third version, respectively. Thus, the first and the third version did not overlap. The second version consisted of items that were either in the first or in the third version, thus enabling the equation of the scores of the three versions. Within each version, the items in which the signs had their normal meaning were assigned to the first part of the test, and the items in which the meaning of the plus and minus signs was reversed were assigned to the second part. Within parts, the items were counterbalanced with respect to the number of changes in counting direction. Complete counterbalancing was not possible for the size of the starting number and the number of stars before a change in counting direction.

The first part of the test contained 12 items and was preceded by an example and two items for practice. The second part consisted of 10 items that were preceded by an example and one practice item. The time permitted for the first part of the test was 12 min, and 10 min were given for the second part. These time limits were imposed because the test was part of a battery of tests administered to groups of children. The time limit for each part of the test was based on a pilot study (de Jong, 1991) of 109 children, who were at the end of Grade 3. Of these children, 95% completed the whole test in the allotted time.

Working Memory Capacity Tests

In accordance with Baddeley's (1986) definition, the tests for working memory capacity required the storage and processing of information and were devised to measure various contents.

Digit Span Test. Forward and backward digit span were established with the Dutch version of the Digit Span Test of the WISC-R (van Haasen, 1986). The test required retention and reproduction of a sequence of digits. The digits were recorded on audiotape with a 1-s interdigit interval. The number of digits in a sequence increased for each successive series. In the forward condition, the digits had to be reported orally in the order of presentation. In the backward condition, the digits had to be reported in reverse order. A score on the test was the sum of the (standardized) score on the forward and the (standardized) score on the backward span.

Group Paced Auditory Serial Addition Test (GPASAT). The GPASAT (de Jong, 1991) is a modification of the PASAT (Gronwall & Sampson, 1974) and is suitable for administration to groups of children. The test required the simultaneous addition and storage of digits. The GPASAT consisted of series of auditorily presented digits. With each presentation of a digit, the child is required to add it to the previous digit and specify the result on an answer sheet before the following digit is presented. A series contained eight digits, and thus required seven answers. A score of one was given when all seven answers were correct. The complete test had 20 series of eight digits. The series were presented at a speed that varied from approximately 1.5 s to 4 s between successive digits. The maximum score on the test was 20.

Following Directions. The Following Directions test required the execution of a series of simple directions, which have been shown to demand working memory capacity (Engle, Carullo, & Collins, 1991). Each item of the test consisted of a picture that was to be marked by the participant according to directions given on audiotape. Directions gave the place and the type of mark (a triangle, a square, a cross, or a circle) that had to be drawn in the picture. The place of the mark was specified by naming an object in the picture and the location around the object (above, beneath, or beside). The number of directions per item varied from two to three. The children were required to wait until they had received all of the directions before they followed any. This requirement made a demand on working memory in that one direction had to be temporarily stored as a new direction was processed. Directions corresponding to the same item were given in one sentence (e.g., "Put a circle above the mushroom with the white dots and beneath the elf"). Here, two directions are given, namely to put a circle above a mushroom and to put a circle beneath the elf. An item was scored as correct if all the directions were properly executed. The test consisted of 32 items and had a maximum score of 32.

Speed Tests

Speed tests were selected that reflected the speed of executing simple mental operations.

Bourdon-Vos Test. The Bourdon-Vos Test (Vos, 1988) is a cancellation test. The test consisted of a sheet with 32 rows of 24 dot patterns each. The number of dots in a pattern varied from three to five. The task was to cancel as quickly as possible all

¹ The number of stars preceding an alternation from forward to backward counting was partly dependent on the number of alternations in an item, because the number of stars per item was fixed. Consequently, both *small* and *large* refer to a different number of stars in items with two, four, and six alternations. In items with two alternations the mean number of stars before a change from forward to backward was 9.5 (small) or 20 (large); in items with four alternations the mean number of stars was 7.4 (small) or 13.8 (large), and in items with six alternations, the mean number of stars was 5.7 (small) or 8.3 (large).

four-dot patterns on the sheet. The score was the average number of seconds used to complete a row.

Coding Test. The Coding Test is part of the Dutch version of the WISC-R (van Haasen, 1986). The test requires the rapid substitution of digits in symbols. The score was the number of correct substitutions completed in 2 min.

Trail Making Test. The Trail Making Test is part of the Halstad-Reitan test battery, and is supposed to measure speed of visual search and mental flexibility (Spren & Strauss, 1991). The first part of the test (Trail A) requires the child to draw consecutive connections by pencil between encircled numbers randomly arranged on a page. In the second part of the test (Trail B) the circles contain letters or numbers, and the child's task is to connect the circles in the appropriate order, alternating between numbers and letters. On both parts the time needed to connect all the circles was scored. The scores on both parts of the test were standardized within the sample. A total score on the test was computed by adding the two standardized scores.

School Achievement Tests

School achievement was assessed with a test for reading comprehension and with one for arithmetic. Both tests are regularly used school achievement tests in The Netherlands.

Reading comprehension. The test (Cito, 1981) consisted of five stories containing 13 to 33 sentences. Each story was followed by several multiple-choice items. The complete test had 25 items. Cronbach's α was .83.

Arithmetic achievement. The test (Cito, 1979) consisted of 40 multiple-choice items. Because some schools indicated that the relevant material had not been covered, eight items were deleted. The remaining 32 items required addition, subtraction, multiplication, and division. Cronbach's α for this test was .83.

Fluid Intelligence (Gf)

The Fluid Intelligence Omnibus Test (FIOT; de Jong, 1991) was used as a measure for verbal and figural reasoning. The test contained 30 items. Most of the items, 23 to be specific, came from the Primary Mental Ability Test 2-4, the Dutch version of the Henmon Nelson Test of Mental Ability, Form A 6-9 (Verbeek & Troch, 1973) and from an adaptation of the Dutch version of the Henmon Nelson Test (Huisman, Stouthart, & Vorst, 1985). As the items of the Dutch version of the Henmon Nelson Test have only four alternatives, the number of alternatives for the selected items of the original Henmon Nelson Test was reduced from five to four. In addition, four items of the FIOT came from the Transitivity Test (Hoeks, 1985) and three items were from a test by Abraham, Clement, Markusse, Ritsma, & Tollenaar (1988).

The FIOT contained 15 items involving verbal reasoning and 15 items about figural reasoning. The verbal reasoning items (see also Carroll, 1993) involved verbal analogies, verbal exclusion, general reasoning (e.g., The father of my mother is the of my sister) and transitivities (e.g., John is smaller than Ann, Jim is taller than Ann. Who is the smallest?). The figural reasoning items included figural analogies, figural exclusion, and number series. Cronbach's α of the complete test was .81.

Procedure

The SCT, the GPASAT, the Following Directions test, the tests for school achievement, and the FIOT were administered to groups and were, with the exception of the FIOT (see below), adminis-

tered to the total sample. For reasons beyond the scope of this article, a random selection of 54 schools was administered the tests in this battery in one order. In the other 57 schools, the tests were given in a different order. The three versions of the SCT were, within classes, randomized over three groups. The FIOT was administered to 507 children from a subgroup of 27 randomly selected schools. Another 555 children from 28 randomly selected schools were administered the verbal reasoning part of the FIOT only. The Digit Span Test and the tests for speed (Bourdon-Vos, Trail Making, and Coding) were administered individually to the subsample only.

The tests for school achievement were administered by the teacher. All other testing was conducted by trained test assistants.

Results

The results are presented in two separate sections. The first section reveals results about the construct representation of the SCT, including its dimensionality and the relationship between item components and item difficulty. The second section deals with the nomothetic span of the SCT. The relationships between the SCT and other tests for working memory capacity are reported, and those between working memory capacity, Gf, school achievement and speed are modeled.

Construct Representation

First, the psychometric characteristics of the SCT were examined. Descriptive statistics of the score distributions of the three versions of the SCT are presented in Table 1. The means and standard deviations of the three versions were similar. In addition, Table 1 shows the percentage of students who completed the test within the allotted time. Unexpectedly, between 43% and 48% of the students did not complete all items on the test. For each separate part of the test these percentages were lower: between 37% and 40% for the first part and between 24% and 28% for the second part. Finally, the correlation between the first and the second part of the test was moderate. Taking the mean correlation between both parts ($r = .597$), and using the Spearman-Brown formula, we found a reliability of the SCT of .75 (see also de Jong, 1991).

The dimensionality of the SCT was examined by item response modeling (Goldstein & Wood, 1989). Thus, the responses for the set of 44 different items, distributed over the three versions of the SCT, could be jointly analyzed. Under the restriction of the model, the items of the SCT can

Table 1
Descriptive Statistics of the Score Distribution of the Star Counting Test (SCT) and the Correlation Between the First and the Second Part

Version	%CT	M	SD	Skewness	Kurtosis	r
SCT 1	55.5	12.40	5.08	-.27	-.68	.59
SCT 2	53.2	12.80	4.90	-.38	-.54	.59
SCT 3	57.1	12.88	4.92	-.41	-.51	.61

Note. %CT = Percentage of students who completed the test.

be scaled on one scale. The Rasch model (Fischer, 1974) was applied to the items of the SCT. We estimated the parameters by using the program RIDA (Rasch Incomplete Data Analysis; Glas, 1989), which computes conditional maximum likelihood estimates. In addition, the program computes a test statistic R_1 , which has a chi-square distribution and that researchers can use to evaluate the fit of the model (for details see Glas, 1988).

The SCT was part of a battery of tests, which was presented in two different orders. Further results concerning the construct representation of the SCT are given separately for each order (to be denoted as Sample 1 and Sample 2). Thus, we were able to examine the generalizability of the results. In addition, by cross-validating the results, we could avoid capitalization on chance with respect to both modification of the Rasch model and the selection of item components that are related to item difficulty.

The Rasch model did not fit the data well (Sample 1: $R_1 = 441.84$, $df = 293$, $p < .01$; Sample 2: $R_1 = 458.13$, $df = 335$, $p < .01$). Consideration of the discrepancies in Sample 1 between the observed and the expected number of students who passed an item revealed that these discrepancies were especially large for the last items in each version of the test. This suggested that the lack of fit of these items to the model could be caused by the fact that some of the students did not finish the test. Because these students automatically did not pass the last items, the observed number of students who did pass these items could be much lower than would be expected if all students had completed each item.

To improve the model fit, we made a distinction between a student who did not pass an item and a student who simply did not reach an item. Following van den Wollenberg (1979; van den Wollenberg & Creemers, 1986), we assumed that students who did not finish the test received a test with a smaller number of items. If the Rasch model applies, the scores of students who have attempted different items can be easily equated. According to this method of scoring, a large number of tests of different length was formed. However, because of limitations of the RIDA computer program, the number of tests of different length had to be reduced to four for each version of the SCT. The four tests consisted of 22 items (the complete test), 19 items (first part, 10 and second part, 9), 15 items (8 and 7) and 6 items (4 and 2), respectively. Rasch analysis of the 12 tests (3 versions $\times$ 4 test lengths) showed that the model fits the data nicely (Sample 1: $R_1 = 823.93$, $df = 838$, $p = .63$; Sample 2: $R_1 = 847.67$, $df = 822$, $p = .27$). This result supports the supposition that the initial lack of model fit was due to the confounding effect of the proportion of students who attempted an item. When this effect was taken into account, the set of 44 items appeared to conform to the Rasch model. This result also implies that the difficulty of an item is similar in all groups. This suggests that an item will not become less difficult if more time is spent, as might have been the case for children who did not complete all items. Also, the fit to the Rasch model suggests that the test is unidimensional. Consequently, the same mechanisms are likely to underlie performance on all items.

To investigate the mechanisms underlying test performance, we conducted regression analyses to determine the relationship between the item components and the item difficulty as estimated under the Rasch model. Regression analyses were performed on the complete set of 44 items. No interaction effects between the components were examined, because no hypotheses were posed in this respect. The standardized regression weights and the squared multiple correlations (R^2) are reported in Table 2. The results show that a considerable amount of the variance in the item difficulties can be explained by the components of the items. For the first sample, R^2 is a little higher than it is for the second sample. All components of the items appeared to exert, as expected, a significant influence on the difficulty of an item. The regression weights are similar in both samples, thus demonstrating the robustness of the results. Coding of each of the components with three categories (number of alternations, size of the starting number, and position of an item in the test) into two dummy variables did not show a curvilinear relationship of these components with item difficulty. The size of the regression weights in Table 2 suggests that the number of changes in counting direction (more changes leads to a higher item difficulty), the size of the starting number (a higher number results in a higher item difficulty), and the position of an item in the test (the closer an item is to the end, the more difficult it is) are more important components than the number of stars before an alternation (a higher number leads to an increase in item difficulty), the meaning of the signs (reversal of the meaning results in a higher item difficulty), and the number of repeated digits (more repeated digits increases the difficulty of an item). To test for these differences, we restricted first the standardized regression weights of the former three components to be equal. Similarly, we restricted the standardized regression weights of the latter three components to be equal. The F tests showed a nonsignificant drop of R^2 in comparison with the model in which all regression

Table 2
Standardized Regression Weights Associated With Predicting Star Counting Test Item Difficulties From Item Components

Item components	Manipulation	Sample 1	Sample 2
Number of alternations	two, four, six	.49**	.40**
Number of stars before alternation	small, large	.21**	.27**
Meaning of the signs	regular, reversed	.26**	.25**
Repeated digits (33, 44, etc.)	three, six	.21*	.25**
Size of the starting number	1–29, 30–57, 58–80	.34**	.35**
Position of item in test	beginning, middle, end	.58**	.48**
	R^2	.77	.69

Note. Number of items is 44.

* $p < .05$. ** $p < .01$.

weights were free: Sample 1, $F(4, 37) = 0.96$, $p > .10$; Sample 2, $F(4, 37) = 0.15$, $p > .10$. Thus, the regression weights of the three most important components and also the regression weights of the three least important components can be considered equal. Next, we restricted the standardized regression weights of all the components to be equal. This led to a significant drop in R^2 in comparison with the former model, Sample 1, $F(1, 41) = 15.94$, $p < .01$; Sample 2, $F(1, 41) = 5.03$, $p < .05$, thus implying that the number of changes in counting direction, the size of the starting number, and the position of an item in the test are the most important item components.

Nomothetic Span of the SCT

The nomothetic span of the SCT involves the relationship between the SCT and other tests for working memory capacity, and the interrelations among working memory capacity, Gf, speed, and school achievement.

The structure of the nomological net around the SCT was examined in the subsample. From this subsample, the scores of 23 children had to be omitted because these children were absent at the time that the tests of school achievement were administered. Scores from 39 children had one or more missing values on the other tests, thus leaving 381 children (175 boys and 206 girls) for the analyses. *T*-tests, adopting a familywise significance level that was controlled at .05 by setting alpha per comparison at .05/11 or .0045, did not show significant differences in the mean scores on the tests that were taken jointly between the group of 381 children used for further analyses and the group of 62 children omitted (smallest $p = .03$).

For each student in the subsample, a Rasch score (SCT-R) on the SCT was computed on the basis of the estimated item parameters under the Rasch model for the total sample (i.e., Sample 1 and Sample 2 combined). The SCT-R of a student was based on the items in one of the 12 (3 versions $\times$ 4 test lengths) tests. Thus, the SCT-R was based only on items that the student had actually attempted or reached.² In addition, a sum score (SCT-S) was computed. SCT-S was a simple count of the number of items that were passed. Because the means and standard deviations on the three versions of the SCT were very similar, we assumed that the three versions of the test were equally difficult and, therefore, that the SCT-S of the three versions would be equivalent.

Because in the SCT-S score students automatically did not pass items that were not reached, the influence of speed on SCT-S should be greater than it is on the SCT-R. The correlation between the SCT-S and the number of items reached (NIR) was .59 ($p < .01$). However, when the score on the SCT was based on the number of items reached, SCT-R, the correlation with NIR dropped to .30 ($p < .01$) but was still statistically significant. The drop in correlation supports the contention that the influence of speed on SCT-S is inflated because many students did not finish all items. The correlation remaining after correction for the number of items reached, suggests that performance on the

test is still affected by speed. Therefore, better performing participants, as measured by the corrected score SCT-R, were more likely to have finished all of the items. Finally, the correlation between SCT-S and SCT-R was .93 ($p < .01$). This indicates that the extra effect of speed on the SCT-S score is very small because SCT-S and SCT-R are highly related.

The correlations between the tests for working memory capacity, Gf, speed, school achievement, and the SCT are presented in Table 3. All tests for speed have been recoded such that a higher score indicates more speed. Several results are of interest here. First, the correlations of the SCT-S and the SCT-R with the other tests are highly similar, thus suggesting once again that the differences between the SCT-S and the SCT-R are practically negligible. The only discrepancies that were found involve the correlations with the speed tests (Bourdon-Vos Test, Coding, and Trail Making), the GPASAT, and the Digit Span Test. One can argue that the GPASAT also requires speed. The somewhat higher correlations of the SCT-S as compared to the SCT-R with the tests involving speed indicates once again that speed is slightly more involved in the SCT-S than in the SCT-R. As the SCT-S and the SCT-R appear to be very similar, further results will concern only the SCT-S (SCT from now on), which is much easier to compute than the SCT-R.

A second point of interest in Table 3 is the differential correlations of the SCT (SCT-S in the table) with the various tests. As predicted, the highest correlations were observed between the SCT and the working memory tests, especially Digit Span and GPASAT. The correlation with Following Directions is somewhat lower. The lowest correlations are found between the SCT and the speed tests (Bourdon-Vos, Coding, and Trail Making). In between are the correlations of the SCT with reasoning (verbal and figural) and with achievement (reading comprehension and arithmetic).

To gain further insight into the nomological net around the SCT, we conducted confirmatory factor analyses. The analyses were done with the program Equations (EQS; Bentler, 1989). Model fit was evaluated with the chi-square statistic and by means of the Non-normed Fit Index (NNFI) and the Comparative Fit Index (CFI; Bentler, 1989; 1990).

First, the structure of the four working memory capacity tests was examined. As expected, a one-factor model appeared to fit the data nicely, $\chi^2(2, N = 381) = 4.11$, $p = .13$ (NNFI = .98; CFI = .99). The GPASAT had the highest loading on the factor, followed by the SCT, the Digit Span Test, and Following Directions (see also Figure 2).

Next, we formulated a four-factor model, similar to the model of Kyllonen and Christal (1990), to determine the interrelations between working memory capacity, speed, Gf, and school achievement (see Figure 2). To account for the missing data on the verbal or the figural part of the

² This Rasch score is similar to a proportion score. The difference is that the Rasch score takes into account the item difficulty of the items that are attempted, whereas with a proportion score one assumes that all item difficulties are equal.

Table 3
Correlations Among the Star Counting Test (SCT) and the Tests for Working Memory Capacity, Speed, Gf, and Achievement

Test	1	2	3	4	5	6	7	8	9	10	11	12
WMC												
1. DST	—											
2. GPASAT	.40**	—										
3. FODI	.29**	.39**	—									
Speed												
4. BOVO	.13**	.36**	.08	—								
5. COD	.16**	.38**	.20**	.44**	—							
6. TRAIL	.15**	.29**	.20**	.38**	.31**	—						
Gf												
7. VERE	.23**	.43**	.40**	.05	.08	.12	—					
8. FIRE	.25*	.34**	.37**	.34**	.10	.29**	.58**	—				
Achievement												
9. RECO	.25**	.36**	.33**	.05	.10	.08	.54**	.46**	—			
10. ARIT	.27**	.45**	.30**	.13*	.11*	.19**	.48**	.38**	.51**	—		
SCT												
11. SCT-S	.40**	.53**	.29**	.23**	.21**	.21**	.29**	.34**	.29**	.35**	—	
12. SCT-R	.36**	.45**	.28**	.16**	.17**	.17**	.30**	.31**	.30**	.34**	—	—

Note. WMC = Working Memory Capacity; DST = Digit Span Test; GPASAT = Group Paced Auditory Serial Addition Test; FODI = Following Directions; BOVO = Bourdon-Vos Test; COD = Coding Test; TRAIL = Trail Making Test; VERE = Verbal Reasoning; FIRE = Figural Reasoning; RECO = Reading Comprehension; ARIT = Arithmetic; SCT-S = Star Counting Test, sum score; SCT-R = Star Counting Test, Rasch score; Gf = Fluid Intelligence.

* $p < .05$. ** $p < .01$.

reasoning test, which were administered only to a random part of the subsample, we performed a multiple group analysis (Bentler, 1989). One group consisted of 84 children (41 boys and 43 girls) who had completed all tests, including both parts of the reasoning test. A second group of 98 children (44 boys and 54 girls) had completed all tests except the figural part of the reasoning test, and the third group contained the 199 children (90 boys and 109 girls) to whom both parts of the reasoning test were not administered. Before the four-factor model was fit to the data, however, we conducted a principal component analysis (PCA) to explore possible alternative models for describing the interrelations between the tests.

PCAs were conducted in the first and the second group. In both groups the eigenvalue-greater-than-1 criterion suggested two factors, whereas the scree criterion indicated that three factors were needed to describe the data. The interpretation of the latter solution was, after a varimax rotation, straightforward and similar in both groups. The three factors can be denoted as Working Memory Capacity, Speed, and a combination of Gf and School Achievement. Gf and school achievement did not form separate factors in either group. Because the number of students in the first and the second group is fairly small, a final PCA was run on the combined first and second group (excluding the figural part of the reasoning test in the first group) of 182 students. The factor loadings of the tests on the first unrotated factor and on the three rotated factors are presented in Table 4. Table 4 reveals that the GPASAT has the highest loading on the first unrotated factor, followed by the SCT and the tests loading on the combined Gf/School Achievement factor.

Thus, PCA suggested a three-factor model as an alternative to our theoretically based four-factor model. Therefore,

both models were subjected to a multigroup confirmatory factor analysis. For both models, the parameters of the model were specified to be equal across groups, thereby we assumed the equivalence of the model in the three groups. The four-factor model appeared to fit to the data, $\chi^2(138, N = 381) = 155.60, p = .15$ (NNFI = .98, CFI = .98). Similarly, however, the three-factor model also fit these data, $\chi^2(141, N = 381) = 159.98, p = .13$ (NNFI = .98, CFI = .98). Note, however, that the particular three-factor model involved (see Table 4) is a special case of the four-factor model. One can obtain the three-factor model from the four-factor model by restricting the intercorrelation between the Gf factor and the School Achievement factor to one, and by requiring the factor intercorrelations of Gf with the other factors (Working Memory Capacity and Speed) to be equal to the factor intercorrelations of School Achievement with these other factors. A chi-square difference test revealed that these extra restrictions in the four-factor model did not lead to a significant decline in model fit, $\Delta\chi^2(3, N = 381) = 4.38, p = .29$.

Although for reasons of parsimony the three-factor model would be preferable, we adopted the four-factor model (i.e., we chose to separate the Gf and School Achievement factors). Theoretically, there is a clear distinction between Gf and School Achievement, a distinction that many factor analytic studies support (Carroll, 1993). The four-factor model also enabled us to compare our results with those of Kyllonen and Christal (1990).

The standardized solution of the parameter estimates of the four-factor model is presented in Figure 2. Two aspects of this solution are of interest. First, Working Memory Capacity is, as expected, highly correlated with Gf and School Achievement. Second, as in the study of Kyllonen

Figure 2. Factor model of the relationships between Working Memory Capacity (WMC), School Achievement (SA), Fluid Intelligence (Gf), and Speed (SP). Rectangles indicate observed variables; circles denote latent factors; single-headed arrows represent factor loadings, and double-headed arrows represent factor intercorrelations. The factor intercorrelations of a four-factor model without the Group Paced Auditory Serial Addition Test (GPASAT) are given in parentheses. SCT = Star Counting Test; DST = Digit Span Test; FODI = Following Directions; RECO = Reading Comprehension; ARIT = Arithmetic; VERE = Verbal Reasoning; FIRE = Figural Reasoning; BOVO = Bourdon-Vos Test; COD = Coding Test; TRAIL = Trail Making Test.

and Christal (1990), the results suggest differential relations of Working Memory Capacity and Gf with School Achievement and Speed. The correlation between Working Memory Capacity and School Achievement seems to be somewhat smaller than that between Gf and School Achievement. However, when we specified the correlations of Working Memory Capacity and Gf with School Achievement to be equal, we found a nonsignificant reduction of model fit, $\Delta\chi^2(1, N = 381) = 1.16, p = .30$. Therefore, a difference between these correlations cannot be demonstrated. Furthermore, as in the Kyllonen and Christal study, the correlation of Speed with Working Memory Capacity is larger than it is with Gf. When we specified these correlations to be equal, we found a significant decline in the fit of the model, $\Delta\chi^2(1, N = 381) = 19.88, p < .01$.

It should be noted that the magnitude of the correlation between Working Memory Capacity and Speed is unexpectedly high. One reason may be that the SCT and the GPASAT are tests with a time limit. On the SCT, some students did not complete all items. The GPASAT requires each computation to be made within a preset amount of time. To examine the influence of these tests on the correlation

between Working Memory Capacity and Speed, we successively excluded the SCT and the GPASAT from the model. The model without the SCT appeared to fit, $\chi^2(110, N = 381) = 133.48, p = .06$ (NNFI = .97, CFI = .97), but the parameter estimates (including the factor intercorrelations) were virtually identical to the complete model in which all tests were included. This indicates once again that the particular time limit imposed on the SCT was negligible. In addition, the model without the GPASAT also fits the data, $\chi^2(110, N = 381) = 128.26, p = .11$ (NNFI = .97, CFI = .97). The factor loadings of the remaining tests in this model were almost identical to those in the complete model. However, the correlation between Working Memory Capacity and Speed dropped from .60 to .49, and the correlation between Working Memory Capacity and Gf increased from .66 to .75 (see also Figure 2). When we specified the correlation of Speed with Working Memory Capacity in this model to be equal to its correlation with Gf, the fit of the model dropped significantly, $\Delta\chi^2(1, N = 381) = 8.70, p < .01$, thus indicating that the correlation of Speed with Working Memory Capacity is larger than its correlation with Gf.

Table 4
Factor Loadings on the First Unrotated Factor and in the Three-Factor Solution After Varimax Rotation

Tests	Unrotated first factor	Varimax solution		
		WMC	SP	Gf/SA
DST	.52	.80	.03	.07
SCT	.64	.67	.24	.21
GPASAT	.81	.54	.39	.44
FODI	.59	.45	.07	.44
BOVO	.42	.06	.83	-.01
TRAIL	.48	.05	.77	.14
COD	.54	.37	.67	.01
ARIT	.62	.06	.12	.81
RECO	.64	.16	.02	.81
VERE	.64	.21	-.01	.79
FIRE ^a	—	—	—	—

Note. WMC = Working Memory Capacity; SP = Speed; Gf/SA = Fluid Intelligence/School Achievement; DST = Digit Span Test; SCT = Star Counting Test; GPASAT = Group Paced Auditory Serial Addition Test; FODI = Following Directions; BOVO = Bourdon-Vos Test; TRAIL = Trail Making Test; COD = Coding Test; ARIT = Arithmetic; RECO = Reading Comprehension; VERE = Verbal Reasoning; FIRE = Figural Reasoning.

^a FIRE was not incorporated into this analysis.

Discussion

One objective of this study was to investigate the construct representation of the SCT. The items of the test can be considered as unidimensional, which implies that the same mechanisms underlie performance on all items of the SCT.

A major mechanism that was hypothesized to be involved in SCT performance was the central executive system of working memory. Three components of the items that were assumed to make demands on the central executive system appeared to be related to the difficulty of an item: the number of alternations of forward and backward counting, the meaning of the signs that indicate forward and backward counting (normal or reversed), and the duration of forward counting before a shift to backward counting is required. The number of alternations was the most important of these components. An increase in the number of alternations led to a greater number of errors on the item. This effect seems to support directly that the central executive is involved in performance on the SCT. The central executive is supposed to be needed for the activation and inhibition of processes in working memory, and this is exactly what the alternation of counting direction was supposed to accomplish.

In addition, mechanisms other than the central executive system contribute to SCT performance. We demonstrated that if the size of the numbers in the counting list increases, the SCT tends to become more difficult. This effect might be due to the inverse relationship of the size of the numbers in the counting list with counting speed. Larger numbers are practiced less, which probably leads to slower access to numbers in long-term memory, as Hitch and McAuley (1991) have suggested, and hence to a slower counting speed. As a result, items with large numbers require more

processing time, and the probability of an error will increase (see Jansen, 1990). The position of an item in the test also had a substantial effect on SCT performance. Errors were made more frequently toward the end of the test. This phenomenon of a gradual decline in test performance is not uncommon (Jansen, 1990; Sanders, 1983). In the SCT, this phenomenon is probably due to the great amount of effort that each item requires, which can be afforded in the beginning but not sustained throughout the test. A practice effect was not detected, either because it was counteracted by effort or fatigue or because several practice items were given in advance. Finally, as we expected, the number of repeated-digit numbers had a small influence on the item difficulty.

All in all, the most important item components appeared to be the number of alternations in the direction of counting, the size of the numbers in the counting list, and the position of an item in the test. The influence of these components on test performance suggests not only that the test requires central executive functioning, but also that it requires counting speed and sustained effort. Thus, the SCT is probably not a completely pure measure of working memory capacity. As has been shown for most current tests of working memory capacity, task-specific processes partly determine test performance. Therefore, as argued by Salthouse and Babcock (1991), multiple measures should be used to reflect working memory capacity. Finally, it should be noted that although experimentally certain mechanisms have been shown to be involved in performance on the SCT, this does not necessarily imply that these mechanisms account for individual differences. To demonstrate the latter, the SCT should be related to separate tasks for each mechanism (Embretson, 1983; Sternberg, 1981).

In this study we also sought to examine the nomothetic span of the SCT. Substantial relationships were found, as we expected, between the SCT and other tests of working memory capacity. Confirmatory factor analysis showed that the various tests for working memory capacity can be described by one common factor. A recurring issue is whether working memory capacity is task specific (e.g., Just & Carpenter, 1992) or whether it reflects a single domain-independent factor (e.g., Kyllonen & Christal, 1990; Swanson, 1992; Turner & Engle, 1989). Although this study was not specifically designed to address this issue, the results tend to favor the latter position. Despite variations in the content of the tests (verbal and numeric), they form a single factor. In addition, the relationship of the working memory capacity tests with reading comprehension was, with the exception of the GPASAT, similar to their relationship with arithmetic achievement.

Another issue that might be raised concerns the necessity of a heavy memory load in the measurement of working memory capacity. Tests for working memory capacity should require the simultaneous processing and storage of information (Baddeley, 1986). Often, complex span tests are used (e.g., Just & Carpenter, 1992; Kyllonen & Christal, 1990; Turner & Engle, 1989), which measure the maximum amount of information that can be stored while processing continues. The tests that were used in this study, however,

vary in the amount of storage and processing required. The SCT and the GPASAT seem to make less heavy demands on the storage component than do complex span tasks. In contrast, the Following Directions and the Digit Span Tests (including forward and backward span) are focused primarily on the storage component, although some simultaneous processing is required in these tests as well. However, despite the variations in storage requirements, the tests form one factor, which suggests that a heavy memory component might not be necessary for the adequate measurement of working memory capacity (see also Baddeley, 1993).

Further examination of the nomothetic span of the SCT focused on the relationship of working memory capacity with Gf, speed and school achievement. As expected, a strong relationship was found between working memory capacity and Gf, although the relationship is somewhat weaker than in a similar study by Kyllonen and Christal (1990). Nevertheless, it is striking that such different factors as working memory capacity and Gf can be so highly correlated. The relationship might be explained by the common requirement of simultaneous storage and processing of information in working memory (Carpenter et al., 1990; Kyllonen & Christal, 1990). Interestingly, as in the Kyllonen and Christal study, working memory capacity is more strongly correlated with speed than is Gf. This demonstrates once again that, despite their high association, working memory capacity and Gf are not identical.

A moderate relationship was observed (after omission of the GPASAT) between working memory capacity and general processing speed. Similar results have been obtained in previous studies (Kyllonen & Christal, 1990; Salthouse & Babcock, 1991). The result tends to support the contention that working memory capacity is influenced by the efficiency of carrying out elementary processing operations (e.g., Salthouse & Babcock, 1991).

Some indication was found for an effect of task-specific processing on SCT performance. The time limit, imposed for practical reasons, appeared to be somewhat too strict, and a substantial number of children did not complete the test. A moderate relationship was observed between the number of items attempted and the participant's score on the SCT, which was corrected for the number of items reached. If we consider the number of items that were attempted to be a measure of task-specific processing speed, the result is in accordance with previous studies, which also indicated a relationship between task-specific processing speed and the performance on working memory capacity measures (e.g., Case et al., 1982; Salthouse & Babcock, 1991). One task-specific process of the SCT that might account for speed differences is, as we mentioned earlier, speed of counting. A slower counting speed might increase the probability of an error. Alternatively, variations in speed might be due to the speed of alternation between counting directions. However, the number of alternations was relatively small as compared with the number of stars that had to be counted. Counting speed, therefore, seems to be a more likely explanation of task-specific processing speed, although, of course, both operations might contribute.

As expected, working memory capacity appeared to be strongly related to school achievement, as measured by reading comprehension and arithmetic. The relationship between working memory capacity and achievement confirms once again that working memory is involved in various everyday cognitive tasks, including reading comprehension (Just & Carpenter, 1992) and arithmetic (Hitch, 1978).

Unexpectedly, we had difficulty separating Gf and school achievement, although many studies have shown that a distinction is warranted (e.g., Carroll, 1993). Reading comprehension and arithmetic can be taken as indicators for crystallized intelligence (Gc; see for example Gustafsson, 1984). One could argue that one of our measures of Gf, namely verbal reasoning, might also require Gc and that this may explain the strong relationship between Gf and school achievement. However, Gf is measured by both verbal and figural reasoning. Therefore, we contend that the latent variable formed by these indicators primarily measures Gf. In contrast, as recently shown by Gustafsson and Undheim (1992), indicators of school achievement can load more heavily on Gf than on Gc (see also Marshalek, Lohman, & Snow, 1983). Accordingly, it is more likely that our measures of school achievement are highly reflective of Gf.

In summary, several results of the current study are in support of the construct validity of the SCT. One major outcome, involving the construct representation of the test, is that the SCT requires the involvement of the central executive system of working memory. With respect to its nomothetic span, the SCT is substantially related to other tests of working memory capacity, and the various working memory capacity tests formed one common factor. Furthermore, as was the case with previous research involving young adults, the results indicate a relatively strong relationship between both individual differences in working memory capacity and Gf and a moderate relationship between working memory capacity and speed.

Finally, we would like to comment briefly on the use of the SCT in educational assessment. In the field of education, children with attentional problems are of great concern. A sound theoretical description of the concept of attention is indispensable to successful diagnosis and treatment of these children. Often however, at least as a first step, children with these problems are identified on the basis of behavior ratings by teachers or parents (e.g., Achenbach, Verhulst, Edelbrock, Baron, & Akkerhuis, 1987; Koot & Verhulst, 1992; Verhulst & Akkerhuis, 1986). Although efficient, behavior ratings have several disadvantages (see de Jong & Das-Smaal, 1990; Taylor, 1987). Moreover, behavior ratings do not account for the cognitive aspects of the concept of attention. In contrast, the SCT is an instrument for the assessment of the control function of attention (that is, the regulation of processes in working memory). The test is based on a clear theoretical framework and this study as well as previous studies (Das-Smaal, de Jong, & Koopmans, 1993; de Jong & Das-Smaal, 1990) supports the construct validity of the SCT. However, like other tests of attention (de Jong & Das-Smaal, 1993) and working memory capacity, the SCT is not a pure measure. Administration of the test should always be supplemented by a more fine-grained

analysis of other factors that might be responsible for an exceptionally low score. In contrast to many tests of attention however, the SCT is suitable for administration to groups of subjects and therefore can be efficiently used in large-scale assessment studies. Practically speaking, therefore, the SCT might be especially useful as an initial screening device for attentional problems.

References

- Abraham, A., Clement, R., Markuse, G., Ritsma, C., & Tollenaar, P. (1988). *Ontwikkeling van een screeningsinstrument voor dyslexie* [The development of a screening instrument for dyslexia] (Rep. No. OP1518). Amsterdam: University of Amsterdam.
- Achenbach, T. M., Verhulst, F. C., Edelbrock, C., Baron, G. D., & Akkerhuis, G. W. (1987). Epidemiological comparisons of American and Dutch children: II. Behavioral/emotional problems reported by teachers for ages 6 to 11. *Journal of the American Academy of Child and Adolescent Psychiatry*, 26, 326–332.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baddeley, A. D. (1993). Working memory or working attention? In A. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control* (pp. 152–170). Oxford: Clarendon Press.
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.
- Bentler, P. M. (1989). *EQS: Structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246.
- Butterfield, E. C., Nielsen, D., Tangen, K. L., & Richardson, M. B. (1985). Theoretically based psychometric measures of inductive reasoning. In S. E. Embretson (Ed.), *Test design* (pp. 77–147). New York: Academic Press.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, 33, 386–404.
- Cito (National Institute for Educational Measurement). (1979). *Rekenoetsen leerjaar 2, 3 en 4 basisonderwijs* [Arithmetic tests for grades 2, 3 and 4 of elementary school]. Arnhem: Author.
- Cito (National Institute for Educational Measurement). (1981). *Begrijpend lezen leerjaar 3, 4 and 5 basisonderwijs* [Reading Comprehension for grades 3, 4 and 5 of elementary school]. Arnhem: Author.
- Crammond, J. (1992). Analyzing the basic cognitive-developmental processes of children with specific types of learning disability. In R. Case (Ed.), *The mind's staircase: Exploring the conceptual underpinnings of human thought and knowledge* (pp. 285–302). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Crawford, J. D. (1991). The relationship between tests of sustained attention and fluid intelligence. *Personality and Individual Differences*, 12, 599–611.
- Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.
- 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, attention regulation and the Star Counting Test. *Personality and Individual Differences*, 14, 815–824.
- de Jong, P. F. (1991). *Het meten van aandacht: de constructie van aandachttests voor kinderen* [Measuring attention: The construction of attention tests for children]. Delft, The Netherlands: Eburon.
- de Jong, P. F., & Das-Smaal, E. A. (1990). The Star Counting Test: An attention test for children. *Personality and Individual Differences*, 11, 597–604.
- de Jong, P. F., & Das-Smaal, E. A. (1993). Factor structure of standard attention tests for children: A distinction between perceptual speed and working memory. *European Journal of Psychological Assessment*, 9, 94–106.
- Embretson (Whitely), S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179–197.
- Engle, R. W., Carullo, J. J., & Collins, K. W. (1991). Individual differences in working memory for comprehension and following directions. *Journal of Educational Research*, 84, 254–262.
- Fischer, G. H. (1974). *Einführung in die Theorie psychologischer Tests*. Vienna, Austria: Verlag Hans Huber.
- Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Synthesis of educational productivity research. *International Journal of Educational Research*, 11, 145–152.
- Geary, D. C., & Widaman, K. F. (1992). Numerical cognition: On the convergence of componential and psychometric models. *Intelligence*, 16, 47–80.
- Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic reasoning tasks. *Memory & Cognition*, 21, 115–124.
- Glas, C. A. W. (1988). The derivation of some tests for the Rasch model from a multinomial distribution. *Psychometrika*, 53, 525–546.
- Glas, C. A. W. (1989). *Contributions to estimating and testing Rasch models*. Unpublished doctoral dissertation, Technical University of Twente, Enschede, The Netherlands.
- Goldstein, H., & Wood, R. (1989). Five decades of item response modelling. *British Journal of Mathematical and Statistical Psychology*, 42, 139–167.
- Gronwall, D. M. A., & Sampson, H. (1974). *The psychological effects of concussion*. New Zealand: University of Auckland Bindery.
- Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.
- Gustafsson, J.-E., & Undheim, J. O. (1992). Stability and change in broad and narrow factors of intelligence from ages 12 to 15 years. *Journal of Educational Psychology*, 84, 141–149.
- Healy, A. F., & Nairne, J. S. (1985). Short-term memory processes in counting. *Cognitive Psychology*, 17, 417–444.
- Hitch, G. J. (1978). The role of short-term working memory in mental arithmetic. *Cognitive Psychology*, 10, 302–323.
- Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties. *British Journal of Psychology*, 82, 375–386.
- Hoeks, J. (1985). *Vaardigheden in begrijpend lezen* [Reading comprehension skills]. Unpublished doctoral dissertation, University of Amsterdam, Amsterdam, The Netherlands.
- Horn, W. F., & Packard, T. (1985). Early identification of learning problems: A meta-analysis. *Journal of Educational Psychology*, 77, 597–607.

- Huisman, H. S., Stouthart, M. E. A., & Vorst, H. C. M. (1985). *Henmon Nelson Omnibus Test*. Amsterdam: University of Amsterdam.
- Jansen, R. (1990). *Mental speed and concentration*. Unpublished doctoral dissertation, Catholic University of Nijmegen, Nijmegen, The Netherlands.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99, 122-149.
- Koot, H. M., & Verhulst, F. C. (1992). Prediction of children's referral to mental health and special education services from earlier adjustment. *Journal of Child Psychology and Psychiatry*, 33, 717-729.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389-433.
- Logie, R. H., & Baddeley, A. D. (1987). Cognitive processes in counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 310-326.
- Marshalek, B. L., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence*, 7, 107-127.
- Morris, R. G., & Baddeley, A. D. (1988). Primary and working memory functioning in Alzheimer-type dementia. *Journal of Clinical and Experimental Neuropsychology*, 10, 279-296.
- Nairne, J. S., & Healy, A. F. (1983). Counting backwards produces systematic errors. *Journal of Experimental Psychology: General*, 112, 37-40.
- Navon, D. (1989a). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system. I. Basic tenets. *European Journal of Cognitive Psychology*, 1, 191-213.
- Navon, D. (1989b). The importance of being visible: On the role of attention in a mind viewed as an anarchic intelligence system. II. Application to the field of attention. *European Journal of Cognitive Psychology*, 1, 215-238.
- Neumann, O. (1987). Beyond capacity: A functional view of attention. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 361-394). Hillsdale, NJ: Erlbaum.
- Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In G. E. Schwartz, R. J. Davidson, & D. Shapiro (Eds.), *Consciousness and self-regulation: Advances in research and theory* (Vol 4, pp. 1-18). New York: Plenum Press.
- Raven, J. C. (1965). *Advanced Progressive Matrices, Sets I and II*. London: H. K. Lewis. (Distributed in the United States by The Psychological Corporation, San Antonio, TX).
- Rowe, K. J. (1991). The influence of reading activity at home on students' attitudes towards reading, classroom attentiveness and reading achievement: An application of structural equation modelling. *British Journal of Educational Psychology*, 61, 19-35.
- Salthouse, T. A. (1992). Influence of processing speed on adult age differences in working memory. *Acta Psychologica*, 79, 155-170.
- Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, 27, 763-776.
- Sanders, A. F. (1983). Towards a model of stress and human performance. *Acta Psychologica*, 53, 61-97.
- Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, England: Cambridge University Press.
- Spren, O., & Strauss, E. (1991). *A compendium of neuropsychological tests: Administration, norms and commentary*. New York: Oxford University Press.
- Stankov, L. (1988). Aging, attention, and intelligence. *Psychology and Aging*, 3, 59-74.
- Sternberg, R. J. (1981). Testing and cognitive psychology. *American Psychologist*, 36, 1181-1189.
- Sternberg, R. J. (1985). *Beyond IQ*. Cambridge, England: Cambridge University Press.
- Swanson, H. L. (1992). Generality and modifiability of working memory among skilled and less skilled readers. *Journal of Educational Psychology*, 84, 473-484.
- Taylor, E. (1987). Cultural differences in hyperactivity. In M. Wolraich, & D. K. Routh (Eds.), *Advances in developmental and behavioral pediatrics* (Vol. 8, pp. 125-150). London, England: JAI Press Inc.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, 28, 127-154.
- van den Wollenberg, A. L. (1979). *The Rasch model and time limit tests*. Unpublished doctoral dissertation, Katholieke University of Nijmegen, Nijmegen, The Netherlands.
- van den Wollenberg, A. L., & Creemers, P. G. (1986). On the predictive validity of a new scoring procedure for time-limit intelligence tests. *Tijdschrift voor Onderwijsresearch*, 11, 53-60.
- van Haasen, P. P. (1986). *WISC-R; Nederlandstalige uitgave. Scoring en normen* [WISC-R; Dutch version. Scoring and norms]. Lisse, The Netherlands: Swets & Zeitlinger.
- Verbeek, J., & Troch, F. (1973). *Henmon Nelson Omnibustest, vorm A 6-9*. Amsterdam: Swets & Zeitlinger.
- Verhulst, F. C., & Akkerhuis, G. W. (1986). Mental health in Dutch children: (III) Behavioral-emotional problems reported by teachers of children aged 4-12. *Acta Psychiatrica Scandinavica*, 74, 1-73.
- Vos, P. G. (1988). *Bourdon-Vos-Test*. Lisse, The Netherlands: Swets & Zeitlinger B.V.
- Wittenborn, J. R. (1943). Factorial equations for tests of attention. *Psychometrika*, 8, 19-35.

Received March 3, 1993

Revision received July 20, 1994

Accepted July 26, 1994 ■