German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD
Working Paper Series

Working Paper No. 49

Improvements and future challenges for the research infrastructure in the field "Measuring cognitive ability"

Roland H. Grabner, Elsbeth Stern

December 2008

Federal Ministry
of Education
and Research

Working Paper Series of the Council for Social and Economic Data (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Contact: Council for Social and Economic Data (RatSWD) | Mohrenstraße 58 | 10117 Berlin | office@ratswd.de

Improvements and future challenges for the research infrastructure in the field "Measuring cognitive ability"

Roland H. Grabner and Elsbeth Stern

Abstract

The assessment of cognitive abilities is critical in large-scale survey studies that aim at elucidating the longitudinal interplay between the individual's cognitive potential and socio-economic variables. The format of such studies calls for assessment methods which can not only be administered economically but also display a high (psychometric) measurement quality. In consideration of recent theoretical and empirical advances in intelligence research, we recommend the implementation of tests drawing on working memory in large-scale survey studies. Working memory is a limited-capacity system for temporary storage and processing of information and currently discussed to be the cognitive key system underlying intellectual abilities. Four types of working memory tests are exemplarily described and critically evaluated with regard to their psychometric quality and the need for further evaluation.

Keywords:
cognitive abilities, intelligence, knowledge, information processing, mental speed, working memory

1. Research questions and theoretical developments

The analyses of gene-environment-interaction and evolution are becoming more and more accepted as a research aim in the social sciences (see Spinath, 2008). The basic argument is that without the “control” of genetic effects one cannot be sure that he or she estimates unbiased socio-economic effects (Guo, 2008, Diewald, 2008). In this context cognitive abilities play an important role. Cognitive abilities are the raw material for developing individual resources and are promoted as well as constrained by the socio-economic context. Research on cognitive abilities has revealed considerable associations between an individual’s cognitive abilities (or:“general intelligence”) and numerous indicators of life success, ranging from educational and vocational performance over delinquency to morbidity and mortality (Jensen, 1998; Deary et al., 2004). The causal nature of most of these correlations is still unknown as well as the mediating role of socio-economic variables. This underlines the importance of including cognitive ability measurements in large-scale survey studies to enhance our knowledge about the longitudinal interplay between individual cognitive resources and socio-economic variables.

1.1. A brief history of cognitive ability assessment

The first systematic approach to objectively measure cognitive abilities can be traced back to Sir Francis Galton at the end of the 19th century (cf. Jensen, 1998). Galton developed a variety of tests to measure elementary mental functions such as sensory discrimination and perception speed, guided by the assumption that differences in intellectual abilities result from a differential efficiency of the central nervous system. Galton’s tests were presented to the public in his Anthropometric Laboratory at the International Health Exhibition in London. The interest into the new anthropometric measurements was enormous; between 1884 and 1890 data of more than 9000 persons were collected. The validity of the tests for measuring cognitive abilities, however, was disappointing. The test results turned out to be only poorly correlated with commonsense criteria of intellectual abilities and educational success.

A more promising approach was pursued by Alfred Binet at the beginning of the 20th century (Binet, 1905). Commissioned by the French Ministry of Public Instruction, he should develop a quick and reliable method of distinguishing mentally retarded children, who could not be expected to profit from normal instruction in school, from those with mere behaviour problems. In contrast to Galton, Binet and his colleague Simon devised a battery of tasks drawing on practical knowledge and skills rather than on elementary mental functions. Children should point at various parts of their body, name objects seen in a picture, give definitions, repeat series of digits or a complete sentence, tell the time of a clock, etc. Besides their focus on rather practical skills, Binet and Simon’s approach was innovative in that they used the children’s age as an external criterion for cognitive abilities. By empirically assigning the tasks to different age groups, their intelligence scales allowed the objective assessment whether a child was advanced or backward for his or her chronological age and, thus, to distinguish mentally retarded children from others. This comparison of mental with chronological age provided the basis for the advent of the intelligence quotient (IQ; Stern, 1912), until it was replaced by the concept of today’s statistical deviation IQ (Wechsler, 1944).

The Binet scales were soon translated and distributed in America and England and became the norm against which later intelligence tests were evaluated. The further development of intelligence tests was strongly related to the question of the structure of cognitive abilities. At a gross level, two different views can be distinguished. Some researchers (e.g., Jensen) emphasized the existence and importance of a general intelligence (g) factor, which was originally discovered by Spearman (1904). If a large and random sample of participants completes a number of diverse cognitive tests, the correlations among the different test scores will be almost entirely positive and, in most of the cases, of moderate size. This means that a person who does well in one test also has a high probability to achieve a good performance level in the other tests. Using statistical methods this correlation pattern can be reduced to one single factor (the g factor) which usually accounts for about 50 % of the entire test variance.

Other researchers (e.g., Thurstone or Gardner), in contrast, questioned the existence of the g factor. This diverging view predominantly resulted from the application of different statistical methods in analysing cognitive test performance data or from the expansion of the intelligence concept to non-academic skills (such as interpersonal and bodily-kinesthetic intelligence; Gardner, 1983).

At present, there is wide consensus on a hierarchical model of cognitive abilities, consisting of three levels of different generality (Carroll, 1993; Gustafsson, 1984): At the top and most general level, Spearman's *g* factor can be found, reflecting the fact that diverse cognitive abilities show near-universal positive correlations.

At the second level, group factors of cognitive abilities such as fluid and crystallised intelligence are located. Fluid intelligence is conceptualised as the ability to solve novel problems and is typically assessed by tasks drawing on abstract reasoning (inductive or deductive) or complex problem solving. Crystallised intelligence reflects the breadth and depth of general knowledge and is usually measured by tests on vocabulary, spelling ability, or general information.

Finally, at the lowest level, there are specific cognitive abilities such as quantitative reasoning (for fluid intelligence) or lexical knowledge (for crystallised intelligence), accounting for variance that is neither attributable to factor *g* nor to the group factors. Although hierarchical models with *g* at the top and second- and third-order factors below might best describe the structure of individual differences in cognitive abilities, it is also widely accepted that most of the predictive value of intelligence tests derives from the *g* factor which is strongly related to fluid intelligence (Brody, 1999; Deary, 1998, Jensen, 1998).

1.2. Bases of cognitive abilities

In the past decades, much research has been conducted to better understand the bases of individual differences in cognitive abilities. At present, two cognitive components are discussed which show consistent associations with intelligence and might, therefore, be considered as potential bases of human intelligence. The first component is *mental speed* (cf. Neubauer, 1995). There is a large body of evidence showing consistent negative associations between intelligence and reaction times in so-called elementary cognitive tasks (ECTs). ECTs are designed to place only minimal requirements on the participant and, thus, are less likely to be influenced by differential strategies or prior knowledge. As an example, in the letter-matching task by Posner and Mitchell (1967), the participants have to judge whether two letters are *semantically* identical or not (e.g., semantically identical: "Aa" or "AA" vs. semantically different: "Ab" or "AB"). In a meta-analysis, Neubauer (1995) reported an average correlation of -.33 between mean reaction times and psychometric intelligence test scores. This suggests that brighter individuals display a higher speed of information processing than less intelligent individuals, probably due to a more efficient functioning of their central nervous systems (Jensen, 1998). A central restriction of ECTs represents the rather low effect sizes of the observed correlations. In most cases, correlations do not exceed absolute values of .30; a recent meta-analysis reports a mean correlation of -.24 (Sheppard & Vernon, 2008). Thus, mental speed usually accounts for scarcely more than 10 % of the variance in intelligence tests.

The second potential basis of individual differences in cognitive abilities is *working memory*. Working memory (WM) can be regarded as a limited-capacity system responsible for temporary storage (or maintenance) and processing of information (Baddeley, 2002, 2003). The inclusion of a processing component distinguishes WM from short-term memory (STM) which only supports temporary storage of information. As an example, in a prototypical STM task (forward span), two to nine words are presented sequentially, and the participants are required to recall the words afterwards in the same order. WM tests usually require the execution of a second, additional task. In the original reading span task, for instance, participants read aloud sentences while trying to remember the last word of each sentence for later recall (Daneman & Carpenter, 1980). Individuals differ in the capacity of WM, and these differences have proven to be related to several higher-order cognitive functions ranging from rather domain-specific skills (like reading comprehension; Daneman & Carpenter, 1980; vocabulary learning; Daneman & Green, 1986; or numeracy; De Rammelaere et al., 1999) to (domain-general) intelligence. The actual size of correlation between WM capacity and intelligence as well as the appropriate statistical approach to determine the true relationship are matters of intensive debate (Ackerman et al., 2005; Beier & Ackerman, 2005; Kane et al., 2005; Oberauer et al., 2005). The current estimates range between about .40 and .80; single previous studies reported even higher correlations (up to .96) which led some authors to conclude that WM may be the psychological mechanism underlying (fluid) intelligence (Kyllonen & Christal, 1990; Colom et al., 2004).

The distinction between storage and processing is also reflected in cognitive theories of WM. Probably the most prominent theory was put forward by Baddeley and colleagues already in the 1970s (Baddeley & Hitch, 1974). According to their tripartite model, WM consists of two "slave systems" which are coordinated and controlled by a third system, the so-called central executive. The slave systems enable the temporary storage of information and are domain-specific: Phonologically coded material (verbal and numerical material) is maintained in the phonological loop, visuo-spatial information in the visuo-spatial sketchpad. The central executive component was considered as an attention control mechanism which is responsible for focussing attention to (task-) relevant information, dividing attention if two tasks are performed, and switching attention between different processes and information (Baddeley, 2002).

There is considerable evidence that the central executive component of WM is domain-independent and drives the relationship between WM capacity and intelligence (e.g., Engle et al., 1999; Kane et al., 2004; but see also Colom et al., 2005). More specifically, Conway and colleagues (2003) regarded the "active maintenance of goal-relevant information in the face of interference" (p. 549) as the critical cognitive basis that is shared between intelligence and WM tasks. Support for their view comes from findings that individuals with high and low WM capacity also differ in the performance of low-level attention-control tasks that place practically no memory demands to the participants. In the anti-saccade task, for example, participants have to make an eye-movement (saccade) in the opposite direction of a visual cue (e.g., a flashing light in the periphery). Since the reflexive response would be to orient towards the cue, the attention control demand consists of suppressing this habitual response. Individuals with higher WM capacity were found to display faster and more correct saccades than individuals with lower WM capacity.

2. Status quo

At present, numerous psychometric "intelligence tests" are available. Virtually all of the currently available market tests do a good job at measuring individual differences in cognitive abilities in that they meet the main criteria that are required for a psychometric test, i.e., objectivity, reliability, and validity.

A test displays *objectivity* if the result is independent of the person who administrates, analyses, and interprets the participant's performance. Objectivity is ensured by standardised instructions during administration as well as by clear-cut instructions of how the test scores are determined and interpreted.

Reliability builds upon objectivity and reflects the measurement precision of a test. Reliability is never perfect (1.0) as the test performance is not only influenced by the true cognitive ability of the person but also by random factors such as momentary fluctuations of attention or mood, fatigue, etc. Usually, intelligence tests display reliabilities around .90, indicating that 10 % of the total variance in test performance is due to random factors (i.e., measurement error) and 90 % reflects true variance in intelligence.

Finally, the *validity* of a test reflects to what extent the test measures the trait or ability that it should measure. The validity of intelligence tests is typically evaluated by relating the performance in the test under investigation to an external criterion, either to the performance in a well-established intelligence test or to criteria such as school grades. The large success of the concept intelligence primarily originates in the high validity of intelligence test performance for a lot of performance indicators in diverse areas of life (cf. Jensen, 1998).

In line with the originally intended purpose of intelligence tests, the strongest associations are found with educational variables. Intelligence correlates with school grades at about .50 and with years of education at about .55 (Neisser et al., 1996). Intelligence can also be regarded as a good predictor of vocational success; in a meta-analysis Schmidt and Hunter

(1998) reported an average validity of .51 for overall job performance. Another quality criterion of psychometric tests is the availability of norms so that the individual test performance can be compared with the performance of an age-matched reference sample. The norms in intelligence tests allow the determination of the IQ, reflecting the standardised position of an individual relative to a reference population with a mean of 100 and a standard deviation of 15.

Given their high reliability and validity, intelligence tests can be definitely regarded as the best choice to assess cognitive abilities. Many of the available market tests do not only provide an estimate of the general intelligence of an individual (the IQ) but also inform about his or her cognitive ability structure. The Berlin Intelligence Structure Test (BIS-T; Jäger et al., 1987), for instance, assesses three content facets (verbal, numerical, spatial-figural) and four operational facets (processing capacity, creativity, memory, and speed) of cognitive abilities with general intelligence as the integral of all ability facets.

The administration of such an intelligence structure test, however, is very costly, predominantly in terms of time. The full version of the BIS-T takes over 2 hours. But even one-dimensional intelligence tests focusing on general intelligence such as the Raven's Advanced Progressive Matrices (Raven, 1958) require a test time of at least 20-30 minutes in their short version. Thus, if we want to disentangle the impacts of cognitive abilities and socio-economic effects on outcomes of human lives there is a strong need for the development of more short cognitive ability assessment procedures that can be applied in large-scale surveys.

Lang and colleagues (2007) recently proposed two ultra-short tests for the measurement of intellectual abilities in the German Socio-Economic Panel (SOEP). One test (the symbol-digit test; SDT) requires the fast assignment of numbers to symbols following a pre-defined number-symbol pairing. In the other test (the Animal Naming Task; ANT), participants have to produce as many animal names as possible within a 90 second time interval. The reliabilities of both tests were reported to be around .90 for the SDT and around .65 for the ANT. Their validities for general intelligence however, were not investigated, but can be expected to be rather low. The SDT draws on mental speed, and the performance in similar task versions was found to be only weakly related to intelligence (Conway et al., 2002). Likewise, the ANT only samples knowledge in a certain domain which turned out to be correlated only between .33 and .39 with broader vocabulary knowledge (Lang et al., 2007).

3. Future developments

In consideration of the recent theoretical insights on the cognitive bases of intelligence and the consistent strong relationship between WM (Working Memory) capacity scores and higher-order intellectual abilities, it appears most promising to further develop short tests drawing on WM or its sub-components. In contrast to intelligence problems, WM tasks typically require only simple cognitive operations whose sequence is highly restricted by the instructions. The difficulty of working-memory tasks arises from the additional load on some facets of the cognitive architecture (Süß et al., 2002). The reading-span task described above, for example, requires continuous updating of the content of WM (with every sentence one new word needs to be memorised) and the maintenance of the words in spite of interference (i.e., reading sentences aloud).

Overall, WM tests offer the following advantages:

(a) Their administration takes shorter time than that of intelligence tests.

(b) Most of these tasks can be implemented in computer-aided testing environments.

(c) According to the current view in research, they tap the central basis of cognitive abilities.

(d) WM tasks are typically less influenced by prior knowledge than intelligence tests.

(e) The limiting factor of WM capacity (central executive) seems to be domain-independent.

To date, the development of WM tests is by far not as advanced as the development of intelligence tests. WM span tasks (such as the reading span task described above) belong to the first WM measures that have been developed and are meanwhile already well-understood which is reflected in the existence of methodological reviews and user's guides (Conway et al., 2005). The psychometric quality of other WM tasks (e.g., focusing on executive processes) is more difficult to evaluate due to the scarcity of studies with larger samples. In the following, an overview of WM tasks that could be employed in the large-scale survey studies is provided.

3.1. Traditional WM span tasks

Since the early reading span task described above, several versions of WM span tasks have been developed. Three key tasks can be identified (Conway et al., 2005; Kane et al., 2004). In the (newer version of the) *reading span task*, the participant is presented with a meaningful or meaningless sentence and a to-be-remembered letter (e.g., "We were fifty lawns out at sea before we lost sight of land. ? X"). Participant's task is to read the sentence, judge whether it makes sense or not, read and remember the letter. The *operation span task* requires judging the correctness of an arithmetic equation and to remember an additionally presented word (e.g., "Is (6 x 2) – 5 = 7? class"). In the *counting span task* participants have to count the number of dark blue circles in displays with other distracting objects (dark blue squares and green circles) and to remember the counted number. All these tasks are designed to force storage of information in the face of processing.

Conway et al. (2005) emphasised three critical task features: First, rehearsal must be avoided by presenting the next stimulus immediately after completion of the preceding one. Second, the timing of the task needs to be adaptive. Both properties are met in current computer versions in which the to-be-remembered stimulus is displayed immediately after completion of the interfering task (e.g., judging the correctness of an equation). Third, the number of stimuli within one item needs to be sufficient. A range from two to five stimuli per item turned out to be adequate for most college students.

The administration of a WM span task with 12 items (with two to five stimuli each) including instruction and practice items takes about 10 minutes. Besides the verbal WM span measures described above, a number of figural-spatial versions have been devised (Kane et al., 2004). As an example, in the symmetry span task, participants have to judge whether a figure in an 8 x 8 matrix is symmetrical or not and to remember the position of a red square in a subsequently presented 4 x 4 matrix.

The reliabilities of WM span tasks are usually in the range between .70 and .90, suggesting good measurement precision for a single test. Their validity for intelligence test performance lies around .50 (Kane et al., 2004).

3.2. Transformation span tasks

In this type of WM tasks, participants are not required to simultaneously store and process information but rather to perform some mental transformation on the stored information. A promising example is the alpha span task, originally developed by Craik (1986). Three to seven words are successively presented to the participant who is required to memorise them. After presenting the last word, the participant has to repeat the first letter of each word in alphabetical order, thus, requiring an alphabetical re-ordering of the memorised words. Süß et al. (2002) presented one item with three words and two items with four, five, six, and seven words each, requiring an estimated test time of about 5 minutes including instruction.

The authors reported a reliability of .81 and a validity for general intelligence of .55. Other studies, however, report much lower validities for similar transformation tasks (e.g., the backward span task requiring the recall of the presented words in reverse order; Engle et al., 1999).

3.3. Dynamic WM tasks

A separate class of WM tasks that are frequently used in neuroscience research require the continuous monitoring and updating of the maintained information. In the prominent n-back task, a list of stimuli (words, numbers, or figures) is successively presented, and the individual has to continuously report whether each stimulus matches the one that had appeared *n* items ago (n-back). In a 2-back task, for instance, participants have to continuously maintain the last 2 stimuli of the list which means that they have to update the content of their WM with every new stimulus and to drop out the least recent one. Even though the n-back task is considered the gold standard in neuroscience research, there is mixed empirical evidence on the question whether this task draws on the same cognitive resources as the well-established WM span tasks (Conway et al., 2005; Kane et al., 2007).

Kane et al. (2007) investigated the construct validity of the n-back task in a sample of 129 young adults and found that the performance in the operation span task and the n-back task was only weakly associated (correlations did not exceed .25). In addition, both tasks accounted for independent variance in general intelligence. These findings suggest that the n-back task does not measure the same WM processes as the operation span task.

3.4. Executive control tasks

Executive processes related to attentional control are central in Baddeley's model of WM and are assumed to play a critical role in the relationship between WM capacity and intelligence. The development of tasks demanding these processes without strong reliance on storage, however, appears to be a great challenge. Süß et al. (2002) as well as Oberauer et al. (2003) have devised tasks requiring task set switching, i.e., the inhibition of an active action schema and the selection of another. In the numerical switching task by Süß et al. (2002), displays with varying number of digits are presented. The participant is required to alternate between reading the digits and counting them; which task was to perform was displayed on the top of the display. In the figural version, a round and an angular figure appears in each display, one left and one right. Participants have to indicate the side of either the angular or the round figure. Finally, in the verbal version, participants have to switch between two semantic categories in determining the presentation side of words. Similar to the transformation span tasks, these tasks can be administered within a few minutes.

Süß et al. (2002) report reliabilities between .78 (numerical) to .94 (verbal and figural) and validities between .33 (figural) and .58 (numerical) for general intelligence. Later research, however, has questioned the construct validity of these tasks as they are only weakly related to traditional WM span tasks (Oberauer et al., 2003, 2005) and reflect processing speed more strongly than reasoning abilities (Süß et al., 2002).

4. Conclusions and recommendations

In the past decades, considerable advances have taken place in understanding the individual differences in cognitive abilities and in the development of psychometric tests for ability assessment. Present research regards WM (Working Memory), reflecting a limited-capacity system supporting temporary storage and processing of information, as the cognitive key system underlying intellectual abilities.

Measures of WM capacity have been found to display substantial correlations with several domain-specific intellectual abilities as well as with intelligence, representing the epitome of domain-general cognitive abilities. Thus, tests assessing WM capacity or executive functions appear to be a more promising method for the cognitive ability assessment in large-scale survey studies than tests focusing on mental speed or surface knowledge in a certain domain.

Several candidate tasks have been described above which can be administered in considerably shorter time than psychometric intelligence tests. In addition, their task characteristics allow the presentation in computer-aided testing environments. The internet seems to offer the ideal infrastructure for the implementation of the cognitive ability screening. The coverage is very high, and it is meanwhile not longer only accessible from the personal computer (at home or at the office) but increasingly also from mobile devices such as netbooks, mobile phones, or personal digital assistants (PDA). So it becomes more and more easy to administer those tests in large-scale surveys.[1]

However, it should be noted that most of these WM tasks are still in the phase of development, and studies with larger samples, which would allow a more accurate evaluation of their reliability and (construct) validity, are very scarce. Thus, some starts would be very helpful. Although the future challenge is to improve the psychometric quality of these tests they should be administered in large-scale surveys. In fact, the data of the large-scale surveys can further contribute to their improvement. The actual reliability of these tests could be accurately quantified and norms for age-matched reference samples, which are presently almost completely missing for WM tests, could be easily established. In addition, the data from large-scale studies can also inform about their validity for indicators of life success. Parallel to these criteria, their validity for intelligence needs to be further investigated.

---

[1] The tests could be offered and advertised, for instance, in virtual social networks such as Facebook.

References

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

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829-839.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*, 85-97.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Beier, M. E., & Ackerman, P. L. (2005). Working memory and intelligence: Different constructs. Reply to Oberauer et al. (2005) and Kane et al. (2005). *Psychological Bulletin, 131*, 72-75.

Binet, A. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, *12*, 191-244. (Translated in 1916 by E. S. Kite in *Development of intelligence in children.* Vineland, NJ: Publications of the Training School at Vineland.)

Brody, N. (1999). What is intelligence? *International Review of Psychiatry, 11,* 19-25.

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

Colom, R., Flores-Mendoza, C., Quiroga, M. A., & Privado, J. (2005). Working memory and general intelligence: The role of short-term storage. *Personality and Individual Differences, 39*(5), 1005-1014.

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

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

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

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*(5), 769-786.

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). Amsterdam: North-Holland.

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

Daneman, M., & Green, I. (1986). Individual-differences in comprehending and producing words in context. *Journal of Memory and Language, 25*, 1-18.

De Rammelaere, S., Stuyven, E., & Vandierendonck, A. (1999). The contribution of working memory resources in the verification of simple mental arithmetic sums. *Psychological Research-Psychologische Forschung, 62*, 72-77.

Deary, I. J. (1998). Differences in mental abilities. *British Medical Journal, 317*, 1701-1703.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The Impact of Childhood Intelligence on Later Life: Following up the Scottish Mental Surveys of 1932 and 1947. *Journal of Personality and Social Psychology, 86*(1), 130-147.

Diewald, M. (2008). Zwillings- und Adoptivkinder-Stichproben für soziologische Analysen? Zum Ertrag verhaltensgenetischer Ansätze für sozialwissenschaftliche Fragestellungen und Erklärungen (Using twin and adoptive child samples for sociological analysis? On the value of behavioral genetics approaches for social scientific issues and interpretations). DIW Research Notes, 27. Berlin: DIW Berlin

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, 128*, 309-331.

Gardner, H. (1983). *Frames of mind.* New York: BasicBooks.

Grigorenko, E. L. (2007). How Can Genomics Inform Education? *Mind, Brain, and Education, 1*, 20-27.

Guo, G. (2008). Introduction to the special issue on society and genetics. *Sociological Methods & Research, 37*, 159-163.

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

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

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: A question of construct validity. *Journal of Experimental Psychology-Learning Memory and Cognition, 33*(3), 615-622.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*, 66-71.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W. & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189-217.

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

Lang, F. R., Weiss, D., Stocker, A. & von Rosenbladt, B. (2007). Assessing Cognitive Capacities in Computer-Assisted Survey Research: Two Ultra-Short Tests of Intellectual Ability in the German Socio-Economic Panel (SOEP). *Schmollers Jahrbuch, 127*, 183-192.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77-101.

Neubauer, A. (1995). Intelligenz und Geschwindigkeit der Informationsverarbeitung. Wien: Springer.

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

Oberauer, K., Suss, H. M., Wilhelm, O., & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*(2), 167-193.

Raven, J. C. (1958). *Advanced Progressive Matrices*. London: Lewis.

Schmidt, F. L., & Hunter, J. E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin, 124*(2), 262-274.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences, 44*(3), 535-551.

Spearman, C. (1904). „General intelligence“, objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Spinath, F. M. (2008). Improvements and Future Challenges in the Field of Genetically Sensitive Sample Designs. RatSWD Working Paper 45/2008.

Stern, W. (1912). Die psychologischen Methoden der Intelligenzprüfung. Leipzig: Barth.

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

Wechsler, D. (1944). *Measurement of adult intelligence*. Baltimore: Williams and Wilkins.