Theory-based construction and validation of a modern computerized intelligence test battery

THEORY-BASED CONSTRUCTION AND VALIDATION OF A MODERN COMPUTERIZED INTELLIGENCE TEST BATTERY

Markus Sommer[1], Martin Arendasy[2], Joachim Häusler[1]

[1]Dr.G.Schuhfried GmbH
Hyrtlstr. 45
2340 Mödling
Austria

[2]Faculty of Psychology
Dept. of Clinical, Biological & Differential Psychology
Liebiggasse 5
A-1010 Vienna
Austria

sommer@schuhfried.at

THEORETICAL INTRODUCTION

In personnel selection the use of standardized intelligence test batteries became increasingly common (Schuler, 2000). The selection of the individual tests is commonly based on a thorough work place analysis and empirical studies on the construct validity and criterion validity of the test batteries. According to Embretson (1983) construct representation comprises an investigation of the nomothetic span and the construct representation of the individual tests which constitute a given test battery. While there is often ample evidence on the nomethetic span and the criterion validity of modern test batteries evidence on construct representation (Embretson, 1983) of psychometric tests is often sparse.

In order to fill this gap a new intelligence test battery (INSBAT) was developed based on the hierarchical G_f-G_c theory (Horn, 1989; Horn & Noll, 1997) and the three-stratum-theory proposed by Carroll (1993; 2003) using various means of theory-guided item construction in order to ensure the construct representation (Embretson, 1983) of the item material used in the various subtests.

The Cattell-Horn-Carroll model (CHC-model) was chosen as a starting point for the construction process since that model of intelligence has been confirmed in a number of previous studies using different test batteries. For instance Brickley et al. (1995) investigated the factorial structure of the Woodcock-Johnson battery using a large sample of respondents aged from six to ninety years using confirmatory factor analysis. The author was not only able to identify eight of the nine broad second order factors proposed by Horn (1989; Horn & Noll, 1997) but also demonstrated the factorial invariance of the model with regard to age. However, contrary to the theoretical assumptions of the G_f-G_c theory Brickley et al. (1995) also identified a g-factor which explains the correlations among the eight second-order factors. These results are also in line with the ones reported by Carroll (2003), who used different subtests. Roberts, Goff, Anjoul, Kyllonen, Pallier und Stankov (2000) investigated the factorial structure of the Armed Service Vocational Aptitude Battery and several subtests taken from the ETS kit of referenced cognitive tests. The authors were able to confirm all second-order factors proposed by Horn (1989; 1991; Horn & Noll, 1997) with the exception of G_q and G_{DS} in two independent samples.

While the construct representation of the individual subtests was investigated in several prior studies evidence on the nomethetic span of the newly developed test battery was

not provided thus far. The present study thus fills this gap by investigating the nomothetic span of the newly developed test battery within the CHC-model

METHOD

Measures

The test battery used in this study comprised four measures of quantitative reasoning (G_q), three measures of fluide intelligence (G_f), two measures of crystalized intelligence (G_c), two measures of short-term memory (G_{stm}) and one measure each for long-term memory (G_{ltm}) and visual processing (G_v) taken from the intelligence-structure-battery (INSBAT: Arendasy et al., 2004). Each subtest used to measure the six stratum two abilities shares the common feature of Rasch-homogenity as indicated by non-significant model fit tests (for a detailed discussion see Arendasy et al., 2004).

Subtest: Algebraic Reasoning (AD)

The subtest 'Algebraic Reasoning' consists of the k=15 algebra word problems. The items were presented in a fixed linear order with no time restriction at the item level or at the level of the subtest itself. The respondents were instructed that they are going to see several algebraic word problems and need to find out the numerical value of the unknown element. The subtest uses a free response item format where the respondent had to type in the numerical value of the unknown element using the calculater answer format of the Vienna test system. The construction of the item material is based on current theoretical models or word algebra problem solving (e.g. Koedinger & Nathan, 2004; Sebrecht, Enright, Bennett & Martin, 1996). The construction rational contributed at R=.88 to the 1PL item difficulty parameters. The reliability of this subtest measured by Cronbach α amounts to .88.

Subtest: Arithmetical Fluency (NF)

The subtest 'Arithmetical Fluency' consists of k=17 items. Each item comprises of several operands and a result value located on the right side of the arithmetic problem. The respondents had to fill in the missing operators in order to complete the given arithmetic problem at hand. In order to do so the respondents had to select one of four basic arithmetic operatos (plus, minus, multiply, divide) to connect each adjuncted operands. The subtest was presented as a linear test with a fixed item order and no time limitation. The construction rational (arendasy et al., 2004) contributed at R=.91 to the 1PL item difficulty parameters.Furthermore, in a study on the development of the computerized adaptive version of this item material Arendasy and Sommer (submitted) report, that the item material does not only fit the Rasch model, but also report a fit of the Linear Logistic Test Model (Fischer, 1995) at α=.01. Taken together these results indicate a high psychometric level of the item material and the presence of construct representation as defined by Embretson (1983). In the present study a Cronbach α value of .88 was observed.

Subtest: Computational Estimation Accuracy (ASF)

The subtest 'Computational Estimation' consists of k=20 computational estimation problems measuring respondents' number sense as part of the second-order factor G_q. Each item consists of multistep and multidigit arithmetic problems together with four answer alternatives. None of the answer alternatives corresponds to the numerically exact solution to the given computational estimation problem. The task of the respondents was to estimate the outcome of the given arithmetic problem and select the answer alternative closest to the exact solution. The construction of the item material is based on the theoretical model of computational estimation problem solving outlined by LeFevre, Greenham and Waheed

(1993). The construction rational derived from this model contributed at R=.86 to the 1PL item difficulty parameters. Furthermore, in a study on the development of an adaptive version of this subtest Arendasy and Sommer (submitted) report, that the item material fits a Linear Logistic Test Model (Fischer, 1995) at α=.01. The measurement accuracy of the linear version used in this study amounts to an Cronbach α value of .71.

Subtest: Arithmetic Competence (AK)

The subtest 'Arithmetic Competence' consists of k=23 multistep arithmetic problems which were constructed based on current theoretical models in mental arithmetic (e.g. Ashcraft, 1995; Campbell & Xu, 2001; for an overview see: Arendasy et al., 2004). The subtest uses a free response item format where the respondent had to type in the numerical value of the solution using the calculater answer format of the Vienna test system and is presented as a linear test with a fixed item order and a time limit of 45 sec. for each individual item. According to the manual 70 percent of the variance of the item difficulty parameter could be explained by construction rational derived from the theoretical model. Cronbach α amounts to .74.

Subtest: Figural-Inductive Reasoning (FID)

The subtest 'Figural-Inductive Reasoning' constitutes a computerized adaptive test and measures figural-inductive reasoning designed to measure inductive reasoning as as part of the second-order factor G_f. The construction of the 3x3 matrices items was based on cognitive models of figural matices problem solving. According to Hornke (2002; Hornke, Etzel & Küppers, 2000) the construction rational contributed at R=.72 to the 1PL item difficulty parameters. The task of the respondent is to select one of eight answer alternatives which completes the given item. The test is administered without a time limit at the item level using the following stopping rules: Minimal number of items administered=10, maximum number of items administered=15, maximum time=30 minutes and standard error of measurement (SEM) =.55. The chosen SEM value results into a reliability coefficient of .70.

Subtest: Numerical-Inductive Reasoning (NID)

The subtest 'Numerical-Inductive Reasoning' consists of k=19 number series items. Each item consists of a series of numbers which are governed by one or more rules. The task of the respondents is to infer the rule and complete the number series by adding another number which follows the construction rule of the number series in question. The subtest was presented as a linear test with a fixed item order and no time limitation. It uses an open response item format which requires the respondent to type in the numerical value of the unknown number using the calculater answer format of the Vienna test system. The construction of the item material is based on a theoretical model of the solution processes used by respondents to solve number series items (Holzman, Pellegrino & Glaser, 1982; LeFevre & Bisanz, 1986; Haudeck, 1973). The construction rational contributed at R=.92 to the 1PL item difficulty parameters. Cronbach α amounts to .90.

Subtest: Verbal-Deductive Reasoning (VDD)

The subtest 'Verbal-Deductive Reasoning' consists of k=23 syllogistic reasoning problems which were designed to measure verbal-deductive reasoning abilities as as part of the second-order factor G_f. The respondents were presented with two premisses and a set of four conclusions which might be drawn from these two premisses. In addition the respondents were also presented with the answer alternative 'None of the conclusions is valid.' to reduce the guessing probability. The task of the responents was to identify to conclusion which can be logically drawn from the given premisses. The construction of the item material was based on previous research conducted by Srp (1993) and Spada and Scheiblechner (1973).

According to the manual the item difficulty parameter vary substantially with the typ of syllogism and the content of the items (for details see: Arendasy et al., 2004). This subtest is presented in as a linear test with a fixed item order and a item time limitation of 45 seconds for each item. the present study a Cronbach α value of .77 was observed.

Subtest: Verbal Comprehension (WS)

The subtest is a computerized adaptive test (CAT) which requires the respondents to define a given word by filling in two missing parts in a cloze using two sets of four distractors each. According to Wagner-Menghin (2004) this subtest constitutues one of the core measures of crystallized intelligence (G_c) as indicated by a serie of correlational studies conducted by the author (for details see: Arendasy et al., 2004). The test is administered without a time limit at the item level using the following stopping rules: Minimal number of items administered=10, maximum number of items administered=25, maximum time=30 minutes and standard error of measurement (SEM) = .56. The chosen SEM value results into a reliability coefficient of .75.

Subtest: Verbal Production (VP)

The subtest consists of k=20 items measuring word fluency as part of the second-order factor G_c. Each item consists of a serie of letters which need to be ensambled into a noun. The subtest uses a free response item format where the respondent has to click the indivudual letters of the serie in the correct order. This subtest is administered as a linear test with a fixed item order and a time limit of 45 sec. for each individual item. The manual reports a Cronbach α value of .72. In the present study the reliability amounts to .69.

Subtest: Visual Short-term Memory (VIK):

The subtest 'Visual Short-term Memory' was presented as a computerized adaptive test. The items construction process was based on a construction rational based on the theorey of mental imagery by Kosslyn (1980) as well as the model of integrative information processing proposed by Hänggi (1989). According to Hornke (2002) the construction rational contributed at R=.94 to the item difficulty parameters. The respondents are see a city map on the screen, which indicates typical sites (e.g. restautant) using symbols. The task of the respondents is to memorize the locationing of these symbols and retrieve them subsequently by marking the location on an empty mapp where they assume the symbols have been presented before.The test is administered without a time limit at the item level using the following stopping rules: Minimal number of items administered=10, maximum number of items administered=15, maximum time=15 minutes and standard error of measurement (SEM) =.55. The chosen SEM value results into a reliability coefficient of .70.

Subtest: Verbal Short-term Memory (VEK):

The subtest 'Verbal-Short-term Memory' was presented as a computerized adaptive test. The items construction process was based on a construction rational derived from the theory of serial learning (Ebbinghaus, 1885) and the dual coding theory by Paivio (1971). According to Hornke (2002) the construction rational contributed at R=.94 to the item difficulty parameters. The respondents see the route of a bus on a map. The bus stops at a predefined number of stations in each item. The stations are presented serially. The task of the respondents is to memorize these bus stations and and retrieve them subsequently in the correct order. The test is administered without a time limit at the item level using the following stopping rules: Minimal number of items administered=10, maximum number of items administered=15, maximum time=15 minutes and standard error of measurement (SEM) =.55. The chosen SEM value results into a reliability coefficient of .70.

Subtest: Long-term Memory (LTM):

The subtest consists of a memorization task where the respondents are required to memorize the various information about eight subjects and a retrieval phase. In the retrieval phase the respondets have to solve k=26 items. In each item a specific information is asked about one or more subjects from the memorization phase. This subtest is administered as a linear test with a fixed item order and a time limit of 30 sec. for each individual item. According to the Arendasy et al. (2004) the 1PL item difficulties vary with the complexity and the amount of the information which has to be retrieved from memory. In the present study the measurement accuracy of this subtest (Cronbach α) amounts to .72.

Subtest: Spatial Comprehension (RV):

The subtest consists of k=17 items measuring mental rotation as part of the second-order factor G_v. Each item consists of one target cube and the left and six comparison cubes on the right side in addition to the answer alternative 'no answer is correct'. The task of the respondents is to mentally rotate the six cubes on the right in order to find out which one is identical to the target cube. In all items one of the six comparison cubes represents the correct solution. The reason for the presentation of the answer alternative 'no answer is correct' is to further reduce guessing probability. The items were constructed based on an explicite construction rational (for details see Gittler, 1992) which has been confirmed in multiple LLTM-analysis. The reliability of this subtest measured by Cronbach α amounts to.84.

Sample

The sample encompasses 101 (51.5%) male and 95 (48.5%) female respondents aged between 17 to 65 years (M=37.74; SD=11.94). A total of 17 (8.5%) respondents completed nine years of school but no vocational training, while 75 (38.3%) respondents also completed a vocational school. 84 (42.9%) respondents had a high school leaving certificate with university entrance permission, and 21 (10.7%) respondents graduated from university or college.

RESULTS

The data were analysed using AMOS 5.0 (Arbuckle, 2003). a hierarchical confirmatory factor analysis is conducted using Maximum Likelihood estimation to calculate the parameters of the theoretically postulated model. all measures meet the standard criteria for univariate normality as indicated by the skew and kurtosis values for the individual measures (c.f. Kline, 1998). Furthermore the data were screened for univatiate and multivariate outliners. Univariate outliners were defined as cases more than 3.5 standard deviations from the mean while multivariate outliners were examined using Mahanalobis' d^2. Since none of the respondents met the criteria for univariate or multivariate outliners all respondents were mantained in this analysis.

Based on the modified G_f-G_c theory (Horn, 1989; Horn & Noll, 1997) and the latest version of the three-stratum-theory (Bickley et al., 1995; Carroll, 2003) it was assumed, that the the observed correlations between the subtests can be explained by six latent factors which correspond to the second-order factors of the modified G_f-G_c theory and the three-stratum-theory and a higher-order g-factor. The model assumes, that the subtests 'Algebraic Reasoning', 'Numerical Fluency', 'Computational Estimation' and 'Arithmetic Competence' load on G_q while the subtests 'Numerical-Inductive Reasoning', 'Figural-Inductive Reasoning' and 'Verbal-Deductive Reasoning load on a G_f-factor. The subtests 'Verbal Comperehension' and 'Verbal Production' were assumed to load on G_c. G_{stm} is marked by the subtests 'Verbal Short-term Memory' and 'Visual Short-term Memory' while the subtests 'Spatial Comprehension' and 'Long-term Memory' load on G_{ltm} and G_v respectively. Since the later two factors are only marked by a single subtest these two subtests were split into

halves using an odd-even split procedure. Furthermore, the correlation of these three latent factors should be explained by a higher-order G-factor (cf. (Bickley et al., 1995; Carroll, 2003).

The fit of the theoretically postulated model to the empirical data set was evaluated using χ^2-value of the likelihood ratio test whereby a non-significant χ^2-value indicates, that the sample covariance matrix and the covariance matrix implied by the theoretically do not differ statistically from each other. Because of the sensitivity of this fit index to sample size additional fit indices are used to evaluate the fit of the theoretically postulated model to the empirical data. The χ^2/df statistic was one of the first fit indices proposed. According to Byrne (1989) a χ^2/df ratio < 2 indicates a good fit of the theoretically postulated model to the empirical data. The CFI compares the theoretically postulated model with the independence model and is able to take the sample size into account and thus avoids the tendency to underestimate the model fit in smaller samples. A CFI-value of .95 or above can be seen as an indicator of a good model fit (Byrne, 2001; Hu & Bentler, 1999; Schuhmaker & Lomax, 2004) while a value of .90 or above indicates and adequate fit (Backhaus et al., 2004). The root mean square error of approximartion (RSMEA) is considered to be one of the most informative fit indices (Byrne, 2001; Browne & Cudeck, 1993; Schuhmaker & Lomax, 2004). The RSMEA takes into account the error of approximation in the population. This discrepancy measure is expressed per degree of freedom and thus also takes into account the complexity of the theoretically postulated model. According to Browne and Cudeck (1993), Schuhmaker and Lomax (2004) and Byrne (2001) RSMEA values < .05 indicate a good model fit wereas values < .08 show an adequate fit of the model. The fit statistics or the theoretically postulated model are presented in table 1.

Table 1: Goodness of fit indices for the hierarchical confirmatory factor analysis

Model	χ^2	df	p	χ^2/df	CFI	RSMEA
theoretical model	153.88	85	<.001	1.81	.94	.06

Even though the χ^2-statistic is significant the χ2/df, the CFI and the RSMEA indicate an adequate fit of the theoretically postulated model to the data. It can thus be concluded, that the theoretically postulated model fits the empirical data reasonable well. Furthermore, all factor loadings reached statistical significance at α=.01. The standardized factor loadings are presented in figure 1.

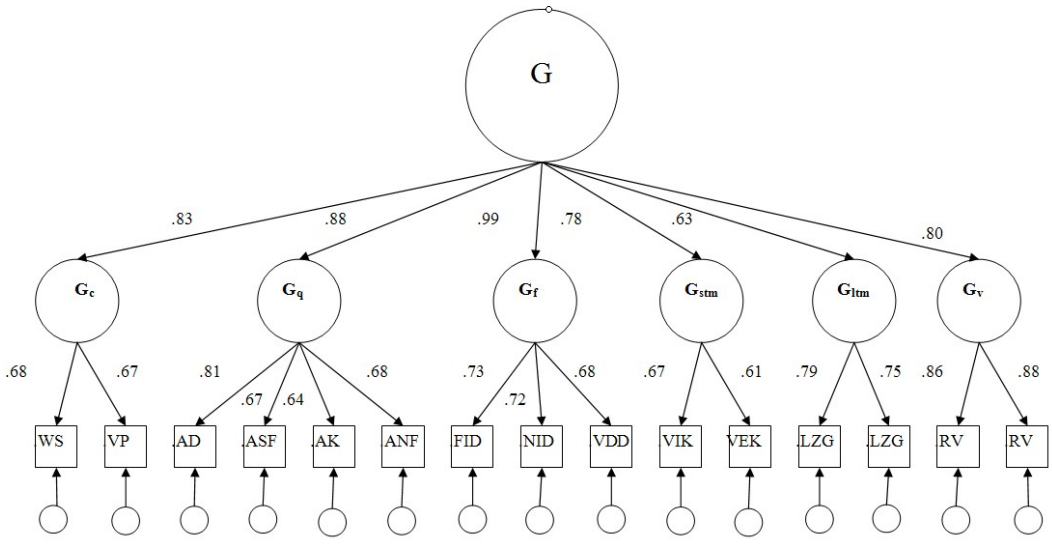

Figure 1: Standardized factor loadings for the theoretically postulated model

As can be seen in figure 1 the subtest 'Algebraic Reasoning' has the highest factor loading on G_q followed by 'Computational Estimation' 'Arithmetical Fluency' and 'Arithmetic Competence'. The factor G_f is primary defined by the subtest 'Figural-Inductive Reasoning' followed by the subtests 'Numerical-Inductive Reasoning' and 'Verbal-Deductive Reasoning', while the G_c loads primary on the subtest 'Verbal Comprehension' followed by the subtest 'Verbal Production'. The factor G_{stm} on the other hand is primary marked by the subtest 'Visual Short-term Memory' followed by the subtest 'Verbal Short-term Memory' while the factor loadings of the two parcels of the subtests 'Long-term Memory' and 'Spatial Comprehension' turned out to be similar in magnitude. The factor loadings of the individual tests on the respective second-stratum factor of the modified G_f-G_c theory (Horn, 1989; Horn & Noll, 1997) and the three-stratum-theory (Bickley et al., 1995; Carroll, 2003) are thus in line with the theoretical assumptions.

With regard to the factor loadings of the g-factor on the six broader stratum two factors the results indicate, that the G-factor is primary marked by G_f. The standardized factor loading turned out to be close to 1 and the error varince did not reach the significance level. This result is similar to the results obtained by several other Gustafsson (1984; Undheim & Gustafsson; 1987) using different test batteries. In order to test whether G_f and the g-factor are indeed indistinguishable the theoretically postulated model was recalculated with the additional restriction that the standardized factor loading from the g-factor to G_f equals 1. The restricted model yeiled a χ^2-statistic of 154.91 with df=84 and p<.001. In order to investigate wether the restricted model fits the data worse than the original theoretically postulated model a $\Delta\chi^2$ test was used which resulted in a value of 1.03 with df=1 and p=.311 indicating, that G_f and the g-factor are indeed indistinguishable.

DISCUSSION

Proir research indicates, that the one-parameter logistic item difficullty estimates vary substantially with the item design rules (for an overview see Arendasy et al., 2004) and thus argue for the construct representation (Embretson, 1983) of the indivudual subtests. Furthermore, the results of the present study indicate, that the nomothetic span of INSBAt can be assumed with regard to the hierarchical G_f-G_c theory (Horn, 1989; Horn & Noll, 1997) and the three-stratum-theory proposed by Carroll (1993; 2003). All subtests significantly load on the corresponding broder second stratum factor. However, these six factors vary in breath. In particular Gv is somewhat narrowly defined in the current version of INSBAT and might be better labeled as 'visualization' (V). The breath of this factor should be increased by adding further subtests measuring related but still different aspects of G_v. The magnitude of factor loadings of the broad second strtaum factors on g is in line with results reported by Gustafsson (1984; Undheim & Gustafsson, 1987) and Brickley et al. (1995) using different test batteries. The results also confirm prior results reported by Gustafsson (1984; Undheim & Gustafsson, 1987) which indicates, that G_f and the g-factor are virtually indistinguishable. Taken together the present results of this study demonstrate, that psychometric tests can be generated using a theory-guided item generation approach for a various intellectual abilities which are relevant to the prediction of educational or job-related success.

REFERENCES

Anastasi, A. (1989). Ability testing in the 1980's and beyond: Some major trends. Public Personnel Management Journal, 18, 471-484.

Arbuckle, J. L. (2003). Amos 5.0 Update to the Amos User's Guide. Smallwaters Corporation, Chicago IL.

Arendasy, M., Hornke, L. F., Sommer, M., Häusler, J., Wagner-Menghin, M., Gittler, G., Heidinger, C. & Herle, M. (2004). Manual Intelligenz-Struktur-Batterie (INSBAT). Mödling: Dr. G. Schuhfried GmbH.

Ashcraft, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. Mathematical Cognition, 1, 3-34.

Austin, J. T. Humphrey, L. G. & Hulin, C. L. (1989). Another view of dynamic criteria: A critical reanalysis of Barrett, Caldwell, Alexander. Personnel Psychology, 42, 593-596.

Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2004). Multivariate Analysemethoden: Eine anwendungsorientierte Einführung (10. Auflage). Berlin: Springer.

Bickley, P. G., Keith, T. Z. & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. Intelligence, 20 (3), 309-328.

Browne, M. W. & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp. 445-455). Newbury Park: Sage.

Byrne, B. M. (1989). A primer of LISREL : Basic applications and programming for confirmatory factor analytic models. New York: Springer.

Byrne, B. M. (2001) Structural equation modeling with AMOS: Basic concepts, application and programming. London: Lawrence Erlbaum.

Campbell, J. I. D. & Xue, Q. (2001). Cognitive arithmetic across cultures. Journal of Experimental Psychology: General, 130, 299-315.

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

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. In H. Nyborg (Ed.), The scientific study of general intelligence: Tribute to Arthur R. Jensen (pp. 5-22). San Diego: Pergamon.

Ebbinghaus, H. (1885). Memory: A contribution to experimental psychology. New York: Teachers College Press.

Embretson, S. E. (1983). Construct validity : Construct representation versus nomothetic span. Psychological Bulletin, 93, 179-197.

Fischer, G. H. (1995). XXXXX. In Fischer, G.H. & Molenaar, I.W. (Eds.). Rasch Models. Foundations, Recent Developments and Applications. New York: Springer.

Gittler, G. (1992). Testpsychologische Aspekte der Raumvorstellungsforschung - Kritik, Lösungsansätze und empirische Befunde. Wien: Habilitationsschrift der Universität Wien.

Greeno, J. G. (1991). Number sense as situated knowing in a conceptual domain. Journal for Research in Mathematics Education, 22, 170-218.

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

Haenggi, D. (1989). Differential aspects of visual short- and long-term memory. European Journal of Cognitive Psychology, 1, 285-292.

Holzman, T. G., Pellegrino, J. W. & Glaser, R. (1982). Cognitive dimensions of numerical rule induction. Journal of Educational Psychology, 74, 360-373.

Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), Intelligence: Measurement, theory and public policy (pp. 29-73). Urbana: University of Illinois Press.

Horn, J. L. & Noll, J. (1997). Human cognitive capabilities: Gf-Gc Theory. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), Contemporary intellectual assessment: theory, tests and issues (pp. 49-91). New York: The Guilford Press.

Hornke, L. F. (2002). Item-generation models for higher order cognitive functions. In S. H. Irvine & P. C. Kyllonen (Eds.), Item generation for test development (pp. 159-178). London: Lawrence Erlbaum.

Hornke, L.F., Etzel, S. & Küppers, A. (2000). Konstruktion und Evaluation eines adaptiven Matrizentests. Diagnostica, 46, 182-188.

Hu, L. T. & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternaties. Structural Equaion Modeling: A Multidisciplinary Journal, 6, 1-55.

Kline, R. B. (1998). Principles and practice of structural equation modeling. New York: Guilford Press.

Koedinger, K. R. & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. Journal of the Learning Sciences, 13 (2), 129-164.

Kosslyn, S. M. (1981). The medium and the message in mental imagery: A theory. Psychological Review, 81, 46-66.

Krause, B. (1985). Zum Erkennen rekursiver Regularitäten. Zeitschrift für Psychologie, 193 (1-2), 71-86.

LeFevre J. & Bisanz J. (1986). A cognitive analysis of number-series problems: Sources of individual differences in performance. Memory and Cognition, 14 (4), 287-298.

LeFevre, J.-A., Greenham, S. L. & Waheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. Cognition and Instruction, 11, 92-132.

Nathan, M. J., Kintsch, W. & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. Cognition and Instruction, 9, 329-389.

Newell, A. (1990). Unified Theories of Cognition. Cambridge, MA: Harvard University Press.

Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rinehart & Winston.

Roberts, R. D., Goff, G. N., Anjoul, F., Kyllonen, P. C., Pallier, G. & Stankov, L. (2000). The Armed Services Vocational Aptitude Battery (ASVAB) - Little more than acculturated learning (Gc)!? Learning and Individual Differences, 12 (1), 81-103.

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, 262-274.

Schumacker, R. E., & Lomax, R. G. (2004). A beginner's guide to structural equation modeling (2nd edition). New Jersey: Lawrence Erlbaum.

Sebrecht, M. M., Enright, M., Bennett, E. & Martin, K. (1996). Using algebra word problems to assess quantitative ability: attributes, strategies and errors. Cognition and Instruction, 14, 285-343.

Spada, H. & Scheiblechner, H. (1973). Stichprobenunabhängige Denkmodelle. Eine Analyse von Denkfehlern beim syllogistischen Schlussfolgern. In Reinert (Hrsg.), Bericht über den 29. Kongress der deutschen Gesellschaft für Psychologie, Kiel 1970. Göttingen: Hogrefe.

Srp, G. (1993). Syllogismen als Aufgaben für einen computerisierten adaptiven Test. Unveröff. Dipl.Arbeit, Universität, Wien.

Tirre, W. C. & Field, K. A. (2002). Structural models of abilities measured by the Ball Aptitude Battery. Educational and Psychological Measurement, 62 (5), 830-856.

Undheim, J. O. & Gustafsson, J. E. (1987). The hierarchical organisation of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL). Multivariate Behavioral Research, 22, 149-171.

Wagner-Menghin, M. (2004). Manual Lexikon-Wissen Test. Mödling: Dr. G. Schuhfried GmbH.

Wittmann, W. & Süß, H.-M. (1997). Challanging G-mania in intelligence research: answers not given, due to questions not asked. International Society for the study of individual differences, 19-23 July, Aarhus, Denmark.