Modeling the construct validity of the Berlin Intelligence Structure Model

O modelo da validade do construto do Modelo Estrutural da Inteligência de Berlim

Heinz-Martin **SÜß**[1]
André **BEAUDUCEL**[2]

Abstract

The Berlin Intelligence Structure Model is a hierarchical and faceted model which is originally based on an almost representative sample of tasks found in the literature. Therefore, the Berlin Intelligence Structure Model is an integrative model with a high degree of generality. The present paper investigates the construct validity of this model by using different confirmatory factor analysis models. The results show that the model assumptions are supported only in part by the data. Moreover, it is demonstrated that there are different possibilities to incorporate the Berlin Intelligence Structure Model assumptions into confirmatory factor analysis models. The results are discussed with regard to the validity of the Berlin Intelligence Structure Model test, and the validity of the model.

Keywords: Construct validity; Factors analysis; Validity of the model.

Resumo

O Modelo Estrutural da Inteligência de Berlim é um modelo hierárquico e facetado que foi originalmente baseado em uma amostra quase representativa de tarefas encontradas na literatura. Portanto, o modelo estrutural da inteligência de Berlim é um modelo integrador com um elevado grau de generalidade. O presente trabalho investiga a validade do construto deste modelo usando diferentes modelos de análise fatorial confirmatória. Os resultados mostram que os pressupostos do modelo suportam apenas em parte pelos dados. Além disso, demonstra-se que há diferentes possibilidades de incorporar os pressupostos do Modelo Estrutural da Inteligência de Berlim em modelos de análise fatorial confirmatória. Os resultados discutidos mostram a validade do teste do Modelo Estrutural da Inteligência de Berlim, bem como a validade do modelo.

Palavras-chave*: Validade do construto; Análise fatorial; Validade do modelo.*

The Berlin Intelligence Structure Model (BIS) of Jäger (1982; 1984) is an integrative, hierarchical and faceted model of intelligence. At the most general level, general intelligence ("g") is assumed

▾ ▾ ▾ ▾ ▾

1 Otto von Guericke University Magdeburg, Institute of Psychology I, Psychological Methods, Assessment, and Evaluation Research. Postfach, 4120, 39106, Magdeburg, Germany. *Correspondência para*/Correspondence to: H.M. SÜß. E-mail: <heinz-martin.suess@ovgu.de>.
2 Rheinische Friedrich-Wilhelms-University Bonn, Insitute of Psychology, Methods, Assessment, & Evaluation. Bonn, Germany.

as the integral part of all ability components, i.e. the positive manifold. On the second level, seven higher order abilities are suggested to belong to two facets (Guttman, 1957): the operational facet distinguishes abilities according to the cognitive processes involved, i.e. reasoning, perceptual speed, memory, and creativity. The content facet distinguishes abilities according to the material that is applied, i.e. verbal, numerical and figurative-spatial intelligence. The cross-classification of the four operational and three content-related components is divided into twelve cells on the third level. However, the cells do not have the status of specific abilities (e.g. numerical reasoning, or verbal creativity) as in Guilford's Structure-of-Intellect model (SOI model) (Guilford, 1967), who postulated an ability factor for every combination of the SOI facets. Instead, performances are assumed on this level indicating that every intellectual performance depends on at least two abilities, an operational and a content ability. Therefore, the third level can be used to classify performance measures (tasks). The well-known Raven matrices (Raven, 1965), for instance, need to be classified into figural-spatial reasoning, indicating that this test is measuring a combination of two specific abilities.

The seven factors of the BIS are described as processing capacity, which corresponds exactly to reasoning (fluid intelligence); creativity, which is close to fluency and flexibility (divergent thinking); memory, which refers to the ability to recall lists and configurations of items after learning them (short-term memory); and perceptual speed, which refers to the quick and accurate performance on simple tasks (mental speed). The three content factors are verbal, numerical, and figural-spatial intelligence. Figure 1 shows a representation of the BIS.

The more general ideas behind the BIS were summarized by Jäger, Süß, and Beauducel (1997) as follows: (1) all intellectual abilities contribute to every intellectual performance, but with different weights. The variance of every intellectual performance can be decomposed according to these abilities; (2) intellectual performance and the ability constructs can be classified according to facets. Two facets,
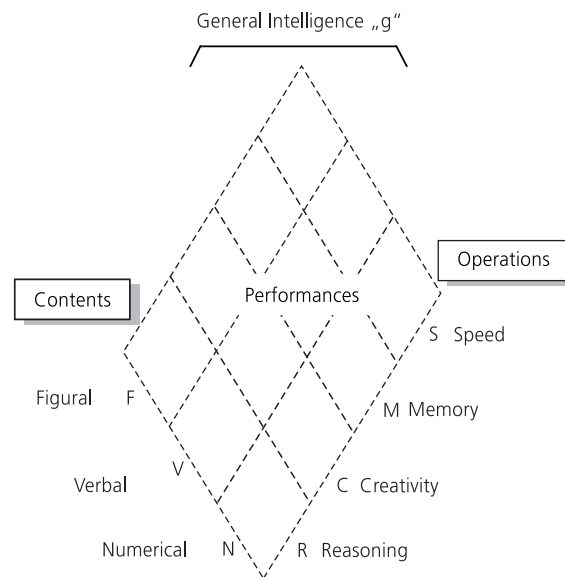

Figure 1. Jäger's (1982) Berlin Intelligence Structure Model.

operations and contents, were specified (bimodality assumption); (3) intelligence constructs are structured hierarchically, i.e. they can be assigned to different levels of generality (hierarchy assumption).

The aims of this paper are (1) to review the development and the validity status of the BIS, and (2) to test the construct validity of the model empirically by means of different confirmatory factor models. Thus, the present paper is devoted to the structural aspect of construct validity, which has also been termed "structural fidelity" by Loevinger (1957).

Model development

The starting point for the development of the BIS was the intention to develop an integrative model, i.e. a model that can explain the differences of competing psychometric models of intelligence published up to the 1970s. The commonalities and differences between these models are described in the three-stage model for the evolution of theories of intelligence by Sternberg (1981). On the first stage, a monistic and a pluralistic view of intelligence are competing with each other assuming a single latent source of person differences (Spearman,

1904; 1927) *versus* multiple independent sources structural bond; (Brown & Thomson, 1921; Thomson, 1950). This conflict was never resolved, but the view of Thomson had only a small impact on intelligence theory and testing. Nevertheless, Thomson's (1950) statement against reifying factors and especially against reifying the "g"-factor as a form of "mental energy" is still important. On the second stage, hierarchical (Cattell, 1971; Royce, 1973; Vernon, 1961) and non-hierarchical (Guilford, 1967; Thurstone, 1938) views are competing with each other. Hierarchical models assume a strong general factor "g" and a different number of broad and specific abilities on two or three lower levels (strata). This view is also the basis of the Three-stratum theory (Carroll, 1993) and the Cattell-Horn-Carroll theory (CHC theory); (McGrew, 1997; 2005). On the other side, a different number of overlapped (correlated) factors are assumed (e.g., primary mental abilities; Thurstone, 1938) but a "g"-factor is not supposed because the factor overlap is seen as of subsidiary interest. This view is also the basis of the extended Gf-Gc theory of Horn (Horn, 1994; Horn & Noll, 1997). The competition on the second stage is resolved on the third stage combining both assumptions in the radex model of Guttman (1957, 1965). Based on Multidimensional Scaling (MDS) or smallest space analysis, this model assumes content-related abilities and a radial expansion of complexity with "g" in the centre with the highest complexity. Like Guttman's radex model, the BIS assumes both a hierarchical and a faceted structure of abilities. However, in contrast to Guttman's radex- or cylindrex-model of intelligence (Guttman & Levy, 1991), the faceted structure of the BIS contains dimensions that can be represented by factors, whereas MDS performed by Guttman only allows for the interpretation of partitions, so that no dimensions are formed that can be directly transformed into individual scores.

Jäger assumed that the differences of the competing models can be explained mainly by the differences of the task samples used for development and validation of the models. Therefore, an almost representative task sample of all psychometric intelligence tasks described in the available research literature at that time was used as the basis for the development of the model. Altogether, about 2000 tasks were sampled. For the empirical investigations that followed the number of tasks in the sample was reduced while the diversity of tasks was maintained. Moreover, the marker tasks of the competing models remained in the sample. Unfortunately, this step is not well documented by Jäger (1982, 1984).

According to these criteria, the task sample was reduced to about 200 tasks, and administered to an age-homogeneous sample of college students (Jäger, 1982). At first, exploratory factor analyses were performed without specific hypotheses concerning a faceted structure. Only four broad factors representing different operations emerged from these analyses (i.e., processing capacity, creativity, memory, and perceptual speed). However, typical content factors such as verbal, numerical and figural-spatial abilities did not emerge. These factors were already well established at that time (Guilford, 1967; Thurstone, 1938) and also part of Jäger's first model (Carroll, 1993; Jäger, 1967). At that point, the idea of a faceted structure of intelligence emerged (Jäger, 1982). The assumption was that the content factors were just masked by the operation factors in exploratory factor analysis. Thus, the emerging factors belonged to the operation facet while a content facet was assumed which contains the verbal, numerical, and figural factors which were presumed to be masked in the analysis. The content factors probably did not occur because the variance due to content was reduced in the sample of participants. This may have happened because it was a homogeneous sample of college students in which the use of words and simple arithmetic operations were probably quite overlearned.

Inspired by the work of Humphreys (1962), Jäger decided to form new variables, called parcels, through the aggregation of tasks by summing up equally-weighted variables. For the identification of the content factors, parcels were formed of tasks that were heterogeneous with respect to the operations but homogeneous with respect to one element of the content facet. In Humphreys (1962), 'homogeneity' and 'heterogeneity' refers to the

15

degree of similarity of the variables forming a parcel (whereas in analysis of variance, 'homogeneity' refers to the similarity of the size of variances). For example, a parcel homogeneous with respect to verbal ability and heterogeneous with respect to the operation facet was formed through the equally weighted aggregation of a verbal reasoning task, a verbal creativity task, a verbal memory task, and a verbal speed task. In that way, verbal ability parcels, numerical ability parcels, and figural ability parcels were formed. Exploratory factor analysis of the content parcels did reveal the corresponding content factors for verbal, numerical, and figural ability (Jäger, 1982; 1984). This result indicates that the content variance was indeed masked by the operation variance and that controlled parceling made it possible to disentangle the relevant content variance. In the same way, parcels which were homogeneous with respect to one element of the operation facet (e.g. processing capacity) and heterogeneous with respect to all elements of the content facet were formed. Exploratory factor analysis of the parcels that were homogeneous with respect to the operations revealed again the four operation factors.

On this basis, a faceted structure containing four operation factors and three content factors was demonstrated. Finally, the aggregation technique was used for the demonstration of a general intelligence factor ("g"). For this, the tasks were aggregated in order to form parcels that were heterogeneous with respect to the operation facet and with respect to the content facet. Exploratory factor analysis of parcels containing one task for every combination of the facet elements revealed a single factor representing general intelligence. General intelligence is relevant in the BIS because it explains the correlation between the operation and content factors which are not regarded as orthogonal.

The use of the parceling technique for validating the BIS will be demonstrated in the empirical part of this paper. The effects of parceling and aggregation in the context of the BIS were described first by Wittmann (1988).

The scope of validity of the BIS

The Berlin Intelligence Structure Model has been successfully validated in numerous studies using different task sets, different methods, and with different populations (apprentices, high-school students, undergraduates, young adults, intellectually gifted people) (Beauducel & Kersting, 2002; Bucik & Neubauer, 1996; Jäger et al., 1997; Jäger, et al., 2006; Pfister & Beauducel, 1993; Süß & Beauducel, 2005; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002).

We would like to point out the replication of the BIS in a the German version of the kit of Factor-Referenced Tests (French, Ekstrom, & Price, 1963) was replicated by Jäger and Tesch-Römer (1988). The KIT is a collection of three marker tasks of each factor of intelligence that have been replicated independently and, on the basis of a literature review, was compiled by French et al. The successful replication study in an independent task sample demonstrated that Jäger's (1982) results did not depend very much on the specific task selection.

Two further replications, one with a Brazilian sample based on a Portuguese version of the BIS kit, a former BIS test (Schmidt, Brocke, Jäger, Doll, & König, 1986) by Kleine and Jäger (1987), and another with a modified and extended Spanish version with Chilean students (Rosas Díaz, 1990), have also to be pointed out. In the empirical part of this paper, the construct validity of the model is further investigated by analyzing a big data set with different methods.

Method

Participants

Data analyses were based on complete data sets from 910 German-speaking high-school students (536 females), who participated voluntarily and received written feedback. Participants' age ranged from 14 to 19 years; the mean age was 16.5 years (Standard Deviation - SD = 1.3 years).

16

Instruments

Each task of the BIS-4 Test is classified into one cell of the model. The verbal, figural, and numerical scales each contain 15 tasks; the speed and the memory scales consist of 9 tasks; the creativity scale consists of 12 tasks; and the reasoning scale contains 15 tasks. The scale of general intelligence comprises all 45 tasks. Next, we give a short description of the tasks. The task descriptions are sorted according to the operational facet, and the content facet is indicated by the character in parenthesis (V for Verbal, F for Figural, and N for Numerical). The task abbreviations in brackets correspond to the German manual test (Jäger et al., 1997).

Speed (V)

Part-whole (TG): In a list of words, every word standing in a part-whole relation to the previous word has to be marked (e.g. tree-leaf). *Word classification* (KW): In a list of words, all words naming plants have to be marked. *Fragmentary words* (UW): The missing letter in fragmentary words has to be completed.

Speed (N)

X-larger (XG): In rows of numbers, all numbers by the amount X larger than the previous number have to be marked. *Divisible by seven* (SI): In lines with binary numbers, all numbers divisible by seven have to be crossed out. *Math operators* (RZ): In simple mathematical equations with plus and minus signs missing, the correct operators have to be inserted.

Speed (F)

Old English (OE): In rows of letters, all letters in the type Old English have to be marked. *Crossing out letters* (BD): In lines of letters, a previously defined letter has to be crossed out. *Digit symbol-coding* (ZS): Several digit-symbol pairs are given. In the following list of numbers, the corresponding symbols have to be added.

Memory (V)

Story (ST): As much information as possible has to be memorized from a short story. At retrieval, questions about story details have to be answered. *Word Memory* (WM): A list of words has to be memorized and reproduced in free order. *Fantasy language* (PS): A list of paired German words and words of a fantasy language are presented. Afterwards, to the given scrambled order of German words, the corresponding fantasy word has to be recognized from several words of the fantasy language.

Memory (N)

Paired associates (ZP): A list of pairs of three-digit numbers has to be learned. At retrieval, the first number of each pair is given in a different order, and the second number has to be recalled and written down. *Two-digit numbers* (ZZ): A sequence of two-digit numbers has to be memorized and reproduced afterwards in free order. *Recognition of Numbers* (ZW): Previously memorized sequences of five-digit numbers have to be recognized in a list of five-digit numbers.

Memory (F)

City map (OG): Buildings marked black on a city map have to be memorized and then crossed out on an unmarked map. *Corporate symbols* (FM): Corporate symbols are given in different characteristic frames and have to be memorized. Afterwards in a scrambled list of symbols, the original frame has to be recognized out of four alternative frames each. *Routes memory* (WE): The route between two places marked in a city map has to be memorized. Afterwards, the route has to be drawn on the unmarked map.

Creativity (V)

Masselon (MA): As many different sentences as possible using three given words have to be written down. *Traits and abilities* (EF): As many different personal characteristics as possible have

17

to be indicated that a person with a defined profession should not have. *Insight Test* (IT): As many different explanations as possible for a social issue have to be found. *Options for utilization* (AM): As many different applications as possible for provided objects have to be found.

Creativity (N)

Telephone numbers (TN): As many different six-digit telephone numbers as possible have to be written down. The numbers should follow different principles that could help remembering them. *Divergent calculation* (DR): A sequence of basic mathematical operations and the result is given. As many different combinations of numbers as possible have to be produced satisfying the equation. *Number equations* (ZG): As many different mathematical operations as possible have to be created using given numbers and basic mathematical operations. *Number pattern* (ZR): Number patterns have to be created. Digits have to be filled in given geometrical schemata resulting in as many different patterns as possible fulfilling as many different mathematical rules as possible.

Creativity (F)

Drawing objects (ZK): As many different objects as possible using given geometrical elements have to be drawn. *Layout* (LO): As many different corporate symbols as possible have to be created for the advertisement of a small shop. *Drawing completion* (ZF): Given incomplete drawings have to be completed resulting in as many different objects as possible. *Object designing* (OJ): Given geometrical figures have to be combined resulting in as many as possible different objects. The objects have to be designated.

For all creativity tasks, scoring is based on the number of different categories of ideas produced, or on the number of ideas produced.

Reasoning (V)

Verbal analogies (WA): Word analogies have to be completed. *Fact-opinion* (TM): Sentences have

to be evaluated as to whether they state a fact or an opinion. *Syllogisms* (SL): Syllogistic inferences from absurd premises have to be evaluated. *Senseless inferences* (SV): The logical validity of deductions from statements about everyday matters has to be evaluated irrespective of their truth or plausibility. *Word knowledge* (WA): The one word not belonging to a set of four words has to be crossed out. The words of successive items were increasingly unfamiliar ones.

Reasoning (N)

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

Reasoning (F)

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

Procedures

The sample was drawn from a pool of several existing studies in which the BIS-4 Test was applied (Jäger et al., 1997). The data were recorded from 1995 to 2001. The test was applied in groups of 5 to 20 subjects. The administration of the fixed sequence of 45 BIS-4 Test tasks with alternating operation and content was fully standardized. The processing time was limited varying between 40 seconds (TG, verbal speed task) and 5 minutes (TL, numerical reasoning). The time to memorize was limited between 30 seconds (WE, figural memory) and 2 minutes (ZP, numerical memory). Overall, the BIS-4 Test took about 2:30 hours, including 2 breaks of 10 minutes each.

Results

Data aggregation and parcel building

Theory-guided aggregation was used to build parcels. The data were aggregated in three ways: (a) aggregation across contents within one operational category, which led to parcels that suppressed content variance as part of the unwanted variance, leaving mainly operational variance; (b) aggregation across operations within one content category, which resulted in parcels that brought content variance to the fore, but let operational variance count as unwanted; and (c) aggregations within the cells of the two-dimensional matrix. Here, both, content and operational variances were wanted variance, and only task-specific variance was suppressed. Aggregated variables were the Z-standardized task scores. Figure 2 shows how the content parcels were built. Because the numbers of variables in the cells were unequal, two variables were averaged to get the same number of tasks in each cell before the content parcels were built. This aggregation resulted in nine content-homogeneous parcels, three for each content factor. Correspondingly, fifteen operation-homogeneous parcels were built by aggregating three equally-weighted tasks (Z-scores) of the same operation but different contents.
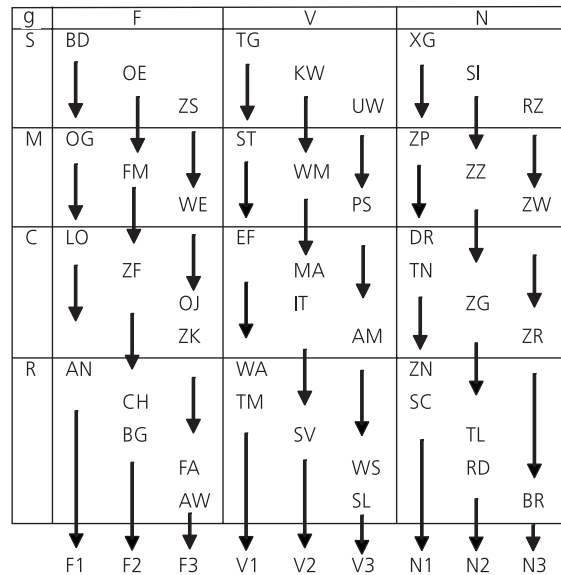

Figure 2. Building of content-homogeneous parcels of the BIS-4 Test.

Note: F1, F2, F3: Figural parcels; V1, V2, V3: Verbal parcels; N1, N2, N3: Numerical parcels. The task abbreviations and short task descriptions see the method section (instruments) of the text.

Confirmatory Factor Analyses with parcels

Two Confirmatory Factor Analyses (CFA), one based on the nine content-parcels and one based on the fifteen operation-homogeneous parcels were calculated with EQS 6.2. General intelligence was included as second-order factor in both models. Figure 3 shows the CFA of the content-homogeneous parcels with three content factors on the second level (The CFA of the operation-homogeneous parcels cannot be presented due to lack of space). Both models have a good data fit (CFA with content parcels: Chi² = 56.390, df = 24, p<0.01; CFI = 0.992; $RMSEA$ = 0.037, 90%CI [0.024, 0.050]; $SRMR$ = 0.020. CFA with operation parcels: Chi² = 256.43, df = 86, p<0.01; CFI = 0.971; $RMSEA$ = 0.047, 90%CI [0.040, 0.053]; $SRMR$= 0.040). The BIS was also confirmed by using MDS with polar partitioning without any mislocation of variables.

Correlation matrix and CFA of cell aggregates

In the next step, the inter-correlation matrix of the BIS cell aggregates was investigated. The 12
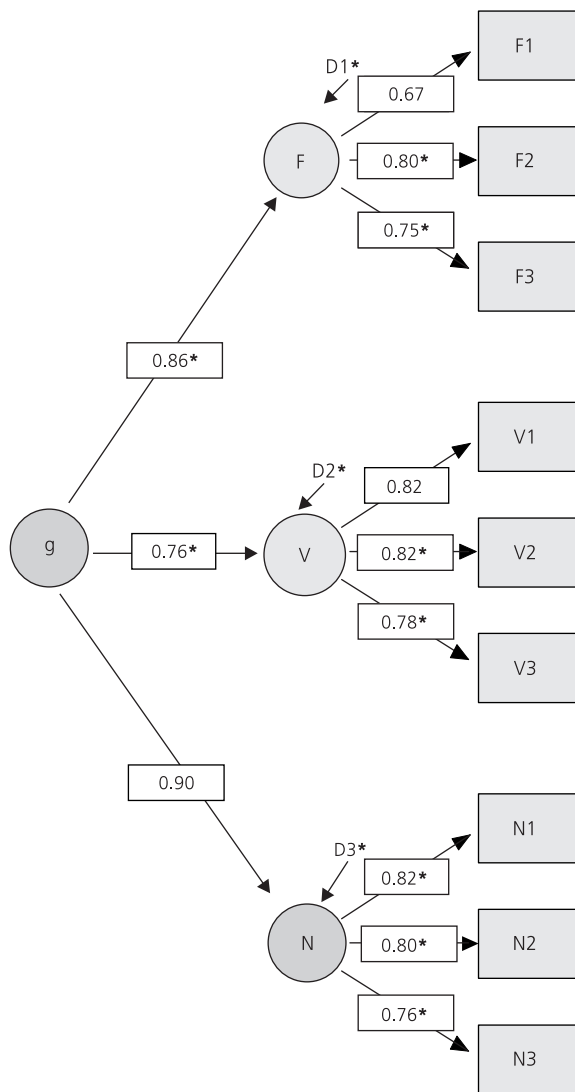
19

Figure 3. Confirmatory Factor Analysis of the content-homogeneous parcels.

Note: g: General intelligence; F: Figural intelligence; V: Verbal intelligence; N: Numerical intelligence; F1, F2, F3: Figural parcels; V1, V2, V3: Verbal parcels; N1, N2, N3: Numerical parcels.

aggregates represent the 12 BIS cells and are based on tasks of the same operation *and* the same content. In this analysis, every combination of facet elements is represented by one variable. According to the contiguity hypothesis of Guttman (1965), cell aggregates sharing the same content should be more strongly correlated than those cells that have neither the operation nor content in common. The common variance of the latter ones is attributed to

the "g"-factor only. The same should be the case for correlations between cell aggregates that share the same operation in comparison with cells that have only "g"-variance in common.

Cells sharing the same operation are more strongly correlated ($Md = 0.46$; range: 0.32 to 0.63) than cells having no facet in common ($Md = 0.25$; range: 0.09 to 0.40). This is also the case for cells sharing the same content ($Md = 0.34$; range: 0.14 to 0.62). However, these correlations are lower than the correlation of cells sharing the same operation, and the variance is much larger. In particular, the four figural cells are rather poorly correlated, especially figural speed with figural reasoning ($r = 0.14$). On the other hand, the correlations of verbal reasoning with numerical speed ($r = 0.40$) and figural reasoning with numerical speed ($r = 0.40$) were more strongly correlated than expected; as in both cases the cells are sharing neither the same operation nor the same content but only "g". Summing up, these results clearly indicate that the model assumptions are only in part supported by the data.

In the next step, CFAs were applied to the BIS cell aggregates testing all BIS assumptions within one model. These CFA models allow the simultaneous investigation of the operation and the content facet as well as the hierarchy assumption. However, in a model, in which "g" is related to both, the content factors and the operation factors, identification problems for some of the loadings generally occur. In the present data, the figural loadings were heterogeneous and in part extremely low when "g" was forced to load both the operation and the content factors. This indicates that the large number of free parameters may limit the possibility to test hierarchically faceted structures within a single analysis with structural equation modeling.

In order to further investigate the reasons for the identification problems, a model in which "g" was based both on the content and the operation facet was estimated with Mplus 6.0. The inter-correlation matrix of the cells was used as the basis for this model. Loadings of all operation and content factors on "g" were estimated in this model

(Figure 4). To prevent problems of model identification, an equality constraint was imposed on the loadings of the factors Speed (S) and Reasoning (R) on "g" (Chi² = 234.530, *df* = 36, *p*<0.0001; *CFI* = 0.948; *RMSEA* = 0.078, 90%CI [0.069, 0.087]; *SRMR* = 0.046; *AIC* = 27401.045; *BIC* = 27603.209). The "g"-factor explains 25.52% of the variance.

The equality constraint for the loadings of S and R on "g" is not very compelling because there is no theoretical justification for the expectation that Speed and Reasoning share exactly the same

amount of "g"-variance. Therefore, a nested factor model was investigated as an additional alternative (Figure 5). In the nested factor model each BIS cell has a loading on the "g"-factor so that each BIS cell has some variance that is accounted for by a content factor, an operation factor, and "g". This allows for the direct investigation of the general idea that the variance of every intellectual performance can be directly decomposed according to all the BIS-abilities, whereas the previous model contained only indirect paths from "g" to the
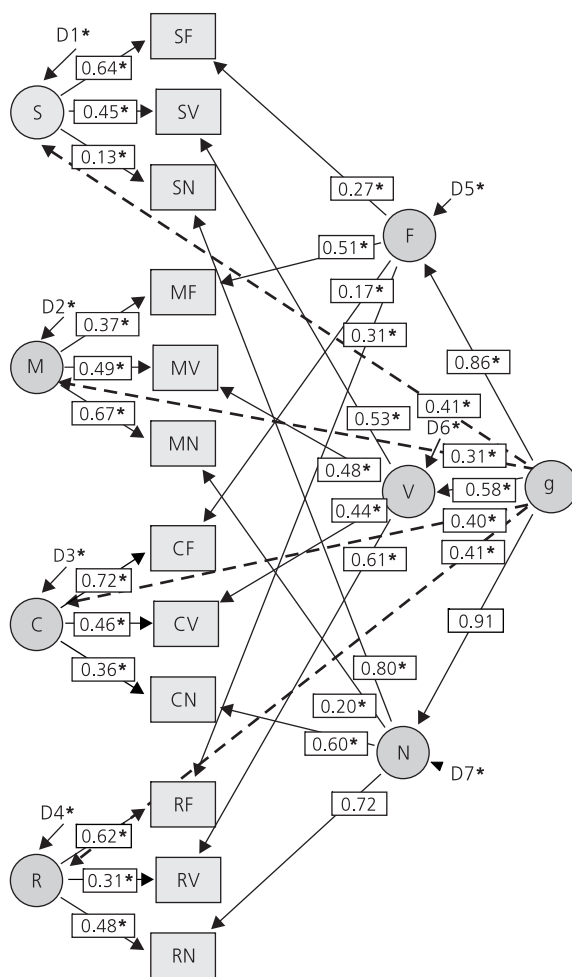

Figure 4. Confirmatory Factor Analyses of the Berlin Intelligence Structure cells.

Note: g: General intelligence; S: Speed; M: Memory; C: Creativity; R: Reasoning; F: Figural intelligence; V: Verbal intelligence; N: Numerical intelligence; SF: Figural Speed; SV: Verbal Speed; SN: Numerical Speed; MF: Figural Memory; MV: Verbal Memory; MN: Memory Numerical; CF: Figural Creativity; CV: Verbal Creativity; CN: Numerical Creativity; RV: Verbal Reasoning; RN: Numerical Reasoning; RF: Figural Reasoning.
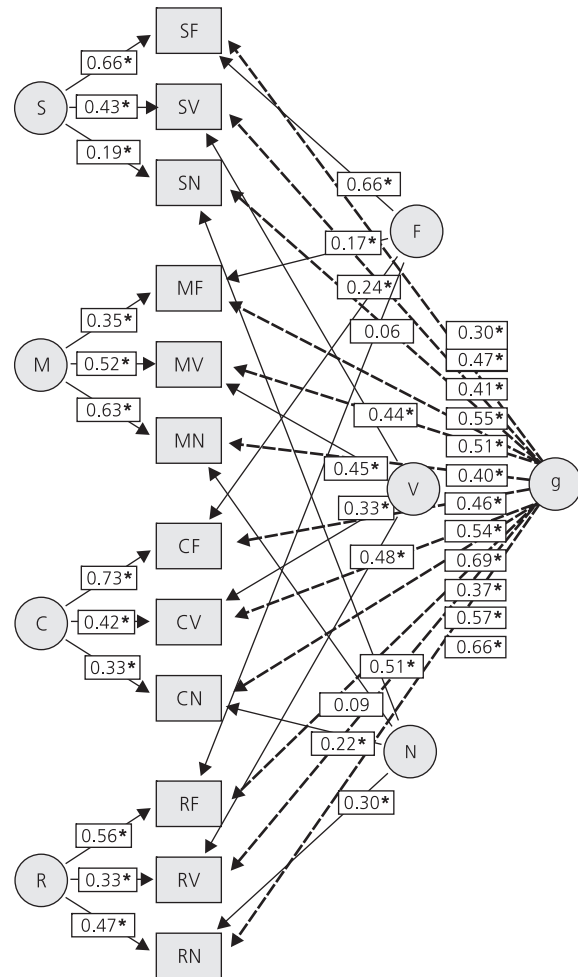

Figure 5. Nested factor model of the Berlin Intelligence Structure cells.

Note: g: General intelligence; S: Speed; M: Memory; C: Creativity; R: Reasoning; F: Figural Intelligence; V: Verbal intelligence; N: Numerical intelligence; SF: Figural Speed; SV: Verbal Speed; SN: Numerical Speed; MF: Figural Memory; MV: Verbal Memory; MN: Numerical Memory; CF: Figural Creativity; CV: Verbal Creativity; CN: Numerical Creativity; RV: Verbal Reasoning; RN: Numerical Reasoning; RF: Figural Reasoning.

intellectual performance variables. However, in a nested factor model, all factors are orthogonal, since the common variances are accounted for by the common "g"-factor. The model fit was similar to the previous model with loadings of the operation factors and content factors on "g" (Chi² = 197.024, df = 30, p<0.0001; CFI = 0.956; $RMSEA$ = 0.078, 90%CI [0.068, 0.089]; $SRMR$ = 0.044; AIC = 27375.539; BIC = 27606.585). The "g"-factor explains 25.61% of the variance. Since the amount of variance explained by "g" is not substantially greater than the previous model, it can be concluded that there is no relevant "g"-variance that is directly related to the task and that nearly all the "g"-variance in the performance variables is indirect.

Since the first model is not directly nested into the present model, a test for the significance of fit improvement was not possible. It should be noted that the loading of the figural reasoning cell on F was not significant and that the loading of the numerical memory cell on N was not significant. All remaining loadings were significant, so that all factors had at least three significant loadings.

Discussion

The results demonstrate that the model assumptions of the BIS are supported only in part by the data. Data analyses based on theory-guided aggregates support the assumptions of the BIS to a great extent. The hierarchy and the bimodality assumption are supported by the data showing the three content factors and the four operation factors in the same data set. The results demonstrate that "g" and reasoning (Horn, 1988) can be separated clearly in a broad test battery. Assuming the identity of both constructs would result in significant worsening of the model fit. This result is in accordance with the Three-Stratum theory of Carroll (1993), the extended Cattell-Horn model (Horn, 1994; Horn & Noll, 1997), and the Cattell-Horn-Carroll theory (CHC theory) of McGrew (1997, 2005) that could also differentiate both factors. However, this result contradicts the assertion of Gustafsson (Gustafsson, 1984, 1999; Gustafsson & Balke, 1993) that within a broad battery of intelligence tasks "g" and Gf cannot be differentiated. The multitude and heterogeneity of the tasks used for the measurement of "g" is a clear consequence of its faceted definition. If a theoretical background is set up to help identify the domain of intelligence tasks (Guttman, 1965), then a concrete task sample can be evaluated against this background. Such kind of analyses show that in many cases construct validity is not given, e.g., the simple identification of Raven's matrices with Gf, or the equation of verbal intelligence with Gc (Süß & Beauducel, 2011). Against this background, many empirical results need to be reinterpreted.

Limitations of the BIS were revealed analyzing the cell level. On this level, Jäger assumed performances not abilities. He assumed that every intellectual performance can be decomposed in at least three variance components: content, operation, and "g". The results show that the variance components of each cell are differently sized (cf. the reliability estimations of Brunner & Süß, 2005). The interesting question is: Does this result question the model? Or is this result only the result of psychometric weaknesses of the specific test, the time limited test administration, and/or a sample effect? The model fits presented in this paper are weaker than those reported earlier for similar models (Süß et al., 2002; Süß & Beauducel, 2005). This could be the result of the more heterogeneous sample, particularly the broader education and age range. The time-limited test application and the sample characteristics together could have increased the speed effects resulting in the strong correlation of "g" and perceptual speed. Moreover, the numerical speed tasks were strongly correlated with verbal and figural reasoning. This indicates that in this sample the basic numerical skills were not overlearned in many cases, changing the character of the tasks from perceptual speed to reasoning tasks. From this view, the demonstrated problems are owing to a great extent to the limited area of validity of this test. The BIS-4 Test was originally developed for persons with medium to higher education level. It is necessary to adapt many tasks before the test is given to broader samples. Nevertheless, it is an illusion that the variance components can be balanced completely as the

22

model suggests. In so far, there will always be a disjunction between theory and empirical results.

Guttman (1957) originally introduced a facet referring to the level of complexity of tests. Guttman predicted that complex tests would be located at the periphery of the radex because complex tests would have fewer components in common with each other than simple tests as they diverge in different directions of complexity. However, Guttman's initial prediction failed in empirical analyses since the complex tests were located in the center of the radex (Schlesinger & Guttman, 1969). Therefore, the complexity facet was replaced by a rule task facet: tests of rule-inference are located in the center of the radex, followed by tests of rule-application; tests for learning or achievement are located at the periphery of the radex (Guttman & Levy, 1991; Schlesinger & Guttman, 1969). The most complex tasks of the BIS are probably the reasoning tasks, the simplest tasks are surely the speed tasks. However, the "g" loadings of the operation factors in Figure 4 and the "g" loadings of the cells in Figure 5 are rather similar. Thus, the present results do not support Guttman's complexity facet.

The analyses presented here also point to the fact that there are different ways to represent faceted hierarchical models within confirmatory factor analysis. On the one hand, identification problems occur when a "g"-factor is modeled at the top of the hierarchy together with operation and content factors loading on "g". These identification issues could be solved by means of an equality constraint on two loadings of operation factors on "g". This was only a slight modification, but this equality constraint is not part of the BIS. Another aspect of this model that is not explicitly part of the BIS is that "g" has only indirect effects on intellectual performance (through the content and operation factors). In order to allow for direct effects of "g" on performance, a nested factor model was investigated. In this model the variances of the cells are decomposed into the variances represented by orthogonal content factors, operation factors, and "g". This type of modeling has greater emphasis on the first BIS assumption (all intellectual abilities contribute to every intellectual performance, but

with different weights) because the loading of the cells on "g" represent the common variances that are represented by the cells, but not by the specific content and operation factor of the cell. However, these analyses revealed that the amount of performance variance directly explained by "g" is not substantially larger than the amount of performance variance that is indirectly explained by "g" through the content and operation factors. One can therefore conclude that there is no relevant "g"-variance that is directly related to the task and that nearly all the "g" -variance in the performance variables is indirect.

From a more general point of view it should be noted that it is impossible to represent the hierarchical relation between "g" and the operation and content factors (third BIS assumption) and the second assumption (intellectual abilities contribute to every intellectual performance with different weights simultaneously) within a single CFA model. This indicates that the large number of parameters to be estimated limits the flexibility of the representation of faceted models by means of CFA models.

In spite of these limitations, the BIS provides a useful framework in intelligence assessment. The tasks of any intelligence test battery can be classified into the BIS to assemble a BIS test. Süß and Beauducel (2011) classified most of the German intelligence tests and some of the well-known English intelligence tests into the BIS (Wechsler, 1997; Woodcock, McGrew & Matther, 2001). This classification can be used for assembling a test battery but need to be validated firstly. Two aspects are of critical importance for assembling a test battery for the BIS: (1) the balance of the content of every operational construct. We recommend at least three tasks (verbal, numerical, figural) for each BIS cell. In our view, it is more important to use several independent measures with limited reliability than only one task with strong reliability. The first option strengthens the validity because test takers need to prove their intelligence in different situations. The second option again strengthens the task reliability under the risk that task specific variance

23

is dominating the ability score; (2) In order to measure "g" and the content-related abilities, a full test battery including tasks for every of the twelve tests is needed. Moreover, it would be an improvement to provide unspeeded measures of reasoning, but this would require additional testing time.

References

Beauducel, A., & Kersting, M. (2002). Fluid and crystallized intelligence and the Berlin Model of Intelligence Structure (BIS). *European Journal of Psychological Assessment, 18*(2), 97-112.

Brunner, M., & Süß, H. M. (2005). Analyzing the reliability of multidimensional measures: An example of intelligence research. *Educational and Psychological Measurement, 65*, 227-240.

Brown, W., & Thomson, G. H. (1921). *The essentials of mental measurement*. Cambridge: Cambridge University Press.

Bucik, V., & Neubauer, A. (1996). Bi-modality in the Berlin Model of Intelligence Structure (BIS): A replication study. *Personality and Individual Differences, 21*(6), 987-1005.

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

Cattell, R. B. (1971). *Abilities*: *Their structure, growth, and action*. Champaign, IL: Institute for Personality and Ability Testing.

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

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

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*(3), 179-203.

Gustafsson, J. E. (1999). Measuring and understanding G: Experimental and correlational approaches. In P. L. Ackermann, P. C. Kyllonen, & R. D. Roberts (Eds.), *Advances in the psychology of human intelligence* (Vol.4, pp.275-289). Washington, DC: American Psychological Association.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*(4), 407-434.

Guttman, L. (1957). Empirical verification of the radex structure of mental abilities and personality traits. *Educational and Psychological Measurement, 17*, 391-407.

Guttman, L. (1965). A faceted definition of intelligence. In R. Eiferman (Ed.), *Studies in psychology, scripta hierosolymitana* (Vol.14, pp.166-181). Jerusalem: The Hebrew University.

Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence, 15*(1), 79-103.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (pp.645-685). New York: Plenum Press.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp.443-452). New York: Macmillan Publishing Company.

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*: *Theories, tests, and issues* (pp.53-91). New York: Guilford Press.

Humphreys, L. (1962). The organization of human abilities. *American Psychologist, 17*(7), 475-483.

Jäger, A. O. (1967). *Dimensionen der intelligenz*. Göttingen: Hogrefe.

Jäger, A. O. (1982). Mehrmodale Klassifikation von Intelligenzleistungen: Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica, 28*(4), 195-225.

Jäger, A. O. (1984). Intelligenzstrukturforschung: konkurrierende modelle, neue entwicklungen, perspektiven. *Psychologische Rundschau, 35*(1), 21-35.

Jäger, A. O., Holling, H., Preckel, F., Schulze, R., Vock, M., Süß, H. M., & Beauducel, A. (2006). *Berliner Intelligenzstruktur-Test für Jugendliche*: *Begabungs- und Hochbegabungsdiagnostik (BIS-HB)*. Göttingen: Hogrefe.

Jäger, A. O., Süß, H. M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test. BIS-Test, Form 4*. Göttingen: Hogrefe.

Jäger, A. O., & Tesch-Römer, C. (1988). Replikation des Berliner Intelligenz-struktur-modells (BIS) in den "Kit of reference tests for cognitive factors" nach French, Ekstrom & Price (1963). Eine Reanalyse der Daten von Scholl (1976). *Zeitschrift für Differentielle und Diagnostische Psychologie, 9*(2), 77-96.

Kleine, D., & Jäger, A. O. (1987). Replikation des Berliner Intelligenzstrukturmodells (BIS) bei brasilianischen Schülern. *Diagnostica, 33*(1), 14-29.

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3*(Suppl.), 635-694.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment*:

Theories, test, and issues (pp.151-179). New York: Guilford Press.

McGrew, K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment*: *Theories, test, and issues* (2ⁿᵈ ed., pp.136-181). New York: Guilford Press.

Pfister, H. R., & Beauducel, A. (1993). *Stability of operation and content facets*: A facet analysis of the Berlin model of intelligence structure BIS. Proceedings of the Fourth International Facet Theory Conference, Prague.

Raven, J. C. (1965). *The standard progressive matrices*: *A non-verbal test of a person's present capacity for intellectual activity.* London: Lewis.

Rosas Díaz, R. (1990). Replikation des Berliner Intelligenzstrukturmodells (BIS) und Vorhersagbarkeit des Studienerfolgs bei chilenischen Studenten (Unpublished doctoral dissertation). Unterrichtswissenschaften, Berlin.

Royce, J. R. (1973). The conceptual framework for a multi-factor theory of individuality. In J. R. Royce (Ed.), *Multivariate analysis and psychological theory* (pp.300-407). New York: Academic Press.

Schlesinger, I. M., & Guttman, L. (1969). Smallest space analysis of intelligence and achievement tests. *Psychological Bulletin, 71,* 95-100.

Schmidt, J. U., Brocke, B., Jäger, A. O., Doll, J., & König, F. (1986). *Entwicklung eines Tests für das Berliner Intelligenzstrukturmodell*. Berlin: Freie Universität.

Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology, 15(2)*, 201-293.

Spearman, C. (1927). *The abilities of men*. London: MacMillan.

Sternberg, R. J. (1981). The evolution of theories of intelligence. *Intelligence, 5*, 209-230.

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

Süß, H. M., & Beauducel, A. (2011). Intelligenztests und ihre Bezüge zu Intelligenz theorien. In L. F. Hornke, M. Amelang, & M. Kersting (Eds.), *Leistungs-, Intelligenz- und Verhaltensdiagnostik* (Enzyklopädie der Psychologie, Serie Psychologische Diagnostik, Vol. 3, pp.97-234). Göttingen: Hogrefe.

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*(3), 261-288.

Thomson, G. H. (1950). *The factorial analysis of human ability* (4ᵗʰ ed.). London: University of London Press.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Vernon, P. E. (1961). *The structure of human abilities*. London: Methuen.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3ᵗʰ ed.). San Antonio, T.X: The Psychological Corporation.

Wittmann, W. W. (1988). Multivariate reliability theory: Principles of symmetry and successful validation strategies. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2ⁿᵈ ed., pp.505-560). New York: Plenum Press.

Woodcock, R. W., McGrew, K., & Matther, N. (2001). *Woodcock-Johnson III Tests of Cognitive Abilities (WJ III COG)*. Itasca, Il: Riverside Publishing.

25