



UNIVERSITÄT
DES
SAARLANDES

**“ DO SPECIFIC ABILITY FACTORS MATTER? ” –
VALIDITÄT SPEZIFISCHER UND ALLGEMEINER
INTELLIGENZFAKTOREN**

Dissertation zur Erlangung des Grades einer Doktorin der Naturwissenschaften
der Fakultät HW (Empirische Humanwissenschaften und
Wirtschaftswissenschaft, Fachrichtung Bildungswissenschaften)
der Universität des Saarlandes

vorgelegt von
Sonja Valerius (Dipl.-Psych.), geb. in Essen

Saarbrücken, 2020

Gutachter: 1. Prof. Dr. Jörn R. Sparfeldt, Universität des Saarlandes, Saarbrücken
2. Prof. Dr. Frank M. Spinath, Universität des Saarlandes, Saarbrücken

Dekan: Prof. Dr. Stefan Strohmeier

Datum der Einreichung: 19.02.2020

Datum der Disputation: 07.07.2020

Für Johannes

*„What are the most important scientific issues in the domain of human intelligence?
Identifying constructs with validity beyond g.“*

*„What are the most promising new ideas and approaches in the study of human intelligence?
Investigating the predictive validity of non-g residuals of tests.“*

Coyle (2014, S. 21)

DANKSAGUNG

Ich danke Herrn Prof. Dr. Jörn Sparfeldt für seine Unterstützung, seine motivierenden Worte und den unerschütterlichen Glauben an ein gutes Ende dieses Großprojektes; Herrn Prof. Dr. Frank Spinath für seine Bereitschaft, als weiterer Gutachter dieser Arbeit zu fungieren.

Vielen Helferinnen und Helfern sei Dank: Dr. Johannes Schult, Dr. Christin Lotz, Dr. Rebecca Schneider und Dr. Nicole Brunnemann für Ratschläge und gemeinsames Grübeln über statistische Tücken der psychometrischen Intelligenzforschung; Forschungspraktikanten, studentischen Hilfskräften, Testleiterinnen und Testleitern namentlich Jessica Arendt, Elwira Essel, Milena Gigelova, Dorothe Katzenberger, Dr. Marie Kees, Sarah Muno, Nils San José, Renske Starkmeth, Hanna Tegeler und Anna Wiebel für ihren fleißigen und zuverlässigen Einsatz rund um Testleitung, Kisten sortieren und Notschokoladenvorräte anlegen.

Katharina Biewener und Meret Wünnemann danke ich fürs Zuhören, eure aufrichtigen und aufmunternden Worte und euren herrlich schwarzen Humor.

Meinem Mann und meiner Familie sage ich herzlich „Danke“: für euren Rückhalt und euer Verständnis, Entlastung und warme Mahlzeiten; Jakob und Helen dafür, dass ihr euch nicht durch dieses „Riesenschiff“ namens Dissertation habt aufhalten lassen, ein nicht wegzudenkender Teil unserer kleinen Familie zu werden.

Zu guter Letzt danke ich vor allem den vielen Schülerinnen und Schülern der teilnehmenden Schulen für ihre Einsatzbereitschaft und die gewissenhafte Bearbeitung der vielen Testaufgaben. Sowie ihren Schulleitungen und Lehrerinnen und Lehrern für ihre flexible und zuverlässige Unterstützung, Organisationsbereitschaft und das Gelingen reibungsloser Testdurchführungen.

INHALTSVERZEICHNIS

Abbildungsverzeichnis	X
Tabellenverzeichnis.....	X
Publikationsliste	XI
Anteilige Beiträge der kumulativen Promotion	XII
Zusammenfassung.....	XIII
1 Einführung	1
2 Theoretischer Rahmen	3
2.1 Intelligenzstruktur: Nested-Factor-Modell vs. Higher-Order Modell	3
2.1.1 Statistische Unterschiede und Einschränkungen.....	4
2.1.2 Theorie und Wesen der Intelligenz(struktur) im NF- vs. HO-Modell	7
2.2 Konsistenz und Relevanz allgemeiner und spezifischer Intelligenzfaktoren	9
2.2.1 Konsistenz allgemeiner und spezifischer Intelligenzfaktoren.....	9
2.2.2 Allgemeine und spezifische Intelligenzfaktoren als Korrelate und Prädiktoren schulischer Leistung.....	11
2.3 Übereinstimmung von Intelligenztestergebnissen im hohen Begabungsbereich	13
3 Empirische Studien	17
3.1 Studie 1: Valerius & Sparfeldt (2014).....	18
3.1.1 Theorie	18
3.1.2 Methode.....	22
3.1.3 Ergebnisse	26
3.1.4 Diskussion	30
3.2 Ergänzende Analysen zu Valerius & Sparfeldt (2014)	34
3.2.1 Ergebnisse der Zusatzanalysen.....	36
3.3 Studie 2: Valerius & Sparfeldt (2015).....	40
3.3.1 Einführung.....	40
3.3.2 Methode.....	43
3.3.3 Ergebnisse	45
3.3.4 Diskussion	46
3.4 Ergänzende Analysen zu Valerius & Sparfeldt (2015)	49
3.4.1 Ergebnisse der Zusatzanalysen.....	50
3.5 Studie 3: Valerius, Lotz & Sparfeldt (eingereicht).....	53
3.5.1 Theorie	53
3.5.2 Methode.....	59

3.5.3	Ergebnisse	62
3.5.4	Diskussion	66
4	Gesamtdiskussion	71
4.1	Zusammenfassung der wesentlichen Befunde.....	71
4.2	Implikationen.....	74
4.2.1	Implikationen für die psychometrische Intelligenzforschung zum Modellvergleich NF- vs. HO-Modell.....	75
4.2.2	Implikationen für die Forschung zur Konsistenz von g und spezifischen Intelligenzfaktoren	79
4.2.3	Implikationen für die Forschung zum Beziehungsgeflecht allgemeiner und spezifischer Intelligenz mit schulischer Leistung	81
4.2.4	Implikationen für Forschung und Praxis zur Hochbegabungsidentifikation mit mehreren Intelligenztests.....	85
4.3	Einschränkungen und Ausblick	88
4.3.1	Stichprobe.....	88
4.3.2	Instrumente.....	89
4.3.3	Datenerhebung	90
4.3.4	Datenanalyse	91
4.4	Fazit	92
	Literatur	93
	Anhang	104
A.	<i>Mplus</i> Syntaxen	104
B.	Eigenständigkeitserklärung	116
C.	Publikationen	117

ABBILDUNGSVERZEICHNIS

Abb. 1: Higher-Order und Nested Factor Modell	4
Abb. 2: Bifactor (S-1) Modell nach Eid et al., (2018)	6
Abb. 3: Klassifikationsfelder der Hochbegabungsidentifikation mit zwei Testverfahren	15
Abb. 4: Das BIS Model (Jäger, 1982)	21
Abb. 5: CFA für alternative Modelle allgemeiner und spezifischer Intelligenzfaktoren	23
Abb. 6: Bifactor (S-1) Modell der Zusatzanalyse zu Valerius & Sparfeldt (2014)	35
Abb. 7: NF-Modell der Intelligenz- und Schulleistungsfaktoren	44
Abb. 8: Bifactor (S-1) Modell der Zusatzanalysen zu Valerius & Sparfeldt (2015)	49

TABELLENVERZEICHNIS

Tab. 1: Testübersicht aus Valerius & Sparfeldt (2014)	24
Tab. 2: Modellfitstatistik aus Valerius & Sparfeldt (2014)	26
Tab. 3: Standardisierte Faktorladungen des kombinierten NF-Modells aus Valerius & Sparfeldt (2014)	28
Tab. 4 g-Faktor Korrelationen aus Valerius & Sparfeldt (2014)	29
Tab. 5 Korrelationen der inhaltspezifischen Intelligenzfaktoren aus Valerius & Sparfeldt (2014)	30
Tab. 6: Modellfitstatistik der Zusatzanalysen zu Valerius & Sparfeldt (2014)	36
Tab. 7: Standardisierte Faktorladungen des Bifactor (S-1) Modells der Zusatzanalysen zu Valerius & Sparfeldt (2014)	38
Tab. 8: Korrelationen der allgemeinen und spezifischen Intelligenzfaktoren im BF (S-1) Modell der Zusatzanalysen zu Valerius & Sparfeldt (2014)	39
Tab. 9: Modellfitstatistik der Zusatzanalysen zu Valerius & Sparfeldt (2015)	50
Tab. 10: Standardisierte Faktorladungen des Bifactor (S-1) Modells der Zusatzanalysen zu Valerius & Sparfeldt (2015)	51
Tab. 11: Stichproben- und Testkennwerte und HB-Anteile aus Valerius, Lotz & Sparfeldt (eingereicht)	62
Tab. 12: Beobachtete und erwartete Anteile übereinstimmend als „hochbegabt“ Klassifizierter aus Valerius, Lotz & Sparfeldt (eingereicht)	63
Tab. 13: RAZ-Indices der Hochbegabtenklassifikation mit zwei Tests aus Valerius, Lotz & Sparfeldt (eingereicht)	65

PUBLIKATIONSLISTE

Die vorliegende Dissertation basiert auf drei wissenschaftlichen Artikeln in peer-reviewed Journals. Die ersten beiden Artikel sind über die entsprechenden Verlagshomepages verfügbar. Der dritte, bislang unveröffentlichte Artikel ist eingereicht und befindet sich derzeit im Begutachtungsprozess.

1. Valerius, S. & Sparfeldt, J. R. (2014). Consistent g- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory Factor Analyses using three test batteries. *Intelligence*, 44, 120–133.
2. Valerius, S. & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell. *Zeitschrift für Pädagogische Psychologie*, 29, 101–108.
3. Valerius, S., Lotz, C. & Sparfeldt, J. R. (eingereicht) „Hochbegabt“ oder „nicht hochbegabt“ – zur Übereinstimmung der Hochbegabungsidentifikation mit SPM, CFT 20-R und KFT-N. *Psychologie in Erziehung und Unterricht*.

Anmerkungen

Wenn ich im Folgenden den Terminus „wir“ verwende, beziehe ich mich auf die Koautoren, die an der Erstellung der jeweiligen Artikel beteiligt waren. Im Allgemeinen habe ich mich beim Verfassen dieser Dissertation um eine gendergerechte Sprache bemüht, dennoch kann es vorkommen, dass aus Gründen der besseren Lesbarkeit, an einigen Stellen verallgemeinernd das generische Maskulinum verwendet wird. Diese Formulierungen umfassen gleichermaßen weibliche und männliche Personen; alle sind damit selbstverständlich gleichberechtigt angesprochen.

ANTEILIGE BEITRÄGE DER KUMULATIVEN PROMOTION

Die anteiligen Beiträge in Bezug auf die drei Studien der vorliegenden kumulativen Promotion waren wie folgt:

Die Hauptfragestellung zum ersten Artikel wurde von Jörn Sparfeldt entwickelt und unter Mitwirkung von mir konkretisiert und konzeptualisiert. Ich erhob die Daten eigenständig und führte die statistischen Analysen durch. Jörn Sparfeldt wirkte bei der Planung der Analysen sowie der Ergebnisinterpretation mit. Die Manuskripterstellung und Manuskriptrevision erfolgte zu großen Teilen durch mich unter Mitwirkung von Jörn Sparfeldt.

Die Idee und Konkretisierung der Forschungsfrage zum zweiten Artikel entwickelte sich im gemeinsamen Austausch mit Jörn Sparfeldt aus der Hauptfragestellung des ersten Artikels heraus. Die statistische Auswertung erfolgte durch mich unter Mitwirkung von Jörn Sparfeldt bei der Auswertungsplanung und Ergebnisinterpretation. Große Teile des Manuskripts wurden durch mich erstellt und revidiert unter Mitwirkung von Jörn Sparfeldt.

Die Forschungsfrage zum dritten Artikel entwickelte ich selbst. Die Idee zur statistischen Auswertung trug Jörn Sparfeldt bei. Die Datenanalysen sowie große Anteile an der Manuskripterstellung erfolgten durch mich unter Mitwirkung von Christin Lotz und Jörn Sparfeldt.

Ich bedanke mich vor allem bei Jörn Sparfeldt sowie den Teams der Bildungswissenschaften der Universität Trier und der Universität des Saarlandes für die Möglichkeit zur Umsetzung dieser Dissertation, den fachlichen Austausch, Unterstützung und wissenschaftliche Expertise.

ZUSAMMENFASSUNG

Die vorliegende Arbeit thematisiert die diagnostische Güte (Validität) von Intelligenztests. Im Einzelnen geht es um Aspekte der Konsistenz der Messung sowohl der allgemeinen Intelligenz und insbesondere der spezifischen Intelligenzfaktoren (verbal, numerisch, figural) neben dem Generalfaktor g (Studie 1), als auch der Kriteriumsvalidität dieser Intelligenzfaktoren zu allgemeinen wie spezifisch sprachlichen und mathematisch-naturwissenschaftlich schulischen Leistungen (Studie 2) latent modelliert über den Nested-Factor-Ansatz (synonym: bifactor-model; Holzinger & Swineford, 1937) im Rahmen konfirmatorischer Faktorenanalysen (CFA). Zeigen die latenten CFA-Analysen der ersten beiden Studien anhand der Daten von $N = 562$ Neuntklässlern aus sechs rheinland-pfälzischen Gymnasien testbatterieübergreifend ein hohes Maß an konsistenter Messung der allgemeinen sowie spezifisch verbalen und numerischen Intelligenzfaktoren sowie deren konvergente und diskriminante Validität für spezifisch sprachliche und mathematisch naturwissenschaftliche wie allgemeine Schulleistung, so wird in Studie 3 auf manifester Ebene geprüft, in wieweit die Ergebnisse zweier konzeptionell ähnlicher Intelligenztests bei der Diagnostik von Hochbegabung übereinstimmen. Diese Dissertation fokussiert somit die testbatterieübergreifende Übereinstimmungsgüte von allgemeinen und spezifischen Intelligenzfaktoren, gemessen mit unterschiedlichen Testbatterien (Studie 1), deren Gültigkeit in Hinsicht auf relevante Außenkriterien (Studie 2) sowie die Klassifikationsübereinstimmung von Intelligenztestergebnissen (Studie 3). Implikationen für Forschung und Praxis werden diskutiert.

1 Einführung

Schon lange ist sich die Forschungsgemeinschaft größtenteils einig: „*g matters*“ (Gottfredson, 1997, S. 79) und bescheinigt in zahlreichen Untersuchungen die psychometrische Existenz, Generalität, Stabilität und Relevanz eines allgemeinen Intelligenzfaktors im Sinne Spearman's „*g*“ (vgl. Deary, 2012; Jensen, 1998; Jensen & Weng, 1994; Lubinski, 2004; Spearman, 1904). Auch nach über 100-jähriger Intelligenzforschung herrscht jedoch wenig Einigkeit hinsichtlich der Faktorstruktur sowie Relevanz spezifischer Intelligenzfaktoren unterhalb oder neben *g*. Galt gegen Ende des letzten Jahrhunderts die weitgehend vorherrschende Meinung, spezifische Intelligenzfaktoren seien nahezu unbrauchbar für die Vorhersage relevanter Kriterien (z.B. Leistungsvariablen in Schule, Studium und Beruf) und als Residualvarianz eines überstrahlenden Generalfaktors vernachlässigbar (vgl. Jensen, 1998; Murphy, 2017) – manch einer bezeichnet diese Epoche, die klar der Vorstellung einer Vorherrschaft von *g* galt, als „*not-much-more-than-g*“-Ära (vgl. z.B. Murphy, 2017) – lässt sich in den letzten Jahren vor allem im Bildungskontext ein Umdenken und wachsendes Interesse an der Erforschung vor allem spezifischer Intelligenzfaktoren neben *g* beobachten (z.B. Beaujean, Parkin & Parker, 2014; Coyle, 2014; Coyle, Snyder Richmond & Little, 2015; Saß, Kampa & Köller, 2017; Schult & Sparfeldt, 2016). Unlängst widmeten die Herausgebenden des *Journal of Intelligence* eine Sonderausgabe dieser Kontroverse, der so genannten „*Great Debate*“ (vgl. Kell & Lang, 2019). Es wird deutlich: das Interesse an der Erforschung sowohl allgemeiner als auch spezifischer Intelligenzfaktoren sowie deren Prädiktionskraft für wichtige Außenkriterien ist relevant und hoch aktuell.

In der vorliegenden Arbeit (insbesondere in Studie 1 und Studie 2) werden Teilaspekte dieser oben genannten – nach wie vor aktuellen – Kontroverse beleuchtet. In Studie 1 untersuchten wir mit dem Nested-Factor-Model Ansatz konfirmatorisch die faktorielle Struktur von allgemeinen sowie spezifisch verbalen, numerischen und figuralen Intelligenzfaktoren und deren testübergreifenden Korrelationen (Kriteriumsvalidität) zu allgemeinen und spezifischen Faktoren dreier unabhängiger Testbatterien (vgl. Valerius & Sparfeldt, 2014). In Studie 2 betrachteten wir die korrelativen Zusammenhänge zwischen einerseits allgemeinem wie spezifischen Intelligenzfaktor(en) mit allgemeinem wie spezifischen Schulleistungsfaktor(en) (Schulnoten) jeweils spezifiziert im Nested Factor Model (Valerius & Sparfeldt, 2015). Lagen die ersten beiden Studien thematisch und methodisch-analytisch nah beieinander und fokussierten die Konsistenz der Messung und Relevanz allgemeiner und spezifischer

Intelligenzfaktoren, gingen wir in der dritten Studie der praxisnahen Frage nach, inwieweit die Klassifikationsaussagen im Rahmen einer Hochbegabungsdiagnostik mit verschiedenen Intelligenztests übereinstimmen. Wenn auch das methodische Vorgehen in den ersten beiden Studien ähnlicher gestaltet wurde (konfirmatorische Faktorenanalysen mit latenten Faktoren) im Vergleich zur dritten Studie (Übereinstimmungsanalysen klassifikatorisch-diagnostischer Aussagen „hochbegabt“ vs. „nicht hochbegabt“ mit manifesten Variablen), so schlägt die übergeordnete Frage nach Validität, Güte und Brauchbarkeit von Intelligenztests als verbindendes Element die Brücke zwischen den einzelnen Fragestellungen dieser Dissertation.

2 Theoretischer Rahmen

In diesem Kapitel spanne ich den theoretischen Rahmen über Forschungsaspekte, die dieser Dissertation zugrunde liegen. Zunächst widme ich mich zum einen dem Modellvergleich von Higher-Order- und Nested-Factor-Modellen und zum anderen, welche Bedeutung die zwei unterschiedlichen Modellvarianten für das Verständnis vom Wesen einer zugrundeliegenden Intelligenz(struktur)theorie haben. Des Weiteren fokussiere ich die Themen der konsistenten Messung und Relevanz (Zusammenhänge zu Schulleistung) allgemeiner und spezifischer Intelligenzfaktoren. Abschließend widme ich mich der Thematik der Übereinstimmungsgüte von Klassifikationsaussagen im Bereich der intellektuellen Hochbegabung.

2.1 Intelligenzstruktur: Nested-Factor-Modell vs. Higher-Order Modell

Das Nested-Factor-Modell (*NF-Modell* [Gustafsson & Balke, 1993]; synonym zu *bifactor-model* [Holzinger & Swineford, 1937]; *general hierarchical factor model* [Yung, Thissen & McLeod, 1999] oder *direct hierarchical model* [Gignac, 2008]) – als Kernelement der vorliegenden ersten zwei Studien – und das Higher-Order Factor Modell (*HO-Modell*; synonym *second-order-model* [vgl. Mansolf & Reise, 2017]) stellen die zwei am häufigsten genutzten Modellvarianten konfirmatorischer Faktoranalysen zur Erforschung der hierarchischen Intelligenzstruktur von Generalfaktor und spezifischen Intelligenzfaktoren dar (vgl. Reeve & Blacksmith, 2009) und sollen daher vorangestellt im Folgenden differenziert dar- und kritisch gegenüber gestellt werden.

Im Rahmen konfirmatorischer Faktorenanalysen stellen das HO-Modell und das NF-Modell zwei unterschiedliche Spezifizierungen der zugrundeliegenden Intelligenzstruktur mit Generalfaktor (*g*) und spezifischen Gruppenfaktoren dar. In einem typischen HO-Modell (s. Abb. 1, S.4, linke Seite) lädt jeder Indikator (Subtest) auf einem spezifischen Gruppenfaktor erster Ordnung und diese Gruppenfaktoren laden wiederum auf dem Generalfaktor zweiter Ordnung. Der Einfluss von *g* auf die Indikatoren wird somit komplett durch die Gruppenfaktoren mediiert (vgl. Gignac, 2008). Im Gegensatz dazu ist der Einfluss von *g* auf die Subtests im NF-Modell (s. Abb. 1, S. 4, rechte Seite) ein direkter: jeder Indikator lädt im NF-Modell sowohl auf dem zugehörigen spezifischen Gruppenfaktor als auch auf dem Generalfaktor. Die Gruppenfaktoren und der Generalfaktor sind orthogonal spezifiziert und somit unabhängig voneinander. HO-Modelle sind sparsamer, dafür restriktiver, da alle

direkten Effekte von g auf die Subtests auf Null fixiert sind. Im Gegensatz dazu sind NF-Modelle komplexer, dafür weniger restriktiv. Obwohl vielleicht nicht offensichtlich, ist das HO-Modell im NF-Modell verschachtelt (vgl. Yung et al., 1999) und wäre mathematisch äquivalent zum NF-Modell, wenn die fixierten Nullladungen des HO-Generalfaktors auf die Subtests freigesetzt würden.

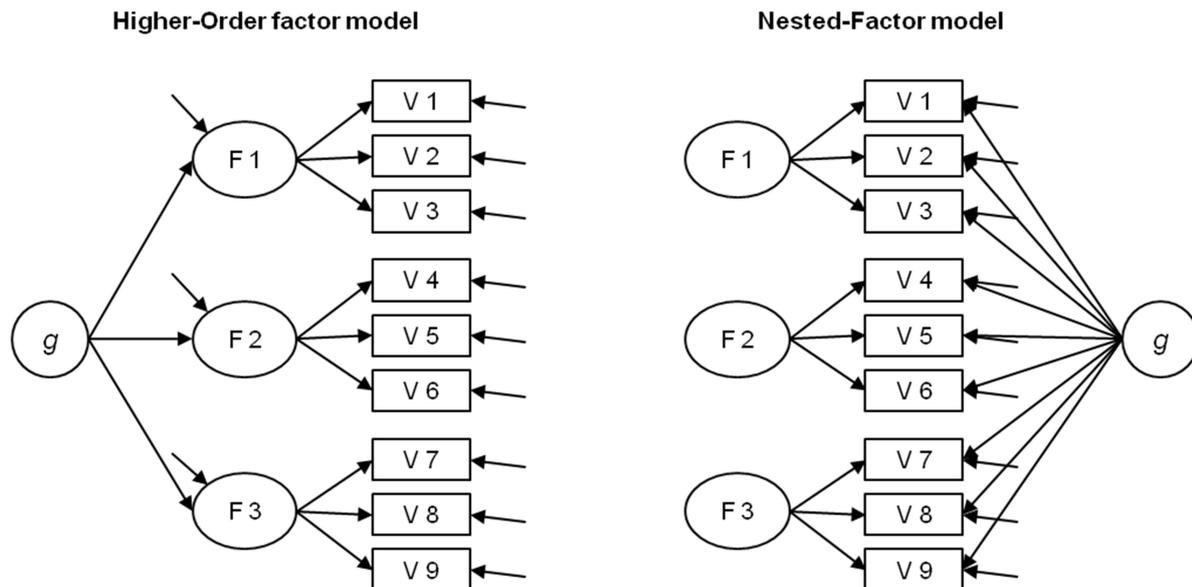


Abb. 1: g = Generalfaktor, F1-F3 = domänenspezifische Faktoren, V1-V9 = manifeste Variablen.

2.1.1 Statistische Unterschiede und Einschränkungen

NF-Modelle schneiden im statistischen Modellvergleich gegenüber HO-Modellen häufig besser ab (vgl. Murray & Johnson, 2013; Brunner, Nagy & Wilhelm, 2012), stehen jedoch in der Kritik, im Vergleich vor allem zum HO-Modell möglicherweise eine misspezifizierte Tendenz zugunsten des NF-Modells zu enthalten (vgl. Maydeu-Olivares & Coffman, 2006; Morgan, Hodge, Wells & Watkins, 2015). Bis dato werden die sogenannten „*proportionality constraints*“, die das HO-Modell im Gegensatz zum NF-Modell enthält (vgl. Brunner et al., 2012; Gignac, 2016; Molenaar, 2016; Schmiedek & Li, 2004; Yung et al., 1999), bzw. so genannte „*tetrad constraints*“ (vgl. Bollen & Ting, 1993 und ausführlicher bei Mansolf & Reise, 2017), als Erklärung für die statistische Überlegenheit des (möglicherweise misspezifizierten) NF-Modells gegenüber dem HO-Modell diskutiert.

Trotz nicht vernachlässigbarer Einschränkungen (wie oben dargestellt) werden gerade im Rahmen der Debatte über die Relevanz und Vorhersagekraft insbesondere spezifischer Intelligenzfaktoren zu Außenvariablen über die des Generalfaktors hinaus (vgl. Kell & Lang, 2019), vor allem in der letzten Dekade um Kriteriumsvariablen oder -faktoren erweiterte NF-Modelle bevorzugt eingesetzt (z.B. Beaujean et al., 2014; Brunner, 2008; Saß et al., 2017; Schult & Sparfeldt, 2016). Den wesentlichen Vorteil des NF-Modells stellt die orthogonale Spezifizierung des Generalfaktors und der spezifischen Faktoren dar und somit die Möglichkeit, in erweiterten NF-Modellen die spezifischen Faktoren distinkt – d.h. unabhängig von g – mit relevanten Außenkriterien wie beispielsweise schulischen oder beruflichen Leistungsvariablen frei von g -Varianz in Beziehung zu stellen (vgl. Reise, 2012). Die spezifischen Faktoren und der Generalfaktor können somit im erweiterten NF-Modell als unabhängige Vorhersagequellen in einem multiplen Regressionsrahmen genutzt werden. Gerade aber diese erweiterten NF-Modelle zur distinkten Vorhersage von Kriteriumsvariablen hielten laut Eid, Krumm, Koch & Schulze (2018) unter bestimmten Bedingungen Fallstricke wie Nichtidentifizierbarkeit und Schätzprobleme bereit. An einer Stichprobe von $N = 219$ Jugendlichen und jungen Erwachsenen ($M = 16$ Jahre; $SD = 1.49$; Range 13 bis 23 Jahre; 53% weiblich) demonstrierten sie in einem um Kriteriumsvariablen (manifeste Noten in Mathematik und Englisch) erweiterten NF-Modell mit einem Generalfaktor und drei orthogonalen spezifischen Intelligenzfaktoren (AN: verbal analogies, NS: number series und UN: unfolding¹) zweierlei: (1) Das erweiterte NF-Modell² mit gleichen Ladungen (Ladungen von 1) auf dem Generalfaktor und den spezifischen Faktoren konvergierte nicht (aufgrund von Identifikationsproblemen), obwohl das NF-Modell (mit ebenfalls gleichen Ladungen) ohne die Kriteriumsvariablen gut fittete ($\chi^2 = 10.121$, $df = 11$, $p = .520$). (2) Die zur Identifikation nötigen Fixierungen der Kovarianzen von entweder Generalfaktor oder spezifischen Faktoren mit den Kriteriumsvariablen auf Null, führten bei jeweils guter Modellpassung zu einem vollständig gegensätzlichen Befundmuster. D.h. bei Fixierung der Kovarianz von spezifischen Faktoren mit den Kriteriumsvariablen auf Null, korrelierte der Generalfaktor signifikant mit der Mathematiknote ($r = .57$) und der Englischnote ($r = .34$). Dies ließe laut Autoren die Interpretation zu, dass nur g zur Vorhersage von Schulnoten relevant sei. Sowie der umgekehrte Befund, dass bei Fixierung der Kovarianz des Generalfaktors mit den Kriteriumsvariablen auf Null, alle spezifischen Faktoren signifikant

¹ Gebildet jeweils aus zwei nach der *odd-even Methode* geparcelten Subtests des Wilde Intelligenz Test 2 (Kersting, Althoff & Jäger, 2008). Ausführlicher zur *odd-even Methode* vgl. Valerius & Sparfeldt (2014, S. 127).

² Alle Faktoren konnten mit den Kriteriumsvariablen korrelieren.

mit der Mathematiknote (NS: $r = .52$, AN: $r = .57$, UN: $r = .45$) und der Englischnote (NS: $r = .32$, AN: $r = .43$, UN: $r = .18$) korrelierten und entsprechend den umgekehrten Schluss zuließen. Wonach zur Vorhersage von Schulnoten nur die spezifischen Faktoren relevant und g irrelevant seien. Eid et al. (2018) resümierten daher: „*applying the bifactor model with criterion variables as covariates can result in invalid conclusions about the importance of general and specific factors*” (S. 17). Alternativ empfehlen Eid et al. (2018) die für Identifikations- und Schätzprobleme weniger anfällige Spezifizierung von „Bifactor (S-1) Modellen“ (s. Abb. 2). Hierbei handelt es sich um eine Variante des klassischen NF-Modells, in dem ein spezifischer Faktor bei der Spezifizierung weggelassen wird, welcher als Referenzfacette für den Generalfaktor betrachtet werden kann. Da die Referenzfacette maßgeblich die Bedeutung des Generalfaktors und der spezifischen Faktoren bestimme, empfehlen Eid et al. (2018), die Wahl der Referenzfacette nach inhaltlich-theoretischen Überlegungen und auf die Forschungsfrage abgestimmt zu treffen. Obgleich sich die erweiterten NF-Modelle der ersten beiden Studien dieser Arbeit hinsichtlich der von Eid et al. (2018) thematisierten Problematik unterscheiden (wie später erläutert), stelle ich im Anschluss an die empirischen Studien (Valerius & Sparfeldt, 2014; Valerius & Sparfeldt, 2015) jeweils eine vergleichende Zusatzanalyse mit spezifizierten Bifactor (S-1) Modellen an.

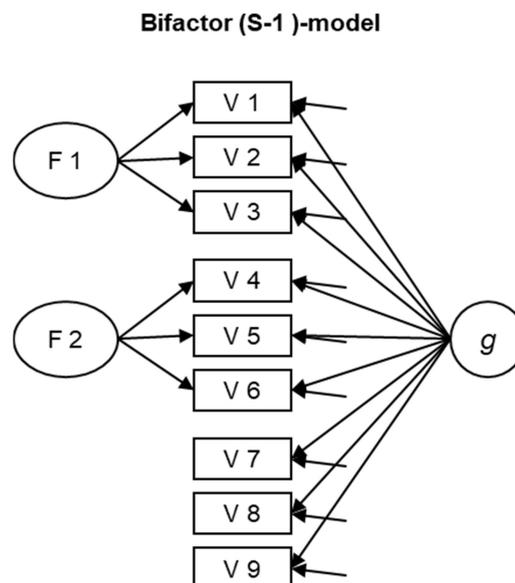


Abb. 2: g = Generalfaktor, F1-F2 = spezifische Faktoren, V1-V9 manifeste Variablen.

2.1.2 Theorie und Wesen der Intelligenz(struktur) im NF- vs. HO-Modell

Bei der Entscheidung für oder gegen ein bestimmtes Analysemodell sollte neben der inhaltlichen Fragestellung vor allem auch die Passung zur zugrundeliegenden Intelligenztheorie sorgfältig berücksichtigt werden (vgl. Kell & Lang, 2018). Wie oben erläutert, stellen das HO- und das NF-Modell unterschiedliche statistische Repräsentationen einer hierarchisch strukturierten Intelligenzstruktur mit g und zusätzlichen Faktoren neben g dar. Ein empirisch umfangreich abgesichertes hierarchisches Intelligenzstrukturmodell stellt die *Cattell-Horn-Carroll* Theorie kognitiver Fähigkeiten (CHC; z.B. McGrew, 2009) als Integral zweier substanziell ähnlicher Intelligenztheorien, der *Gf-Gc* Theorie fluider und kristalliner Intelligenz (Cattell, 1941; Horn & Cattell, 1966) und der *Three-Stratum* Theorie (Carroll, 1993), dar. Das CHC-Modell umfasst drei Schichten (s.g. „Strata“) kognitiver Fähigkeiten, die unterschiedlich weit gefasst werden. Stratum III: General ability [ähnlich zu g]; Stratum II: acht „breitere“ oder spezifischere Fähigkeiten (Fluide Intelligenz [Gf], Kristalline Intelligenz [Gc], Allgemeine Gedächtnisfähigkeit [Gy], Breite visuelle Wahrnehmung [Gv], Breite auditive Wahrnehmung [Gu], Breite Abruffähigkeit [Gr], Breite kognitive Schnelligkeit [Gs] und Verarbeitungsgeschwindigkeit [Gt]); und Stratum I: ca. 70 Primärfaktoren, die „nähere“ aufgabenspezifische Fähigkeiten umfassen. Das CHC-Modell entspricht der Vorstellung vieler moderner Intelligenzforschenden (vgl. Neisser et al., 1996, S. 81), dennoch ist nicht ganz klar, wie Carroll die Faktoren in seinem Modell konzeptualisierte: als HO-Modell oder als NF-Modell? In den meisten Fällen wurde das CHC-Modell als HO-Modell spezifiziert, ließ sich jedoch ebenso erfolgreich als NF-Modell darstellen (vgl. Beaujean et al., 2014). Verschiedene Schriften von Carroll analysierend argumentiert Beaujean (2015) darüber hinaus, dass Carroll explizit annahm, dass ein NF-Modell das Wesen der Intelligenz besser erfasse und seine Theorie am besten repräsentiere als das weitaus häufiger angenommene HO-Modell.

Der Einfluss des Generalfaktors auf die einzelnen (Sub-) Tests ist im NF-Modell ein direkter und entspricht einer eher breiten (und ursprünglich von Spearman, 1904 angenommenen) Konzeptualisierung von g , da alle Subtests auf g laden. Durch die übergeordnete Konzeptualisierung von g im HO-Modell ist dessen Einfluss auf die (Sub-) Tests mediiert durch die untergeordneten spezifischen Gruppenfaktoren und somit ein indirekter (vgl. Gignac, 2008). Beaujean (2015, S. 128) argumentiert: die Konzeptualisierung der spezifischen (Gruppen-) Faktoren unterscheidet sich ebenfalls: im NF-Modell spiegeln sie, vereinfacht gesprochen, den gemeinsamen Varianzanteil einer Gruppe von Subtests wider, der

übrig bleibt, nachdem der Varianzanteil von g geklärt wurde (s.g. „*top-down*“ Prozess, vgl. Carroll & Schweiker, 1951; Gustafsson & Balke, 1993). Daher sind g und die spezifischen Faktoren unabhängig voneinander. Im HO-Modell hingegen spiegeln die spezifischen Gruppenfaktoren – vereinfacht ausgedrückt – wider, was eine Gruppe von Subtests gemeinsam hat, sodass bei ausreichend vorhandener gemeinsamer Varianz der Gruppenfaktoren g gebildet werden kann (s.g. „*bottom-up*“ Prozess, vgl. Carroll & Schweiker, 1951; Gustafsson & Balke, 1993). Wenn g gebildet ist, kann analysiert werden, welche Varianzanteile in den spezifischen Gruppenfaktoren auf g und welche auf spezifische Anteile (Residuen) zurückzuführen sind. Die Gruppenfaktoren und g sind demnach im HO-Modell nicht unabhängig voneinander. In der Anwendung bedeutet dies Folgendes: Im HO-Modell würde man annehmen, dass die kognitive Leistung beim Lösen einer (Intelligenztest-) Aufgabe (bspw. verbale Analogien) auf eine eher „breitere“ (sensu CHC-Theorie, McGrew, 2009) Fähigkeit (bspw. Reasoning) zurückzuführen ist, die wiederum beeinflusst wird durch eine übergeordnete allgemeine Fähigkeit (g) und systematische aufgabenspezifische Fähigkeit (bspw. Wortschatz, Umgang mit Sprache) plus unsystematische Residualvarianz. Im NF-Modell würde man annehmen, dass die kognitive Leistung zum Lösen der Aufgabe durch eine allgemeine Fähigkeit (g), komplexe Aufgaben zu lösen beeinflusst wird und zusätzlich durch eine unabhängige, für die Aufgabe relevante, eher „breitere“ spezifische Fähigkeit (z.B. verbales Reasoning) neben unsystematischer Residualvarianz.

Zusammengefasst lässt sich festhalten: Eine einheitliche, empirisch gut abgesicherte Leitlinie oder Empfehlung innerhalb der aktuellen psychometrischen Intelligenzforschung scheint es hinsichtlich des klassischen HO- vs. NF-Modell Vergleichs aktuell (noch) nicht zu geben. Auch die Klärung hinsichtlich der Bedingungen für problemfreies Modellieren mit erweitertem NF-Modell mit externen Variablen/Faktoren, steht erst am Anfang. Eines scheint jedoch sicher: Bei der Wahl des NF-Modells sollte nicht der statistische Fit als einziges Kriterium herangezogen werden, sondern auch von Forschungsfrage und Messmodell (vgl. Murray & Johnson, 2013) abhängen. Der Nutzen des NF-Modells besteht vor allem darin, Zusammenhänge der spezifischen Gruppenfaktoren – neben denen des Generalfaktors – zu externen Faktoren/Variablen, distinkt zu untersuchen, d.h. frei von g -Varianz (vgl. Chen, West & Sousa, 2006). Dies war ein wesentlicher Aspekt und der ausschlaggebende Grund für die Verwendung des NF-Modells zur Beantwortung der Forschungsfragen der ersten beiden Studien dieser Arbeit. Abgesehen von den methodischen Vorteilen der NF-Modellierung ist

jedoch die psychologisch-inhaltliche Interpretation spezifischer NF-Fähigkeitsfaktoren unabhängig vom Generalfaktor nicht abschließend geklärt.

2.2 Konsistenz und Relevanz allgemeiner und spezifischer Intelligenzfaktoren

Forschungsgeschichtlich lässt sich aktuell ein Richtungswechsel in der psychometrischen Intelligenzforschung beobachten. Lange Zeit stand der Generalfaktor im Sinne Spearman's g (Spearman, 1904) im Fokus und spezifische Intelligenzfaktoren unterhalb oder neben g galten nahezu konsensuell – mit wenigen Ausnahmen – als vernachlässigbar, beziehungsweise von wenig inkrementellem Nutzen. Rückblickend bezeichnet Murphy (2017) die Vorherrschaft von g als voreilig und bekräftigt *“that measures of specific abilities can make important contributions even if general measures are taken into account”* (Murphy, 2017, S. 4). Auch Coyle (2014) argumentiert ähnlich, indem er annimmt, dass die Erforschung der prädiktiven Validität der non- g Residuen von Tests eine der vielversprechendsten neuen Ideen in der Erforschung menschlicher Intelligenz darstelle (S. 21). Die ersten beiden Studien dieser Arbeit setzen genau daran an und fokussieren neben der Konsistenz der Messung und Relevanz von g vor allem die Konsistenz und Relevanz der spezifischen Intelligenzfaktoren neben dem Einfluss des Generalfaktors. Im Folgenden stelle ich die ersten beiden Studien dieser Arbeit in den Kontext bestehender Forschung.

2.2.1 Konsistenz allgemeiner und spezifischer Intelligenzfaktoren

Sowohl die Forschung als auch die angewandte Psychologie ist wesentlich darauf angewiesen, dass Intelligenztests unabhängig voneinander stabile und verlässliche Ergebnisse liefern (Konsistenz der Messung).

Ein Fokus dieser Arbeit liegt klar auf dem Aspekt der konvergenten Validität mit unterschiedlichen Testbatterien. Aufbauend auf zwei Studien von Johnson und Mitarbeitenden (Johnson, Bouchard, Krueger, McGue & Gottesman, 2004; Johnson, te Nijenhuis & Bouchard, 2008), die mit (sehr) hohen Korrelationen der g -Faktoren ($.77 \leq r \leq 1.00$) unabhängig voneinander entwickelter Intelligenztestbatterien die Konsistenz der Messung von g belegen, wurde die Fragestellung der ersten Studie dieser Arbeit zur konvergenten Validität allgemeiner und spezifischer Intelligenzfaktoren entwickelt.

Johnson und Mitarbeitende (2004, 2008) analysierten faktorenanalytisch die Daten von $N = 436$ Erwachsenen, die drei unabhängige Testbatterien bearbeiteten (*Comprehensive Ability Battery*, CAB, Hakstian & Cattell, 1975 [14 tests]; *Hawaii Battery*, HB, DeFries et al., 1974, inklusive Raven's Progressive Matrices, 1941 [17 tests]; *Wechsler Adult Intelligence Scale*, WAIS, Wechsler, 1955 [11 tests]). Die Fit-Indices der drei Modelle (pro Testbatterie) jeweils mit second-order g -Faktor an der Spitze, waren mindestens akzeptabel – in Einklang mit Hu und Bentlers (1999) *cut-off* Kriterien (RMSEA [CAB/HB/WAIS] = .031/.050/.061). Der Fit des kombinierten Modells, welches die drei testbatteriespezifischen hierarchischen Modelle mit testbatteriespezifischem g -Faktor an der Spitze enthielt, war ebenfalls akzeptabel (RMSEA = .069); die Korrelationen der korrespondierenden g -Faktoren reichte von $r = .99$ bis 1.00. Zusätzlich zeigte dieses testbatteriespezifische second-order Model einen besseren Fit als (a) ein Modell mit batteriespezifischen g -Faktoren ohne ein zwischengelagertes Hierarchielevel spezifischer Gruppenfaktoren (RMSEA = .104; Johnson et al., 2004, p. 104) und (b) einen vergleichbaren Fit für ein Model mit gleicher first-order Struktur aber mit nur einem einzigen g -Faktor (Johnson et al., 2004, p. 104; RMSEA=.069). Konsequenterweise schloss Deary (2012), dass “*the individual differences in g were identical from the three different batteries*” (p. 457). In einer Replikationsstudie reanalysierten Johnson et al. (2008) eine Datenmatrix von de Wolf and Buiten (1963) die auf Daten von $N = 500$ 16-jährigen niederländischen Marinerekruten beruhte. Johnson et al. (2008) nutzten für ihre Analysen 44 Tests aus fünf Batterien: (1) die *Battery of Royal Dutch Navy* (RDN; 8 Subtests), (2) eine Adaptation der Testbatterie des *Twente Instituut voor Bedrijfspsychologie* (TIB; 13 Subtests), (3) den *Cattell Culture Test* (CCFT; 4 Subtests), (4) die *General Aptitude Test Battery* (GATB; 12 Subtests), und (5) den *Groningse Intelligentie Test* (GIT; 7 Subtests). Konfirmatorische Faktorenanalysen erzielten gute bis akzeptable Fit-Statistiken in separaten Analysen pro Testbatterie mit testbatteriespezifischem second-order g -Faktor (RMSEA [RDN/TIB/CCFT/GATB/GIT] = .071/035/.000/.046/.040), genauso wie das kombinierte Modell mit allen Testbatterien (RMSEA = .073). Erneut waren die Korrelationen der testbatteriespezifischen g -Faktoren sehr hoch und reichten von $r = .77$ für CCFT mit GATB bis 1.00 für TIB mit GATB und respektive GIT. Dadurch, dass jeder Test restringiert wurde und nur auf demjenigen Testbatterie-Faktor laden durfte, von dem er stammte, ergaben sich teilweise g -Faktor-Korrelationen, die die statistische Grenze von 1.00 überschritten. Mit der Absicht diese Korrelationen unterhalb des Wertes von 1.00 zu bringen, erlaubten Johnson et al. (2008) in einer nachträglichen Modellanpassung, dass Residuen und First-order Faktoren batterieübergreifend korrelieren durften. Johnson et al. (2004, 2008) fanden zusätzlich

substanzielle und systematische Korrelationen zwischen Subtests und first-order Faktoren, die nicht durch g aufgeklärt wurden, vernachlässigten jedoch deren genauere Betrachtung.

Dies wiederum taten Floyd, Reynolds, Farmer & Kranzler (2013). Sie fanden in 5 Stichproben von ($n = 83$ bis $n = 200$; $N_{ges} = 684$) Kindern und Jugendlichen (6 bis 17 Jahre alt), die jeweils zwei unabhängige Intelligenztestbatterien bearbeiteten, neben sehr hohen g -Faktor Korrelationen ($.89 \leq r \leq 1.00$) zusätzlich hohe Korrelationen (einiger) spezifischer CHC-Stratum II Faktoren über die Testbatterien hinweg (Comprehension Knowledge: $.81 \leq r \leq .98$; Short Term Memory $.84 \leq r \leq .88$; Processing Speed: $.90 \leq r \leq .97$; Fluid Reasoning: $r = .94$). Als konfirmatorisch faktorenanalytisches Modell wählten die Autoren – wie bereits Johnson et al. (2004, 2008) – das HO-Modell und nutzten die Residuen der spezifischen Stratum II Faktoren als Kovariaten. Die Ergebnisinterpretation ist jedoch – neben dem Problem einzelner kleiner Stichproben und damit verbundenen Signifikanzeinschränkungen – dahingehend erschwert, dass die Autoren ebenfalls wie Johnson et al. (2004, 2008) Subtest-Residuen korrelierten und somit die Korrelation der höhergeordneten Stratum II und III Faktoren beeinflussten.

Der Frage nach konsistenter Messung von spezifischen Intelligenz-Gruppenfaktoren neben g gingen wir im Rahmen der ersten Studie nach. Wie bereits angedeutet, bot sich der NF-Modell Ansatz an, da er die distinkte Erfassung von g und spezifischen Faktoren unabhängig voneinander ermöglicht. Der Ansatz, durch unabhängige Testbatterien gewonnene Daten, mithilfe des NF-Modells hinsichtlich dieser Fragestellung zu analysieren, wurde bis zum Erscheinen des ersten Artikels in keiner weiteren Arbeit (soweit mir bekannt) verfolgt.

2.2.2 Allgemeine und spezifische Intelligenzfaktoren als Korrelate und Prädiktoren schulischer Leistung

Die allgemeine Intelligenz (im Sinne Spearman's g) gilt weithin als (ein) wesentlicher Prädiktor für schulische Leistungen (z.B. Jensen, 1998; Roth, Becker, Romeyke, Schäfer, Domnick, & Spinath, 2015). Über den inkrementellen Nutzen spezifischer Intelligenzfaktoren gibt es in diesem Zusammenhang hingegen weniger Einigkeit. Die Hypothese einer „Vorherrschaft von g “ (vgl. Coyle, 2018a) wird mehrheitlich gestützt. Darin wird angenommen, dass g die Vorhersagekraft von Tests maßgeblich bestimmt und die spezifischen Faktoren begrenzte oder vernachlässigbare Prädiktionskraft besitzen (vgl. auch Rost, 2013). Einige wenige – auch aktuellere – Studien befassen sich jedoch mit der

systematischen Erforschung des Zusammenhanggefüges von Intelligenz und Schulleistung neben oder unterhalb von g . Auf die NF-Modellierung griffen dabei beispielsweise die Arbeiten von Beaujean et al. (2014), Brunner (2008), Gustafsson & Balke (1993) sowie Schult & Sparfeldt (2016) zurück. In den Arbeiten von beispielsweise Coyle, Purcell, Snyder & Kochunov (2013) sowie Coyle et al. (2015) prädierten die Autoren im HO-Strukturgleichungsmodell über die g -unabhängigen, latenten Residuen der Gruppenfaktoren abhängige schulisch/akademische Variablen. Beide Ansätze haben gemein, dass die spezifischen Gruppenfaktoren im NF-Modell sowie die g -unabhängigen Residuen der spezifischen Gruppenfaktoren (bspw. im HO-Modell) beide eine Quelle gemeinsamer Varianz (der entsprechenden Subtests) orthogonal zum g -Faktor darstellen (vgl. Gignac, 2016; Yung, et al., 1999) und somit eine zusätzliche, spezifische Erklärungsquelle schulisch/akademischer Leistung über den Einfluss des g -Faktors hinaus liefern. Die genannten Studien belegen übereinstimmend mindestens konvergente und – teilweise darüber hinaus – divergente Beziehungsmuster der spezifischen Gruppenfaktoren mit spezifischen Schulleistungsfaktoren oder -variablen. Dies zeigte sich darin, dass beispielsweise verbale respektive numerische Intelligenzfaktoren (bzw. verbale respektive numerische g -unabhängige Residuen) positive Zusammenhänge zu sprachlichen respektive mathematisch-naturwissenschaftlichen Schulleistungen und keine bis negative Zusammenhänge zu mathematischer respektive sprachlichen Schulleistungen aufwiesen.

Zusammengefasst kann festgehalten werden, dass die konsistente Messung von HO- g (vgl. Johnson et al., 2004, 2008) und darüber hinaus HO-spezifischen Gruppenfaktoren (vgl. Floyd et al. 2013) mit unterschiedlichen Testbatterien (sehr) gut gelang. Eine Prüfung mit NF-Modell-Ansatz – im Sinne von Methodeninvarianz – wurde bisher vernachlässigt, obwohl sich der NF-Ansatz mit unkorrelierten Faktoren und der Möglichkeit einer distinkten und unabhängigen Erfassung der einzelnen Faktoren, anböte. Darüber hinaus zeigte sich ein sich wiederholendes Befundmuster: spezifische (Intelligenz-) Testfaktoren lieferten neben oder zusätzlich zu g inkrementelle Validität zur Aufklärung des Beziehungsmusters von Intelligenz und Schulleistung. Die distinkte Betrachtung der Zusammenhänge von g und spezifischen Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungsfaktoren unter Verwendung des NF-Modells (Studie 2 der vorliegenden Arbeit) wurde zwar in einigen weiteren Studien realisiert (z.B. Brunner, 2008; Gustafsson & Balke, 1993; Schult & Sparfeldt, 2016), jedoch enthielten diese zusätzlich zu Intelligenzsubtests weitere Studierfähigkeits- oder Schulleistungstests als Indikatoren für die spezifischen und/oder

allgemeinen Faktoren, was möglicherweise eine Überschätzung des Beziehungsmusters zur Folge hat (vgl. Süß, 2001).

2.3 Übereinstimmung von Intelligenztestergebnissen im hohen Begabungsbereich

Latent modelliert (und damit messfehlerbereinigt) korreliert die allgemeine Intelligenz (g) gemessen mit unabhängigen Intelligenztestbatterien sehr hoch (vgl. Johnson et al. 2004, 2008). Gemessen wird somit zu großen Teilen Gleiches. Auf manifester Ebene fallen die Korrelationen von Gesamtergebnissen unabhängiger Testbatterien als Indikatoren für g messfehlerbedingt in der Regel etwas niedriger aus. In pädagogisch-psychologischer Forschung und Anwendung werden Intelligenztests aufgrund ihrer hohen Prädiktionskraft für schulische Leistungen (vgl. z.B. Roth et al., 2015) häufig eingesetzt und sollten somit stabile und verlässliche Ergebnisse liefern. Vor allem dann, wenn ein Intelligenztest als Klassifikationsinstrument genutzt wird, beispielsweise um anhand von festgelegten Grenzwerten eine besonders hohe oder niedrige Begabung festzustellen, und entsprechende bildungsrelevante Implikationen darauf folgen. Intelligenztests nehmen zur Identifikation von intellektueller Hochbegabung einen zentralen Platz ein (vgl. Preckel, 2010; Preckel & Vock, 2013; Rost, 2009; Rost & Buch, 2018) und ab konventionell festgelegten Grenzwerten von PR 98 (IQ 130) – gelegentlich auch liberaler gefasst mit PR 95/90 (IQ 125/120) – werden Testpersonen als „hochbegabt“ klassifiziert. Werden zur Identifikation von Hochbegabung nun zwei (oder mehr) unabhängigen Intelligenztests eingesetzt – was einem relativ gängigen Vorgehen entspricht, sind die statistisch erwarteten Anteile derjenigen, die in beiden Testverfahren Werte über der entsprechenden Grenzmarke erzielen, limitiert (vgl. Lohman 2009; Lohman & Korb, 2006). Die Autorenschaft konnte eindrucksvoll aufzeigen, dass bei einer manifesten Testinterkorrelation von $r = .85$ statistisch betrachtet nur knapp die Hälfte (49%) derjenigen, die in einem ersten Intelligenztest die Grenzmarke von PR 98 überschreiten, auch ein entsprechend hohes Ergebnis im zweiten Test erzielen können. Bei liberaleren Grenzmarken (z.B. PR 95/90) und höheren respektive niedrigeren Testinterkorrelationen steigen respektive sinken erwartungsgemäß die erwarteten Anteile (vgl. Lohman & Korb, 2006, Tab. 1, S. 457).

Für die Überprüfung der Testergebnisse über das gesamte Begabungsspektrum konnten Floyd, Clark & Shadish (2008) an Daten von sechs Stichproben (Kinder im Schulalter und Studierenden; ausführliche Stichprobenbeschreibung bei Floyd, Bergeron, McCormack,

Anderson, Hargrove-Owens, 2005), die im Rahmen von Validitätsstudien zu verschiedenen Intelligenztestbatterien (*Kaufman Assessment Battery for Children – Second Edition*, KABC-II: Kaufman & Kaufman, 2004; *Kaufman Adolescent and Adult Intelligence Test*, KAIT: Kaufman & Kaufman, 1993; *Wechsler Intelligence Scale for Children – Third Edition*, WISC-III: Wechsler, 1991; *Wechsler Intelligence Scale for Children – Fourth Edition*, WISC-IV: Wechsler, 2003; *Wechsler Adult Intelligence Scale – Third Edition* WAIS-III: Wechsler, 1997; *Woodcock-Johnson Psychoeducational Battery – Third Edition*, WJ III: Woodcock, McGrew & Mather, 2001; *Differential Ability Scales*, DAS: Elliott, 1990) erhoben wurden, zeigen, dass im paarweisen Vergleich 74% der Teilnehmenden ähnliche IQ-Werte (d.h. die intraindividuellen Differenzen der IQ-Werte der Probanden lagen unterhalb der festgelegten Grenzmarke von 10 IQ-Punkten) in jeweils zwei verschiedenen Testverfahren erzielten. Wurde zur Bestimmung der kritischen Differenz als Maß für (Un-) Ähnlichkeit der IQ-Wert-Differenzen das aktuelle Konfidenzintervall herangezogen, lag die Ähnlichkeitsquote bei 66%. Trotz überwiegend (im Rahmen der jeweiligen Konfidenzintervalle) übereinstimmender Testergebnisse über unterschiedliche Testbatterien hinweg, konnte für ein Viertel bis ein Drittel der Teilnehmenden kein hinreichend übereinstimmendes Ergebnis in einem zweiten Intelligenztest erzielt werden.

Ein möglicher Ansatz zur Bestimmung der Übereinstimmungsgüte beim Einsatz zweier Intelligenztests kann durch die Betrachtung der Gesamttrefferquote erfolgen. Kommen im Rahmen der Hochbegabungsklassifikation zwei Intelligenztests zum Einsatz, entsteht unter Berücksichtigung der klassifikatorischen Grenzwertziehung „hochbegabt“ ab beispielsweise $IQ \leq 130$ ein Vier-Felder-Klassifikationsschema (vgl. Abb. 3, S. 15: Feld a: Schülerinnen und Schüler, die in beiden Tests übereinstimmend als „hochbegabt“ klassifiziert wurden; Felder b und c: Schülerinnen und Schüler, die nur in einem der beiden Tests – also nicht übereinstimmend – als „hochbegabt“ klassifiziert wurden; Feld d: Schülerinnen und Schüler, die in beiden Tests als übereinstimmend „nicht hochbegabt“ klassifiziert wurden). Die Gesamttrefferquote beschreibt den Anteil der übereinstimmend positiv Klassifizierten („hochbegabt“ in beiden Tests; Feld a) und übereinstimmend negativ Klassifizierten („nicht hochbegabt“ in beiden Tests; Feld d) an der Gesamtanzahl ($[a+d]/N$). Aufgrund der sehr hohen Grenzwertsetzung im Hochbegabungsbereich und der somit resultierenden unproportionalen Verteilung in den vier Zellen übersteigt in der Regel der Anteil der in beiden Tests als übereinstimmend „nicht hochbegabt“ Klassifizierten (Feld d), den Anteil der als übereinstimmend „hochbegabt“ Klassifizierten (Feld a) um ein Vielfaches. Folglich fällt

die Gesamttrefferquote zwar relativ hoch aus, sagt jedoch eher etwas über die Übereinstimmungsgüte der Nicht-Hochbegabungsklassifikation als über die Hochbegabungsklassifikation aus. Um dieser Problematik entgegenzuwirken, verwendeten wir zur Klassifikationsgütebestimmung der Hochbegabungsdiagnostik einen bisher nicht verwendeten Ansatz mittels RATZ-Index (*Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote*; vgl. Beißert, Hasselhorn & Lösche, 2014; ausführlicher s. Abschnitt 3.5.1, S. 53 ff.). Dieser Index wurde ursprünglich entwickelt zur Prüfung der Übereinstimmungsgüte von Klassifikationsaussagen (Ergebnis liegt über/unter einem kritischen Wert) gewonnen mit einem Screeningverfahren (z.B. Vorläuferkompetenzen des Lesens und Schreibens) verglichen mit Störungsbefunden (z.B. Feststellung einer Lese-Rechtschreib-Schwäche) aus nachgelagert eingesetzten Testverfahren (z.B. Lesetests, Rechtschreibtests) zu einem späteren Zeitpunkt (vgl. Marx, 1992; Marx, Jansen & Skowronek, 2000). Der RATZ-Index bietet somit die Möglichkeit, die Übereinstimmungsgüte zweier Testverfahren zur Klassifikation von Hochbegabung, zu bestimmen, da er zu quantifizieren vermag, wie viele der als „hochbegabt“ klassifizierten, auch in einem zweiten Test als „hochbegabt“ eingestuft wurden. Bislang wurde in keiner weiteren Studie zur Hochbegabungsidentifikation – soweit mir bekannt – die Verwendung des RATZ-Indexes gewählt und stellt somit einen neuartigen Ansatz in diesem Forschungsbereich dar.

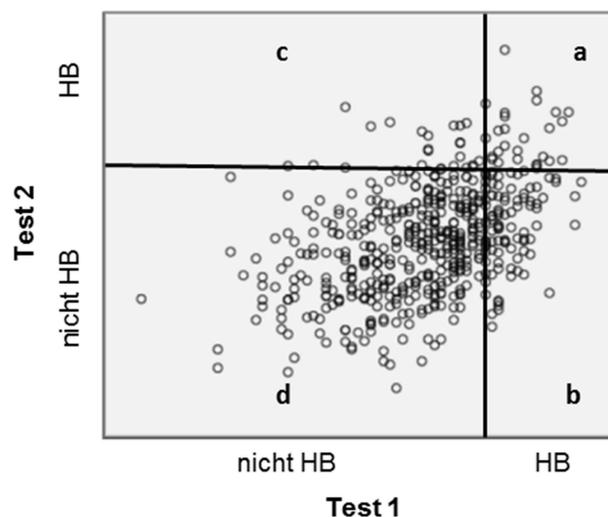


Abb. 3: Klassifikationsfelder der Hochbegabungsidentifikation mit zwei Testverfahren. Das Ergebnis von Test 1 ist an der x-Achse abgetragen, das Ergebnis von Test 2 an der y-Achse. Die Flächenanteile bezeichnen (a) die konsistent Hochbegabten, (d) die konsistent durchschnittlich Begabten, (b) und (c) jeweils nur durch einen Test als hochbegabt Diagnostizierte. HB = hochbegabt.

Zusammengefasst kann festgehalten werden, dass obwohl Intelligenztestergebnisse latent modelliert sehr hoch (manifest in der Regel niedriger) korrelieren (Johnson et al., 2004, 2008), bedeutet es nicht automatisch, dass zum einen die Intelligenztestwerte in zwei Testverfahren im allgemeinen (vgl. Floyd et al, 2008) sowie Klassifikationsaussagen (bspw. „hochbegabt“ vs. „nicht hochbegabt“) mit zwei (oder mehreren) Intelligenztests zu übereinstimmenden Ergebnissen kommen (vgl. Lohman 2009; Lohman & Korb, 2006). Im Rahmen der dritten Studie dieser Arbeit, näherten wir uns dieser Fragestellung.

Insgesamt gesehen bauen die drei Studien der vorliegenden Dissertation insofern aufeinander auf, dass zum einen betrachtet wird, wie konsistent unabhängige Intelligenztests die allgemeine Intelligenz sowie spezifische Fähigkeiten messen (Studie 1) und welche Relevanz diese für schulische Leistungen haben (Studie 2). Darüber hinaus wird anwendungsnah geprüft, inwieweit übereinstimmende Klassifikationsaussagen im Bereich der Hochbegabung mit unabhängigen Intelligenztests gelingen (Studie 3).

3 Empirische Studien

Die drei vorliegenden Studien bilden das Kernstück dieser Dissertation. Der Fokus und Zusammenhang der ersten beiden Studien ist relativ eng hinsichtlich Fragestellung (Validitätsanalysen) und Methode (NF-Modellierung im CFA-Rahmen). Konkret wurden die konvergente (und divergente) Validität allgemeiner sowie spezifischer Intelligenzfaktoren unabhängig voneinander entwickelter Testbatterien untersucht (Studie 1), sowie die Kriteriumsvalidität von allgemeinen und spezifischen Intelligenzfaktoren mit allgemeinen und spezifischen schulischen Leistungsfaktoren (Studie 2). In beiden Studien griffen wir im Rahmen konfirmatorischer Faktorenanalysen auf den NF-Modell-Ansatz zurück.

Im Rahmen der dritten Studie widmeten wir uns der Betrachtung der Übereinstimmungsgüte von Klassifizierungsaussagen im Bereich der intellektuellen Hochbegabung, die durch jeweils zwei unabhängig voneinander entwickelte, jedoch konzeptionell ähnliche Intelligenztests (figurales reasoning) getroffen wurden (Studie 3). Ließen sich die statistischen Analysen in den ersten beiden Studien auf der latenten Betrachtungsebene lokalisieren, fand die statistische Datenanalyse in der dritten Studie auf manifester Ebene statt.

Im Folgenden werden die Studien zusammengefasst dargestellt. Grundlage der Zusammenfassungen bildeten die jeweiligen veröffentlichten (resp. eingereichten [Studie 3]) Zeitschriftenartikel.

3.1 Studie 1: Valerius & Sparfeldt (2014)

Valerius, S. & Sparfeldt, J. R. (2014). Consistent *g*- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory factor analyses using three test batteries. *Intelligence*, 44, 120-133.

3.1.1 Theorie

Einführung

Aktuell herrscht weitestgehend Einigkeit über die Vorstellung eines Generalfaktors im Sinne Spearman's *g* (Spearman, 1904) an der Spitze von hierarchischen Intelligenzstrukturmodellen (vgl. Carroll, 1993; Deary, 2012; Jensen, 1998; Neisser et al., 1996). Weniger Konsens besteht hingegen hinsichtlich der Anzahl, der Inhalte und der Struktur spezifischer Intelligenzfaktoren unterhalb oder neben *g*. *Nested-Factor Modelle* (synonym *Bifactor Modelle* (vgl. Chen et al., 2006; Holzinger & Swineford, 1937; Reise, 2012) stellen neben *Higher-Order-* oder *Generalfaktormodellen* (ohne Gruppenfaktoren) eine empirisch und theoretisch gut interpretierbare alternative Modellvariante hierarchischer Intelligenzstrukturtheorien dar (vgl. Brunner et al., 2012; Gignac, 2008; Gustafsson & Balke, 1993; Reise, 2012), besonders dann, wenn *g* und spezifische Intelligenzfaktoren von Interesse sind (Chen et al., 2006). Bezüglich der Intelligenztestung mit verschiedenen Testbatterien fanden aktuelle Studien (sehr) hohe Korrelationen der testbatteriespezifischen *g*-Faktoren (Johnson et al., 2004, 2008) und belegen somit die Konsistenz der Messung der zugehörigen *g*-Faktoren. Jedoch wurde die systematische Betrachtung der (zusätzlichen) Konsistenz der spezifischen Intelligenzfaktoren vernachlässigt. Diese Aspekte zusammenbringend, ist die Absicht der vorliegenden Untersuchung dreierlei: (1) der Vergleich verschiedener Modellspezifikationen innerhalb der korrelativen Struktur kognitiver Fähigkeiten, (2) die Analyse der Konsistenz der *g*-Faktoren unterschiedlicher Testbatterien und (3) die Konsistenzanalyse verschiedener spezifischer Intelligenzfaktoren unterhalb/neben *g*.

Konsistenz von g

Die angewandte Psychologie ist maßgeblich darauf angewiesen, dass Intelligenz mit unterschiedlichen Testbatterien konsistent gemessen wird. Verschiedene Forschende näherten sich dieser Fragestellung in unterschiedlicher Herangehensweise und inspizierten dazu

beispielsweise die *g*-Ladungen unterschiedlicher Tests (vgl. Thorndike, 1987; Vernon, 1989) oder verglichen verschiedene faktorenanalytische Methoden, um *g* zu extrahieren (vgl. Jensen & Weng, 1994; Ree & Earles, 1991; s.a. Floyd, Shands, Rafael, Bergeron, & McGrew, 2009; Major, Johnson, & Bouchard, 2011). Zwei Studien befassten sich damit, inwieweit *g*-Faktorwerte von den spezifischen Aufgaben und Fähigkeiten, die sie erfassen, abhängen (Johnson et al., 2004, 2008). In konfirmatorischen Faktorenanalysen fanden sie (sehr) hohe Faktorkorrelationen ($.77 \leq r \leq 1.00$) der *g*-Faktoren unabhängig voneinander entwickelter Testbatterien. Die Ergebnisse der Johnson et al. Arbeiten (2004, 2008) liefern somit Evidenz zum einen für die Existenz eines allgemeinen Intelligenzfaktors und zum anderen dessen Konsistenz und Akkuratessse seiner Messung (vgl. Johnson et al., 2008, S. 91). Nichtsdestotrotz zeigten sich substantielle und systematische Korrelationen zwischen Subtests und first-order Gruppenfaktoren, die nicht auf den Einfluss des *g*-Faktors zurückzuführen waren.

Alternative Konzeptionen: Higher-Order, Nested-Factor und Generalfaktorenmodell

Während es an der psychometrischen Existenz, Generalität, Stabilität und Relevanz von *g* wenig Zweifel gibt (z.B. Deary, 2012; Gottfredson, 1997; Jensen, 1998; Jensen & Weng, 1994; Lubinski, 2004), herrscht nach wie vor Uneinigkeit hinsichtlich der Faktorstruktur unterhalb oder neben *g*. Im Rahmen von konfirmatorischen Faktorenanalysen werden verschiedene Modellspezifikationen diskutiert, die die Varianz-Kovarianz-Matrix hierarchisch strukturierter Fähigkeitskonstrukte am besten erklären. Vor allem das HO-Faktormodell (s. linke Seite der Abb. 1, S. 4) und das NF-Modell (s. rechte Seite der Abb. 1, S. 4) stellen neben dem Generalfaktormodell (GF-Modell), das ohne eine weitere Hierarchieebene auskommt, zwei unterschiedliche statistische Repräsentationen hierarchisch strukturierter Intelligenztheorien mit zusätzlichen Faktoren neben *g* und spezifischen Aufgaben oder Subtests (z.B. Carroll, 1993; McGrew, 2009) dar. In HO-Modellen – so die Annahme – werden ähnliche Subtests durch einen übergeordneten first-order Faktor (s.g. Gruppenfaktor) beeinflusst, während diese Gruppenfaktoren wiederum durch einen übergeordneten second-order *g*-Faktor an der Spitze beeinflusst werden. Im NF-Modell hingegen sind die domänenspezifischen Intelligenzfaktoren mit einem allgemeinen Faktor (*g*) genested, wobei *g* und die domänenspezifischen Intelligenzfaktoren sowohl untereinander als auch gegenseitig als unkorreliert angenommen werden.

Higher-Order versus Nested-Factor Modelle

Obwohl in ihrer Annahme von hierarchischer Intelligenzstruktur sehr verschieden, liegen die Unterschiede des HO- und NF-Modells aus statistischer Perspektive gar nicht so weit auseinander. Beide Modelle wären mathematisch äquivalent, wenn man im HO-Modell direkte Effekte von g auf jeden einzelnen (Sub)Test ergänzen würde (Yung et al., 1999). Da in einem „Standard“ HO-Modell diese Effekte üblicherweise auf Null fixiert sind, stellt das HO-Modell eine restringierte Version des NF-Modells dar. Das NF-Modell ist somit weniger restringiert und das HO-Modell ist sparsamer. Im statistischen Modellvergleich schneidet das NF-Modell häufig besser ab (vgl. Murray & Johnson, 2013; Brunner et al. 2012), jedoch wird ihm unterstellt, einen statistischen bias zugunsten des NF-Modells zu enthalten (vgl. Maydeu-Olivares & Coffman, 2006). Dem gegenüber werden die so genannten „*proportionality constraints*“, die das HO-Modell im Gegensatz zum NF-Modell enthalte, kritisch diskutiert (vgl. Brunner et al., 2012; Schmiedek & Li, 2004; Yung et al., 1999). Die statistische Überlegenheit des NF-Modells sollte folglich nicht allein ausschlaggebender Grund für die Wahl der Modelliervariante darstellen, vielmehr besteht der Vorteil aufgrund der Unkorreliertheit der Faktoren in der unabhängigen und distinkten Erfassung domänenspezifischer *und* allgemeiner Intelligenzfaktoren und deren entsprechender Korrelate (vgl. Chen et al., 2006). In Abwesenheit substanzieller empirischer Gründe für oder gegen eines dieser Modelle, sollte die Wahl vor allem in Abhängigkeit der Fragestellung und des Messmodells getroffen werden (vgl. Murray & Johnson, 2013). Verschiedene Studien belegen die mindestens akzeptable bis gute Passung einer NF-Struktur an die Daten gewonnen mit verschiedenen Wechsler Intelligenztestbatterien (vgl. Brunner et al., 2012; Gignac, 2005, 2006b). Für den deutschsprachigen Raum fanden Brunner & Süß (2005, 2007) eine angemessene Passung von NF-Modellen an die Daten des Berliner Intelligenzstruktur Tests, Form 4 (BIS-4; Jäger, Süß & Beauducel, 1997). Das Klassifikationsschema des Berliner Intelligenzstruktur Modells (BIS; Jäger, 1982; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; s. Abb. 4, S. 21) bietet zudem die Möglichkeit, Intelligenztestaufgaben neben den Operationsfacetten (Merkfähigkeit, Verarbeitungskapazität [ähnlich Reasoning], Einfallsreichtum und Bearbeitungsgeschwindigkeit) vor allem nach Inhaltsfacetten (Verbal, Numerisch und Figural) zu klassifizieren. Eine verbale Analogie-Aufgabe eines Intelligenztests ließe sich beispielsweise in das BIS-Feld einordnen, welches sich aus der Kreuzung der Inhaltsfacette „Verbal“ und Operationsfacette „Verarbeitungskapazität“ (Reasoning) ergibt.

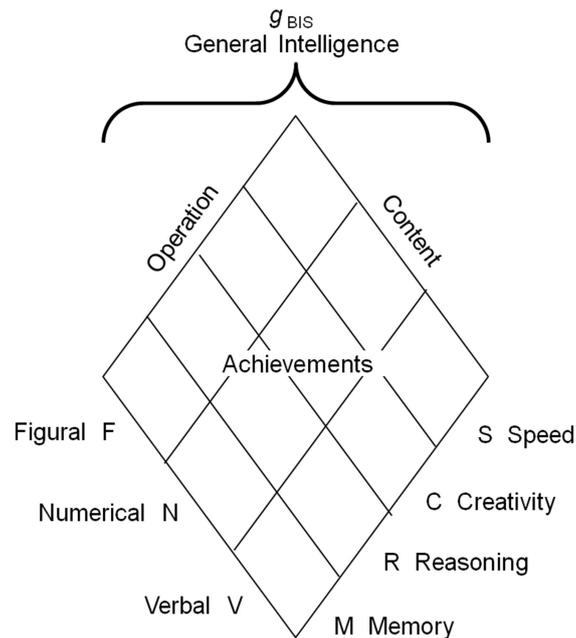


Abb. 4: Das BIS Model (Jäger, 1982)

Die vorliegende Studie

In den beiden Johnson-Arbeiten (Johnson et al., 2004, 2008) belegten die Autorin und Autoren zwar die Konsistenz von g , vernachlässigten jedoch die systematische Betrachtung der von ihnen gefundenen substanziellen, batterieübergreifenden Korrelationen spezifischer first-order Gruppenfaktoren. Die NF-Modellierung stellt zur Beantwortung dieser Fragestellung eine geeignete Methode dar, da die Konsistenz der Messung von g und spezifischen Intelligenzfaktoren simultan und distinkt untersucht werden kann. Die Ziele der vorliegenden Studie sind dreierlei: (1) vorangehenden Arbeiten folgend (z.B. Brunner et al., 2012; Gignac, 2005, 2006a,b; Gustafsson & Balke, 1993) verglichen wir im Rahmen konfirmatorischer Faktorenanalysen verschiedene Modellierungen (GF-Modell, HO-Modell, NF-Modell) hierarchisch strukturierter Intelligenzmodelle, wobei das GF-Modell als eine Art „Baseline Modell“ fungierte. (2) Konzeptionell orientiert an den Johnson et al. Arbeiten (2004, 2008) gingen wir der Frage nach konsistenter Messung von batteriespezifischer g -Faktoren anhand von aktuellen Daten nach. (3) Darüber hinaus untersuchten wir die Konsistenz spezifischer Intelligenzfaktoren von drei unabhängigen Testbatterien in einer NF-Modell-Struktur.

3.1.2 Methode

Instrumente und Ablauf

$N = 562$ Neuntklässler aus sechs rheinland-pfälzischen Gymnasien (50% weiblich, $M_{Alter} = 15.6$ Jahre, $SD = 0.45$, Teilnahmequote 91%) bearbeiteten 26 Intelligenzsubtests aus bekannten und häufig eingesetzten Intelligenztestbatterien (Berliner Intelligenzstruktur Test Form 4 [BIS-4], Grundintelligenztest Skala 2 [CFT 20-R; Weiß, 2006], Standard Progressive Matrices [SPM; Raven, 1941; Heller, Kratzmeier & Lengfelder, 1998] und Kognitiver Fähigkeitstest 4.-12. Klasse, Revision [KFT 4-12-R; Heller & Perleth, 2000]) zusammengefasst zu drei unterschiedlichen Testheften, welche in randomisierter Reihenfolge an jeweils drei Testtagen während der Unterrichtszeit in 90-Minutenblöcken administriert wurden. Eine genaue Beschreibung der eingesetzten (Sub-) Tests sowie deren Zuordnung zu einer spezifischen Inhaltsfacette (Verbal [V], Numerisch [N], Figural [F]) findet sich in Tabelle 1 auf Seite 24. Die 26 Subtests bildeten drei (neue) Testbatterien (BIS-Batterie bestehend aus 15 BIS-4 Subtests der Kurzform; CogAT[KFT]-Batterie bestehend aus sechs KFT-Subtests der Kurzform; Fluid-Batterie bestehend aus vier CFT-Subtests der Kurzform sowie dem vollen Set [A-E] der SPM).

Datenanalyse

Bezogen auf die erste Fragestellung führten wir konfirmatorische Faktorenanalysen (CFA) mithilfe des Softwareprogramms Mplus (Muthén & Muthén, 1998–2007) durch, mit der Absicht, verschiedene hierarchische Modellspezifikationen kognitiver Fähigkeiten zu evaluieren und statistisch zu vergleichen (s. Abb. 5, S. 23): a) ein HO-Modell mit testbatteriespezifischen first-order Gruppenfaktoren (V, N, F) und testbatteriespezifischen g-Faktoren (HO-Modell), b) ein NF-Modell mit testbatteriespezifischen g-Faktoren und testbatteriespezifischen Intelligenzfaktoren (V, N, F) (NF-Modell), c) ein GF-Modell mit einem allgemeinen g-Faktor für alle Subtests (GF g general) und d) ein GF-Modell mit testbatteriespezifischen g-Faktoren (GF g battery-specific). Die Zuordnung der Indikatoren (nichtstandardisierte Subtestwerte) zu den jeweiligen verbalen, numerischen und figuralen testbatteriespezifischen first-order Gruppenfaktoren (im HO-Modell) respektive testbatteriespezifischen Intelligenzfaktoren (im NF-Modell) erfolgte anhand der Vorgaben der jeweiligen Testmanuale.

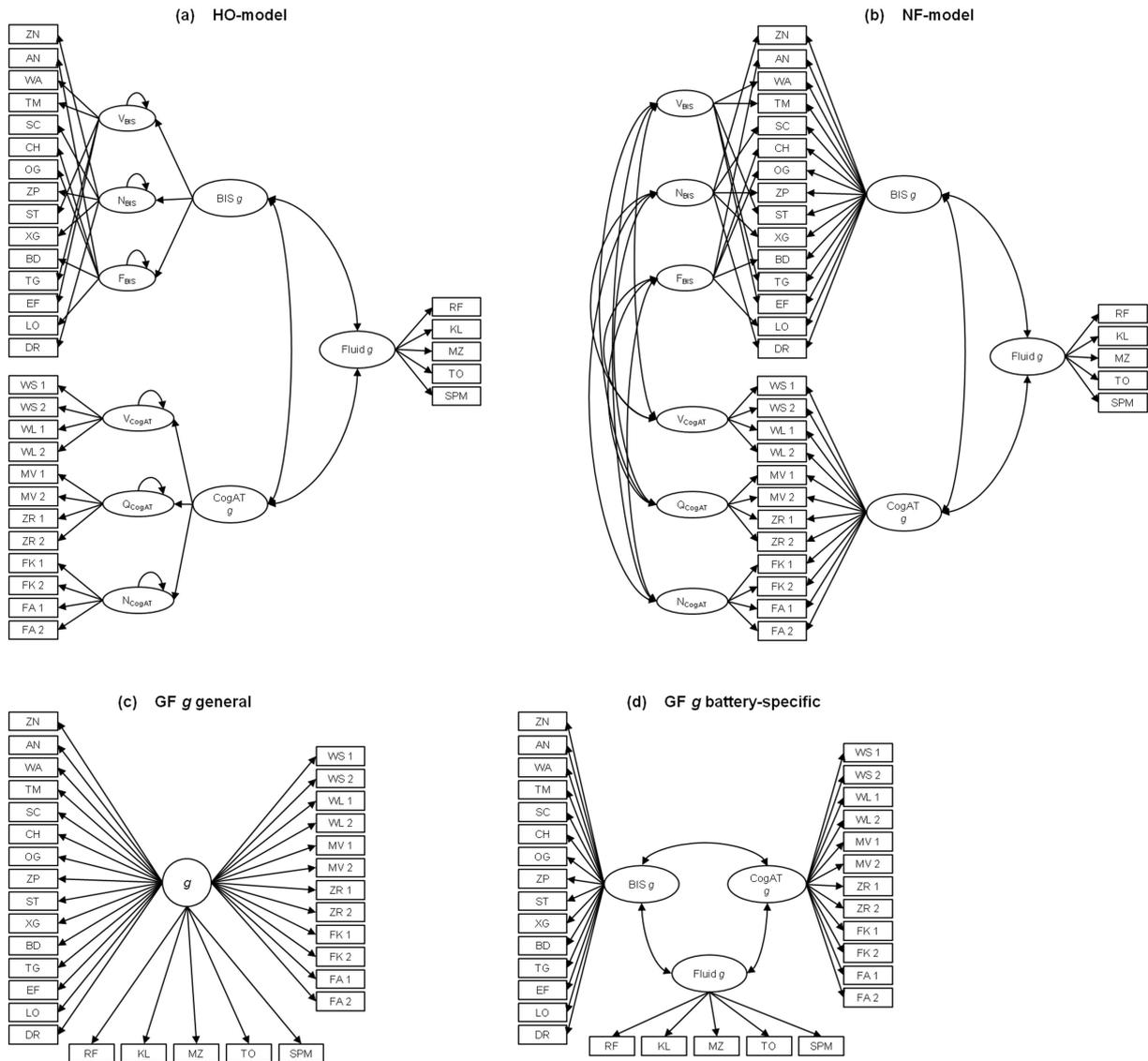


Abb. 5: Konfirmatorische Faktorenanalysen für alternative Modelle allgemeiner und spezifischer Intelligenzfaktoren in drei Testbatterien.

g = Generalfaktor, BIS = Berliner Intelligenzstruktur Test, Form 4, CogAT[KFT] = Kognitiver Fähigkeitstest für 4-12 Klasse, Fluid = Grundintelligenztestskala 2 mit Raven's Standard Progressive Matrices. VBIS = Verbal Inhaltsfaktor sensu Jägers BIS-Model (1982), NBIS = Numerischer Inhaltsfaktor, FBIS = Räumlich/figuraler Inhaltsfaktor. VCogAT = Verbales Reasoning, QCogAT = Quantitative (numerisches) Reasoning, NCogAT = Nonverbal (figural) Reasoning. Testbeschreibung der manifesten Variablen (Subtests) s. Tabelle 1, S. 24. Der besseren Lesbarkeit wegen wurden die Fehlerterme (alle unkorreliert) nicht dargestellt.

Tabelle 1

Tests der drei Testbatterien ergänzt um Testbeschreibungen, Zuordnung zur jeweiligen Inhaltsfacette, Testzeiten, Anzahl der Items, Mittelwerte (M), Standardabweichungen (SD) und Intraclass-Korrelationen (ICC).

Test (Abkürzung.)	Beschreibung	Inhaltsfaktor	Testzeit	Itemanzahl	M	SD	ICC
<i>Berliner Intelligenzstruktur Test, Form 4 (BIS-4) Kurzversion</i>							
1.	Orientierungs-Gedächtnis (OG)	Wiedererkennen von Gebäuden auf einem Stadtplan	F	1:30 + 1:40	27	15.78	4.24 .06
2.	Zahlenreihen (ZN)	Vervollständigen von Zahlenreihen	N	3:40	9	5.25	2.17 .10
3.	Eigenschaften-Fähigkeiten (EF)	Eigenschaften einer bestimmten Berufsgruppe generieren	V	2:30	f. A..	10.14	3.71 .07
4.	Analogien (AN)	Vervollständigen einer Reihe von figürlichen Mustern	F	1:45	8	3.97	1.63 .16
5.	X-Größer (XG)	Anstreichen von Zahlen, die um X größer sind als ihr Vorgänger	N	1:00	44	22.03	8.11 .06
6.	Wortanalogien (WA)	Identifizieren analoger Wortpaare	V	1:30	8	3.30	1.67 .07
7.	Layout (LO)	Zeichnen grafischer Label	F	3:00	f. A.	4.27	1.48 .09
8.	Zahlen-Paare (ZP)	Merken und wiedergeben von Zahlenpaaren	N	2:00 + 2:00	12	7.07	2.54 .08
9.	Tatsache-Meinung (TM)	Entscheidung über Tatsache oder Meinung sprachlicher Aussagen	V	1:00	16	9.30	3.17 .07
10.	Buchstaben-Durchstreichen (BD)	Ausstreichen von Buchstaben in einer Anordnung Verschiedener	F	0:50	130	55.36	10.88 .09
11.	Schätzen (SC)	Schätzen komplexer Rechnungen	N	2:45	7	4.15	1.70 .07
12.	Sinnvoller Text (ST)	Wiedergabe von Informationen aus einem Text	V	1:00 + 2:00	22	10.60	3.51 .05
13.	Divergentes Rechnen (DR)	Generieren arithmetischer Gleichungen mit vorgegebenen Elementen	N	1:50	f. A.	11.40	4.22 .05
14.	Charkow (CH)	Vervollständigen und Generieren von Figuren in einer Folge	F	3:00	6	2.49	1.44 .06
15.	Teil-Ganzes (TG)	Ausstreichen von Wörtern in einer Folge	V	0:40	22	14.15	3.25 .05
<i>Fluid Batterie mit Grundintelligenztestskala 2 (CFT 20-R), Kurzversion und Raven Standard Progressive Matrices (SPM)</i>							
16.	Reihenfortsetzen (RF)	Identifikation des nächsten figürlichen Elements einer Reihe	F	4:00	15	12.79	1.53 .08
17.	Klassifikation (KL)	Identifikation des Elements, das in einer Reihe nicht dazugehört	F	4:00	15	10.59	2.17 .03
18.	Matrizen (MZ)	Vervollständigen einer Matrix sich verändernder Muster	F	3:00	15	11.86	1.93 .06
19.	Topologien (TO)	Identifizieren des topologisch äquivalenten, grafischen Musters	F	3:00	11	7.05	2.01 .04
20.	SPM Set A – E	Vervollständigen einer Matrix sich verändernder Muster	F	45:00	60	50.98	4.29 .09
<i>Kognitiver Fähigkeits-Test für 4.-12- Klassen, Revision (KFT 4-12 R), Kurzversion</i>							
21.	Wortschatz (WS)	Identifikation eines Wortes mit gleicher oder ähnlicher Bedeutung	V	7:00	25	17.53	3.01 .06
22.	Wortanalogien (WL)	Identifikation analoger Wortpaare	V	7:00	20	12.07	2.77 .06
23.	Mengenvergleiche (MV)	Vergleich von größer/kleiner Relationen numerischer Elemente	N	10:00	25	16.98	3.51 .11
24.	Zahlenreihen (ZR)	Vervollständigen von Zahlenreihen	N	9:00	20	17.51	2.51 .04
25.	Figurenklassifikation (FK)	Identifikation der passenden Figur	F	9:00	25	21.79	2.66 .07
26.	Figurenanalogien (FA)	Identifikation analoger Figurenpaare	F	8:00	25	20.62	3.20 .03

Anmerkung. f. A. = freies Antwortformat: Die Teilnehmenden generieren freie Antworten, sodass eine Itemanzahlbestimmung nicht möglich ist.

Um lokale Unterparametrisierung und damit korrespondierende Identifikationseinschränkungen für die jeweils zwei Indikatoren der inhaltsspezifischen latenten Faktoren (V, N, F) der CogAT-Batterie zu vermeiden, bildeten wir für jeden CogAT-Subtest jeweils zwei *odd-even Parcel*³. Um zu berücksichtigen, dass diese jeweils zwei Parcel vom gleichen Subtest stammen, setzten wir deren nichtstandardisierten Ladungen auf den korrespondierenden latenten Faktoren gleich. Jede Faktorvarianz in jedem Modell wurde auf 1 gesetzt. Zur Berücksichtigung der hierarchischen Struktur der Daten (Klassencluster) nutzten wir nach Inspektion der Intraclass-Korrelationen die programminterne Methode „*type complex*“, die Mplus (Muthén & Muthén, 1998-2007) zur Verfügung stellt. Dem Umstand fehlender Datenwerte (Mean = 2.5%; Median = 2.5%; Min = 2.5%; Max = 6.2%) begegneten wir durch die Verwendung des Full Information Maximum Likelihood (FIML) Algorithmus. Zur Parameterschätzung verwendeten wir den robusten Maximum-Likelihood Algorithmus (MLR). Zusätzlich zum χ^2 -test orientierten wir uns zur Modellevaluation an den Richtlinien von Hu & Bentler (1998, 1999) mit *cut-off* Werten nahe bei .08 für den Standardized Root Mean Square of Residuals (SRMR; Bentler, 1995) zusammen mit dem Root Mean Square Error of Approximation (RMSEA; Steiger, 1990) mit Werten nahe .06. Darüber hinaus betrachteten wir die standardisierten Faktorladungen mit Werten größer oder gleich .30 als substantiell (vgl. Carroll, 1993; McDonald, 1999). Zusätzlich verglichen wir die verschiedenen Modelle anhand von Schwarz' (1978) Bayesian Information Criterion (BIC) und Akaike's (1974) Information Criterion (AIC) dahingehend, Veränderungen der Fit-Werte zu vergleichen und Modelle mit niedrigeren Werten zu präferieren. Obwohl eher eine Ad-hoc-Richtlinie zur Bewertung der Modellfit-Unterschiede und daher mit Vorsicht zu handhaben, schlug Chen (2007) vor, dass das komplexere Modell gewählt werden sollte, wenn die RMSEA-Differenz (Δ RMSEA) .015 überschreitet.

Bezogen auf die zweite Fragestellung zur Konsistenzanalyse der g-Faktoren von unterschiedlichen Testbatterien, betrachteten wir die batterieübergreifenden g-Faktor-Korrelationen für die verschiedenen Modelle – einen mindestens akzeptablen Modellfit vorausgesetzt.

Hinsichtlich der dritten Fragestellung zur Konsistenz spezifischer Intelligenzfaktoren neben g inspizierten wir die batterieübergreifenden Korrelationen der inhaltsspezifischen Intelligenzfaktoren des NF-Modells. Dabei stützten wir uns auf Kriterien für eine konvergente

³ Ausführlicher vgl. Valerius & Sparfeldt (2014, S. 127)

und diskriminante Validierung, die im Rahmen von Multitrait-Multimethod Analysen eingeführt wurden (Campbell & Fiske, 1959). Aufgrund des Fehlens allgemein anerkannter wissenschaftlicher Kriterien hinsichtlich der absoluten Größe der entsprechenden Korrelationskoeffizienten orientierten wir uns an folgenden Kriterien: Zur Bewertung der konvergenten Validität, mussten die batterieübergreifenden, inhaltspezifischen Faktorkorrelationen bedeutsam von Null verschieden und von „großer“ Effektgröße sein (d.h. $r \geq .50$; vgl. Cohen, 1988). Diskriminante Validität sahen wir dann als gegeben, wenn die konvergenten Korrelationen die verbleibenden (diskriminanten) batterieübergreifenden Korrelationen der nicht-korrespondierenden spezifischen Inhaltsfaktoren numerisch überragten. Die modellbasierte Reliabilität der spezifischen Inhaltsfaktoren berechneten wir mittels Omega hierarchisch (ω_h) (vgl. Brunner et al., 2012).

3.1.3 Ergebnisse

Modellevaluation

Wie Tabelle 2 zu entnehmen, waren die Modellfitstatistiken für alle vier Modelle mindestens akzeptabel bis gut. Mit wenigen Ausnahmen waren die meisten standardisierten Faktorladungen in allen vier Modellen statistisch signifikant ($p < .05$). Die Faktorladungen waren in den meisten Fällen – ebenfalls mit wenigen Ausnahmen – für mindestens einen Faktor auf dem sie luden mit Werten $\geq .30$ substantiell (s. Tabelle 3, S. 28). Das NF-Modell zeigte die niedrigsten AIC- und BIC-Werte, gefolgt vom HO-Modell; beide GF-Modelle zeigten deutlich höhere BIC-Werte.

Tabelle 2

Fitindices für die alternativen Modelle: Higher-Order-Modell (HO), Nested-Faktor-Modell (NF) und beide Generalfaktormodelle (GF: g general und g battery-specific, s. Abb. 5, S. 23)

Model	χ^2	df	p	SRMR	RMSEA	BIC	AIC
HO	1203.807	461	<.001	.069	.054	76057.075	75628.257
NF	899.468	431	<.001	.055	.044	75898.279	75339.515
GF (g general)	1530.051	464	<.001	.071	.064	76423.261	76007.437
GF (g battery-specific)	1471.523	461	<.001	.069	.062	76370.952	75942.134

Anmerkung. SRMR = Standard Root Mean Square of Residuals, RMSEA = Root Mean Square Error of Approximation, BIC = Bayesian Information Criterion, AIC = Akaike Information Criterion.

In Bezug auf Δ RMSEA zeigte das NF-Modell einen niedrigeren Wert als das HO-Modell (Δ RMSEA = .010) und beide GF-Modelle (Δ RMSEA = .018 und .020). Das HO-Modell unterschied sich von den anderen Modellen hinsichtlich Δ RMSEA nicht um mehr als .015. Die SRMR, AIC- und BIC-Werte deuteten auf eine (numerisch) bessere Passung des NF-Modells gegenüber den anderen Modellen hin. Δ RMSEA – als Daumenregel und damit mit Vorsicht zu betrachten – deuteten auf eine bessere Passung des NF-Modells im Vergleich zu beiden GF-Modellen (als „Baseline Modelle“ ohne zusätzliches Hierarchielevel inhaltsspezifischer Intelligenzfaktoren zu betrachten) hin. Das HO-Modell hingegen wies hinsichtlich Δ RMSEA auf keinen deutlich besseren Fit sowohl gegenüber dem NF- als auch den beiden GF-Modellen hin.

Tabelle 3

Standardisierte Faktorladungen im kombinierten Nested-Faktor-Modell (NF-Modell)

Test	BIS-4				CogAT				Fluid
	λ_{VBIS}	λ_{NBIS}	λ_{FBIS}	λ_{BIS_g}	λ_{VCogAT}	λ_{QCogAT}	λ_{NCogAT}	λ_{CogAT_g}	λ_{Fluid_g}
TM	.45*			.25*					
EF	.25*			.08					
WA	.28*			.36*					
TG	.45*			.14					
ST	.37*			.18*					
ZN		.34*		.45*					
XG		.51*		.27*					
ZP		.34*		.14*					
SC		.26*		.42*					
DR		.22*		.17*					
OG			.31*	.35*					
AN			.06	.55*					
LO			.32*	.09					
BD			.39*	.14*					
CH			-.14*	.59*					
WS 1					.51 ^{a*}			.31*	
WS 2					.55 ^{a*}			.19*	
WA 1					.43 ^{b*}			.42*	
WA 2					.42 ^{b*}			.39*	
MV 1						.06 ^c		.48*	
MV 2						.06 ^c		.56*	
ZR 1						.66 ^{d*}		.47*	
ZR 2						.63 ^{d*}		.40*	
FK 1							.49 ^{e*}	.51*	
FK 2							.53 ^{e*}	.53*	
FA 1							.23 ^f	.67*	
FA 2							.26 ^f	.58*	
RF									.52*
KL									.44*
MZ									.41*
TO									.42*
SPM									.62*
ω_h	.39	.29	.12	.67	.46	.25	.24	.85	.60

Anmerkung. Bei den Angaben mit hochgestellten Buchstaben (a-f) wurden die unstandardisierten Faktorladungen (mit jeweils gleichem Buchstaben) gleichgesetzt. Die modellbasierten Reliabilitätsschätzer Omega hierarchisch (ω_h) für die inhaltspezifischen Intelligenzfaktoren sind in der untersten Reihe abgetragen.

* $p < .05$.

Korrelationen der *g*-Faktoren

Die *g*-Faktoren der drei verschiedenen Testbatterien innerhalb des NF-Rahmens ergaben sehr hohe, statistisch bedeutsame ($p < .05$) Koeffizienten ($.91 \leq r \leq 1.00$; s. Tabelle 4). Die modellbasierten Reliabilitätsschätzungen ω_h der allgemeinen Faktoren erreichten .67, .85 und .60 für BIS, CogAT und Fluid. Die *g*-Faktor-Korrelationen waren von vergleichbarer Größe innerhalb des HO-Rahmens ($.92 \leq r \leq .99$). Doch auch für das GF-Modell erreichten dessen testbatteriespezifischen *g*-Faktoren batterieübergreifende Korrelationen ($.80 \leq r \leq .91$).

Tabelle 4

Batterieübergreifende Korrelationen der *g*-Faktoren in den unterschiedlichen Modellen: Higher-Order-Modell (HO), Nested-Faktor-Modell (NF) und Generalfaktormodell mit testbatteriespezifischen *g*-Faktoren (GF)

Modell	CogAT <i>g</i>			Fluid <i>g</i>		
	HO	NF	GF	HO	NF	GF
Bis <i>g</i>	.92*	.91*	.80*	.99*	1.00*	.91*
CogAT <i>g</i>				.95*	.92*	.88*

* $p < .05$.

Korrelationen der inhaltspezifischen Intelligenzfaktoren

Die Konsistenz der *g*-unabhängigen inhaltspezifischen Intelligenzfaktoren (V, N, F) betrachteten wir im Rahmen des NF-Modells. Das Korrelationsmuster (s. Tabelle 5, S. 30) ergab konvergente statistisch bedeutsame ($p < .05$) Korrelationen für die verbalen ($r = .83$) und numerischen ($r = .46$) Inhaltsfaktoren. Die konvergenten Korrelationen waren (in absoluter Höhe) jeweils numerisch größer als deren korrespondierenden diskriminanten Korrelationskoeffizienten und belegten somit konvergent-divergente Validität für die verbalen und numerischen Inhaltsfaktoren. Bezüglich der figuralen Inhaltsfaktoren fielen sowohl die konvergenten Korrelationskoeffizienten numerisch etwas niedriger ($r = .22$; n.s.) als auch einige divergente Korrelationskoeffizienten in absoluter Höhe numerisch etwas höher aus, was auf einen Mangel an konvergent-divergenter Validität für die figuralen Inhaltsfaktoren hindeutet. Die modellbasierten Reliabilitätsschätzer ω_h der inhaltspezifischen Faktoren reichten von .12 (für Figural der BIS-Batterie) bis .46 (für Verbal der CogAT-Batterie) (s. unterste Zeile von Tabelle 3, S. 28).

Tabelle 5

Batterieübergreifende Korrelationen der inhaltspezifischen Intelligenzfaktoren innerhalb des kombinierten NF-Modells

Faktor	V_{CogAT} <i>verbal</i>	Q_{CogAT} <i>numerisch</i>	N_{CogAT} <i>figural</i>
V_{BIS}	.83*	.01	.03
N_{BIS}	-.41*	.46*	-.25
F_{BIS}	-.47*	.00	.22

* $p < .05$.

3.1.4 Diskussion

Die Hauptbefunde der vorliegenden Studie sind dreierlei: anhand aktueller Daten einer größeren Stichprobe gymnasialer Neuntklässler, die insgesamt 26 Tests aus unterschiedlichen unabhängigen Testbatterien bearbeiteten, zeigte sich (1) der Modellvergleich verschiedener hierarchisch strukturierter Intelligenzmodelle ergab in CFA-Analysen a) mindestens akzeptable absolute Fitstatistiken für die vier analysierten Modelle (HO-Modell, NF-Modell und die zwei GF-Modelle), und b) eine numerisch bessere Anpassung an die Daten im NF-Modell. (2) Die Korrelationen der drei testbatteriespezifischen g -Faktoren waren (in allen Modellen, die testbatteriespezifische g -Faktoren spezifizierten) sehr hoch, was dafür spricht, dass die testbatteriespezifischen g -Faktoren (nahezu) austauschbar waren. (3) Darüber hinaus belegten die Korrelationen der g -unabhängigen inhaltspezifischen Faktoren im NF-Modell konvergent-divergente Validität für die verbalen und numerischen spezifischen Intelligenzfaktoren.

Bezüglich der ersten Fragenstellung, die sich mit der spezifischen Struktur der Intelligenz befasst, wurden verschiedene theoretisch abgeleitete hierarchische Intelligenzmodelle verglichen. Entsprechend der Mehrzahl theoretischer und empirischer Annahmen (erstmalig von Spearman, 1904 erwähnt und von zahlreichen Forschern fortgesetzt z.B. Carroll, 1993; Deary, 2012; Jensen, 1998; McGrew, 2009) konvergierten alle unsere CFA-Modelle in der Vorstellung eines Generalfaktors an der Modellspitze. In Replikation früherer Befunde (z.B. Brunner et al., 2012) mit Daten, die auf einer deutlich unterschiedlichen Stichprobe von Schülerinnen und Schülern und konzeptionell unterschiedlichen Intelligenztests beruht, ergaben unsere Analysen zumindest akzeptable absolute Fitstatistiken für alle vier Modellspezifikationen. Die weitverbreitete Annahme eines allgemeinen Faktors, der kognitiven Fähigkeiten zugrunde liegt, konnte somit abermals gestützt werden. Darüber

hinaus verbesserte sich der Modellfit sukzessive von einer eher allgemeinen (GF-Modell) zu einer eher differenzierten Intelligenzstruktur (HO- und NF-Modell) mit zusätzlich spezifischen (verbalen, numerischen und figuralen) Intelligenzfaktoren neben g . Der sukzessiv steigende Modellfit von GF- über HO- hin zu NF-Modellen, steht in Einklang mit früheren Analyseergebnissen der Wechsler Skalen für Kinder und Erwachsene (vgl. Brunner et al., 2012, Gignac, 2005, 2006b) und stärkt zudem die mehrheitlich anerkannte Vorstellung, dass sowohl allgemeine als auch spezifische Intelligenzfaktoren an der Lösung von kognitiven Denkaufgaben beteiligt sind (vgl. Carroll, 1993; McGrew, 2009; Spearman, 1904). Dennoch gilt es zu berücksichtigen, dass die statistische Überlegenheit des NF-Modells aufgrund nicht zu leugnender Einwände (z.B. Maydeu-Olivares & Coffman, 2006) nicht als einziges Kriterium bei der Wahl des Analysemodells herangezogen werden sollte. Vielmehr sollte die Modellwahl maßgeblich von der Forschungsfrage und dem Messmodell abhängen (vgl. Murray & Johnson, 2013). Neben überzeugende theoretische und empirische Argumente waren für die vorliegende Studie (Fragestellung 2 und 3) vor allem die Orthogonalität der Faktoren im NF-Modell wesentlich und ausschlaggebend für die Modellwahl, da alle Intelligenzfaktoren im NF-Modell unabhängig und distinkt voneinander batterieübergreifend in Beziehung gestellt werden konnten. Die gelegentlich aufkommende Frage innerhalb der psychometrischen Intelligenzforschung, ob das NF-Modell (v.a. aufgrund der Orthogonalität der Faktoren) das Wesen der Intelligenz tatsächlich angemessen abbilden könne, war zwar nicht Gegenstand der vorliegenden Untersuchung, sollte jedoch nicht unerwähnt bleiben. Eine Beantwortung dieser Frage ist nach wie vor nicht abschließend geklärt.

Hinsichtlich der zweiten Forschungsfrage, die sich mit der Konsistenz der g -Faktoren verschiedener Testbatterien befasst, korrelierten die testbatteriespezifischen g -Faktoren erwartungsgemäß in allen Modellen (NF-, HO-, GF-Modell) (sehr) hoch. Dies ist im Einklang mit früheren Forschungsbefunden und bekräftigt: (a) die (weitestgehende) Testunabhängigkeit und somit Konsistenz von g (Johnson et al., 2004, 2008) und (b) in großem Maße dessen Invarianz gegenüber der faktorenanalytischen Methode (z.B. Jensen & Weng, 1994). Eine besondere Beachtung verdienen die ebenfalls (sehr) hohen Korrelationen der g -Faktoren der Fluid-Batterie mit sowohl den BIS-Batterie- als auch CogAT [KFT]-Batterie-spezifischen g -Faktoren ($.88 \leq r \leq 1.00$). Im Gegensatz zu Johnson et al. (2008), die ihre niedrigste g -Faktor-Korrelation der CCFT-Batterie mit der GATB-Batterie ($r = .77$) mit einer zu geringen Aufgabenvielfalt der CCFT-Batterie erklärten, enthielt unsere Fluid-Batterie zusätzlich zu den vier CFT-Subtests das reliable Komplet-Set der SPM (Raven, 1941), welche als hoch g -

geladen bekannt sind (vgl. Jensen, 1998) und die g -Faktor-Korrelationen gestärkt haben könnten. Eine weitere Erklärung für die (sehr) hohen g -Faktor-Korrelationen der Fluid-Batterie zu den anderen Batterien könnte sein, dass wir unser g sehr „figural“ operationalisierten – worauf die hohen Ladungen der figuralen Subtests auf den HO- g -Faktoren ($\lambda > .84$) hindeuten. Dies ist vor allem für die nächste Fragestellung von Bedeutung.

Im Rahmen der dritten Fragestellung untersuchten wir die konvergent-divergente Korrelationsstruktur der inhaltspezifischen Intelligenzfaktoren über die Korrelationen der batteriespezifischen g -Faktoren hinaus und fanden vor allem für die spezifisch verbalen ($r = .83$) und numerischen ($r = .46$) Intelligenzfaktoren im NF-Modell batterieübergreifend konvergente und divergente Beziehungsmuster. Die konvergente Korrelation der spezifisch figuralen Intelligenzfaktoren ($r = .22$) war statistisch nicht bedeutsam und zudem in absoluter Höhe nicht diskriminativ zu den übrigen, nicht korrespondierenden, spezifischen Intelligenzfaktoren. Wie schon erwähnt, könnte die figurale Färbung unserer g -Faktoren die vergleichsweise niedrigen batterieübergreifenden Korrelationen der spezifisch figuralen NF-Intelligenzfaktoren erklären, da ein größerer Teil reliabler figuraler Varianz bereits in den figural getönten g -Faktoren gebunden war.

Wie bereits in Vorgängeruntersuchungen (z.B. Brunner et al., 2012; Reise, 2012) fielen die modellbasierten Reliabilitätsschätzungen der inhaltspezifischen Faktoren formelbedingt relativ niedrig aus⁴. Die Interpretation der Subskalen als präzise Indikatoren eigenständiger Konstrukte scheint somit auf den ersten Blick limitiert (Reise, 2012, S. 691). Jedoch sollte zusätzlich nicht außer Acht gelassen werden, dass die Reliabilität (neben Weiteren) bloß *einen* Aspekt zur Qualitätsbeurteilung psychometrischer Maße darstellt. Entsprechend können die konvergenten Korrelationen der spezifisch verbalen und numerischen Intelligenzfaktoren als Untergrenze für die Reliabilität dieser Faktoren betrachtet werden.

Die Befunde sind vor dem Hintergrund einer leicht varianzeingeschränkten Stichprobe (gymnasiale Neuntklässler) zu betrachten. Mit dem Ziel Konfundierungen mit potenziellen Störvariablen zu vermeiden, legten wir die Untersuchungsanlage so an. Obwohl Varianzeinschränkungen die Korrelationshöhe beeinflussen können, erzielten wir Korrelationen, die aufgrund früherer Studienbefunde zu erwarten waren. Die Teilnahme an der Studie hing wesentlich von Entscheidungen der Schulleitungen, Lehrkräfte sowie

⁴ Ausführlicher dazu vgl. Valerius & Sparfeldt (2014, S. 127ff und S. 131)

Erziehungsberechtigten der teilnehmenden Jugendlichen ab und war insofern nicht vollständig randomisiert. Die hohe Teilnahmequote schützt die Befunde jedoch vor substanziellem Bias im Sinne von Selbstselektion innerhalb der Zielgruppe. Vor allem untersuchungspraktische Gründe limitierten die Anzahl der eingesetzten Tests, was unter anderem die Aufsplittung der KFT-Subtests erforderlich machte. Ein Ziel weiterführender Forschung könnte somit sein, die Befunde in internationalem Rahmen mit heterogenen Stichproben und einer größeren Aufgabenvielfalt zu stabilisieren.

Zusammengefasst sind die Hauptbefunde dieser Studie: (1) im Rahmen von CFA-Analysen hierarchisch strukturierter Intelligenzmodelle stellt das NF-Modell eine nützliche und fruchtbare Alternative zu GF- und HO-Modellen dar, besonders dann, wenn sowohl der Generalfaktor als auch spezifische Intelligenzfaktoren von Interesse sind (konzeptionell repliziert von z.B. Brunner et al., 2012; Gignac, 2005, 2006b; Gustafsson & Balke, 1993). (2) Die Generalfaktoren korrelierten batterieübergreifend sehr hoch und belegen somit dessen Konsistenz (weitestgehend) unabhängig vom Messinstrument (konzeptionell repliziert von Johnson et al., 2004, 2008). (3) Zusätzlich zu und über frühere Ergebnisse hinaus, zeigte sich die Konsistenz der Messung testbatteriespezifischer verbaler und numerischer Intelligenzfaktoren über unabhängige Testbatterien.

3.2 Ergänzende Analysen zu Valerius & Sparfeldt (2014)

Wie in Abschnitt 2.1.1 (S. 4 ff. dieser Arbeit) ausführlicher beschrieben, kann es trotz guter Fitstatistik unter bestimmten Bedingungen (z.B. gleichgesetzte Ladungen, Nichtidentifizierbarkeit durch zugefügte manifeste Variablen) in einem um Kriteriumsvariablen erweiterten NF-Modell zu Modellmissfit kommen (z.B. Eid et al., 2018, Koch, Holtmann, Bohn & Eid, 2018). Als Alternative schlagen Eid et al. (2018) vor, Bifactor (S-1) Modelle (s. Abb. 2, S. 6) zu spezifizieren, welche im Vergleich zum klassischen NF-Modell weniger anfällig für Schätzprobleme seien. Der aktuelle Artikel von Eid und Mitarbeitenden (2018) wurde veröffentlicht, nachdem der eben vorgestellte *Intelligence*-Artikel erschienen war und obwohl sich die Bedingungen des kombinierten NF-Modells der ersten Studie (Valerius & Sparfeldt, 2014) von denen Eid et al.'s (2018) unterschieden, überprüfte ich nachträglich die Stabilität unserer Befunde indem ich für die zwei Testbatterien (CogAT, BIS) Bifactor (S-1) Modelle im kombinierten NF-Modell spezifizierte (s. Abb. 6, S. 35). Zwar setzten wir im Originalmodell aus Valerius & Sparfeldt (2014) ähnlich wie Eid et al. (2018) die unstandardisierten Faktorladungen der KFT-Subtest-Parcel auf den zugehörigen spezifischen Intelligenzfaktoren gleich (der üblichen Praxis bei geparcelten Subtesthälften [*odd-even-split*] entsprechend), jedoch enthielt unser kombiniertes NF-Modell im Unterschied zu Eid et al. (2018) keine manifesten Kriteriumsvariablen – was mitunter zu Schätzproblemen bei um Kriterien erweiterten NF-Modellen führen kann (vgl. Koch et al., 2018). Stattdessen spezifizierten wir latente (spezifische und allgemeine) Intelligenzfaktoren, die wir batterieübergreifend miteinander in Beziehung stellten. Da wir die Varianzen aller Faktoren auf 1 setzten, trat zudem kein Problem von Nichtidentifizierbarkeit (wie bei Eid et al., 2018) auf.

In den Zusatzanalysen wählte ich für das erweiterte Bifactor (S-1) Modell – im Folgenden *kombiniertes Bifactor (S-1) Modell* (kombiniertes BF [S-1]) genannt – als Referenzfacette für die batteriespezifischen *g*-Faktoren (BIS-*g* und CogAT-*g*) diejenigen Subtests aus, die im Original NF-Modell (vgl. Valerius & Sparfeldt, 2014 oder Abb. 5, S. 23 dieser Arbeit) die spezifisch figuralen Intelligenzfaktoren pro Testbatterie bildeten und zwar aus folgenden Gründen: a) nicht alle Subtests, die modellkonform (BIS-Modell; Jäger, 1982) für den figuralen BIS-Inhaltsfaktor spezifiziert wurden, luden darauf statistisch bedeutsam positiv ($p < .05$; λ_{FBIS} für AN = .06 [n.s.]; BD = .39; CH = -.14; LO = .32; OG = .31; vgl. Tabelle 3, S. 28), was auf Schätzprobleme des Modells hindeuten könnte; b) die spezifisch figuralen

Intelligenzfaktoren wiesen im kombinierten NF-Modell keine signifikanten Zusammenhänge (mit einer Ausnahme BIS-figural mit CogAT-verbal $r = -.47, p < .05$; vgl. Tabelle 5, S. 30) zu den batterieübergreifenden spezifischen Faktoren auf (CogAT-figural mit: BIS-verbal $r = .03$, BIS-numerisch $r = -.25$ und BIS-figural $r = .22$; vgl. Tabelle 5, S. 30); und c) vor allem die „reasoninglastigen“ figuralen Subtests im CogAT (im Gegensatz dazu enthielten die figuralen Subtests im BIS-4 neben Verarbeitungskapazität [ähnlich Reasoning] auch weitere

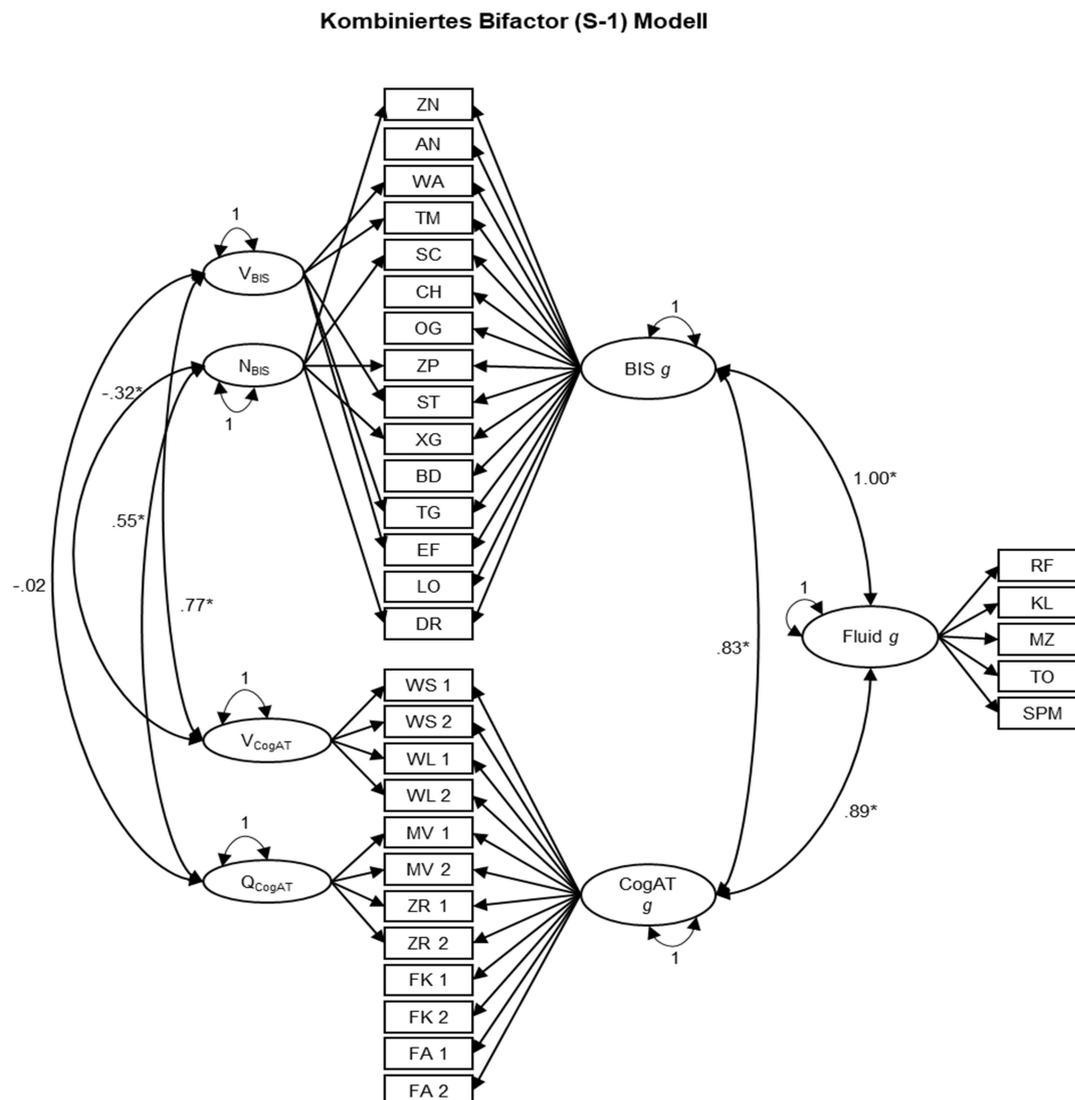


Abb. 6: Kombiniertes Bifactor (S-1) Modell mit hier fehlendem figuralem Inhaltsfaktor (vgl. Abb. 5, S. 23) als Referenzfacette für g .

g = Generalfaktor, BIS = Berliner Intelligenzstruktur Test, Form 4, CogAT [KFT] = Kognitiver Fähigkeitstest für 4-12 Klasse, Fluid = Grundintelligenztestskala 2 mit Raven's Standard Progressive Matrices. VBIS = Verbal Inhaltsfaktor sensu Jägers BIS-Modell (1982), NBIS = Numerischer Inhaltsfaktor, VCogAT = Verbales Reasoning, QCogAT = Quantitative (numerisches) Reasoning. Testbeschreibung der manifesten Variablen (Subtests) s. Tabelle 1 bei Valerius & Sparfeldt (2014; S. 125). Der besseren Lesbarkeit wegen wurden die Fehlerterme (alle unkorreliert) nicht dargestellt.

Operationsfacetten⁵) können mit bekanntlich hohen *g*-Ladungen als guter Indikator für *g* betrachtet werden (z.B. Jensen, 1998).

Zwar ist zu erwarten, dass das Befundmuster stabil bleibt, dennoch sollen die im Folgenden berichteten Zusatzanalysen, die Ergebnisse der ersten Studie hinsichtlich potenziell unentdeckten Modellmissfits prüfen.

3.2.1 Ergebnisse der Zusatzanalysen

Modellevaluation

Das zusätzlich spezifizierte kombinierte Bifactor (S-1) Modell fittete ähnlich gut wie das kombinierte Original NF-Modell aus Valerius & Sparfeldt (2014) (vgl. Tabelle 6, S. 36). Die statistischen Kennwerte SRMR, RMSEA und AIC fielen mit Ausnahme des BIC-Wertes im Original NF-Modell etwas niedriger aus, Δ RMSEA lag jedoch unterhalb des kritischen Differenzwertes von .015 (vgl. Chen, 2007), was darauf hindeutet, dass die Modelle ähnlich gut konvergieren.

Tabelle 6

Fit Indices für alternative Modelle: kombiniertes Nested-Faktor Model (komb. NF; aus Valerius & Sparfeldt, 2014) und kombiniertes Bifactor (S-1) Modell

Model	χ^2	<i>df</i>	<i>p</i>	SRMR	RMSEA	BIC	AIC
Kombiniertes NF	899.468	431	<.001	.055	.044	75898.279	75339.515
Kombiniertes BF (S-1)	984.237	439	<.001	.056	.047	75428.634	75952.746

Anmerkung. SRMR = Standard Root Mean Square of Residuals, RMSEA = Root Mean Square Error of Approximation, BIC = Bayesian Information Criterion, AIC = Akaike Information Criterion.

⁵ Aufgrund der Gestaltung der BIS-4-Subtest anhand der BIS-Raute (s. Jäger, 1982; oder Abb. 4, S. 21) enthielten die BIS-Subtests, die den spezifisch figuralen Inhaltsfaktor bildeten, neben den figuralen Inhalten nicht nur Verarbeitungskapazität (Reasoning) sondern auch weitere Operationsfacetten (Merkfähigkeit, Einfallsreichtum und Bearbeitungsgeschwindigkeit).

Hinsichtlich der statistischen Signifikanz der standardisierten Faktorladungen zeigten sich keine Unterschiede in den beiden Modellen (s. Tabelle 7, S. 38). Zu 46% fielen die standardisierten Faktorladungen im Bifactor (S-1) Modell (BF [S-1] Modell) in Relation zum Original NF-Modell numerisch etwas höher und somit substanzieller aus. Zusätzlich kamen im BF (S-1) Modell keine negativen Ladungen vor, wie es vergleichsweise für den spezifisch figuralen BIS-Faktor im kombinierten NF-Modell der Fall war.

Konsistenz der allgemeinen und spezifischen Intelligenzfaktoren

Wie in Tabelle 8 (S. 39) dargestellt, ergab das batterieübergreifende Korrelationsmuster im BF (S-1) Modell – ähnlich dem Original NF-Modell – signifikante konvergente Beziehungen zwischen den spezifisch verbalen ($r = .77$ vs. $r = .83$ im kombinierten Original NF-Modell aus Valerius & Sparfeldt, 2014, vgl. Tabelle 5, S. 30) und numerischen ($r = .55$ vs. $r = .46$ im Originalmodell, vgl. Tabelle 5, S. 30) Intelligenzfaktoren und ähnlich divergente Zusammenhänge zwischen spezifisch verbalen und numerischen Faktoren ($-.32 \leq r \leq -.02$ vs. $-.47 \leq r \leq .03$ im Originalmodell, vgl. Tabelle 5, S. 30). Die Korrelationen der batteriespezifischen g-Faktoren blieben auch im kombinierten BF (S-1) Modell signifikant, sanken jedoch leicht mit Ausnahme von Fluid-g mit BIS-g ($r = 1.00$) im Vergleich zum Original NF-Modell leicht ($r_{\text{BIS-g/CogAT-g}} = .83$ vs. $.91$ im Original-Modell; $r_{\text{Fluid-g/CogAT-g}} = .89$ vs. $.92$ im Original-Modell; vgl. Tabelle 4, S. 29).

Tabelle 7

Standardisierte Faktorladungen für alternative Modelle: kombiniertes Bifactor (S-1) Modell und kombiniertes NF-Modell (aus Valerius & Sparfeldt, 2014; Werte in Klammern)

Test	BIS-4				CogAT				Fluid
	λ_{VBIS}	λ_{NBIS}	λ_{FBIS}	λ_{BIS_g}	λ_{VCogAT}	λ_{QCogAT}	λ_{NCogAT}	λ_{CogAT_g}	λ_{Fluid_g}
TM	.51* (.45*)			.23* (.25*)					
EF	.21* (.25*)			.10 (.08)					
WA	.34* (.28*)			.34* (.36*)					
TG	.38* (.45*)			.14 (.14)					
ST	.30* (.37*)			.19* (.18*)					
ZN		.36* (.34*)		.42* (.45*)					
XG		.52* (.51*)		.23* (.27*)					
ZP		.34* (.34*)		.16* (.14*)					
SC		.27* (.26*)		.41* (.42*)					
DR		.24* (.22*)		.16* (.17*)					
OG			(.31*)	.37* (.35*)					
AN			(.06)	.55* (.55*)					
LO			(.32*)	.11 (.09)					
BD			(.39*)	.16* (.14*)					
CH			(-.14*)	.58* (.59*)					
WS 1					.56 ^{a*} (.51 ^{a*})			.26* (.31*)	
WS 2					.49 ^{a*} (.55 ^{a*})			.18* (.19*)	
WA 1					.38 ^{b*} (.43 ^{b*})			.40* (.42*)	
WA 2					.49 ^{b*} (.42 ^{b*})			.35* (.39*)	
MV 1						.14 ^c (.06 ^c)		.40* (.48*)	
MV 2						.14 ^c (.06 ^c)		.49* (.56*)	
ZR 1						.74 ^{d*} (.66 ^{d*})		.45* (.47*)	
ZR 2						.57 ^{d*} (.63 ^{d*})		.41* (.40*)	
FK 1							(.49 ^{e*})	.63* (.51*)	
FK 2							(.53 ^{e*})	.64* (.53*)	
FA 1							(.23 ^f)	.73* (.67*)	
FA 2							(.26 ^f)	.65* (.58*)	
RF									.51* (.52*)
KL									.44* (.44*)
MZ									.40* (.41*)
TO									.41* (.42*)
SPM									.63* (.62*)

Anmerkung. Die Faktorladungen mit gleichem hochgestelltem Buchstaben (a-f) wurden gleichgesetzt (*odd-even-split parcel*). Beschreibung der Tests s. Tabelle 1, S. 24.* $p < .05$

Zusammengefasst zeigen die Ergebnisse der Zusatzanalysen im kombinierten Bifactor (S-1) Modell, dass das Befundmuster des kombinierten NF-Modells aus Valerius & Sparfeldt (2014) erwartungsgemäß weitestgehend stabil blieb. Die diesbezügliche Methodeninvarianz der Befundinterpretation deutet an, dass die von Eid et al. (2018) thematisierten NF-Kritikpunkte hier nicht bzw. wenig bedeutsam waren. Es wurde deutlich, dass sich sowohl anhand der statistischen Kennwerte SRMR, RMSEA und AIC als auch anhand der Höhe/Substanz der standardisierten Faktorladungen, keine Überlegenheit des in den Zusatzanalysen spezifizierten Bifactor (S-1) Modells gegenüber dem Original-NF-Modell aus Valerius & Sparfeldt (2014) interpretieren ließ (mit Ausnahme einer negativen Faktorladung auf den figuralen Inhaltsfaktor, die im BF [S-1] Modell nicht mehr auftauchte).

Tabelle 8

Korrelationen der allgemeinen und spezifischen Intelligenzfaktoren im kombinierten BF (S-1) Modell

Faktor	V_{CogAT} <i>verbal</i>	Q_{CogAT} <i>numerisch</i>	CogAT-g	Fluid-g
V_{BIS}	.77*	-.02		
N_{BIS}	-.32*	.55*		
BIS-g			.83*	1.00*
CogAT-g				.89*

* $p < .05$.

Das Korrelationsmuster der spezifischen verbalen und numerischen Intelligenzfaktoren blieb im BF (S-1) Modell weitestgehend stabil, teilweise wurden die g-Faktor-Korrelationen batterieübergreifend leicht geschwächt, die Korrelation der spezifisch numerischen Faktoren leicht gestärkt.

Ergänzend zu den Befunden der ersten Studie zur Konsistenz der Messung von sowohl allgemeinen wie auch inhaltspezifischen Intelligenzfaktoren verschiedener Intelligenztestbatterien – und somit deren konvergente und divergente Validität – überprüften wir in der nachfolgenden, zweiten Studie den Zusammenhang der allgemeinen sowie spezifisch verbalen, numerischen und figuralen Intelligenzfaktoren zu allgemeinen, wie auch spezifischen sprachlichen und mathematisch/naturwissenschaftlichen Schulleistungsfaktoren – modelliert über den Nested-Factor Ansatz. Fokus der zweiten Studie lag somit – ergänzend zu Studie 1 – auf der kriteriumsbezogenen Validität der allgemeinen und spezifischen Intelligenzfaktoren für relevante Außenkriterien (Schulleistungsfaktoren).

3.3 Studie 2: Valerius & Sparfeldt (2015)

Valerius, S. & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell. *Zeitschrift für Pädagogische Psychologie*, 29, 101-108.

3.3.1 Einführung

Die Bedeutung der allgemeinen Intelligenz im Sinne von Spearman's g (Spearman, 1904) als Korrelat und Prädiktor schulischer Leistungen ist vielfach belegt (z. B. Gottfredson, 1997; Jensen, 1998; Rost, 2013). Im Mittel korrelieren allgemeine Intelligenz und Schulleistungen um $.5 \leq r \leq .6$ (Helmke & Schrader, 2010, S. 91). Weniger Einigkeit herrscht hingegen darüber, ob und wenn ja, welche Intelligenzgruppenfaktoren systematisch mit spezifischen Schulleistungen in einzelnen Schulfächern oder Schulfachgruppen zusammenhängen. Im Folgenden soll es daher über die Beziehungen allgemeiner Intelligenz mit allgemeinen Schulleistungen hinaus insbesondere um differenzielle Zusammenhänge spezifischer Intelligenzfaktoren mit spezifischen Schulfachleistungen gehen (vgl. z. B. Brunner, 2006; Coyle et al., 2013; Freund, Holling & Preckel, 2007; Gustafsson & Balke, 1993) – bis hin zu ggf. g -unabhängigen, korrespondierenden Begabungs- und Schulleistungsschwerpunkten. Auch wenn theoretisch differenzierte Zusammenhänge der (z.B. hinsichtlich der Inhalte oder Materialarten) korrespondierenden Intelligenz- und Schulleistungsfaktoren erwartet werden können, haben sich bisher nur wenige Studien der entsprechenden Forschungsfrage systematisch genähert; auf zentrale Arbeiten gehen wir nachfolgend ein. Dabei unterscheiden wir zwischen Arbeiten, die – wie in der vorliegenden Untersuchung – den Nested-Factor-Modellansatz verfolgten und Arbeiten, die einen anderen Analyseansatz verfolgten. Die NF-Modellierung kam in den Studien von z.B. Beaujean et al. (2014), Brunner (2006; 2008), Gustafsson & Balke (1993) zum Einsatz; einen jeweils anderen Analyseansatz wurde in den Untersuchungen von z.B. Calvin, Fernandes, Smith, Visscher, & Deary, (2010) sowie Coyle et al. (2013) realisiert.

Gustafsson & Balke (1993) spezifizierten ein NF-Modell der kognitiven Leistungen (13 Begabungssubtests, 3 Schulleistungstests [Mathematik, Schwedisch, Englisch]) und ein NF-Modell der Schulnoten in 17 Fächern. In Regressionsanalysen klärte der allgemeine kognitive

Leistungsfaktor (g) 28 % der Varianz des allgemeinen Schulleistungsfaktors auf. Die Varianzaufklärung für den spezifisch numerischen, kognitiven Leistungsfaktor betrug 56% durch den spezifisch numerischen Leistungsfaktor (neben 17% Varianzaufklärung durch kognitiv- g) und der spezifisch sprachliche Schulleistungsfaktor wurde zu 52% durch den spezifisch kristallinen und 30% durch kognitiv- g aufgeklärt. Das somit differenzierte Beziehungsgeflecht der g -unabhängigen Intelligenzfaktoren mit den spezifischen Schulleistungsfaktoren scheint lediglich dadurch getrübt, dass die Zusammenhänge möglicherweise überschätzt wurden, da die „kognitiven Leistungen“ neben Intelligenzsubtests auch drei Schulfachleistungstests umfassten (vgl. Süß, 2001).

Basierend auf den PISA-2000-Daten spezifizierte Brunner (2008; vgl. auch 2006) im NF-Modell allgemeine (g) sowie spezifisch verbale und numerische kognitive Leistungen. Es ergaben sich – neben den Zusammenhängen von kognitivem g und umgepolten Schulnoten (Deutsch: $r = .15$; Mathematik: $r = .22$) – für die g -unabhängigen spezifisch verbalen und numerischen Intelligenzfaktoren konvergente (spezifisch verbaler Fähigkeitsfaktor mit der umgepolten Deutschnote $r = .13$; spezifisch numerischer Fähigkeitsfaktor mit der Mathematiknote $r = .22$) und divergente ($-.13 \leq r \leq -.10$) Validitätshinweise. Die Interpretation der Befunde wird jedoch ähnlich wie bei Gustafsson & Balke (1993) dadurch erschwert, dass zusätzlich zu Intelligenzsubtests Kompetenztests im Lesen und Mathematik in die Bildung der kognitiven Fähigkeitsfaktoren einfließen.

Beaujean et al. (2014) konnten anhand der amerikanischen Standardisierungsstichprobe der Wechsler Skalen für Kinder zeigen, dass im NF-Modell, ergänzt um ein Pfadmodell, der Intelligenz g -Faktor Pfadkoeffizienten von $\beta = .81$ für Lesen und $\beta = .83$ für Schreiben aufwies und darüber hinaus der g -unabhängige Sprachverständnis-Intelligenzfaktor Lesen statistisch prädizierte ($\beta = .21$). Die Autorin und Autoren versäumten jedoch die Betrachtung divergenter Zusammenhänge nicht korrespondierender (z.B. mathematisch/naturwissenschaftlicher) Schulleistungsfaktoren.

In den nachfolgend erläuterten zwei Studien wurde zwar nicht die NF-Modellierung verwendet, die konvergente und divergente Validität allgemeiner Intelligenz- und Schulleistungsvariablen wurde jedoch belegt und soll daher diesbezüglich dargestellt werden. In der Untersuchung von Calvin et al. (2010, Tab. 2) verwies das berichtete Beziehungsmuster – über g hinausgehend – auf konvergente Validität der spezifisch

verbalen/numerischen Begabungsschwerpunkte mit Englisch- bzw. Mathematik-Schulleistungen ($r = .23/.13$) und auf divergente Validität verbaler/numerischer Intelligenz mit jeweils entsprechend entgegengesetzten schulischen Leistungen ($-.08 \leq r \leq -.06$).

Coyle et al. (2013, S. 118) modellierten einen gemeinsamen g -Faktor aus fünf Gruppenfaktoren: einem Studierfähigkeitsfaktor (verbale und numerische Bereichskennwerte aus zwei Studierfähigkeitstests), einem verbalen, einem numerischen, einem speed- und einem shop-Faktor (mechanisch-technisches Verständnis) eines Intelligenzstrukturtests. Die g -unabhängigen Residuen der numerischen bzw. verbalen Studierfähigkeitstests korrelierten mit den konvergenten numerischen bzw. verbalen Faktoren positiv ($.25 \leq r \leq .33$) und mit den divergenten verbalen bzw. numerischen Faktoren negativ ($-.34 \leq r \leq -.22$); die Zusammenhänge mit dem speed- und shop-Faktor schwankten um Null ($-.11 \leq r \leq .09$). Das Muster zeigt somit deutlich differenzielle Zusammenhänge, allerdings erschwert der g -Faktor als Konglomerat aus Intelligenz und Studierfähigkeit die Interpretation.

Zentrale Voraussetzungen zur Ermittlung differenzieller Zusammenhänge von Intelligenz und Schulleistung liegen in der psychometrischen Güte der Erfassung und der angemessenen Modellierung beider Variablen. Trotz der bekannten Kritikpunkte und Schwächen von Zensuren als Schulleistungsindikatoren werden die psychometrischen Eigenschaften häufig als mindestens ausreichend beurteilt; im deutschen Schulsystem „besitzen sie logische Validität, weil der Schulerfolg von Amts wegen allein durch sie definiert ist“ (Tent & Birkel, 2010, S. 953). Dimensionsanalytisch weisen Sekundarstufenzeugnisse in der Regel, so Tent und Birkel (2010), einen Fremdsprachenfaktor (auf den auch Deutsch lädt), einen mathematisch-naturwissenschaftlichen und einen Sachfächerfaktor auf. Bezogen auf die Güte der Intelligenzerfassung werden nach Rost (2013, S. 229) häufig Reliabilitäten von mindestens $r_{tt} = .90$ erreicht; neben vielfältigen kriteriumsbezogenen Validitätshinweisen (vgl. zusammenfassend Rost, 2013) zeigten sich sehr hohe konvergente Zusammenhänge zwischen den g -Faktoren unabhängig voneinander entwickelter Testbatterien (z. B. $r \geq .99$ bei Johnson et al., 2004; $r \geq .77$ bei Johnson et al., 2008). Valerius & Sparfeldt (2014) fanden in einem kombinierten NF-Modell über die sehr hohen testübergreifenden g -Faktor-Korrelationen ($r \geq .91$) dreier unterschiedlicher Testbatterien hinaus ein konvergent/divergentes Beziehungsmuster der g -unabhängigen spezifisch verbalen ($r = .83$) und numerischen ($r = .46$) Intelligenzfaktoren. Dies spricht für die Reliabilität und die kriteriumsbezogene Validität

(andere Testbatterie als Kriterium) der g -unabhängigen verbalen und numerischen Intelligenzgruppenfaktoren. Allerdings betrachteten die Autorin und der Autor keine weiteren Kriteriumsvariablen (wie beispielsweise Schulleistungen).

Daher soll im vorliegenden Beitrag die Forschungsfrage zur Klärung des differenzierten Zusammenhanggefüges von Intelligenz und Schulleistung beantwortet werden. Dazu spezifizierten wir ein kombiniertes NF-Modell mit einerseits NF-Begabungsteil und andererseits einem NF-Schulleistungsteil. Der Vorteil der NF-Modellierung liegt klar in der distinkten Betrachtung des differenzierten Zusammenhanggefüges vor allem der spezifischen Faktoren frei von g -Varianz (Chen et al., 2006). Wir erwarteten entsprechend im kombinierten NF-Modell der Intelligenz und Schulleistungen – neben substanziellen Korrelationen des g -Faktors mit dem allgemeinen Schulleistungsfaktor – besonders für die spezifischen, g -unabhängigen verbalen und numerischen Intelligenzgruppenfaktoren ein konvergent-divergentes Beziehungsmuster mit den korrespondierenden, von der allgemeinen Schulleistung unabhängigen, spezifisch sprachlichen und mathematisch/naturwissenschaftlichen Schulleistungsfaktoren.

3.3.2 Methode

Stichprobe

Für die Analysen wurde auf die bei Valerius und Sparfeldt (2014) betrachtete Stichprobe zurückgegriffen. Insgesamt nahmen $N = 562$ Schülerinnen und Schüler (49.6 % weiblich; 4.4 % ohne Angabe; 15.6 Jahre, $SD = 0.45$) der neunten Klassenstufe aus sechs rheinland-pfälzischen Gymnasien an der Erhebung teil.

Instrumente und Durchführung

Geschulte Testleitende administrierten die Kurzform des Kognitiven Fähigkeitstests für 4. bis 12. Klassen (KFT 4-12 R; Heller & Perleth, 2000) mit je zwei verbalen, numerischen und figuralen Subtests sowie die Kurzform des Berliner Intelligenzstrukturtests, Form 4 (BIS 4; Jäger et al., 1997) mit je fünf verbalen, numerischen und figuralen Subtests, in (zwischen den Schulklassen) randomisierter Reihenfolge in je 90-Minuten-Blöcken während der regulären Unterrichtszeit an (nahezu) aufeinander folgenden Tagen. Die Schulleistungen in Deutsch, Mathematik, Englisch, Französisch, Physik und Geschichte wurden über die selbstberichteten

Zensuren des letzten Halbjahreszeugnisses erfasst (vgl. zur Akkuratess selbstberichteter Zensuren: Sparfeldt, Buch, Rost & Lehmann, 2008).

Datenanalyse

Zur Prüfung der Zusammenhänge von allgemeinen und spezifischen Intelligenzfaktoren sowie Schulleistungen spezifizierten wir im Rahmen konfirmatorischer faktorenanalytischer Modelle ein kombiniertes NF-Modell (s. Abb. 7, S. 44) der Intelligenz und Schulleistung.

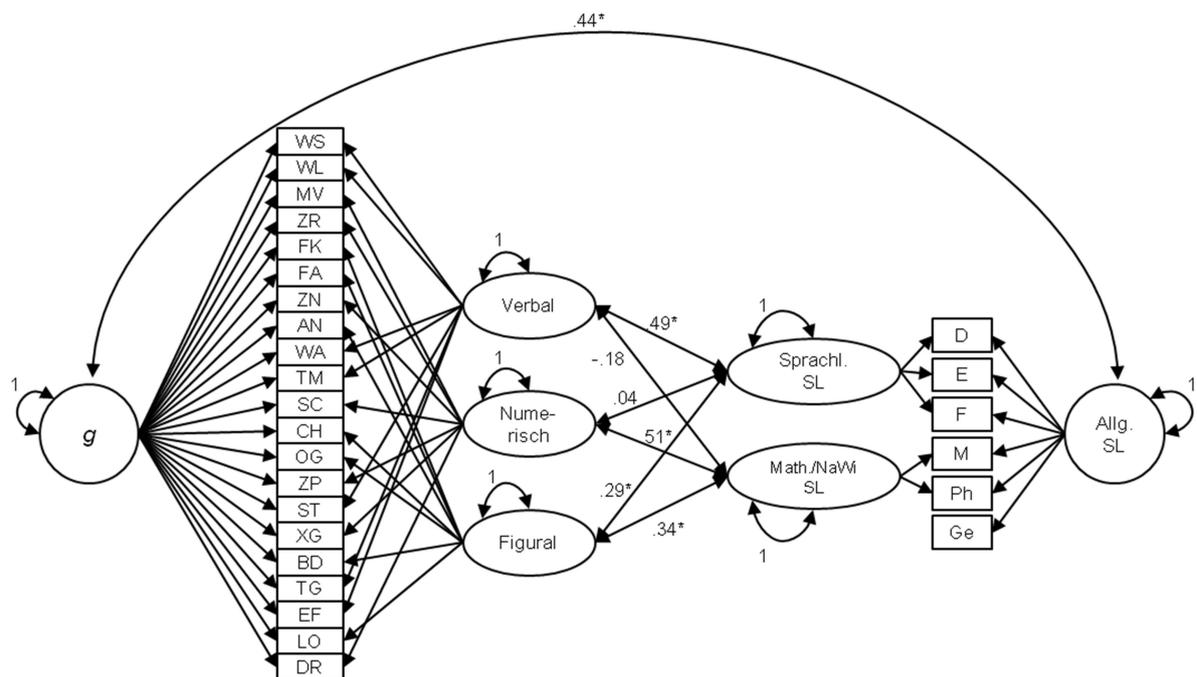


Abb. 7: Zusammenhänge zwischen Intelligenzfaktoren und Schulleistungsfaktoren im Nested-Factor-Modell.

Anmerkungen: *g* = Generalfaktor; Verbal = Verbaler Intelligenzfaktor; Numerisch = Quantitativ/Numerischer Intelligenzfaktor; Figural = Figuraler Intelligenzfaktor; Allg. SL = Allgemeiner Schulleistungsfaktor; Sprachl. SL = Sprachlicher Schulleistungsfaktor; Math./NaWi SL = Mathematisch/Naturwissenschaftlicher Schulleistungsfaktor; WS = Wortschatztest; WL = Wortanalogien; MV = Mengenvergleich; ZR = Zahlenreihen; FK = Figurenklassifikation; FA = Figurenanalogien; ZN = Zahlenreihen; AN = Analogien; WA = Wortanalogien; TM = Tatsache-Meinung; SC = Schätzen; CH = Charkow; OG = Orientierungsgedächtnis; ZP = Zahlenpaare; ST = Sinnvoller Text; XG = X-Größer; BD = Buchstaben Durchstreichen; TG = Teil-Ganzes; EF = Eigenschaften-Fähigkeiten; LO = Layout; DR = Divergentes Rechnen; D = Deutschzensur; E = Englischzensur; F = Französischzensur; M = Mathematikzensur; Ph = Physikzensur; Ge = Geschichtszensur. Der Übersicht halber werden die Fehlerterme (alle unkorreliert) nicht abgebildet. * $p < .05$

Der durch alle Intelligenzsubtests indikatorierte Generalfaktor der Intelligenz wurde neben den spezifisch verbalen, numerischen und figuralen Gruppenfaktoren spezifiziert, die durch die jeweiligen Subtests indikoriert wurden. Diese vier Intelligenzfaktoren durften nicht korrelieren. Analog dazu spezifizierten wir den allgemeinen Schulleistungsfaktor, der durch alle Fachzensuren indikoriert wurde und einen spezifisch sprachlichen (Deutsch, Englisch,

Französisch) und spezifisch mathematisch/ naturwissenschaftlichen (Mathematik, Physik) Gruppenfaktor. Auch diese drei Schulleistungsfaktoren durften nicht korrelieren (siehe Abb. 7, S. 44). Die Fachzensuren wurden umgepolt (höhere Ausprägung = bessere Leistung) und am jeweiligen Klassenmittelwert zentriert. Die modellbasierte Reliabilität der Faktoren beschreibt Omega hierarchisch (ω_h ; vgl. Brunner et al., 2012). Der Anteil fehlender Werte bei den Intelligenztests aufgrund der testheftweisen Datenerhebungen an verschiedenen Tagen fiel mit 2.5 % bis 6.2 % (Median 2.5 %) gering aus. Für die Zensuren war der Anteil fehlender Werte mit 3.4 % (Englisch) bis 9.4 % (Französisch) etwas höher (Median 3.7 %), was vor allem auf Fremdsprachenwahlen zurückzuführen ist. Zur Berücksichtigung fehlender Werte wurden die Analysen mittels FIML-Algorithmus in Mplus geschätzt. Die Mehrebenenstruktur der Daten (Schülerinnen und Schüler in Klassen) berücksichtigten wir durch die programminterne Mplus-Option „type = complex“, welche bei hierarchischen Datensätzen die Standardfehler korrigiert. Zur Modellevaluation zogen wir entsprechend den Empfehlungen von Hu und Bentler (1998) den Standardized Root Mean Square of Residuals (SRMR) zusammen mit dem Root Mean Square Error of Approximation (RMSEA) heran und berichten ergänzend den Comparative Fit Index (CFI). Zur Modellbewertung orientierten wir uns an Hu und Bentler (1999): Fit-Werte bei $SRMR \leq .08$ und $RMSEA \leq .06$ zeigen demnach einen guten Modell-Fit; CFI Werte um .95 verweisen auf einen (mindestens) akzeptablen Fit.

3.3.3 Ergebnisse

Modelltest

Der Modellfit des Gesamtmodells war (weitgehend) akzeptabel ($\chi^2 = 818.831$, $df = 292$, $RMSEA = .06$, $SRMR = .07$, $CFI = .83$). Die standardisierten Regressionsgewichte (Faktorladungen) der Zensuren auf dem allgemeinen ($.63 \leq \lambda \leq .74$, $\lambda_{\text{Median}} = .66$) und spezifisch sprachlichen ($.34 \leq \lambda \leq .45$, $\lambda_{\text{Median}} = .39$) bzw. mathematisch/naturwissenschaftlichen Schulleistungsfaktor ($.43 \leq \lambda \leq .46$, $\lambda_{\text{Median}} = .45$) erreichten durchweg substanzielle Werte und differierten jeweils statistisch bedeutsam von Null ($p < .05$). Bezogen auf den Intelligenz-Messmodellteil unterschieden sich die standardisierten Regressionsgewichte der Intelligenzsubtests auf den spezifischen Intelligenzgruppenfaktoren bis auf eine Ausnahme (BIS-Wortanalogien: $\lambda = .18$) ebenfalls jeweils statistisch bedeutsam von Null und waren überwiegend substanziell (Verbal: $.18 \leq \lambda \leq .54$, $\lambda_{\text{Median}} = .34$; Numerisch: $.28 \leq \lambda \leq .58$, $\lambda_{\text{Median}} = .43$; Figural: $.20 \leq \lambda \leq .53$, $\lambda_{\text{Median}} = .35$).

Dies traf auch für die meisten Ladungen auf dem Generalfaktor zu ($-.14 \leq \lambda \leq .66$, $\lambda_{\text{Median}} = .27$). Die modellbasierten Reliabilitätsschätzungen erreichten für den Intelligenzgeneralfaktor $\omega_h = .64$ und für die spezifisch verbalen bzw. numerischen bzw. figuralen Intelligenzfaktoren $\omega_h = .39$ bzw. $.57$ bzw. $.44$ sowie für den allgemeinen Schulleistungsfaktor $\omega_h = .87$ und für die spezifisch sprachlichen bzw. mathematisch/naturwissenschaftlichen Schulleistungsfaktoren $\omega_h = .19$ bzw. $.24$.

Zusammenhänge allgemeiner und spezifischer Intelligenz- und Schulleistungsfaktoren

Bezogen auf die inhaltliche Frage des Zusammenhangsmusters im Strukturmodellteil zeigte sich erstens, dass intelligentere Schülerinnen und Schüler fächerübergreifend im Mittel bessere Schulleistungen erreichten (Korrelation des allgemeinen Schulleistungsfaktors [Allg. SL] mit dem Intelligenz-g-Faktor: $r = .44$; vgl. Abb. 7, S. 44). Zweitens fiel das Beziehungsmuster der Intelligenz- und Schulleistungsgruppenfaktoren differenzierter aus: Schülerinnen und Schüler mit – von der allgemeinen Begabung unabhängigen – verbalen (bzw. numerischen) Stärken wiesen einen damit korrespondierenden und von der allgemeinen Schulleistungshöhe unabhängigen sprachlichen (bzw. mathematisch-naturwissenschaftlichen) Schulleistungsschwerpunkt auf. Die Korrelationskoeffizienten zwischen dem Intelligenz-g-unabhängigen spezifischen verbalen bzw. numerischen Intelligenzfaktor und den korrespondierenden Schulleistungs-g-unabhängigen spezifisch sprachlichen ($r = .49$; vgl. Abb. 7, S. 44) bzw. mathematisch/naturwissenschaftlichen ($r = .51$; vgl. Abb. 7, S. 44) Schulleistungsfaktoren wiesen entsprechend auf konvergente Validität hin. Zudem fielen diese konvergenten Korrelationskoeffizienten numerisch höher aus als die jeweils divergenten Korrelationen der spezifischen Faktoren ($|r| \leq .18$; vgl. Abb. 7, S. 44). Für die spezifisch figuralen Intelligenzfaktoren zeigten sich bedeutsame Beziehungen sowohl zum spezifisch sprachlichen ($r = .29$; vgl. Abb. 7, S. 44) als auch mathematisch/naturwissenschaftlichen ($r = .34$; vgl. Abb. 7, S. 44) Schulleistungsfaktor, obgleich diese jeweils numerisch geringer als die Beziehungen zu den entsprechenden konvergenten, spezifischen Intelligenzfaktoren ausfielen.

3.3.4 Diskussion

In dieser Arbeit untersuchten wir das differenzielle Beziehungsmuster allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungsfaktoren im

kombinierten NF-Modell. Neben der mittelhohen Korrelation des Intelligenz-g-Faktors mit dem allgemeinen Schulleistungsfaktor ergaben sich für die verbalen und numerischen Intelligenzgruppenfaktoren konvergent-divergente Zusammenhänge mit den spezifisch sprachlichen und mathematisch/naturwissenschaftlichen Schulleistungsfaktoren. Der figurale Intelligenzfaktor stand in bedeutsamer Beziehung sowohl zum spezifisch sprachlichen als auch zum spezifisch mathematisch/naturwissenschaftlichen Schulleistungsfaktor; diese Zusammenhangskoeffizienten waren numerisch geringer als die zu den entsprechend konvergenten Intelligenzfaktoren. Dieses differenzielle Beziehungsmuster der Intelligenzgruppenfaktoren repliziert im Wesentlichen die differenziellen Befunde von Gustafsson & Balke (1993), Brunner (2008) sowie Coyle und Kollegen (2013), wobei in der vorliegenden Studie deutlicher zwischen einerseits Intelligenzsubtests und andererseits Schulleistungen unterschieden wurde. Zwar konnten Beaujean et al. (2014) bereits die Zusammenhänge eines spezifisch verbalen Intelligenzgruppenfaktors (Sprachverständnis) und Schulleistungen im Lesen und Schreiben unabhängig vom Intelligenzgeneralfaktor im NF-Modell dokumentieren, doch ergänzen die vorliegenden Befunde das Bild eines bedeutsamen konvergenten Zusammenhangs des verbalen Intelligenzgruppenfaktors mit dem korrespondierenden sprachlichen Schulleistungsfaktor: zum einen um den (konvergenten) Zusammenhang des numerischen NF-Inhaltsfaktors mit dem mathematisch/naturwissenschaftlichen NF-Schulleistungsfaktor und zum anderen um die jeweils divergenten Korrelationen dieser Faktoren. In Übereinstimmung mit den genannten Studien haben wir die allgemeinen und spezifischen Faktoren im NF-Modell unabhängig voneinander modelliert, was den Vorteil aufweist, dass die jeweiligen spezifischen Stärken und Schwächen (Intelligenz- bzw. Schulleistungs-Gruppenfaktoren) unabhängig von der allgemeinen Intelligenzhöhe bzw. Schulleistungshöhe sind. Somit resultierten hier für den sprachlichen *und* mathematisch/naturwissenschaftlichen Bereich entsprechende konvergente *und* divergente Validitätshinweise korrespondierender Begabungs- und Schulleistungsschwerpunkte. Die substanziellen Korrelationen des spezifisch figuralen Begabungsgruppenfaktor sowohl mit sprachlicher als auch mathematisch/numerischer Schulleistung spricht zum einen für die nicht-triviale Bedeutung figuraler Fähigkeiten, hält jedoch noch offene Fragen (z.B. nach eindeutig korrespondierenden Schulfächern) für weitere Untersuchungen bereit.

Die prognostische Bedeutung von Begabungsschwerpunkten belegen Lubinski & Benbow (2006) im längsschnittlichen Design an einer Stichprobe Hoch- und Höchstbegabter: die im frühen Jugendalter erfassten Begabungsschwerpunkte (verbal bzw. numerisch bzw. figural) korrespondierten beispielsweise mit der fachlichen Zuordnung der späteren High-school-Lieblingskurse, der Domäne des Hochschul-Studienfachs und dem Feld der beruflichen Tätigkeit im Alter von 33 Jahren. Auch die gelegentlich eher vernachlässigten figural-räumlichen Fähigkeiten (erfasst mit 13 Jahren) zeigten – über verbale und numerische Fähigkeiten hinaus – Vorhersagebedeutung beispielsweise für innerhalb der mehr als 30 folgenden Lebensjahre verfassten wissenschaftlichen Publikationen und Patente im korrespondierenden Feld (Kell, Lubinski, Benbow & Steiger, 2013).

Vor dem Hintergrund der Untersuchungsanlage und der Ergebnisse lassen sich einige Perspektiven für zukünftige Forschung ableiten: Zur Absteckung des Generalisierungsrahmens wäre es z.B. denkbar, entwicklungspsychologisch zu untersuchen ab welcher Altersstufe bereits eine ähnlich differenzierte Intelligenzstruktur samt Kriteriumsbeziehungen sichtbar wird. Weiterhin wäre zu klären, welche Bedeutung die Befunde für die praktische Diagnostik und Beratung in der pädagogisch-psychologischen Anwendungspraxis haben. Zu beachten ist jedoch, dass die Voraussetzungen für eine individualdiagnostische Intelligenzprofilinterpretation in der Regel nicht ausreichend erfüllt sind und bisherige Studien ein solches Vorgehen nicht stützen (vgl. zusammenfassend Rost, 2013, S. 223–227 und 340–342).

Trotz noch offener Fragen, liefern die hier berichteten Befunde erste konvergent-divergente Validitätshinweise spezifischer Begabungs- und Schulleistungsstärken und -schwächen, die über die mittelhohe Korrelation der allgemeinen Intelligenz mit allgemeinen Schulleistungen hinausgehen.

3.4 Ergänzende Analysen zu Valerius & Sparfeldt (2015)

Zur Überprüfung von potenziellen Schätzproblemen des erweiterten NF-Modells und möglichen Konsequenzen für die Ergebnisinterpretation (vgl. Eid et al., 2018; Koch et al. 2018) führte ich für die Fragestellungen der zweiten Studie (Valerius & Sparfeldt, 2015) zusätzliche Analysen mit einem wie von Eid et al. (2018) empfohlenen Bifactor (S-1) Modell (s. Abb. 8, S. 49). Dazu spezifizierte ich für den Intelligenzteil des kombinierten NF-Modells eine Referenzfacette für den g -Faktor aus den Subtests, die im Original NF-Modell (vgl. Valerius & Sparfeldt, 2015) den spezifisch figuralen Intelligenzfaktor bildeten. Neben theoretischen Überlegungen zur Wahl des figuralen Intelligenzfaktors als Referenzfacette (ausführlicher s. Abschnitt 3.2, S. 34), zeigten die empirischen Befunde der zweiten Studie einen bedeutsamen positiven Zusammenhang des figuralen Intelligenzfaktors sowohl mit dem spezifisch sprachlichen als auch mathematisch/naturwissenschaftlichen Schulleistungsfaktor auf (vgl. Abb. 7, S. 44), was für einen gleichgerichteten oder eher „allgemeinen“ Einfluss des figuralen Intelligenzfaktors auf die spezifischen Schulleistungsfaktoren spricht und somit die Eignung als Referenzfacette für g empirisch unterstützt.

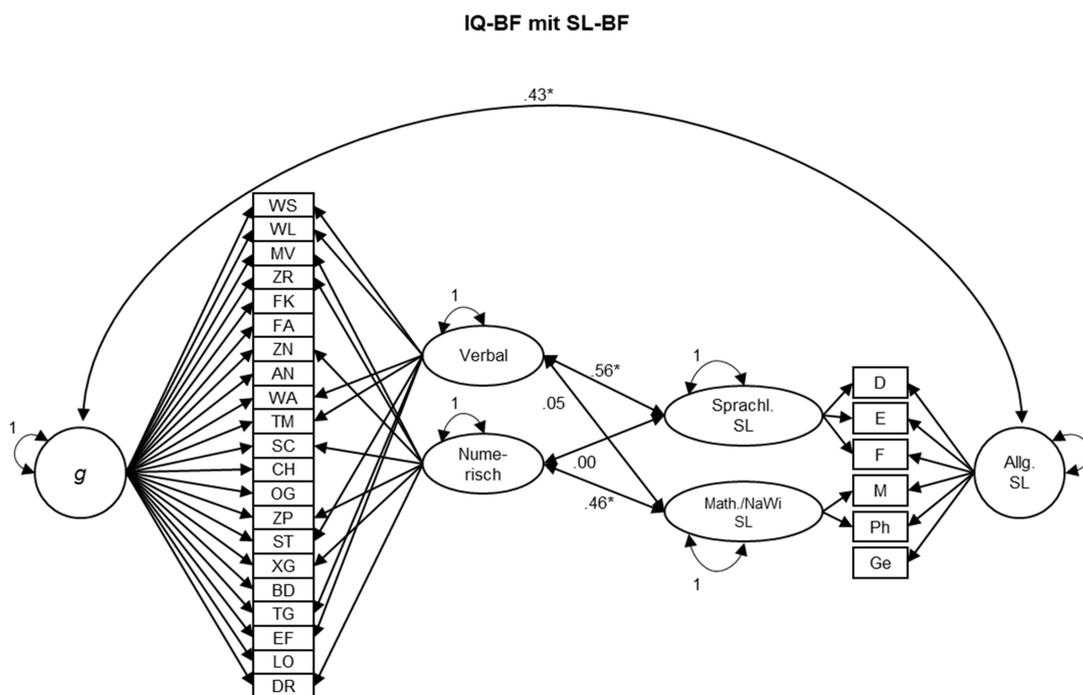


Abb. 8 Kombiniertes Bifactor (S-1) Modell für Intelligenz und Schulleistung. Der ursprünglich spezifisch figurale Intelligenzfaktor des Originalmodells aus Valerius & Sparfeldt (2015, S. 104) dient als Referenzfacette für g . $p < .05$.

Der NF-Schulleistungsteil des kombinierten und erweiterten NF-Modells aus Valerius & Sparfeldt (2015) enthielt schon im Original eine Referenzfacette (Geschichtsnote) für den allgemeinen Schulleistungsfaktor (vgl. Abb. 7, S. 44) und stellte somit bereits ein Bifactor (S-1) Modell im Sinne Eid et al.'s (2018) dar. Zur Unterscheidung der beiden zu vergleichenden Alternativmodelle verwende ich im Folgenden die Bezeichnungen *Intelligenz-Bifactor (S-1) mit Schulleistung-Bifactor (S-1) Modell (IQ-BF mit SL-BF Modell)*; für das Modell der Zusatzanalyse) sowie *Intelligenz-Nested Factor mit Schulleistung-Bifactor (S-1) Modell (IQ-NF mit SL-BF Modell)*; entspricht dem Original aus Valerius & Sparfeldt, 2015).

3.4.1 Ergebnisse der Zusatzanalysen

Modellevaluation

Das alternativ spezifizierte *IQ-BF mit SL-BF Modell* fittete ähnlich gut wie das Original *IQ-NF mit SL-BF-Modell* (aus Valerius & Sparfeldt, 2015) ohne figurale Referenzfacette ($\chi^2 = 875.565$, $df = 301$). Wie in Tabelle 9 (S. 50) berichtet, lagen die Fitstatistikwerte (RMSEA, BIC und AIC) mit Ausnahme von SRMR im *IQ-BF mit SL-BF Modell* numerisch leicht höher als im kombinierten Original Modell (*IQ-NF mit SL-BF*). Δ RMSEA (vgl. Chen, 2007) lag bei .001, was darauf hindeutet, dass die beiden Modelle sich nicht bedeutsam unterschieden.

Tabelle 9

Fit Statistiken für alternative Modelle: Vergleich von *IQ-BF mit SL-BF Modell* und *IQ-NF mit SL-BF Modell* (aus Valerius & Sparfeldt, 2015)

Modell	χ^2	df	p	SRMR	RMSEA	BIC	AIC
<i>IQ-NF mit SL-BF</i>	818.831	292	<.001	.074	.057	63060.530	62571.070
<i>IQ-BF mit SL-BF</i>	875.565	301	<.001	.066	.058	63076.812	62626.336

Anmerkung. SRMR = Standard Root Mean Square of Residuals, RMSEA = Root Mean Square Error of Approximation, BIC = Bayesian Information Criterion, AIC = Akaike Information Criterion.

70% der standardisierten Faktorladungen, die im Original *IQ-NF mit SL-BF Modell* nicht signifikant auf den zugehörigen Faktoren luden, wurden im *IQ-BF mit SL-BF Modell*

signifikant (vgl. Tabelle 10, S. 51). 54% der standardisierten Faktorladungen zeigten im zusätzlich analysierten *IQ-BF mit SL-BF Modell* numerisch höhere Werte im Sinne von „substanzieller“ im Vergleich zum Original *IQ-NF mit SL-BF Modell*. Drei standardisierte Faktorladungen wiesen im Original-Modell negative Ladungen oder Ladungen von Null auf dem *g*-Faktor auf (λ_{IQ-g} für BD = -.09, LO = .00, ZP = -.14; vgl. Tabelle 10, S. 51). Diese wurden durch die veränderte Spezifizierung des *Intelligenz-BF (S-I) Modells* positiv und mit einer Ausnahme auch signifikant (λ_{IQ-g} für BD = .14, LO = .13, ZP = .07 [n.s.]; $p < .05$; vgl. Tabelle 10, S. 51).

Tabelle 10

Standardisierte Faktorladungen des *IQ-BF mit SL-BF Modells* im Vergleich zu *IQ-NF mit SL-BF Modell* (aus Valerius & Sparfeldt, 2015; Werte in Klammern)

	Intelligenzfaktoren				Schulleistungsfaktoren		
	λ_{IQ-V}	λ_{IQ-N}	λ_F	λ_{IQ-g}	λ_{SL-V}	λ_{SL-N}	λ_{SL-g}
TM	.50* (.40*)			.28* (.34*)			
EF	.19* (.30*)			.11 (.01)			
WA	.33* (.18)			.38* (.51*)			
TG	.36* (.54*)			.14 (.04)			
ST	.33* (.48*)			.19* (.10)			
V1	.48* (.34*)			.32* (.48*)			
V2	.39* (.23*)			.50* (.66*)			
ZN		.44* (.49*)		.33* (.27*)			
XG		.62* (.58*)		.12* (.06)			
ZP		.24* (.35*)		.07 (-.14)			
SC		.34* (.43*)		.35* (.22*)			
DR		.23* (.28*)		.13* (.03)			
Q1		.38* (.43*)		.49* (.49*)			
Q2		.32* (.42*)		.45* (.35*)			
LO			(.20*)	.13* (.00)			
BD			(.35*)	.14* (-.09)			
OG			(.42*)	.32* (.08)			
AN			(.31*)	.48* (.35*)			
CH			(.27*)	.53* (.46*)			
N1			(.50*)	.66* (.44*)			
N2			(.53*)	.76* (.55*)			
D					.42* (.34*)		.63* (.65*)
E					.39* (.39*)		.66* (.66*)
F					.31* (.45*)		.68* (.63*)
M						.40* (.43*)	.68* (.65*)
P						.43* (.46*)	.73* (.72*)
G							.70* (.74*)

Anmerkung. Beschreibung der Tests s. Tabelle 1, S. 24.

* $p < .05$

Zusammenhänge der Intelligenz- und Schulleistungsfaktoren

Der allgemeine Intelligenzfaktor korrelierte im zusätzlich spezifizierten *IQ-BF mit SL-BF Modell* bedeutsam positiv mit dem allgemeinen Schulleistungsfaktor ($r = .46$; vgl. Abb. 8, S. 49). Die konvergente Beziehung des spezifisch verbalen Intelligenzfaktors mit sprachlichem Schulleistungsfaktor korrelierte wie bereits im Originalmodell aus Valerius & Sparfeldt (2015) ebenfalls bedeutsam positiv ($r = .56$; vgl. Abb. 8, S. 49). Der spezifisch numerische Intelligenzfaktor korrelierte mit dem mathematisch/naturwissenschaftlichem Schulleistungsfaktor ähnlich wie im Originalmodell aus Valerius & Sparfeldt (2015) statistisch bedeutsam zu $r = .46$ (vgl. Abb. 8, S. 49). Die divergenten Beziehungen der spezifischen Intelligenzfaktoren mit den entsprechend gegensätzlich korrespondierenden spezifischen Schulleistungsfaktoren (verbaler respektive numerischer Intelligenzfaktor mit mathematisch/numerischem respektive sprachlichem Schulleistungsfaktor) korrelierten im zusätzlich spezifizierten *IQ-BF mit SL-BF Modell* ähnlich wie im Originalmodell aus Valerius & Sparfeldt (2015) um Null ($.00 \leq r \leq .05$; vgl. Abb. 8, S. 49).

Zusammengefasst lässt sich festhalten, dass die Zusatzanalysen mit für den Intelligenzteil spezifizierten Bifactor (S-1) Modell auch für die zweite Studie (Valerius & Sparfeldt, 2015) keine nennenswerten Abweichungen der Interpretation der zentralen Ergebnissen des Originalmodells aufzeigen und somit die Hypothese einer Erklärung aufgrund von Missspezifikationen weitestgehend ausgeschlossen werden kann. Unter Betrachtung der Fitstatistikwerte RMSEA, BIC und AIC zeigten sich keine bedeutsamen Unterschiede des Originalmodells (*IQ-NF mit SL-BF*) in Relation zum zusätzlich spezifizierten *IQ-BF mit SL-BF Modell*. Das Befundmuster der standardisierten Faktorladungen fiel jedoch im Gegensatz zum zusätzlichen *IQ-BF mit SL-BF Modell* insgesamt etwas passender aus als im Originalmodell, da keine negativen oder Nullladungen auf den Generalfaktor der Intelligenz auftraten und über die Hälfte der nichtsignifikanten Ladungen signifikant von Null differierten und/oder in ihrer absoluten Höhe stiegen. Die Korrelationen der allgemeinen und spezifischen Intelligenz- und Schulleistungsfaktoren veränderten sich in den Zusatzanalysen hingegen nur wenig, was als Methodeninvarianz und Stabilitätshinweis für die in Studie 2 gefundenen Zusammenhänge allgemeiner und spezifischer Intelligenz- mit Schulleistungsfaktoren gewertet werden kann.

3.5 Studie 3: Valerius, Lotz & Sparfeldt (eingereicht)

Valerius, S., Lotz, C. & Sparfeldt, J. R. (eingereicht). Hochbegabt oder nicht hochbegabt? Begabungsdiagnostik mit verschiedenen Intelligenztests. *Psychologie in Erziehung und Unterricht*.

3.5.1 Theorie

Im Zuge einer Beschäftigung mit „Hochbegabung“ in Forschung (vgl. Preckel & Krampen, 2016) und Praxis (vgl. Konferenz der Kultusminister [KMK], 2009, 2015) begegnet man schnell der Frage, wie Hochbegabte gefunden bzw. identifiziert werden können. Trotz unterschiedlicher Modellvorstellungen (z.B. Sternberg & Davidson, 2005; Preckel & Vock, 2013) definiert die aus guten Gründen verbreitete intelligenzbasierte Hochbegabungskonzeption intellektuelle Hochbegabung als sehr hohe Ausprägung der allgemeinen Intelligenz (z.B. Rost, 2009). Auch in mehrdimensionalen Hochbegabungsmodellen bildet eine hohe allgemeine Intelligenz in aller Regel das oder zumindest ein Kernmerkmal. Folglich nehmen Intelligenztests im Rahmen der Identifikation intellektuell Hochbegabter den oder zumindest einen zentralen Platz ein (vgl. Preckel, 2010; Preckel & Vock, 2013; Rost, 2009; Rost & Buch, 2018). Im Anschluss an die prinzipielle Entscheidung für eine Intelligenzdiagnostik ist dann ein spezifischer Intelligenztest auszuwählen. In der Regel sollten Hochbegabungsdiagnosen jedoch möglichst wenig vom eingesetzten Intelligenztest abhängen. Somit werfen Nicht-Übereinstimmungen und ein testabhängig unterschiedliches klassifikatorisch-diagnostisches Fazit („hochbegabt“ oder „nicht hochbegabt“) grundsätzlichere Fragen auf. Daher untersuchten wir, ob und inwieweit Hochbegabungsdiagnosen intelligenztestabhängig sind. Zur Bestimmung der korrelativen und klassifikatorischen Übereinstimmung betrachteten wir drei verbreitete Intelligenztests mit Aufgaben zum figuralen reasoning.

Intellektuelle Hochbegabung

Es existieren verschiedene Modelle zur intellektuellen Hochbegabung, wobei die weit überdurchschnittliche allgemeine Intelligenz als Kernmerkmal in aller Regel vertreten ist (z.B. Rost & Buch, 2018). Im Gegensatz zur eindimensionalen, intelligenzbasierten

Hochbegabungsdefinition bleibt in mehrdimensionalen Hochbegabungskonzeptionen die Grenzsetzung pro Variable sowie Kombinationsregel der zugrundeliegenden Variablen unklar (vgl. Preckel & Vock, 2013; Rost & Buch, 2018). Sowohl inhaltlich-psychologische, als auch diagnostische Argumente sprechen für eine intelligenzbasierte, eindimensionale Konzeption intellektueller Hochbegabung (vgl. Rost, 2009), um die es auch im Folgenden gehen soll. In Forschung und Praxis hat sich ein mindestens zwei Standardabweichungen über dem Mittelwert liegender Intelligenzgrenzwert eingebürgert. Somit gelten Personen mit einem IQ ab 130, was einem Prozentrang (PR) von 98 entspricht, als „hochbegabt“; Personen mit niedrigerer Intelligenz ($IQ \leq 130$ bzw. $PR \leq 98$) gelten entsprechend als „nicht hochbegabt“. Gelegentlich werden auch liberalere *cut-off*-Werte gewählt: z.B. $IQ \geq 125$ (bzw. $PR \geq 95$) oder $IQ \geq 120$ (bzw. $PR \geq 90$).

Intelligenztests in der (Hoch-)Begabungsdiagnostik

Im Rahmen der Hochbegabungsdiagnostik gelten Intelligenztests als Methode der Wahl. Doch unterscheiden sich gängige Intelligenztests, womit die Gefahr einer sogenannten testimmanenten Hochbegabungsdefinition (vgl. Preckel, 2010) verbunden sein könnte. Intelligenztests unterscheiden sich u.a. in Bezug auf (z.B. Schmidt-Atzert & Amelang, 2012, S. 203) (a) den zu erfassenden Intelligenzaspekt (z.B. allgemeine Intelligenz im Sinne eines Generalfaktors *g* vs. Intelligenzgruppenfaktor vs. Intelligenzstruktur; Aspekte des Aufgabenmaterials, z.B. sprachfrei/kulturfair und Ausmaß der Schul- bzw. Bildungsabhängigkeit), (b) Aspekte der Durchführung der Intelligenztestung (u.a. Einzel- vs. Gruppentestung, *speed*- vs. *power*-Test, Computer- vs. Papier-und-Stift-Test, Bearbeitungsdauer) sowie (c) der Zielgruppe (z.B. Eignung für den hohen Begabungsbereich bzw. spezifische [Sub-]Populationen und/oder Alters- bzw. Klassenstufen). Neben der Passung zur wissenschaftlichen und/oder praktisch-diagnostischen Fragestellung sollten folglich insbesondere theoretische, psychometrisch-diagnostische und (forschungs-)praktische Argumente bei der Entscheidung für einen konkreten Intelligenztest (oder mehrere) herangezogen und offengelegt werden.

Für eine Hochbegabungsdiagnostik bei Kindern und Jugendlichen im Schulalter stehen einige prinzipiell brauchbare Intelligenztests zur Verfügung (z.B. Preckel, 2010; Preckel & Vock,

2013). Neben überzeugenden Ausprägungen der gängigen Gütekriterien (insbesondere Objektivität, Reliabilität sowie Validität), sollten für Intelligenztests im Rahmen einer Hochbegabungsdiagnostik ausreichend viele schwierige Aufgaben zur Vermeidung von Deckeneffekten sowie eine angemessene und zeitnahe Normierung vorliegen (z.B. Rost, 2013). Insbesondere bei einer Verwendung veralteter Normen kann es andernfalls zu einer teilweise deutlichen IQ-Über- bzw. Verschätzung kommen (vgl. Flynn, 1987; Rost, 2013; Pietschnig & Voracek, 2015; Wai & Putallaz, 2011). Häufig kommen in der Praxis zwei oder mehr Tests zum Einsatz (z.B. um zu einer belastbareren Einschätzung zu kommen oder im Falle wiederholter Testungen). Damit stellt sich die Frage, ob und inwieweit diese zu vergleichbaren Ergebnissen kommen.

Korrelative Übereinstimmung zweier Intelligenztestverfahren

Kommen im Rahmen einer Hochbegabungsdiagnostik mindestens zwei Testverfahren zum Einsatz, lässt sich deren korrelative Übereinstimmung bestimmen. Die Korrelationshöhe zweier unabhängiger Intelligenztests wird u.a. durch ihre Ähnlichkeit beeinflusst. Dabei sind höhere Korrelationen zu erwarten, wenn sich beide Tests auf denselben Intelligenzfaktor beziehen (z.B. figurale Verarbeitungskapazität im Sinne des Berliner Intelligenzstrukturmodells [Jäger, Süß & Beauducel, 1997] oder *reasoning* im Sinne des Cattell-Horn-Carroll-Modells [McGrew, 2009] oder einen in beiden Tests vergleichbar zusammengesetzten Intelligenz-Generalfaktor [*g*; z.B. Jensen, 1998; Rost, 2013]). Im Gegensatz dazu fallen Korrelationskoeffizienten von Testverfahren, die unterschiedliche Intelligenzfaktoren erfassen, in der Regel etwas niedriger aus. Aus empirischer Sicht beeinträchtigen zudem Objektivitäts-, Reliabilitäts- und Validitätseinschränkungen eines Tests oder beider Tests die Korrelationshöhe. Darüber hinaus kann es im oberen Begabungsbereich aufgrund von Deckeneffekten zu Varianzeinschränkungen kommen, welche die Korrelationshöhe zusätzlich beeinflussen.

Korrelieren die latenten *g*-Faktoren unterschiedlicher Intelligenztestbatterien üblicherweise sehr hoch ($r \approx .90$; z.B. Johnson et al., 2004, 2008; Valerius & Sparfeldt, 2014), fallen die manifesten Korrelationen häufig niedriger aus mit Werten von $.50 \leq r \leq .60$. Im Rahmen eines Berliner Hochbegabtenprojekts ermittelte Weiß (2006) an einer Stichprobe von $N = 860$ Viertklässlern Testinterkorrelationen von $.44 \leq r \leq .55$ zwischen dem Intelligenzwert des

Strukturtests PSB-R 4-6 (Lukesch, Kormann & Mayrhofer, 2002) und der Kurzform des ausschließlich Aufgaben zum figuralen *reasoning* umfassenden CFT 20-R (Weiß, 2006). Außerdem korrelierten HAWIK-IV (Daseking, Petermann & Petermann, 2007) und IDS (Grob, Meyer & Hagmann-von Arx, 2009) in einer Begabten- und Normalbegabtenstichprobe (jeweils $N = 77$ Sechs- bis Zehnjährige) zu $r = .73$ in der Normalbegabtenstichprobe bzw. zu $r = .51$ in der Begabtenstichprobe (Hagmann-von Arx, Meyer & Grob, 2008).

Umfassen die Intelligenztestverfahren figurale *reasoning*-Aufgaben, werden in der Regel Korrelationen mindestens gleicher Größenordnung berichtet: So korrelierten drei Testbatterien mit figuralen reasoning-Aufgaben (CogAT-N, Lohmann & Hagen, 2001; SPM, Raven, 1941; NNAT, Naglieri, 1997) bei $N = 1198$ US-amerikanischen Grundschulkindern zwischen $.60 \leq r \leq .66$ (Lohman, Korb & Lakin, 2008). Bei Jugendlichen aus Hochbegabtenklassen ($N = 98$; 9. und 10. Jahrgangsstufe) hingen APM (Raven, 1962) und der nonverbale Teil des KFT 4-13 (Heller, Gaedike & Weinländer, 1985) zu $r = .51$ zusammen (Heller et al., 1985). Bei Schülerinnen und Schülern der Sekundarstufe ($N = 91$) korrelierten SPM und CFT 20 zu $r = .55$ (Heller, Kratzmeier & Lengfelder, 1998).

Übereinstimmung der Hochbegabungsklassifikation in zwei Testverfahren

Eine etwas andere Perspektive fokussiert auf die Übereinstimmung der klassifikatorisch-diagnostischen Ergebnisse zweier Intelligenzmessungen (hochbegabt vs. nicht hochbegabt; z.B. in zwei verschiedenen Intelligenztests zu einem Zeitpunkt *oder* in einem identischen, zu verschiedenen Zeitpunkten administrierten Test). Der erwartete Anteil derjenigen, die in zwei Intelligenztestungen konsistent als hochbegabt klassifiziert werden, kann mit Hilfe der Verteilungsfunktion der bivariaten Standardnormalverteilung berechnet werden (vgl. Hanses, 2009; Lohman & Korb, 2006). Hierbei variiert der Anteil übereinstimmend in beiden Tests als hochbegabt Klassifizierter in Abhängigkeit von der Testinterkorrelation und den Selektionsquoten (*Grenzwertsetzung*). So konnte Hanses (2009) anhand der Daten des Marburger Hochbegabtenprojekts zeigen, dass 50% der in der dritten Klassenstufe Hochbegabten (*cut-off*: $IQ \geq 130$) in Klassenstufe 9 erneut als hochbegabt mit $IQ \geq 130$ re-identifiziert wurden; bei etwas liberalerer Grenzmarke in der 9. Klassenstufe ($IQ \geq 125$) erreichte die Re-Identifikationsquote 71 %. Die über die Verteilungsfunktion der bivariaten

Standardnormalverteilung *a priori* geschätzten Anteile lagen in ähnlicher Höhe (50 % bzw. 72 % bei $r_{it} = .85$).

Lohman und Korb (2006, S. 457; vgl. Lohman, 2009) konnten zeigen, dass statistisch betrachtet eine Nicht-Übereinstimmung der Hochbegabungsklassifikation in zwei Testverfahren bei üblichen Testinterkorrelationen und Grenzwertsetzungen keinesfalls ungewöhnlich ist. So ist bei einem *cut-off*-Wert in einer Testung von $PR > 90/95/98$ und einer Testinterkorrelation von $r = .50$ zu erwarten, dass 32 %/24 %/17 % auch in einer zweiten Testung einen Wert von $PR > 90/95/98$ erreichen (bzw. z. B. $r = .70$ und identische *cut-off*-Werte in beiden Testungen von $PR > 90/95/98$: 47 %/39 %/31 %). In einer größeren Stichprobe (Lohman et al., 2008) erreichten von den mit drei Testverfahren untersuchten $N = 1064$ Schülerinnen und Schülern $n = 146$ (Raven) bzw. $n = 51$ (NNAT) bzw. $n = 26$ (CogAT-N) Stanine-Werte von 9 im entsprechenden Test und $n = 9$ in allen drei Tests einen entsprechend hohen Wert. Beachtenswert ist zudem, dass – u. a. bei unterschiedlichen *cut-off*-Werten oder unterschiedlich „strengen“ Normen der Intelligenztests – die Anteile der in beiden Tests als hochbegabt eingestuft in Abhängigkeit von der Betrachtungsrichtung variieren können (z. B. könnten 30 % von den in Test 1 Hochbegabten auch in Test 2 hochbegabt sein, während nur 10 % von den in Test 2 Hochbegabten auch in Test 1 hochbegabt sind). Zudem schlägt der Effekt des Messfehlers bzw. der Regression zur Mitte umso stärker zu Buche, je weiter der individuelle Testwert vom Mittelwert abweicht (vgl. Lohman & Korb, 2006; Lohman, 2009).

Klassifikationsgüte und der RATZ-Index

Grundlage der Beschreibung der Güte der klassifikatorischen Übereinstimmung von zwei Intelligenztests zur Hochbegabtenidentifikation sind die verschiedenen Klassifikationsfelder (s. Abb. 3, S. 15). Schülerinnen und Schüler in Klassifikationsfeld „a“ erzielten in beiden Testverfahren Ergebnisse oberhalb des Grenzwerts (valide positiv). In Feld „d“ kommen die übereinstimmend als nicht hochbegabt Klassifizierten zu liegen (valide negativ). Schülerinnen und Schüler, die jeweils in einem, nicht jedoch dem anderen der beiden Tests ein Ergebnis im Hochbegabungsbereich erreichten, finden sich in den Feldern „b“ und „c“. Die Gesamttrefferquote (GQ) beschreibt den Anteil der valide positiv und negativ Klassifizierten an der Gesamtanzahl ($[a+d]/N$). Die Selektionsraten (SR) bezeichnen jeweils den Anteil

derjenigen, die mittels eines Tests als hochbegabt klassifiziert wurden (für Test 1: $[a+b]/N$; für Test 2: $[a+c]/N$). Die Maximaltrefferquote ($MQ = 1 - [b-c]/N$) beschreibt die Obergrenze der GQ. Im Rahmen einer Hochbegabungsdiagnostik bedingen die sehr hohen Grenzmarken in üblichen Stichproben mit geringem Hochbegabtenanteil deutlich unproportionale Zellaufteilungen. In der Regel übersteigt dann die Anzahl der übereinstimmend als nicht hochbegabt Klassifizierten (d) die Anteile in den übrigen Zellen deutlich. Folglich fallen GQs dann zwar hoch aus, sagen jedoch eher etwas über die Güte der übereinstimmenden Klassifizierung Nicht-Hochbegabter als Hochbegabter aus.

Zur Beschreibung der Klassifikationsgüte in der Hochbegabungsdiagnostik kann alternativ der RAZ-Index (*Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote*) herangezogen werden (vgl. Beißert, Hasselhorn & Lösche, 2014). Verglichen wird wie angedeutet der relative Anstieg der Trefferquote gegenüber der (Zufalls-)Trefferquote (ZQ) im Verhältnis zum maximal möglichen Anstieg ($ZQ = 100 * \frac{[a+b] * [a+c]}{N^2} + \frac{[b+d] * [c+d]}{N^2}$; $RAZ = \frac{[GQ - ZQ]}{[MQ - ZQ]} * 100$). Bislang fehlen jedoch inferenzstatistische Verfahren zur Absicherung von Unterschieden der RAZ-Indices. Marx und Lenhard (2010) schlugen daher in Analogie zur Interpretation von Effektstärken vor, den theoretischen Bereich des Indexes etwa zu dritteln: RAZ-Werte < 34 gelten damit als *nicht zufriedenstellende*, Werte zwischen 34 und 66 als *gute* und Werte > 66 als *sehr gute* Übereinstimmung. Der RAZ-Index kann also die klassifikatorische Übereinstimmung von Intelligenztests bei der Hochbegabungsdiagnostik quantifizieren. Uns ist keine Studie zur Hochbegabungsdiagnostik bekannt, in der der RAZ-Index berechnet wurde. Schlagheck und Petermann (2006) berichteten jedoch Befunde für eine Inanspruchnahmestichprobe einer kinderpsychologischen Ambulanz, die zur Abklärung einer Hochbegabungsvermutung ($N = 86$) den HAWIK III (Tewes, Rossmann & Schallberger 1999) und den AID 2 (Kubinger & Wurst, 2000) in unterschiedlicher Reihenfolge bearbeiteten (vgl. zur Kritik: Sparfeldt & Rost, 2008), aus denen sich nachträglich der RAZ-Index berechnen ließ (RAZ = 69; sehr gute Übereinstimmung der Hochbegabungsdiagnose [$IQ \geq 125$]).

Fragestellungen

Vor dem Hintergrund der in der Einleitung angesprochenen Inhalte und Fragen sollte in unserer Studie die korrelative und klassifikatorische Übereinstimmung verschiedener Intelligenztests im Zuge einer Hochbegabungsdiagnostik betrachtet werden. Bisherige Untersuchungen zur Übereinstimmung von Intelligenztests bezogen sich zumeist nur auf korrelative Zusammenhänge. Es mangelt jedoch an Studien, die empirisch die klassifikatorische Übereinstimmung von mindestens zwei Intelligenztests bei der Hochbegabtenidentifikation an einer größeren Stichprobe in den Blick nehmen. Die vorliegende Studie untersuchte daher die Übereinstimmung von drei weit verbreiteten und konzeptuell ähnlichen Intelligenztests bzw. Intelligenztestteilen mit ähnlichem Messanspruch, nämlich figuralem *reasoning* als sehr gutem Proxi der allgemeinen Intelligenz *g* (CFT 20-R, SPM und KFT-N). Die folgenden spezifischen Fragestellungen sollten beantwortet werden: (1) Wie hoch korrelieren die Ergebnisse in diesen drei konzeptionell ähnlichen Intelligenztests? (2) Wie viele Hochbegabte finden sich in diesen drei konzeptionell ähnlichen Intelligenztests in Abhängigkeit vom Grenzwert bei Verwendung der testspezifischen Normen ($IQ \geq 130/125/120$)? (3) Wie gut stimmen die Klassifikationen als „hochbegabt“ bzw. „nicht hochbegabt“ in jeweils zwei konzeptionell ähnlichen Testverfahren überein, wenn man (a) die prozentualen Übereinstimmungen und (b) den RAZ-Index betrachtet?

3.5.2 Methode

Stichprobe

Insgesamt nahmen $N = 562$ Schülerinnen (49,6 %) und Schüler (45,9 %; 4,5 % ohne Geschlechtsangabe) der 9. Klassenstufe von sechs rheinland-pfälzischen Gymnasien teil (23 Klassen; Alter: $M = 15.6$ Jahre, $SD = .45$). Wir wählten eine Gymnasialstichprobe, um einen erwartungsgemäß höheren Anteil intelligenterer Schülerinnen und Schüler zur Beantwortung der vorliegenden Fragestellungen der Hochbegabungsklassifizierung zu nutzen. Die Teilnahmequote lag bei 91 %.

Instrumente und Ablauf

Folgende drei Intelligenztests bzw. Intelligenztestteile zur Erfassung von figuralem *reasoning* kamen zum Einsatz: (1) Kurzform der Grundintelligenztestskala 2 (CFT 20-R; Weiß, 2006) mit vier figuralen Subtests (Reihenfortsetzen, Klassifikationen, Matrizen und Topologien), (2) Standard Progressive Matrices (SPM; Raven, 1941; Heller et al., 1998), (3) Kurzform des Kognitiven Fähigkeitstest für 4. bis 12. Klassen (KFT 4-12+R, Heller & Perleth, 2000). Um eine möglichst hohe konzeptuelle Ähnlichkeit des Messanspruchs der drei Intelligenztests (figurales *reasoning*) zu gewährleisten, verwendeten wir für die Analysen ausschließlich die nonverbal-figuralen KFT-Subtests (KFT-N; Figurenklassifikation, Figurenanalogien). Die drei Intelligenztests wurden in klassenweise randomisierter Reihenfolge an drei verschiedenen Schultagen von geschulten Testleiterinnen und Testleitern während der regulären Unterrichtszeit administriert.

Datenanalyse

Im Anschluss an die Darstellung deskriptiver Befunde berechneten wir zur Beantwortung der ersten Forschungsfrage nach den korrelativen Zusammenhängen die bivariaten Korrelationen der Rohwertsummen von CFT 20-R, SPM und KFT-N.

Die zweite Fragestellung bezog sich auf Anzahl bzw. Anteile Hochbegabter in den drei Intelligenztests in Abhängigkeit vom gesetzten Grenzwert. Hierfür bestimmten wir zunächst für jede Schülerin und jeden Schüler in jedem der drei Tests unter Verwendung der Testnormen die entsprechenden IQ-Werte (testnormspezifische Standardwerte⁶) und dann die prozentualen Anteile Hochbegabter für verschiedene *cut-off* Grenzwerte ($IQ \geq 130/125/120$). Dabei griffen wir beim CFT 20-R auf die Klassenstufennormen aller Schularten (Weiß, 2006) zurück. Bei den SPM nutzten wir die Altersnormen (vgl. Heller et al., 1998). Beim KFT-N verwendeten wir die Normen der 9. Klassenstufe aller Schularten (Heller & Perleth, 2000). Zusätzlich zur Hochbegabungsklassifizierung anhand der testnormspezifischen Standardwerte (IQ) klassifizierten wir die Teilnehmenden anhand der Kennwerteverteilung der vorliegenden

⁶ Nicht bearbeitete Subtests führten aufgrund des geringen Anteils (CFT: 2,5 %, SPM: 4,1 %, KFT-N: 5,7 %) zum fallweisen Ausschluss von den Analysen. Nicht bearbeitete Einzelitems wurden – der üblichen Auswertungspraxis von Intelligenztests folgend – mit keinem Gutpunkt bewertet.

Stichprobe (stichprobenspezifische Grenzwerte). Dafür klassifizierten wir anhand des an unserer Stichprobe ermittelten Prozentrangs (PR) der testspezifischen Rohwertsummenverteilung die Teilnehmenden pro Test als „hochbegabt“ ($\geq cut-off$) oder „nicht hochbegabt“ ($< cut-off$). In Anlehnung an häufig verwendete Grenzmarken intellektueller Hochbegabung wählten wir drei verschiedenen hohe *cut-off*-Werte auf der Prozentrangskala, die den erwähnten IQ-Grenzmarken entsprechen ($PR \geq 98$ bzw. $PR \geq 95$ bzw. $PR \geq 90$).

Zur Beantwortung der dritten Fragestellung bestimmten wir (a) die klassifikatorische Übereinstimmung einer Hochbegabungsdiagnose mit zwei Testverfahren. Dabei variierten wir die Vergleichsgruppe (testnormspezifische Standardisierung, stichprobenspezifische Grenzwerte), die Grenzwerte ($IQ \geq 130/125/120$ bzw. $PR \geq 98/95/90$) sowie die Testreihenfolge bzw. Betrachtungsrichtung (einerseits Anteil an Hochbegabten in Test A, die auch in Test B hochbegabt sind; andererseits Anteil an Hochbegabten in Test B, die auch in Test A hochbegabt sind). Zusätzlich schätzten wir den erwarteten Anteil als übereinstimmend hochbegabt Klassifizierter über die Verteilungsfunktion der bivariaten Normalverteilung unter Berücksichtigung der Testinterkorrelation (vgl. zur Formel z. B. Hanses, 2009, S. 98). Systematisch verglichen wir die prozentualen Anteile Hochbegabter miteinander – jeweils unter dem Aspekt der unterschiedlichen Vergleichsgruppen, der Grenzwerte und Testreihenfolgen bzw. Betrachtungsrichtungen sowie im Vergleich zu den berechneten erwarteten Hochbegabtenanteilen. Zur Einschätzung der Bedeutsamkeit der Anteilsdifferenzen orientierten wir uns an den von Cohen (1988) vorgeschlagenen Empfehlungen für die Effektstärke h ($h \geq 0.2/.5/.8$ entspricht einem kleinen/mittleren/großen Effekt).

Zusätzlich gingen wir bei Fragestellung 3 (b) der Klassifikationsgüte zweier Tests in der Hochbegabungsdiagnostik mittels RATZ-Index nach (vgl. Beißert et al., 2014). Wir ermittelten den RATZ-Index, indem wir jeweils paarweise zwei der drei Testverfahren betrachteten. Die Berechnung der RATZ-Indices sowie die Interpretation der Güte der Hochbegabtenklassifizierung orientierte sich an dem beschriebenen Vorgehen (vgl. Marx et al., 2000; Marx & Lenhard, 2010); dabei berücksichtigen wir für die Hochbegabungsklassifikation ebenfalls die erwähnten Grenzwerte und Testreihenfolgen bzw. Betrachtungsrichtungen für die beiden Vergleichsgruppen.

3.5.3 Ergebnisse

Erwartungsgemäß lagen die IQ-Mittelwerte ($103 \leq IQ \leq 112$) in den drei Verfahren für die vorliegende Gymnasialstichprobe jeweils numerisch oberhalb des Populationsmittelwerts und die Standardabweichungen ($12,1 \leq SD \leq 13$) numerisch unterhalb der Populationsstandardabweichung (s. Tabelle 11, S. 62). In der vorliegenden Stichprobe erreichten 1,2 % der Schülerinnen und Schüler im SPM mit 59 Rohwertpunkten die Testdecke der theoretisch maximal möglichen Rohwertpunktsumme. Ähnlich viele Schülerinnen und Schüler (1,4 %) erzielten mit 50 Rohwertpunkten die Testdecke des KFT-N. Hingegen erreichte im CFT 20-R kein Proband die maximale Rohwertsumme von 56. Die zur Beantwortung von Fragestellung 1 nach der korrelativen Übereinstimmung berechneten Korrelationskoeffizienten waren – Cohen (1988) folgend – von mittlerer bis großer Effektgröße: $r_{\text{CFT 20-R/KFT-N}} = .43$, $r_{\text{CFT 20-R/SPM}} = .35$, $r_{\text{SPM/KFT-N}} = .52$.

Tabelle 11

Prozentualer Anteil Hochbegabter bei Verwendung unterschiedlicher Hochbegabungs-Grenzmarken ($IQ \geq 130/125/120$) in den drei Testverfahren, ergänzt um Stichproben- und Testkennwerte.

	IQ \geq 130	IQ \geq 125	IQ \geq 120	<i>M</i>	<i>SD</i>	<i>r_{tt}</i>
KFT-N	4.5	16.2	18.5	112.6	12.4	.86
CFT 20-R	5.3	14.2	22	109.5	13.0	.76
SPM	2.6	7.2	7.2	103.2	12,1	.80

Anmerkung. Häufigkeitsangaben in Prozent. *r_{tt}*: split-half-Reliabilität; verwendete Testnormen: KFT-N (Jahrgangsstufe 9 aller Schularten; Heller & Perleth, 2000); CFT 20-R (Jahrgangsstufe 9 aller Schularten; Weiß, 2006); SPM (Altersnormen; Heller et al., 1998).

Bezogen auf die zweite Fragestellung nach den prozentualen Anteilen Hochbegabter ergaben sich bei Verwendung der testnormspezifischen Standardwerte und der Grenzwertsetzung „IQ \geq 130“ im CFT (5,3 %) und KFT-N (4,5 %) ähnliche und numerisch höhere Anteile als im SPM (2,6 %). Bei zunehmender Liberalisierung der Grenzmarke stieg erwartungsgemäß der Hochbegabtenanteil im CFT (5,3 %/14,2 %/22,0 %) und KFT-N (4,5 %/16,2 %/18,5 %), hingegen im SPM nur zum Teil (2,6 %/7,2 %/7,2 %). Im SPM zeigte sich auch bei den Grenzmarken „IQ \geq 125“ und „IQ \geq 120“ ein geringerer Anteil Hochbegabter als im KFT und CFT.

Fragestellung 3 (a) zielte auf die klassifikatorische Übereinstimmung Hochbegabter mit verschiedenen Tests unter Berücksichtigung (1) verschiedener Grenzwerte und (2) Testreihenfolgen bzw. Betrachtungsrichtungen unter Verwendung der stichprobenspezifischen Grenzwerte (mittlerer Teil von Tabelle 12, S. 63) sowie der testnormspezifischen Standardwerte (rechter Teil der Tabelle 12), ergänzt um Vergleiche mit den statistisch erwarteten Anteilen (linker Teil von Tabelle 12).

Tabelle 12

Beobachteter und erwarteter Anteil (%) als übereinstimmend hochbegabt Klassifizierter unter Variation von Testreihenfolge (Betrachtungsrichtung), Grenzwertziehung ($IQ \geq 130/125/120$ bzw. $PR = 98/95/90$) und Standardisierungsgruppe (Testnormspezifischer Standardwert vs. stichprobenspezifischer Grenzwert).

	Erwartet			Stichprobenspezifischer Standardwert			Testnormspezifischer Standardwert		
	KFT-N	CFT	SPM	KFT-N	CFT	SPM	KFT-N	CFT	SPM
$IQ \geq 130$									
<i>bzw.</i>									
$PR = 98$									
KFT-N		.17	.19		.20	.27		.12	.13
CFT			.11	.20		.24	.35		.14
SPM				.25	.25		.57	.29	
$IQ \geq 125$									
<i>bzw.</i>									
$PR = 95$									
KFT-N	.23		.22		.18	.32		.30	.13
CFT		.23	.16	.16		.17	.30		.09
SPM				.33	.21		.60	.33	
$IQ \geq 120$									
<i>bzw.</i>									
$PR = 90$									
KFT-N	.31		.33		.42	.54		.45	.21
CFT		.31	.24	.24		.40	.43		.22
SPM				.32	.41		.54	.55	

Anmerkung. In den Zellen stehen die Anteile (%) der Hochbegabten im ersten Test (Spalte 1), die im 2. Test (Zeile 2) ebenfalls als hochbegabt klassifiziert wurden.

Ablesebeispiel: Bei einem *cut-off*-Wert von $IQ \geq 130$ betrug der a priori erwartete Anteil der im KFT-N Hochbegabten, die im SPM ebenfalls als hochbegabt klassifiziert wurden (übereinstimmend als hochbegabt Klassifizierte) 19 %. 13 % derjenigen, die im KFT-N einen $IQ \geq 130$ erreichten, erzielten auch im SPM einen $IQ \geq 130$. In umgekehrter Betrachtungsrichtung erzielten 57 % derjenigen, die im SPM einen $IQ \geq 130$ erreichten, auch im KFT-N einen $IQ \geq 130$.

Mit Fokus auf die Grenzwerte (1) zeigte sich für die klassifikatorische Hochbegabungs-Übereinstimmung in zwei Tests, dass sich die Übereinstimmungs-Anteile beim Vergleich der Grenzmarke $IQ \geq 130$ auf $IQ \geq 125$ bzw. $PR > 98$ auf $PR > 95$ in beiden Standardisierungsgruppen nicht bedeutsam ($|h| \leq .20$) änderten (Ausnahme: testnormspezifischer Standardwert – von im KFT-N mit $IQ \geq 130$ Hochbegabten waren 12 % auch im CFT mit $IQ \geq 130$ hochbegabt, während von den im KFT-N mit $IQ \geq 125$ Hochbegabten 30 % auch im CFT mit $IQ \geq 125$ hochbegabt waren; $h = .45$). Mit der Liberalisierung der Hochbegabungs-Grenzmarke von $IQ \geq 125$ auf $IQ \geq 120$ bzw. $PR > 95$ auf $PR > 90$ erhöhte sich jedoch in beiden Standardisierungsgruppen in der Regel der Anteil der auch im „zweiten“ Test Hochbegabten ($.20 \leq h \leq .53$; zwei Ausnahmen: stichprobenspezifischer Grenzwert – von im SPM mit $PR > 95$ Hochbegabten waren 33 % auch im KFT-N mit $PR > 95$ hochbegabt, mit 32 % anteilig vergleichbar viele der im SPM mit $PR > 90$ Hochbegabten waren auch im KFT-N mit $PR > 90$ hochbegabt, $h = .02$; testnormspezifischer Standardwert – von im SPM mit $IQ \geq 125$ Hochbegabten waren 60 % auch im KFT-N mit $IQ \geq 125$ hochbegabt, mit 54 % waren anteilig vergleichbar viele der im SPM mit $IQ \geq 120$ Hochbegabten auch im KFT-N mit $IQ \geq 120$ hochbegabt, $h = -.12$).

Bezogen auf die klassifikatorische Hochbegabungs-Übereinstimmung in zwei Tests mit Fokus auf der Testreihenfolge bzw. Betrachtungsrichtung (2) ergaben sich unter Verwendung der stichprobenspezifischen Grenzwerte – abgesehen von zwei Ausnahmen – durchgängig vernachlässigbare Anteilsdifferenzen ($|h| \leq .20$): Während 24 % von denjenigen, die im CFT einen Wert von $PR > 90$ erzielten, auch im KFT-N einen Wert von $PR > 90$ erreichten, erzielten umgekehrt 42 % von denjenigen, die im KFT-N einen Wert von $PR > 90$ erreichten, im CFT ebenfalls einen Wert von $PR > 90$ ($h = .39$). Weiterhin erzielten 32 % von denjenigen, die im SPM einen Wert von $PR > 90$ erreichten, auch im KFT-N einen Wert von $PR > 90$; umgekehrt erzielten 54 % von denjenigen, die im KFT-N einen Wert von $PR > 90$ erreichten, im SPM ebenfalls einen Wert von $PR > 90$ ($h = .45$).

Bei Verwendung der testnormspezifischen Standardwerte ergaben sich die folgenden Ergebnisse: Bei Grenzmarke $IQ \geq 130/125/120$ erreichten 57 %/60 %/54 % derjenigen, die im SPM über der entsprechenden Grenzmarke lagen, im KFT-N ebenfalls ein Testergebnis im korrespondierenden Grenzwertbereich. Bei umgekehrter Betrachtung erreichten bei Grenzmarke $IQ \geq 130/125/120$ nur 13 %/13 %/21 % der KFT-N-Hochbegabten im SPM

ebenfalls ein Ergebnis im korrespondierenden Begabungsbereich ($h = .97/1.03/.70$). Bei Grenzmarke $IQ \geq 130/125/120$ erreichten 29 %/33 %/55 % derjenigen, die im SPM als hochbegabt klassifiziert wurden, im CFT ebenfalls ein Testergebnis im korrespondierenden Bereich; wurde der CFT vorangehend zur Klassifizierung herangezogen, erzielten im SPM bei Grenzmarken $IQ \geq 130/125/120$ lediglich 14 %/9 %/22 % Werte im korrespondierenden Begabungsbereich ($h = .37/.61/.69$). Außerdem erzielten 35 % von denjenigen, die im CFT bei der strengsten Grenzmarke ($IQ \geq 130$) ein Hochbegabungsergebnis erreichten, auch im KFT-N einen entsprechend hohen Wert; umgekehrt erzielten 12 % von denjenigen, die im KFT-N einen Wert von $IQ \geq 130$ erreichten, im CFT ebenfalls einen Wert von $IQ \geq 130$ ($h = .56$).

Einen Überblick über die erwarteten – also anhand der bivariaten Verteilungsfunktion unter Berücksichtigung der jeweiligen Testinterkorrelation errechneten – Anteile Hochbegabter liefert Tabelle 12 (linker Teil). Auch diese Anteile fielen numerisch eher gering aus und lagen teilweise oberhalb und teilweise unterhalb der empirischen Anteile der drei Testverfahren (Tabelle 12, mittlerer und rechter Teil).

Tabelle 13

Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote (RATZ-Index) für die eingesetzten Testverfahren unter Variation der Grenzmarke und Standardisierungsgruppe.

		SPM	CFT 20-R
$IQ \geq 130/PR > 98$	CFT 20-R	25/37	
	KFT-N	50/26	24/29
$IQ \geq 125/PR > 95$	CFT 20-R	25/20	
	KFT-N	50/25	19/19
$IQ \geq 120/PR > 90$	CFT 20-R	51/28	
	KFT-N	45/33	29/30

Anmerkung. Der vordere Kennwert der Zellen beschreibt den RATZ-Index der Klassifizierungsübereinstimmung, der mittels der testnormspezifischen Standardwerte ($IQ \geq 130/125/120$) gewonnen wurde. Der hintere RATZ-Indexwert unterliegt der Klassifizierung anhand der stichprobenspezifischen Grenzwerte ($PR > 98/95/90$).

Bezogen auf die Ergebnisse zum RATZ-Index (Fragestellung 3b) differierten die Werte in Abhängigkeit von der Standardisierungsgruppe: Wurden testnormspezifische Standardwerte herangezogen, lagen die RATZ-Indizes für die Testkombination SPM/KFT-N jeweils im „guten“ Bereich ($\text{RATZ} \geq 45$; vgl. Tabelle 13, S. 65). Wurde die Klassifizierung anhand der stichprobenspezifischen Grenzwerte vorgenommen, zeigte sich ausschließlich für die Testkombination SPM/CFT bei $\text{PR} > 98$ eine „gute“ Übereinstimmung ($\text{RATZ} = 37$). Alle weiteren RATZ-Indices lagen unterhalb der vorgeschlagenen Grenze von „ $\text{RATZ} = 34$ “.

3.5.4 Diskussion

Die vorliegende Untersuchung zielte auf die Übereinstimmung der Identifikation intellektuell Hochbegabter beim Einsatz dreier konzeptionell ähnlicher Intelligenztestverfahren. Die Testinterkorrelationen lagen erwartungsgemäß in mindestens mittlerer Höhe (Fragestellung 1). Im Vergleich zu den Normstichproben zeigte sich mit Ausnahme des SPM ein erwartungsgemäß höherer Anteil Hochbegabter für die vorliegende gymnasiale Stichprobe (Fragestellung 2). Die zentrale Frage nach der Übereinstimmungsgüte der Hochbegabtenidentifikation mit zwei konzeptionell ähnlichen Testverfahren (Fragestellung 3) ergab Folgendes: Erstens erhöhte sich mit wenigen Ausnahmen bei Liberalisierung der Grenzmarke in beiden Standardisierungsgruppen die Klassifizierungsübereinstimmung von der Grenzmarke $\text{IQ} \geq 125/\text{PR} > 95$ hin zu $\text{IQ} \geq 120/\text{PR} > 90$ bedeutsam. Zweitens war die Testreihenfolge bzw. Betrachtungsrichtung in Abhängigkeit von der Testnormstrenge (insbesondere des SPM) bedeutsam für die Hochbegabungsübereinstimmung. Die Betrachtung der RATZ-Indices ergab, dass die klassifikatorische Güte unter Verwendung der Testnormen lediglich für die Testkombination von SPM und KFT-N über alle Grenzmarken als „gut“ bezeichnet werden kann, unter Verwendung der stichprobenspezifischen Grenzwerte nur für die Kombination von SPM und CFT (ausschließlich bei Grenzmarke $\text{PR} > 98$).

Bezogen auf die erste Fragestellung nach den korrelativen Übereinstimmungen der drei Testverfahren resultierten erwartungsgemäß Interkorrelationen in mindestens mittlerer Höhe. Numerisch korrelierten SPM und KFT-N am höchsten ($r = .52$), CFT und KFT-N etwas niedriger ($r = .43$) sowie SPM und CFT numerisch noch niedriger ($r = .35$). Bei der Interpretation der Höhe unserer Korrelationskoeffizienten ist zu beachten, dass die Standardabweichungen in der betrachteten Gymnasialstichprobe (mit für die Beantwortung

der weiteren Fragestellungen sinnvollerweise höheren Hochbegabtenanteilen als in der Population) etwas niedriger als Populationsstreuungen ausgefallen sind. Eine statistische Aufwertung der von uns ermittelten Korrelationskoeffizienten um diese Varianzeinschränkungen hätte entsprechend einen Korrelationsanstieg zur Folge. Insgesamt gesehen korrespondieren unsere Ergebnisse mit den im Theorieteil berichteten Befunden vorangehender Forschung bzw. liegen im unteren Bereich.

Die im Rahmen der zweiten Fragestellung betrachteten absoluten Hochbegabtenanteile fielen mit Ausnahme der SPM in der vorliegenden gymnasialen Stichprobe erwartungsgemäß insgesamt etwas höher als in heterogeneren Stichproben aus und stiegen erwartungsgemäß (nahezu durchgängig) mit einer Liberalisierung der Grenzmarke an. Ausmaß und Anstieg der Hochbegabtenanteile im CFT und KFT-N fielen im Gegensatz zum (strengerem) SPM ähnlich hoch aus. Die Hochbegabtenanteile lagen im SPM vergleichsweise niedriger. Die SPM-Normierung aus dem Jahr 1996/1997 (Heller et al., 1998) scheint somit den Anteil Hochbegabter für die Altersstufe 14 bis 15-Jähriger eher zu unterschätzen, strenger in Bezug auf Hochbegabung zu testen sowie schlechter als die des KFT-N und des CFT 20-R im hohen Begabungsbereich zu differenzieren.

Der zentralen Frage, wie übereinstimmend die Identifikation intellektuell Hochbegabter mit konzeptionell ähnlichen Testverfahren gelingt, gingen wir im Rahmen der dritten Fragestellung nach. Bezogen auf die Grenzwertziehung veränderte sich die Hochbegabungsübereinstimmung beim Einsatz eines weiteren konzeptionell ähnlichen Tests mit sehr wenigen Ausnahmen sowohl unter Verwendung der testnormspezifischen Standardwerte als auch der stichprobenspezifischen Grenzwerte von Grenzmarke $IQ \geq 130/PR > 98$ zu $IQ \geq 125/PR > 95$ nicht bedeutsam. Hingegen ließ sich mit Liberalisierung der Grenzmarke von $IQ \geq 125/PR > 95$ auf $IQ \geq 120/PR > 90$ ebenfalls in beiden Standardisierungsgruppen – mit wenigen Ausnahmen – ein bedeutsamer Anstieg der als übereinstimmend hochbegabt Klassifizierten finden. Im Rahmen des Marburger Hochbegabten Projekts wurden im Zuge der Re-Testung (3. Klassenstufe vs. 9. Klassenstufe) 50 % als Hochbegabt re-identifiziert ($IQ \geq 130$), jedoch 71 % bei Liberalisierung der Grenzmarke ($IQ \geq 125$) in der 9. Klassenstufe (vgl. Hanses, 2009). Wir betrachteten nicht einen Vergleich über ein längeres Zeitintervall (dort: 6 Jahre), sondern verschiedene, konzeptuell ähnliche Testverfahren zu einem Zeitpunkt. Nichtsdestotrotz fanden wir ähnlich

hohe Übereinstimmungsraten in der vorliegenden Untersuchung unter Verwendung der testspezifischen Standardwerte für die Testkombination SPM/KFT-N (über alle Grenzmarken) und SPM/CFT (bei Grenzwert $IQ \geq 120$). Unter Verwendung der stichprobenspezifischen Grenzwerte erreichten in der vorliegenden Untersuchung nur KFT-N/SPM (bei $PR > 90$) entsprechend hohe Werte.

Bezüglich der Testreihenfolge bzw. Betrachtungsrichtung war ein Reihenfolgeeffekt – mit wenigen Ausnahmen – vorrangig unter Verwendung der testnormspezifischen Standardwerte zu finden. Insbesondere für den SPM ließen sich über alle Grenzmarken hinweg mehr Hochbegabte sowohl im KFT-N als auch im CFT re-identifizieren, wenn die entsprechenden Schülerinnen und Schüler vorher die korrespondierenden Grenzmarken im SPM überschritten hatten. Umgekehrt betrachtet, fanden sich deutlich weniger re-identifizierte Hochbegabte im SPM, wenn sie zuvor die korrespondierende Grenzmarke in KFT-N oder CFT überschritten hatten. Bei der Interpretation ist zu beachten, dass weniger Personen im SPM als hochbegabt klassifiziert wurden (s. Fragestellung 2); daher überrascht es kaum, dass diese wenigen im SPM Hochbegabten auch in einem zweiten Test mit höherer Wahrscheinlichkeit über der entsprechenden Grenzmarke lagen. Der Einfluss der Testreihenfolge bzw.

Betrachtungsrichtung auf die Re-Identifikationsgüte von Hochbegabung unter Verwendung unterschiedlich „strenger“ Normen deckt sich mit den Befunden von Lohman et al. (2008). Die in der vorliegenden Untersuchung gefundenen Anteile in zwei Testverfahren als übereinstimmend hochbegabt Identifizierter entsprachen mit wenigen Ausnahmen den (unter Berücksichtigung von Testinterkorrelation und Verteilungsfunktion) erwarteten Anteilen.

Die Ergebnisse der RATZ-Berechnung verweisen darauf, dass eine (in Anlehnung an Marx und Lenhard, 2010) als „gut“ zu bezeichnende, übereinstimmende „hochbegabt“-Klassifikation fast ausschließlich mittels SPM und KFT-N über alle Grenzmarken hinweg gelang (Ausnahmekombination SPM mit CFT bei Grenzmarke $PR > 98$). Die Übereinstimmung gelang insgesamt besser, wenn die Klassifizierung anhand der testnormspezifischen Standardwerte vorgenommen wurde. Beachtenswert erscheint zudem, dass ein deutlicher Reihenfolgeeffekt insbesondere unter Verwendung der testspezifischen Normwerte (vor allem für den eher „strenger“ SPM) gefunden wurde.

Doch unsere Ergebnisse beschränken sich auf den Altersbereich der ca. 14- und 15-Jährigen. Zwar erschwert zudem unsere gymnasiale Stichprobe eine Verallgemeinerung auf

z. B. hinsichtlich Bildungsgrad und Alter heterogenere Gruppen, ermöglicht jedoch die Betrachtung einer vergleichsweise größeren Gruppe von Hochbegabten. Die eingesetzten Tests erfassen einen vergleichbaren Intelligenzfaktor (*figurales reasoning*). Dieser Intelligenzfaktor bildet nicht das gesamte Intelligenztestspektrum oder einen Intelligenzgeneralfaktor im Sinne eines „guten g“ (Jensen & Weng, 1994) ab; diese Aufgabenarten laden jedoch hoch auf dem Intelligenz-Generalfaktor und ermöglichen somit eine gute Abschätzung der allgemeinen Intelligenz (vgl. bspw. Jensen, 1998). Auch wenn wir daher eine diesbezügliche Generalisierung unserer Befunde annehmen, erscheint im Rahmen künftiger Forschung die Betrachtung weiterer – konzeptionell ähnlicher sowie unähnlicher – Testverfahren lohnenswert. Ebenso könnte eine Berücksichtigung anderer Altersgruppen und größerer Stichproben die Befunde im Hinblick auf eine Generalisierbarkeit weiten. Selbstverständlich sind bei der Interpretation diagnostischer Ergebnisse die jeweiligen Konfidenzintervalle zu berücksichtigen. Entsprechend könnte man ergänzend bei unseren Befunden ähnlich wie Floyd et al. (2008) Konfidenzintervalle bestimmen. Da für die Berechnung der Übereinstimmung weitere Entscheidungen (z. B. über die zur Berechnung heranzuziehende Reliabilitätsschätzung, das spezifisch zu verwendende Konfidenzintervall) nötig wären, die eine Ergebnisinterpretation erschwert hätten, verzichteten wir darauf.

Zusammenfassend fällt auf, dass eine übereinstimmende Diagnose als „hochbegabt“ sogar mit zwei konzeptionell ähnlichen Testverfahren nur eher mäßig gelingt (mit Ausnahme der Kombination SPM/KFT-N). Dies ist theoretisch bedeutsam, da damit die relativ enge Verbindung des Konstrukts „Hochbegabung“ zum spezifisch eingesetzten Messinstrument bzw. Intelligenztest angesprochen ist (vgl. die bekannte Intelligenzdefinition von Cohen, 1923). Die Übereinstimmungsgüte scheint stark von der Normierungsstichprobe bzw. Normierungsstrenge (vgl. Lohman et al., 2008) abzuhängen. Bei der Testauswahl sollte demnach neben einer sorgfältigen Auswahl hinsichtlich eines geeigneten Testkonzepts, überzeugenden Hinweisen zu den Gütekriterien, einer Vermeidung von Deckeneffekten vor allem auch auf eine ausreichend hohe Qualität der Normen geachtet werden. Besonders relevant erscheint uns dies im Hinblick auf die Anwendungspraxis, wenn Entscheidungen (bspw. zur Aufnahme in Begabtenförderprogramme) allein auf einem Intelligenztest fußen – samt der unseren Ergebnissen zufolge häufig eher mäßigen Generalisierbarkeit auf das Diagnoseergebnis in einem weiteren Intelligenztest. Zwar können zusätzlich zu Intelligenztests eingesetzte weitere Instrumente (bspw. Verhaltensbeobachtungen,

Arbeitsproben, Zeugnisse) eine sinnvolle Ergänzung darstellen, um eine Passung zum entsprechenden Förderprogramm besser abschätzen zu können; allerdings ergeben sich daraus weitere Fragen wie die nach der (häufig geringeren) psychometrischen Qualität dieser weiteren Verfahren sowie der Kombinationsregel zur Befundintegration. Als Fazit kann festgehalten werden, dass einerseits die mit verschiedenen Intelligenztests erfasste Intelligenz zwar manifest mindestens mittelhoch korreliert. Doch zeigen die Ergebnisse darüber hinaus, dass eine übereinstimmende Hochbegabungsdiagnose in zwei konzeptionell ähnlichen Testverfahren nicht zwangsläufig ausreichend gut gelingt.

4 Gesamtdiskussion

Die drei vorliegenden Studien dieser Dissertation zielen darauf ab, die Konsistenz der Messung von allgemeiner und spezifischer Intelligenz genauer zu untersuchen sowie deren Relevanz in Bezug auf allgemeine und spezifische Schulleistung aufzuzeigen. Vor allem die Betrachtung der inkrementellen Validität spezifischer Intelligenzfaktoren in diesem Beziehungsgeflecht gliedert sich ein in eine hochaktuelle und kontroverse Debatte der psychometrischen Intelligenzforschung über die Prädiktionskraft allgemeiner und spezifischer Intelligenzfaktoren (vgl. Kell & Lang 2019). Darüber hinaus wird die klassifikatorische Übereinstimmungsgüte beim praktischen Einsatz von mehreren unabhängigen Intelligenztests zur Hochbegabtenidentifikation genauer beleuchtet.

Im folgenden Kapitel werden die wesentlichen Befunde dieses Dissertationsprojektes zusammengefasst und deren Implikationen für Forschung und Praxis diskutiert. Eine kritische Einordnung der Befunde, sowie der Studiendesigns, wird erörtert.

4.1 Zusammenfassung der wesentlichen Befunde

Die ersten beiden Studien belegen mit zum Teil sehr hohen Korrelationen die Konsistenz der Messung und somit Generalität von g . Darüber hinaus kann jedoch auch den spezifischen Intelligenzfaktoren neben oder zusätzlich zu g mit substanziellen, konvergent/divergenten Beziehungen sowohl über Testbatterien hinweg, als auch in Bezug zu spezifischen Schulleistungsfaktoren inkrementelle Validität (d.h. jede statistisch bedeutsame Korrelation verweist auf ein Inkrement) und Relevanz zugesprochen werden. Vor allem der Nested-Factor-Ansatz bietet sich an, dieses Beziehungsgeflecht distinkt zu erforschen. Obgleich Intelligenztests (sehr) stabil messen (latent modelliert und messfehlerbereinigt), zeigte sich, dass die Klassifikation von Hochbegabung (manifest erfasst) mit zwei konzeptionell ähnlichen Testverfahren nicht zwangsläufig treffsicher gelingt.

Konkret konnten wir in Studie 1 dreierlei zeigen:

- 1) Die von vielen Forschenden angenommene und empirisch gut belegte Intelligenzstruktur mit Generalfaktor und zusätzlichen spezifischen Intelligenzfaktoren (vgl. Carroll, 1993; Jensen, 1998; McGrew, 2009; Spearman, 1904) ließ sich mit

unterschiedlichen konfirmatorisch faktorenanalytischen Modellvarianten erneut belegen. Die statistische Überlegenheit des NF-Modells gegenüber dem HO-Modell (vgl. Murray & Johnson, 2013; Brunner et al. 2012) zeigte sich wie in vielen Vorgängeruntersuchungen auch in dieser Studie. Mit Hinblick auf den gelegentlich formulierten Verdacht auf Modellmissspezifikation von NF-Modellen trotz guten Modellfits (vgl. Morgan et al., 2015; Maydeu-Olivares & Coffman, 2006), zeigten die nachträglich durchgeführten Zusatzanalysen anhand von Eid et al. (2018) vorgeschlagenen Bifactor (S-1) Modellen, dass die o.g. Kritikpunkte für die Befunde dieser Dissertation wenig bedeutsam sind. Denn das zusätzlich analysierte, kombinierte Bifactor (S-1) Modell fittete ähnlich gut wie das kombinierte Original NF-Modell und das Befundmuster blieb mit wenigen Ausnahmen weitestgehend stabil. Die zusätzlichen Analysen können dahingehend interpretiert werden, dass das kombinierte Original NF-Modell aus Valerius & Sparfeldt (2014) eine gute Anpassung an die Daten sowie ein stabiles Befundmuster liefert.

- 2) In Übereinstimmung mit bisheriger Forschung (vgl. Johnson et al., 2004, 2008; Floyd et al., 2013) belegten wir mit (sehr) hohen g -Faktorkorrelationen über unabhängige Testbatterien hinweg die (weitestgehend) testunabhängige Konsistenz von g im Sinne von Spearman's *Indifference of the Indicator* (Spearman, 1927, S. 197), die besagt, dass fast alle Intelligenztestaufgaben trotz unterschiedlicher Erscheinungsform, in unterschiedlich starker Ausprägung die allgemeine Intelligenz im Sinne des Generalfaktors g messen. Darüber hinaus belegten die Befunde die Unabhängigkeit des Generalfaktors von der faktorenanalytischen Extraktionsmethode und somit die Generalität von g .
- 3) Darüber hinaus konnten wir batterieübergreifend substantielle konvergente und divergente Beziehungen der spezifisch verbalen und numerischen Intelligenzfaktoren nachweisen. Die Ergebnisse ergänzen die Befunde vorheriger HO-Modell orientierter Forschung (vgl. Floyd et al., 2013) dahingehend, dass unter Verwendung des NF-Modells zumindest der spezifisch verbale Faktor batterieübergreifend eine ähnlich hohe Korrelation aufwies ($r = .83$; vgl. Tabelle 5, S. 30) wie die batterieübergreifenden „breiten“ Stratum II Faktoren bei Floyd et al. (2013: $.81 \leq r \leq .98$ Comprehension Knowledge; $.84 \leq r \leq .88$ Short Term Memory; $.90 \leq r \leq .97$ Processing Speed; $r = .94$ Fluid Reasoning). Die figuralen Faktoren wiesen (mit einer

Ausnahme für V_{CogaT} mit $F_{\text{BIS}} r = .47$; vgl. Tabelle 5, S. 30) keine bedeutsamen Beziehungen zu den verbalen und numerischen Faktoren der anderen Testbatterie auf. In der nachträglichen Zusatzanalyse mit Bifactor (S-1) Modell (vgl. Eid et al., 2018) wurde u.a. aus diesem Grund der spezifisch figurale Faktor (bzw. die Subtests die darauf luden) für die jeweilige Testbatterie als Referenzfacette für g genutzt. Wie bereits angedeutet änderten sich die Korrelationen im Bifactor (S-1) Modell nur geringfügig in ihrer absoluten Höhe, fielen jedoch im Vergleich zum kombinierten Original NF-Modell tendenziell etwas niedriger aus. Die Zusatzanalysen mit dem Bifactor (S-1) Modell können somit dahingehend interpretiert werden, dass a) die gelegentlich genannten Kritikpunkte am NF-Modell (vgl. z.B. Eid et al., 2018; Koch et al., 2018; Maydeu-Olivares & Coffman, 2006) in der vorliegenden Studie wenig relevant sind und b) die Invarianz der Befunde gegenüber der faktorenanalytischen Modelliervariante nachgewiesen ist und c) zeigen, dass das Original NF-Modell (mit überwiegend [etwas] höheren Korrelationen und [leicht] besserem Modellfit) eine vergleichsweise gute Anpassung an die Daten liefert.

In Studie 2 konnten wir ergänzend zu den Befunden von Valerius & Sparfeldt (2014) am gleichen Datensatz zeigen, dass im kombinierten NF-Modell der Intelligenz und Schulleistung in Übereinstimmung mit vielen empirischen Befunden (vgl. Gottfredson, 1997; Jensen, 1998; Roth et al., 2015) der allgemeine Intelligenzfaktor mit dem allgemeinen Schulleistungsfaktor bedeutsam zusammenhing. Über diese Zusammenhänge der allgemeinen Intelligenz- und Schulleistungsfaktoren hinaus zeichnete sich durch die statistisch bedeutsamen Korrelationen der spezifisch verbalen und numerischen Intelligenzfaktoren zu den sprachlichen und mathematisch/naturwissenschaftlichen Schulleistungsfaktoren zusätzliche inkrementelle Validität ab. Dieses Beziehungsmuster, sowohl konvergenter als auch divergenter Zusammenhänge, der jeweils korrespondierenden verbalen/numerischen, spezifischen Faktoren mit sprachlicher und mathematisch/naturwissenschaftlicher Schulleistung (Zensuren) konnte kurz darauf (vgl. Schult & Sparfeldt, 2016) sowie aktuell (vgl. Coyle, 2018b) an einem größeren US-amerikanischen Datensatz (NLSY) repliziert werden. Der figurale Intelligenzfaktor korrelierte bei Valerius & Sparfeldt (2015) numerisch etwas niedriger, jedoch ebenfalls bedeutsam sowohl mit der sprachlichen als auch der mathematisch/naturwissenschaftlichen Schulleistung. Dies war mitunter ausschlaggebend

dafür, den figuralen Faktor im nachträglich spezifizierten kombinierten Bifactor (S-1) Modell zum Zweck der Prüfung von potenziell möglicher Modellmisspezifikation in Anlehnung an die Empfehlung von Eid et. al. (2018) als Referenzfacette für g zu wählen. Wie bereits in den Zusatzanalysen mit dem Bifactor (S-1) Modell für die erste Studie, zeigte sich auch hier, dass sich der Modellfit nur minimal änderte und das Befundmuster weitestgehend stabil blieb. Die Ergebnisse der nachträglich durchgeführten Zusatzanalysen zu Studie 2 mit dem Bifactor (S-1) Modell können wiederholt dahingehend interpretiert werden, dass das Original NF-Modell aus Valerius & Sparfeldt (2015) eine angemessene Passung an die Daten liefert und sich die Befunde invariant und somit stabil gegenüber der zugrundeliegenden Modellstruktur (im NF-Rahmen) zeigen.

In Studie 3 gingen wir der zentralen Frage nach, wie gut die Hochbegabungsidentifikation mithilfe dreier konzeptionell ähnlicher Intelligenztests gelingt. Es zeigte sich, dass bei wie erwartet mittlerer Testinterkorrelation, die Übereinstimmungsgüte in Abhängigkeit der Testreihenfolge bzw. Betrachtungsrichtung sowie der Testnormstrenge in Einklang mit den Befunden von Lohman et al. (2008) bedeutsam variierte. Darüber hinaus wurde deutlich, dass die klassifikatorische Güte, bestimmt mittels RAZ-Index, über alle Grenzmarken hinweg nur für die Testkombination SPM und KFT-N als „gut“ bezeichnet werden kann. Die Befunde der dritten Studie zeigen zusammenfassend, dass trotz mittlerer Testinterkorrelation die übereinstimmende Diagnose „hochbegabt“ mit zwei konzeptionell ähnlichen Intelligenztests nicht zwangsläufig treffsicher gelingt. Im Hinblick auf eine testimmanente Hochbegabungsdefinition (vgl. Preckel, 2010) erscheint es durchaus kritisch zu betrachten zu sein.

4.2 Implikationen

Im folgenden Abschnitt werden die Implikationen für Forschung und Praxis herausgearbeitet, die aus den Ergebnissen dieser Dissertation abgeleitet werden können. Dabei werden Aspekte des Modellvergleichs von NF- und HO-Modellen, die Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit relevanten Außenkriterien im Bildungskontext, sowie Aspekte der Hochbegabungsidentifikation mit verschiedenen Intelligenztests hervorgehoben.

4.2.1 Implikationen für die psychometrische Intelligenzforschung zum Modellvergleich NF- vs. HO-Modell

Vor allem im Bildungskontext wird das (um Kriterienvariablen erweiterte) NF-Modell zunehmend häufiger von unterschiedlichen Forschendengruppen eingesetzt, wenn es um Fragestellungen geht, die den Zusammenhang von allgemeinen und spezifischen Intelligenzfaktoren mit relevanten Außenkriterien fokussieren und dabei besonderer Wert auf die distinkte Betrachtung der einzelnen Faktoren unabhängig voneinander gelegt wird (vgl. z.B. Beaujean et al., 2014; Saß et al., 2017; Schult & Sparfeldt, 2016). Eid et al. (2018) bezeichnen das NF-Modell als weit verbreitetes und attraktives Analysemodell zur Erfassung von Zusammenhängen allgemeiner und spezifischer Faktoren und betonen zudem den Nutzen des Modells aufgrund der Unkorreliertheit der Faktoren und eindeutigen Varianzaufteilung bei der Vorhersage von Kriteriumsvariablen. Eben diese Unabhängigkeit bzw. Unkorreliertheit der Faktoren stellt jedoch Dreh- und Angelpunkt vieler Diskussionen im Modellvergleich von HO- vs. NF-Modell dar. HO-Befürwortende argumentieren häufig, dass im Gegensatz zum NF-Modell die theoretische Passung des HO-Modells zu vielen konventionellen hierarchischen Intelligenztheorien, vor allem zu Theorien, die eine positive Korreliertheit der spezifischen Gruppenfaktoren annehmen (z.B. Carroll, 1993), erleichtert. Umgekehrt bemängeln NF-Verfechtende, wie es zu erklären sei, dass im HO-Modell die Varianzanteile, die auf g und spezifische Gruppenfaktorresiduen zurückzuführen sind, in jedem Indikator innerhalb eines Gruppenfaktors perfekt proportional sind (s.g. *proportionality constraints*).

Eine Reihe aktueller Studien, die nach dem Erscheinen unserer ersten beiden Artikel veröffentlicht wurden, beschäftigten sich intensiv mit der Frage unter welchen Bedingungen das NF-Modell besser fittet als das HO-Modell (Gignac, 2016; Mansolf & Reise, 2017; Molenaar, 2016; Morgan et al., 2015) und sollen daher im Folgenden ausführlicher dargestellt werden. Gignac (2016) argumentiert beispielsweise, dass in dem Ausmaß, in dem die Faktorladungen im Second-order Modell (HO) den *proportionality constraints* unterliegen (vgl. Brunner et al., 2012; Schmiedek & Li, 2004; Yung et al., 1999), das NF-Modell besser fittet als das Second-Order HO-Modell. Dazu simulierte er 12 Korrelationsmatrizen derart, dass die *proportionality-constraints*-Verletzung progressiv stieg und zeigte, dass die Chi-Square-Statistik zugunsten des NF-Modells anstieg als eine Funktion der Disproportionalität (vgl. Gignac, 2016). Molenaar, (2016) argumentierten jedoch, dass Gignacs (2016) Befunde

auch als vorhersagbare Konsequenz von Modellmisspezifikation im Rahmen von Strukturgleichungsmodellen zu erklären sei und nicht einzig auf die Disproportionalitätseinschränkung des HO-Modells zurückzuführen sei.

Durch mathematische Nachweise und Simulationen konnten Mansolf & Reise (2017) zeigen, dass die statistische Überlegenheit des NF-Modells gegenüber dem HO-Modell letztlich auf ein Set bestimmter „*tetrad constraints*“ (ausführlicher vgl. Bollen & Ting, 1993), die das HO-Modell im Gegensatz zum NF-Modell enthalte, zurückzuführen sei. Relative Einigkeit herrscht somit dahingehend, dass verschiedene statistische Einschränkungen (d.h. „*constraints*“), die im HO-Modell enthalten sind, eine Fit-Statistik geleitete Favorisierung des NF-Modells begünstigen (Gignac, 2016, Mansolf & Reise, 2017; Molenaar, 2016;). Als Ursache für einen Modellmissfit benennen darüber hinaus Mansolf & Reise (2017) wie auch Gignac (2016) so genannte „*unmodellerte Komplexitäten*“ (gemeint sind Kreuzladungen und korrelierte Residuen), die reale Daten in der Regel mit sich brächten und zwangsläufig zu Modellmissfit führten, wenn „reine“ HO- oder NF-Modelle (d.h. ohne Berücksichtigung der Komplexitäten) spezifiziert würden. So resümieren Mansolf & Reise: „*Finally, it is not disproportionality of loadings per se that causes fit indices to favor a bifactor when there are unmodeled complexities, it is their effect on tetrads [...] this kind of unmodeled complexity leads to violation of tetrad constraints and a preference in fit for the pure bifactor model over the pure second order model* (2017, S. 127).

Zwar zeigen die Fit-Statistiken vieler Studien – wie auch in dieser Dissertation (vgl. Valerius & Sparfeldt, 2014) – eine numerische Überlegenheit des NF-Modells gegenüber dem HO-Modell (z.B. Gignac, 2016, 2008; Brunner et al. 2012), jedoch weisen viele Arbeiten darauf hin, dass NF-Modelle nicht allein aufgrund einer rein statistischen Überlegenheit gegenüber dem HO-Modell gewählt werden sollten, sondern die Entscheidung für das NF-Modell maßgeblich von der zu beantwortenden Forschungsfrage abhängen sollte (vgl. z.B. Brunner et al., 2012; Gignac, 2016; Maydeu-Olivares & Coffman, 2006; Morgan et al., 2015; Murray & Johnson, 2013). Wie auch in dieser Dissertation (vgl. Valerius & Sparfeldt, 2014) fitten die analog spezifizierten HO-Modelle in vielen der eben genannten Studien in Anlehnung an konventionelle Standards (Hu & Bentler, 1999) im Vergleich zu den NF-Modellen ebenfalls mindestens akzeptabel und belegen somit adäquate Modellpassung beider Modelliervarianten. Mit etwas Abstand betrachtet, erscheint die aktuell geführte Modellvergleich-Kontroverse

somit wie eine Art Glaubenskrieg im Rahmen der psychometrischen Intelligenzforschung. Beide Modelle können je nach Fragestellung oder Auffassung der zugrundeliegenden Intelligenzstrukturtheorie sowohl Nutzen als auch Einschränkungen mit sich bringen: Tut man sich schwer mit dem Gedanken, unabhängige Faktoren zu spezifizieren (verstehet sie als voneinander losgelöste kognitive Prozesse) und kann gleichzeitig über die Schwierigkeiten hinwegsehen, die die Abhängigkeiten im HO-Modell mit sich bringen (z.B. *proportionality constraints*), so wird die Wahl auf das HO-Modell fallen. Akzeptiert man jedoch eine eher breite Konzeptualisierung von *g* mit direktem Einfluss des Generalfaktors auf jedem Indikator (vgl. Gignac, 2008) und fürchtet sich nicht vor potenziell möglicher Modellmisspezifikation oder Überparametrisierung (vgl. Maydeu-Olivares & Coffman, 2006; Murray & Johnson, 2013), kann das NF-Modell – vor allem wenn die Erweiterung um Kriteriumsvariablen von Relevanz ist – eine sinnvolle Alternative darstellen. Liegen zum statistischen Modellvergleich vielfältige und aktuelle Studien vor (z.B. Brunner et al., 2012; Gignac, 2016; Molenaar, 2016; Morgan et al., 2015; Murray & Johnson, 2013), wird die theoretische Frage nach den zugrundeliegenden kognitiven Prozessen nur in sehr wenigen Studien ausführlicher behandelt (vgl. z.B. Beaujean, 2015). In den meisten Fällen thematisieren Autorinnen und Autoren verschiedener Studien diese Frage eher am Rande (vgl. z.B. Gignac, 2016). Nur einige wenige Studien befassen sich empirisch und systematisch mit der Frage nach der Passung von NF-Modell zu anerkannten Intelligenzstrukturtheorien (vgl. z.B. Beaujean et al., 2014; Brunner & Süß, 2005, 2007). In der Bewusstwerdung, dass das NF-Modell zunehmend an Popularität gewinnt (vgl. Reise, 2012), sollte ein Ziel weiterführender Forschung demnach sein, zum einen die Passung des NF-Modells zu anerkannten Intelligenzstrukturtheorien weiter zu fokussieren sowie sich systematisch und empirisch der Frage zu nähern, wie sich ein Verständnis von unabhängigen Intelligenzfaktoren (im NF-Modell) mit den zugrundeliegenden kognitiven Prozessen auf neuronaler/zerebraler Ebene in Einklang bringen lässt. Eine gedankliche Öffnung der psychometrischen Intelligenzforschung in Richtung weiterer Forschungsrichtungen (wie beispielsweise Kognitionswissenschaften; EEG-, MEG-fMRT-Forschung etc.) könnten an dieser Stelle eine Weiterentwicklungsoption darstellen.

Einige aktuelle Arbeiten im bildungswissenschaftlichen Kontext lassen den Modellvergleich von HO- und NF-Modell bereits hinter sich und beschäftigen sich vielmehr mit der Frage, wie der für viele Intelligenz- und bildungsrelevante Fragestellungen durchaus attraktive NF-Ansatz optimiert genutzt werden kann, damit bekannte Probleme (wie bspw.

Modellmisspezifikation, Nichtidentifizierbarkeit, Überparametrisierung) vermieden werden können (vgl. Eid et al., 2018; Koch et al., 2018). Als Empfehlung aus diesen Arbeiten, regen die Autorinnen und Autoren an, als eine mögliche Alternative zum klassischen (um Kriteriumsvariablen erweiterten) NF-Modell, Bifactor (S-1) Modelle zu spezifizieren (vgl. Eid et al., 2018). Ein Ansatz, der bereits in dieser Dissertation in nachträglichen Zusatzanalysen erfolgreich umgesetzt wurde. Da der NF-Ansatz im Rahmen der psychometrischen Intelligenzforschung zunehmend an Bedeutung gewinnt, erscheint es von großer Relevanz, in diesem Bereich weiter zu forschen, um die Fallstricke der NF-Modellierung zu kennen und berücksichtigen zu können.

Gignac (2016, S. 65 ff.) versucht durch ein konstruktiv-versöhnliches Resümee die Kontroverse zum Modellvergleich ebenfalls zu überwinden, indem er vorschlägt hybride NF-Modelle, in denen spezifische Gruppenfaktoren korrelieren dürfen (vgl. Rindskopf & Rose, 1988) oder hybride HO-Modelle mit direkten Effekten von g zu einigen Subtests (vgl. Gignac, 2014) zu spezifizieren. Eine empirische Prüfung der Tauglichkeit eines solchen Ansatzes ist bislang jedoch noch offen.

Zusammenfassend fällt auf, dass die Kontroverse zum Modellvergleich von HO- vs. NF-Modell vor allem in den letzten 10 bis 15 Jahren neu entfachte, mittlerweile jedoch vor allem im bildungswissenschaftlichen Kontext weitestgehend überwunden wurde. Was sich darin zeigt, dass das NF-Modell mit der Spezifizierung unkorrelierter Faktoren als attraktives Analysemodell für viele Fragestellungen, die die Beziehungen allgemeiner und vor allem spezifischer Faktoren zu relevanten Außenkriterien fokussieren, vielfach dem HO-Ansatz vorgezogen wird. Bereits einige brauchbare Ansätze existieren, mit den Problemen der NF-Modellierung umzugehen (z.B. Eid et al., 2018; Koch et al., 2018). Weitere Ideen, die unterschiedlichen Sichtweisen durch beispielsweise hybride NF- oder HO-Modelle aneinander anzunähern, wurden ebenfalls formuliert (vgl. Gignac, 2016). Schon einmal konnte eine sehr kontrovers geführte Debatte innerhalb der psychometrischen Intelligenzforschung durch die Integration zweier ähnlicher jedoch teilweise unterschiedlicher Vorstellungen von Intelligenzstruktur zu einer Theorie sehr erfolgreich beendet werden: Die Zusammenführung der *Three-Stratum-Theorie kognitiver Fähigkeiten* (Carroll, 1993) mit der *Theorie fluider und kristalliner Intelligenz* (Horn & Cattell, 1966) zur *Cattell-Horn-Carroll (CHC) Theorie*, führte zu hohem Konsens innerhalb der Forschendengemeinde und setzte

einen Punkt hinter eine Jahrzehnte andauernde Diskussion über die Struktur und Beschaffenheit des Intelligenzkonstrukts. Wie sich die Diskussion mit Hinblick auf den kontroversen HO- vs. NF-Modellvergleich verhält, bleibt abzuwarten. Verschiedene (Denk-) Ansätze existieren bereits (z.B. Eid et al., 2018; Gignac, 2016) und es erscheint lohnenswert, diese weiterzuentwickeln und darüber hinaus neue Ansätze zum Umgang mit den NF-Modellschwierigkeiten zu entwerfen, da NF-Modelle v.a. aufgrund der orthogonalen Spezifizierung der Faktoren eine distinkte Analyse aller Faktoren unbeeinflusst voneinander ermöglichen und das Bild über die Struktur kognitiver Fähigkeiten und deren Korrelaten zu schärfen vermögen.

4.2.2 Implikationen für die Forschung zur Konsistenz von g und spezifischen Intelligenzfaktoren

Die sehr hohen g-Faktor Korrelationen der unterschiedlichen Testbatterien in der für diese Dissertation erhobenen, unabhängigen Stichprobe gehen einher mit vorangehender Forschung und belegen erneut die Konsistenz von g (vgl. Floyd et al., 2013; Johnson et al., 2004, 2008). Obwohl sich die in dieser Arbeit verwendeten Testbatterien hinsichtlich Subtestumfang und Facettenreichtum durchaus unterschieden, können die sehr hohen g-Faktor Korrelationen somit als Unterstützung für Spearman's „*Indifference of the Indicator*“ (Spearman, 1927, S. 197) herangezogen werden. Das batterieübergreifend konvergent/divergente Beziehungsmuster der spezifisch verbalen und numerischen NF-Inhaltsfaktoren stützt ebenfalls die Annahme, dass unabhängig voneinander entwickelte Intelligenztests die spezifischen Fähigkeitsfaktoren in (relativ) ähnlicher Weise und distinkt voneinander erfassen. Die Ergebnisse sind (zumindest für Generalfaktor und spezifisch verbalen Faktor) in Einklang mit den Befunden von Floyd et al. (2013). Zwar korrelierten die Autoren batterieübergreifend keine spezifischen Inhaltsfaktoren im Sinne des Berliner Intelligenzstruktur Modells (Jäger, 1982) sondern „breite“ Stratum II Faktoren im Sinne der CHC-Theorie (McGrew, 2009), dennoch zeigte sich ein sehr ähnliches Befundmuster konsistenter Messung von g und darüber hinaus zusätzlich spezifischen latenten Fähigkeitsfaktoren. Zusammengefasst ergänzen sich die Arbeiten von Floyd et al. (2013) und dieser Dissertation (Studie 1) dahingehend, dass sich unabhängig von der faktorenanalytischen Modellvariante (HO Modell bei Floyd et al., 2013 vs. NF-Modell bei Valerius & Sparfeldt, 2014) ein ähnliches Befundmuster zeigte: Die Eignung von

Intelligenztests zur konsistenten Messung von g und zusätzlich zur Konsistenz einer Reihe von spezifischen Faktoren unabhängig von g . Sowohl Forschung als auch die angewandte Psychologie sind maßgeblich darauf angewiesen, dass die zur Verfügung stehenden Messinstrumente verlässlich und akkurat messen. Es erscheint lohnenswert, die Konsistenz der Messung aller (v.a. spezifischer) Faktoren verschiedener differentieller Testbatterien weiter zu erforschen. Das Interesse an der Güte der Messung spezifischer Intelligenzfaktoren sollte im Rahmen des aktuellen Trends der psychometrischen Intelligenzforschung hinsichtlich der Relevanz spezifischer Fähigkeiten (vgl. Kell & Lang, 2019) von großer Bedeutung sein, da es die Grundlage für darauf aufbauende Untersuchungen liefert. Die Überprüfung relevanter Gütekriterien (in diesem Fall der konvergenten Validität) ist nicht nur für neu entwickelte Testverfahren angezeigt, sondern auch für welche, die gleichermaßen für Belange der Forschung und Praxis in regelmäßiger Anwendung stehen.

Ergebnisse latenter Analysen liefern ein um Messfehler reduziertes Befundmuster und können nicht unbedingt unmittelbar auf die Einzelfalldiagnostik übertragen werden. Dennoch entsteht unter Praktizierenden im Rahmen von Bildungsberatung (bspw. [schul]psychologische Beratung) ein ungebrochenes Interesse daran, spezifische Intelligenzfacetten (z.B. Bereichswerte, Subskalen) zusätzlich zum Gesamtscore eines Tests im Sinne von individuellen Stärken und Schwächen zurückzumelden. Eine als hilfreich angenommene Methode stellt die Profilinterpretation dar, die jedoch an sehr viele Voraussetzungen geknüpft ist (z.B. geringe Interkorrelation der Subtests [$r \leq .30$], hohe Reliabilität und Validität der Subskalen, Profilreliabilität und Profilvalidität; vgl. zusammenfassend Rost, 2013, S. 340), welche nicht von allen Intelligenztests erfüllt werden. Trotz überwiegend ernüchternder Befunde zur Profilanalyse erscheint das Forschungsinteresse zum Nutzen von Subskalenwerten als Ergänzung zum Gesamtestwert nicht erloschen (z.B. Liu, Qian, Luo & Woo, 2018; Meijer, Boevé, Tendeiro, Bosker & Albers, 2017; Sinharay, 2010). Die Entwicklung hoch reliabler und valider Subskalen mit ausreichendem Aufgabenumfang sollte somit Ziel weiterer Testentwicklung sein und bleiben.

4.2.3 Implikationen für die Forschung zum Beziehungsgeflecht allgemeiner und spezifischer Intelligenz mit schulischer Leistung

Die Befunde dieser Dissertation zum Zusammenhang allgemeiner und spezifischer Intelligenz- und Schulleistungsfaktoren (Studie 2, Valerius & Sparfeldt, 2015) sind in Einklang mit vorheriger und aktueller Forschung (z.B. Beaujean et al., 2014; Brunner, 2008; Coyle, 2018b; Schult & Sparfeldt, 2016; Wee, 2018). Neben dem umfassend abgesicherten Befund, dass die allgemeine Intelligenz als einer der wichtigsten Prädiktoren für schulische Leistung anerkannt wird (vgl. Jensen 1998, Roth et al., 2015), kann zusätzlich festgehalten werden: spezifische Fähigkeitsfaktoren sind neben dem Generalfaktor (auch) wichtig und geeignet zur Vorhersage spezifischer Schulleistung. Der Nutzen des erweiterten NF-Modells (oder Bifactor [S-1] Modells; vgl. Eid et al., 2018) zur Erforschung dieses Sachverhalts spielt hierbei eine wesentliche Rolle. Verschiedene Arbeiten, die kurz vor oder nach der Veröffentlichung der zweiten Studie dieser Dissertation (vgl. Valerius & Sparfeldt, 2015) erschienen sind, näherten sich der Fragestellung über eine Modellierung mit latenten Faktoren in unterschiedlicher Weise. Coyle et al. (2015) prädizierten zum Beispiel im HO-Modell für die Daten der SAT-(Scholastic Aptitude Test) Validierungsstichprobe sowohl konvergente Zusammenhänge von SAT Mathe-Residuen zu STEM (Science, Technology, Engineering und Math) Noten (grade point averages: GPAs) als auch divergente Zusammenhänge zu verbalen/geisteswissenschaftlichen (z.B. Englisch, Kunst, Geschichte, Fremdsprachen) GPAs. Während verbale SAT-Residuen (Lesen und Schreiben) ein entgegengesetztes Zusammenhangsmuster aufwiesen.

Ähnlich wie Coyle et al. (2013) griffen auch Schult & Sparfeldt (2016) auf die Daten der National Longitudinal Survey of Youth 1997 (NLSY97; $n = 2155$) zurück. Im Gegensatz zu Coyle et al. (2013, 2015), die für ihre Analysen g -unabhängige Residuen verwendeten, korrelierten Schult & Sparfeldt (2016) im NF-Modell orthogonale spezifische Testfaktoren (verbal, numerisch und shop⁷) mit spezifischen sprachlichen und mathematischen Schulleistungsfaktoren (Noten/GPA) zusätzlich zu den korrelierten Generalfaktoren (Intelligenz- g [bestehend aus allen ASVAB Subtests], Test- g [bestehend aus allen SAT und

⁷ Der Shop Faktor wurde gebildet aus technischen Subtests (AI: Wissen über Autos, EI: Verstehen von elektrischen Prinzipien, MC: Verstehen von mechanischen Prinzipien und SI: Wissen über Werkzeuge) der Armed Service Vocational Aptitude Battery (ASVAB), einem Intelligenzstrukturtest.

ACT Subtests] und GPA-g [bestehend aus allen Noten]). Zur Schätzung der spezifisch verbalen und numerischen Testfaktoren wurden die verbalen und mathematischen SAT- und ACT-Subtests zusammen mit den bereichsspezifischen verbalen und numerischen Intelligenzsubtests des ASVAB herangezogen. Es zeigte sich neben hohen Korrelationen der Generalfaktoren ($r > .54$) ein konvergent-divergentes Beziehungsmuster stärker für die mathematischen ($r_{konv.} = .75$; $r_{div.} = -.25$) als für die verbalen/sprachlichen ($r_{konv.} = .11$; $r_{div.} = -.27$) Test- und Studierfähigkeitsgruppenfaktoren mit Zensuren (GPA). Dadurch, dass zur Schätzung der spezifischen verbalen und numerischen Testgruppenfaktoren, Subtests herangezogen wurden, die sowohl dem Intelligenzstrukturtest (ASVAB) als auch den Studierfähigkeitstests (SAT und ACT) entstammten, wird ähnlich wie bei Coyle et al. (2013, 2015), die einen gemeinsamen g-Faktor als Konglomerat aus Intelligenz und Schulleistung bzw. Studierfähigkeit bildeten, die Interpretation der Zusammenhänge erschwert.

Wee (2018) nutzte im Rahmen von NF-Modellierung zwei Analysemethoden (*structural equation modeling* [SEM] und *relative importance analysis*) zur Untersuchung der Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren (Wilde Intelligenz Test) mit allgemeinen und spezifischen Schulleistungen (Zensuren) anhand der im Rahmen der Sonderausgabe des *Journal of Intelligence* (Kell & Lang, 2019) zur Verfügung gestellten Stichprobe von $N = 219$ Jugendlichen und jungen Erwachsenen aus Deutschland aus den 1950er Jahren. Sie fand heraus, dass in den NF-SEM-Analysen der allgemeine Intelligenzfaktor den besten (und einzigen) Prädiktor für einen allgemeinen Schulleistungsfaktor (abgeleitet aus gemeinsamer Varianz von Fachnoten) darstellte (44 % Varianzaufklärung), hingegen in der NF-relative importance analysis der verbale Reasoningfaktor (41% Varianzaufklärung) die Prädiktionskraft des allgemeinen Intelligenzfaktors (20%) für den allgemeinen Schulleistungsfaktor überbot. Darüber hinaus zeigte sich der verbale Reasoningfaktor als bester Prädiktor für alle spezifischen Schulnoten (31% Varianzaufklärung für Mathematik, 38% für Deutsch, 51% für Englisch und 55% für Sport). Dennoch variierte das Befundmuster über die analytischen Ansätze hinweg, denn im SEM-Modell erzielte weder der allgemeine noch der verbale Reasoningfaktor zusätzliche bedeutsame Varianzaufklärung in der spezifischen Schulleistung (d.h. Fachnoten-Residuen) nachdem die Beziehung zwischen allgemeinem Intelligenz- und allgemeinem Schulleistungsfaktor berücksichtigt wurde. Wee (2018) schlussfolgerte daher, dass sowohl allgemeine wie auch spezifische Intelligenzfaktoren als sinnvolle Prädiktoren für schulische

Leistung herangezogen werden können, aber die Interpretation hinsichtlich ihres Nutzens können selbst innerhalb eines gemeinsamen faktorenanalytischen Rahmens (NF-Modellierung) entscheidend vom verwendeten Analyseansatz abhängen.

Jedoch wurden in nur sehr wenige Arbeiten (z.B. Wee, 2018) die Intelligenzfaktoren „pur“ spezifiziert, das heißt ohne zusätzliche Indikatoren aus Schulleistungstests (vgl. Gustafsson & Balke, 1993), Kompetenztests (z.B. Brunner, 2008) oder Studierfähigkeitstests (z.B. Coyle, 2018b; Schult & Sparfeldt, 2016), was die Interpretation der Ergebnisse erschwert. Zusätzlich zeigte sich, dass die Modellierung klassischer und um Kriterien erweiterter NF-Modelle Schwierigkeiten mit sich bringt: die Bandbreite reicht von z.B. Schätzproblemen und Nichtidentifizierbarkeit (vgl. Eid et al., 2018) über Schwierigkeiten, wenn manifeste Kriterienvariablen hinzugefügt werden (z.B. Koch et al., 2018) bis hin zu uneindeutigen Befundmustern in Abhängigkeit des Analyseansatzes innerhalb der NF-Modellierung (z.B. Wee, 2018). Der Vorschlag, Bifactor (S-1) Modelle zu verwenden (vgl. Eid et al., 2018), stellt einen Anfang dar und ebnet den Weg für weitere Forschung. Diesen Weg weiterzugehen erscheint äußerst sinnvoll, da der NF-Ansatz an sich für viele Fragestellungen – nicht nur innerhalb der psychometrischen Intelligenzforschung – von großem Nutzen erscheint, wenn die Beziehung spezifischer Faktoren zusätzlich und unabhängig vom Generalfaktor von Interesse ist. Um vor allem auch berechtigte, kritische Einwände auszuräumen, erscheint es äußerst wertvoll, einen Weg zu finden, mit den Schwierigkeiten des NF-Modells umzugehen.

Ein weiterer Aspekt, der teilweise nur randständig behandelt wird, ist der Versuch einer Zusammenführung von Theorie und Empirie zum Wirkgefüge spezifischer Intelligenzfaktoren und spezifischer akademischer Leistung unabhängig von dem Einfluss von *g*. Coyle (2018a) diskutiert in diesem Rahmen unter anderem die *Investment Theorie der Intelligenz* (Cattell, 1987), in der er annimmt, dass ein unterschiedliches Maß an Investition von Zeit und Einsatz der fluiden Intelligenz (*Gf*) in bestimmten Bereichen die Herausbildung von kristalliner Intelligenz (*Gc*), d.h. den Wissenserwerb in speziellen Bereichen beeinflusst. Coyle (2018a, 2018b) erklärt die Befunde seiner Studien damit, dass Personen, die beispielsweise viel kognitive Energie in STEM-Bereiche (Science, Technologies, Engineering & Math) investieren, ihre mathematischen Fähigkeiten stärken, was wiederum zu einer Stärke in Mathe und Präferenzen für STEM-Fächer und Berufe führe. Umgekehrt verhalte es sich für die gesteigerte Investition in geisteswissenschaftlichen Bereichen (enthält

Englisch, Kunst, Geschichte, Fremdsprachen, Philosophie und Theologie), was zu gesteigerten verbalen Fähigkeiten führe und wiederum sprachliche Leistung und Präferenzen in den geisteswissenschaftlichen Fächern und der Berufswahl hervorrufe. Im Grunde proklamiert er damit einen sich selbst verstärkenden Kreislauf spezifischer Interessen, Fertigkeiten und Fähigkeiten unabhängig von *g*. Dieser Denkansatz erscheint im Hinblick auf weitere Forschung frucht- und erweiterbar. Verschiedene Aspekte sind denkbar: Wie wirkt sich unterschiedlich starkes Investment auf die Entwicklung von Expertise und Spezialisierung im längsschnittlichen Design aus (Berufswahl, Forschungsinteressen)? Welchen Einfluss hat das Geschlecht auf die Ausbildung spezifischer Stärken und Schwächen? Ab welchem Alter bilden sich Neigungsprofile (i.S.v. Stärken und Schwächen) aus? Wie verändert sich der Einfluss spezifischer Faktoren im Vergleich zum Generalfaktor im Alter? Das Themenfeld spezifischer Fähigkeitsfaktoren ist weit. Forschende, die sich aufmachen dieses Forschungsfeld zu bearbeiten, finden aktuell gute Gegebenheiten vor: die Anzahl der auf diesem Feld Forschenden ist hoch, die nötigen statistischen Werkzeuge (unterschiedliche Analysemethoden) werden kontinuierlich auf Tauglichkeit geprüft (modellvergleichende Analysen; z.B. HO- vs. NF-Vergleich) angepasst und verbessert (z.B. Bifactor (S-1) Modelle, vgl. Eid et al., 2018), die Forschungsergebnisse stoßen auf breites Interesse und Befunde werden diskutiert und integriert (vgl. Kell & Lang, 2019), was eine gemeinsame Weiterentwicklung in diesem Forschungsfeld begünstigt.

Zusammenfassend lässt sich festhalten, dass sowohl allgemeine als auch spezifische Fähigkeiten wichtig sind zur Vorhersage schulischer Leistung. Dies deckt sich mit Befunden aktueller Forschung (z.B. Coyle, 2018b, Schult & Sparfeldt, 2016; Wee, 2018). Zwar wird der Nutzen des NF-Modells innerhalb der Bildungsforschung als geeignetes Analysemodell zur Erforschung allgemeiner und spezifischer Intelligenz- und akademischer Leistungsfaktoren weitestgehend anerkannt (v.a. aufgrund der distinkten Erfassung der unabhängigen Faktoren), dennoch besteht weiterer Forschungsbedarf, um diese Analysemethode gegenüber uneindeutiger Befundlagen (z.B. Wee, 2018) zu stärken. Ein bislang erfolgreich eingeschlagener Weg scheint die Modellierung von Bifactor (S-1) Modellen zu sein (vgl. Eid et al., 2018). Ein Weg bei dem es lohnt, ihn weiter zu verfolgen.

4.2.4 Implikationen für Forschung und Praxis zur Hochbegabungsidentifikation mit mehreren Intelligenztests

Die Befunde der dritten Studie der vorliegenden Dissertation zur Güte der Hochbegabungsklassifikation mithilfe zweier konzeptionell ähnlicher Intelligenztests zeigen deutlich, dass trotz mindestens mittelhoher Testinterkorrelation, die Hochbegabungsklassifizierung nicht zwangsläufig sicher gelingt. Die Ergebnisse der dritten Studie sind unter Verwendung eines bisher nicht verwendeten Analyseansatzes (RATZ-Index) in Einklang mit früheren Forschungsergebnissen (bspw. Lohman, 2009, Lohman & Korb, 2006, Lohman et al. 2008).

Der Befund, dass kognitive Fähigkeitstests in Populationen mit höherer Intelligenz weniger stark korreliert sind und weniger stark auf g laden als in Populationen mit niedriger Intelligenz wurde erstmals von Spearman (1925, 1927) als „*law of diminishing returns*“⁸ beschrieben. Im Speziellen wird in verschiedenen Arbeiten zu *Differenzierungstheorien* der Intelligenz (z.B. Deary, Egan, Gibson, Brand, Austin & Kellaghan, 1996; Reinert, Baltes & Schmidt, 1965) die Annahme vertreten, dass kognitive Fähigkeiten mit höherem Fähigkeitsniveau differenzierter (und weniger g -geladen) werden und mit einer Fähigkeitsspezialisierung einhergehen. Studien zu Differenzierungstheorien wurden in methodisch unterschiedlicher Weise über die bisherige Zeitspanne der Intelligenzforschung mehrfach repliziert (z.B. Blum & Holling, 2017; Detterman & Daniel, 1989; Reinert et al., 1965; Spearman, 1925). Recht aktuell konnten Blum & Holling (2017) zeigen, dass der gefundene (negative) Zusammenhang von Fähigkeitsniveau und Testinterkorrelation linear (vs. U-förmig vs. kurvilinear) ist, d.h. ein kontinuierliches Absinken der Testinterkorrelationen sowie Verringern der g -Saturierung mit kontinuierlich steigendem Fähigkeitsniveau zu beobachten ist. Die Autoren fanden zusätzlich in ihrer Metaanalyse heraus, dass in Stichproben, deren durchschnittliche Intelligenz zwei Standardabweichungen

⁸ Spearman (1925) bezog sich dabei auf ein fundamentales Prinzip aus der Wirtschaft, das besagt, wenn die Menge des Inputs eines Produktionsprozesses kontinuierlich zunimmt und alle anderen Produktionsfaktoren konstant bleiben, die Wachstumsrate der Produktion schließlich abnehmen wird. Analog dazu beschreibt er den Einfluss des g -Faktors wie folgt: „*The correlations always become smaller - showing the influence of g to grow less – in just the classes of person which, on the whole, possess this g more abundantly.*“ (Spearman, 1925, S. 438).

(d.h. 30 IQ-Punkte) über dem Mittel lagen, die zu erwartenden Korrelationskoeffizienten durchschnittlich um etwa .15 Punkte absanken (Blum & Holling, 2017, S. 65). Die in der vorliegenden dritten Studie dieser Dissertation erfassten Testinterkorrelationen lagen in der (tendenziell etwas intelligenteren) gymnasialen Stichprobe ($103.2 \leq M_{IQ} \leq 112.6$) mit $.35 \leq r \leq .52$ teilweise an der unteren Grenze des aus vorangehender Forschung zu erwartenden Wertebereichs, was sich mit dem oben beschriebenen Phänomen erklären ließe.

Ein Absinken der *g*-Saturierung von Intelligenztestaufgaben mit steigendem Fähigkeitsniveau und einer damit verbundenen Fähigkeitsspezialisierung im hohen Begabungsbereich (vgl. Blum & Holling, 2017) legt die Notwendigkeit einer facettenreichen Intelligenz erfassung im Rahmen der Hochbegabungsdiagnostik nahe (Lubinski, 2009; Lubinski, Webb, Morelock & Benbow, 2001, S. 726). Vor allem im (extrem) hohen Begabungsbereich erwiesen sich die im frühen Jugendalter erfassten Begabungsschwerpunkte (verbal vs. mathematisch vs. figural) – neben der allgemeinen Intelligenz – als relevant und vorhersagebedeutsam für Leistungen in der entsprechenden Domäne (d.h. ein mathematischer Begabungsschwerpunkt prädizierte z.B. eine naturwissenschaftliche Fächerwahl, Anzahl von Patenten und das Ausüben von Ingenieursberufen; ein verbaler Begabungsschwerpunkt prädizierte z.B. eine geisteswissenschaftliche Fächerwahl, Anzahl publizierter Bücher und journalistische Berufswahl; vgl. Lubinski, 2009). In vielen Studien von Lubinski und Mitarbeitenden ziehen die Autorinnen und Autoren für ihre Analysen im hohen Begabungsbereich Neigungsprofile (s.g. *tilts*) heran (z.B. Kell et al., 2013; Lubinski, 2009; Lubinski et al., 2001; Lubinski & Benbow, 2006). Methodisch betrachtet errechnen sich Neigungsprofile aus der intraindividuellen Differenz z.B. verbaler und mathematischer Subtestwerte standardisierter Tests. Ein mathematischer Neigungsschwerpunkt resultiert demnach, wenn mathematische Subtestwerte höher ausfallen als die verbalen Subtestwerte und ein verbaler Neigungsschwerpunkt entsteht, wenn verbale Subtestwerte höher ausfallen als die mathematischen Subtestwerte. Es ist zusätzlich festzulegen, ab wann eine intraindividuelle Subskalendifferenz als bedeutsam, d.h. als verbale oder mathematische Neigung, zu bezeichnen ist.

Vor dem Hintergrund einer stärkeren Differenzierung im hohen Begabungsbereich (z.B. Blum & Holling, 2017; Deary et al., 1996; Reinert et al., 1965) stellt sich für die Klassifikationsgütebestimmung von Hochbegabungsdiagnosen mit unterschiedlichen

Intelligenztests (dritte Studie) die Frage, ob sich differenzierte Intelligenzmaße im Sinne von Subskalen spezifischer (z.B. verbaler oder numerischer) Intelligenzfaktoren oder gar Neigungsprofile gegenüber der Verwendung von Globalmaßen der Intelligenz (z.B. Gesamtwert) eignen. So könnte ein Ansatz weitergehender Forschung sein, zu prüfen, ob die mit einem Subtest gefundene „sprachliche (Teil-) Begabung“ (z.B. erfasst über einen verbalen Reasoning Subtest einer Intelligenztestbatterie) oder „mathematische (Teil-) Begabung“ (z.B. erfasst über einen numerischen Reasoning Subtest) sich auch mit den spezifischen verbalen oder numerischen Subskalen einer weiteren unabhängigen Testbatterie diagnostizieren lässt. Die Erforschung der Übereinstimmungsgüte von Begabungsschwerpunkten oder Neigungsprofilen mit verschiedenen Intelligenztests, setzt selbstverständlich ein sorgsames methodisches Vorgehen im Umgang mit den Herausforderungen von Profilerinterpretation (vgl. Rost, 2013) sowie im Umgang mit Differenzen (vgl. Stelzl, 1982) voraus.

Wenn weitreichende Klassifikations- oder Selektionsentscheidungen (z.B. im Rahmen von Förderentscheidungen für Hoch- oder Minderbegabte, im klinischen Kontext; Mündigkeitsentscheidungen etc.) aufgrund von IQ-Testergebnissen getroffen werden, steht eine sorgfältige Testauswahl und Ausrichtung auf die Fragestellung an oberster Stelle. Vor dem Hintergrund einer zunehmenden Differenziertheit mit steigendem Fähigkeitsniveau (z.B. Blum & Holling, 2017), sollten zur Diagnostik von Hochbegabung vor allem Intelligenztests eingesetzt werden, die die Intelligenz sowohl im Sinne eines „guten g“ (Jensen & Weng 1994), als auch facettenreich (vgl. Lubinski, 2009, Lubinski et al., 2001) zu erfassen vermögen.

In Einklang mit vorheriger Forschung (vgl. Lohman et al., 2008) ist ein Befund der dritten Studie, dass die Betrachtungsrichtung oder Testreihenfolge einen bedeutsamen Einfluss auf die Übereinstimmungsgüte der Hochbegabungsdiagnose ausübt in Abhängigkeit der (Strenge der) Normen. Neben überzeugenden Hinweisen zu den gängigen Gütekriterien (Objektivität, Reliabilität und Validität) sollte demnach im Rahmen der Hochbegabungsdiagnostik vor allem auch auf die Aktualität der Normen geachtet werden.

Die Forschungsbefunde zur Klassifikationsgüte der Hochbegabungsdiagnose mit zwei oder mehreren Intelligenztests sind insgesamt ernüchternd. Dennoch stehen Praktizierende (nicht nur im Rahmen der Hochbegabungsdiagnostik) tagtäglich vor der Herausforderung der

Befundintegration und -interpretation, wenn sie zwei oder mehrerer Intelligenztests eingesetzt haben. Rost (2013, S. 235 ff.) schlägt daher beim Einsatz zweier (oder mehrerer) Intelligenztests im Rahmen einer Hochbegabungsdiagnostik vor, das diagnostische Fazit „hochbegabt“ oder „nicht hochbegabt“ über die Integration beider Testergebnisse zu einem Messwert unter Berücksichtigung der Testinterkorrelation, zu ziehen.

4.3 Einschränkungen und Ausblick

In diesem Abschnitt werden verschiedene Aspekte des Untersuchungsdesigns sowie der eingesetzten Testverfahren, die die Interpretation der Ergebnisse dieser Dissertation möglicherweise einschränken, kritisch erörtert. Weiterhin werden daraus resultierende Ideen für Folgeuntersuchungen in Ausblick gestellt.

4.3.1 Stichprobe

Die Daten stammen aus einer relativ großen Stichprobe von $N = 562$ Schülerinnen und Schülern der 9. Klassenstufe aus sechs rheinland-pfälzischen Gymnasien erhoben in den Monaten Mai bis Juli 2010. Die Beschränkung auf eine hinsichtlich Bildungsgrad und Alter homogene Stichprobe hatte das Ziel, Konfundierungen mit potenziellen Störvariablen zu vermeiden, sowie den Nutzen einen vergleichsweise größeren Anteil Hochbegabter (aufgrund des hohen Bildungsgrades) für die dritte Studie zu erheben. Dies beeinflusst jedoch möglicherweise die Interpretation der Ergebnisse in zweierlei Hinsicht: 1) die Stichprobe ist leicht varianzeingeschränkt (Vergleich mit den Korrelationskoeffizienten der KFT-Normstichprobe ergaben kleinere bis mittlere Effekte; Median $q = .16$). Die Höhe der Korrelationskoeffizienten der batterieübergreifenden g -Faktoren (vgl. Valerius & Sparfeldt, 2014) entsprach jedoch der erwarteten Größenordnung vorangehender Forschung (Johnson et al. 2004, 2008). Da Varianzeinschränkungen in der Regel mit niedrigeren Korrelationen einhergehen, können die Ergebnisse dieser Dissertation als konservative Schätzung bezeichnet werden. 2) Die Ergebnisse sind nur bedingt übertragbar auf andere Altersstufen, andere Bildungsgrade und andere Kulturkreise (national sowie international). Eine

Ausweitung oder Replikation der Ergebnisse mit heterogeneren, größeren und internationalen Stichproben könnte helfen, das Befundmuster vor allem hinsichtlich der spezifischen Intelligenzfaktoren zu stabilisieren. Dies könnte vor allem deswegen relevant sein, da in Stichproben mit höherem Fähigkeitsniveau (wie auch der vorliegenden), die Fähigkeitsstruktur differenzierter wird (vgl. Blum & Holling, 2017; Lubinski) und somit eine stichprobenunabhängige Überprüfung spezifischer Intelligenzfaktoren (unabhängig von g) zum Absichern von Stichprobeneffekten erfordert (vgl. Blum & Holling, 2017, S. 66). Eine Erweiterung des Altersspektrums (jüngere und ältere Probanden) könnte die Forschungsmöglichkeit eröffnen, ab welcher Altersstufe sich Stärken und Schwächen herausbilden (vgl. *Differenzierungstheorie* des Alters; z.B. Blum & Holling, 2017), oder wie sich das Verhältnis des Einflusses von g und insbesondere der spezifischen Faktoren auf relevante Leistungsvariablen mit dem Alter (und ggf. ansteigender Spezialisierung) verändert.

4.3.2 Instrumente

Die Intelligenztestauswahl erfolgte anhand verschiedener Kriterien. Überzeugende Hinweise auf relevante Gütekriterien (v.a. Objektivität, Reliabilität und Validität) vorausgesetzt, sollten die eingesetzten Intelligenztestverfahren ein möglichst breites Spektrum gängiger, aktuell normierter Intelligenztestbatterien darstellen und rein pragmatisch im Gruppensetting (SchülerInnen in Klassen) sowie in Papier- Stift-Form in maximal 90 Minuten administrierbar sein. Darüber hinaus sollte die Bildung aller testspezifischen Skalenwerte inklusive eines Gesamtskalenwertes möglich sein. Entsprechend entschieden wir uns somit gegen eine forschungsgeleitete Subtestauswahl (bspw. Subtests mit höchster g -Ladung) sondern griffen (aus pragmatischen Gründen) in den meisten Fällen auf die reliablen und valider Kurzskalen zurück, die laut jeweiligem Testmanual zur Verfügung standen. Ausnahme bildete das Komplett-Set der Raven Matrizen (SPM), welches mit vollem Aufgabensatz administriert wurde. Die anschließenden Analysen basierten somit auf einer relativ breiten und heterogenen Auswahl an 26 (Sub-) Tests. Zur Bildung der testbatteriespezifischen inhaltspezifischen Intelligenzfaktoren (verbal, numerisch, figural) standen somit in den HO- und NF-Modellen jeweils zwei (KFT 4-12) bis fünf (BIS-4) Subtest als Indikatoren zur Verfügung. Dies erforderte für die KFT-Subskalen aus modellbedingten Identifikationsgründen eine Aufspaltung der Skalensubtest in weitestgehend gleiche Parcel mithilfe der Odd-Even-

Methode. Bei einer Fortführung der Analyse von v.a. konvergent-divergenter Beziehungsmuster der spezifischen Intelligenzfaktoren könnte die Operationalisierung mit einer Ergänzung um eine größeren Anzahl und höhere Bandbreite inhaltspezifischer Subtests, den Generalisierungsrahmen der Befunde abzustecken helfen.

Zur Erfassung der schulischen Leistung wählten wir selbstberichtete Zensuren des letzten Halbjahreszeugnisses. Dies entspricht einem durchaus üblichen und forschungspragmatischen Vorgehen im Rahmen bildungsrelevanter Fragestellungen. Nichtsdestotrotz wird dieser Ansatz häufig kritisiert. Eine für die deutsche Bildungslandschaft sowie die Stichprobe der vorliegenden Dissertation relevante Studie lieferten Sparfeldt, Buch, Rost & Lehmann (2008). Anhand von $N = 540$ Gymnasiastinnen und Gymnasiasten der 10. Klassenstufe konnten die Autorin und die Autoren zeigen, dass die selbstberichteten Zensuren in insgesamt elf Fächern sehr hoch (z.B. Mathematik $r = .94$, Deutsch $r = .91$, Englisch $r = .90$, Physik $r = .90$, Geschichte $r = .90$) mit den erhaltenen Zeugniszensuren korrelierten. Dies bestätigt laut Autorenschaft „*die forschungsökonomische Praxis, Zensuren im Selbstbericht zu erheben.*“ (Sparfeldt et al. 2008, S. 55).

4.3.3 Datenerhebung

Die Datenerhebung fand an drei separaten Terminen (jeweils 90 Minuten) während der regulären Unterrichtszeit statt. Die Tests wurden aufgeteilt auf drei Testhefte (Testheft 1: CFT 20-R, BIS-4; Testheft 2: SPM; Testheft 3: KFT 4-12 R). Die Reihenfolge der einzelnen Tests innerhalb der Testhefte variierte nicht, jedoch die Reihenfolge der Testheftadministration über die Probandengruppen (meist klassen- oder schulstufenweise) hinweg. Die Teilnahme der minderjährigen Schülerinnen und Schüler hing von den Entscheidungen der jeweiligen Schulleitungen, Klassenleitungen sowie Eltern und Erziehungsberechtigten ab. Die Teilnahmequote war sehr hoch (91%), was vor einer Verzerrung der Ergebnisse aufgrund einer systematischen Selbstselektion schützte. Die geschulten Testleiter und Testleiterinnen wurden zufällig den verschiedenen Schulen und Klassen zugeteilt. Eine vollständige Randomisierung der Datenerhebung war somit aus pragmatischen Gründen nicht möglich, wurde jedoch unter den gegebenen Bedingungen (z.B. Berücksichtigung der Klassenverbandstruktur; Kombination von Testverfahren zu einem Testheft, die zusammen in

90 Minuten bearbeitet werden konnten) sorgfältig durchdacht und weitestgehend realisiert, um einer systematischen Ergebnisverzerrung vorzubeugen.

4.3.4 Datenanalyse

Zur Analyse des Zusammenhanggefüges spezifischer und allgemeiner Intelligenzfaktoren mit relevanten Außenkriterien (Faktoren anderer Testbatterien, schulische Leistungsfaktoren) im Rahmen der ersten beiden Studien dieser Dissertation wählten wir den – sowohl aufgrund einer generellen Eignung aber auch im Speziellen für die Fragestellungen geeigneten (v.a. im Hinblick auf die spezifischen Faktoren) – NF-Modellansatz aus. Trotz hinreichender Belege für eine gute Modellpassung in verschiedenen Arbeiten nicht nur im Bildungskontext, bleibt die NF-Modellierung unter verschiedenen Aspekten nicht unkritisiert. Einem großen Kritikpunkt an potenziell unentdeckten Modellmissfits (vgl. z.B. Maydeu-Olivares & Coffman, 2006; Morgan et al., 2015) lässt sich laut Eid et al. (2018) u.a. durch eine Variation (Bifactor [S-1] Modelle) des NF-Ansatzes begegnen. Zur Absicherung der Befunde der ersten beiden Studien führten wir daher zusätzliche Analysen mit Bifactor (S-1) Modellen durch. Als weiterer Kritikpunkt am NF-Modell wird mitunter die psychologisch-inhaltliche Interpretation der spezifischen Faktoren als unabhängig vom Generalfaktor genannt. Wenn auch dieser Aspekt an einigen Stellen (eher theoretisch) diskutiert wird (vgl. z.B. Beaujean, 2015) zeigt sich an dieser Stelle ein großer Forschungsbedarf zur Klärung des Verständnisses von „Intelligenz(struktur)“ beziehungsweise den zugrundeliegenden allgemeinen sowie spezifischen kognitiven Prozessen.

Eine weitere Frage werfen die Faktorreliabilitäten Omega hierarchisch (ω_h) im NF-Modell auf. Diese fielen in Übereinstimmung mit weiteren Untersuchungsbefunden (z.B. Brunner et al., 2012) formelbedingt für die spezifischen Gruppenfaktoren relativ niedrig aus, da zu ihrer Berechnung die teilweise niedrigen Faktorladungen der spezifischen Faktoren den g-Faktorladungen im Nenner der Formel gegenübergestellt wurden. Trotz relativ niedriger Faktorreliabilität ($\omega_h = .39$ bzw. $.46$) korrelierten die spezifisch verbalen Intelligenzfaktoren batterieübergreifend hoch ($r = .83$), was gegen massive Reliabilitätseinschränkungen spricht. Die Reliabilität stellt lediglich *ein* Kriterium zur Beurteilung der Skalengüte dar. Die konvergenten Beziehungen der verschiedenen allgemeinen wie auch spezifischen Testskalen belegen batterieübergreifend (hohe) konvergente Validität.

4.4 Fazit

Zusammengenommen belegen die Befunde dieser Dissertation Folgendes: Intelligenztests messen mit hoher Konsistenz und Verlässlichkeit die allgemeine Intelligenz und darüber hinaus mindestens einige spezifische Faktoren (v.a. verbale Intelligenz). Die Zusammenhänge (verbaler und numerischer) spezifischer Intelligenzfaktoren unabhängiger Testbatterien sind auf latenter und messfehlerbereinigter Ebene neben denen der allgemeinen Intelligenzfaktoren in konvergenter Übereinstimmung und divergenter Abgrenzung zueinander bedeutsam.

Dass die allgemeine Intelligenz mit allgemeiner Schulleistung wesentlich zusammenhängt, konnte abermals bestätigt und darüber hinaus um konvergent/divergente Beziehungsmuster spezifisch verbaler und numerischer Intelligenz- mit spezifisch sprachlichen und mathematisch-naturwissenschaftlichen Schulleistungsfaktoren unabhängig vom allgemeinen Intelligenzniveau ergänzt werden. Vor allem der Nested-Factor Ansatz (bzw. in Variation: Bifactor [S-1] Modell) bietet eine gewinnbringende Methode, dieses Beziehungsgeflecht allgemeiner sowie spezifischer Intelligenz- und Schulleistungsfaktoren distinkt zu untersuchen.

Trotz sehr konsistenter Messung der allgemeinen Intelligenz mit unabhängigen Testbatterien (auf latenter Ebene), gelingen in der Anwendungspraxis Klassifikationsaussagen zur Hochbegabungsidentifikation mit unabhängigen Testbatterien nicht automatisch übereinstimmend gut.

LITERATUR

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.
- Beaujean, A. A. (2015). John Carroll's View on Intelligence: Bi-Factor vs. Higher-Order Models. *Journal of Intelligence*, *3*, 121-136.
- Beaujean, A. A., Parkin, J. & Parker, S. (2014). Comparing Cattell-Horn-Carroll factor models: Differences between bifactor and higher order factor models in predicting language achievement. *Psychological Assessment*, *26*, 789–805.
- Beißert, H., Hasselhorn, M. & Lösche, P. (2014). Möglichkeiten und Grenzen der Frühprognose von Hochbegabung. In M. Stamm (Hrsg.), *Handbuch Talententwicklung* (S. 415-425). Bern, Schweiz: Huber.
- Bentler, P. M. (1995). *EQS: A structural equations program manual*. Encino, CA: Multivariate Software, Inc.
- Blum, D. & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence*, *65*, 60-66.
- Bollen, K. A. & Ting, K-F. (1993). Confirmatory Tetrad Analysis. *Sociological Methodology*, *23*, 147-175.
- Boring, E. G. (1923). Intelligence as the tests test it. *New Republic*. *36*, 35-37.
- Brunner, M. (2006). *Mathematische Schülerleistung: Struktur, Schulformunterschiede und Validität*. Unveröffentlichte Dissertation, Humboldt-Universität zu Berlin. Verfügbar unter <http://edoc.hu-berlin.de/dissertationen/brunner-martin-2006-02-08/PDF/brunner.pdf>.
- Brunner, M. (2008). No g in education? *Learning and Individual Differences*, *18*, 152–165.
- Brunner, M. & Süß, H.-M. (2005). Analyzing the reliability of multidimensional measures: an example from intelligence research. *Educational and Psychological Measurement*, *65*, 227–240.
- Brunner, M. & Süß, H.-M. (2007). Wie genau können kognitive Fähigkeiten gemessen werden?: Die Unterscheidung von Gesamt- und Konstruktreliabilitäten in der Intelligenzdiagnostik für den Berliner Intelligenzstrukturtest. *Diagnostica*, *53*, 184-193.
- Brunner, M., Nagy, G. & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*, 796–846.

- Brunner, M., Gogol, K. M., Sonnleitner, P., Keller, U., Kraus, S. & Preckel, F. (2013). Gender differences in the mean level, variability, and profile shape of student achievement: Results from 41 countries. *Intelligence*, *41*, 378–395.
- Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
- Calvin, C. M., Fernandes, C., Smith, P., Visscher, P. M. & Deary, I. J. (2010). Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England. *Intelligence*, *38*, 424–432.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- Carroll, J. B. & Schweiker, R. F. (1951). Factor analysis in educational research. *Review of Educational Research*, *21*, 368-388.
- Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, *38*, 592.
- Cattell, R. B. (1987). *Intelligence: Its Structure, Growth and Action*. Amsterdam, NL: North-Holland.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464-504.
- Chen, F. F., West, S. G. & Sousa, K. H. (2006). A Comparison of Bifactor and Second-Order Models of Quality of Life. *Multivariate Behavioral Research*, *41*, 189-225.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York: Psychology Press.
- Coyle, T. R. (2014). Predictive Validity of Non-g Residuals of Tests: More Than g. *Journal of Intelligence*, *2*, 21-25.
- Coyle, T. R. (2018a). Non-g Factors Predict Educational and Occupational Criteria: More than g. *Journal of Intelligence*, *6*, 43.
- Coyle, T. R. (2018b). Non-g residuals of group factors predict ability tilt, college majors, and jobs: A non-g nexus. *Intelligence*, *67*, 19-25.
- Coyle, T. R., Purcell, J. M., Snyder, A. C. & Kochunov, P. (2013). Non-g residuals of the SAT and ACT predict specific abilities. *Intelligence*, *41*, 114–120.
- Coyle, T. R., Snyder, A. C., Richmond, M. C. & Little, M. (2015). SAT non-g residuals predict course specific GPAs: Support for investment theory. *Intelligence*, *51*, 57-66.
- Daseking, M., Petermann, U. & Petermann, F. (2007). Intelligenzdiagnostik mit dem HAWIK-IV. *Kindheit und Entwicklung*, *16*, 205-259.

-
- Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*, 453-482.
- Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence*, *23*, 105-132.
- DeFries, J. C., Vandenberg, S. G., McClearn, G. E., Kuse, A. R., Wilson, J. R., Ashton, G. G. & Johnson, R. C. (1974). Near identity of cognitive structure in two ethnic groups. *Science*, *183*, 338-339.
- Detterman, D. K. & Daniel, M. H. (1989). Correlations of Mental Tests with Each Other and with Cognitive Variables are Highest for Low IQ Groups. *Intelligence*, *13*, 349-359.
- de Wolf, C. J. & Buiten, B. (1963). Een factorenanalyse van vier testbatterijen [A factor analysis of four test batteries]. *Nederlands tijdschrift voor de Psychologie*, *18*, 220-239.
- Eid, M. Krumm, S., Koch, T. & Schulze, J. (2018). Bifactor Models for Predicting Criteria by General and Specific Factors: Problems of Nonidentifiability and Alternative Solutions. *Journal of Intelligence*, *6*, 42.
- Elliott, C. D. (1990). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.
- Floyd, R. G., Bergeron, R., McCormack, A. C., Anderson, J. L. & Hargrove-Owens, G. L. (2005). Are Cattell-Horn-Carroll Broad Ability Composite Scores Exchangeable Across Batteries? *School Psychology Review*, *34*, 329-357.
- Floyd, R. G., Clark, M. H. & Shadish, W. R. (2008). The Exchangeability of IQs: Implications for Professional Psychology. *Professional Psychology: Research and Practice*, *39*, 414-423.
- Floyd, R. G., Shands, E. I., Rafael, F. A., Bergeron, R. & McGrew, K. S. (2009). The dependability of general-factor loadings: The effects of factor-extraction methods, test battery composition, test battery size, and their interactions. *Intelligence*, *37*, 453-465.
- Floyd, R. G., Reynolds M. R., Farmer, R. L. & Kranzler, J. H. (2013). Are the General Factors from Different Child and Adolescent Intelligence Tests the Same? Results from a Five-Sample, Six-Test Analysis. *School Psychology Review*, *42*, 383-401.
- Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171-191.
- Freund, P. A., Holling, H. & Preckel, F. (2007). A multivariate, multilevel analysis of the relationship between cognitive abilities and scholastic achievement. *Journal of Individual Differences*, *28*, 188-197.
- Gignac, G. E. (2005). Revisiting the factor structure of the WAIS-R: Insights through nested factor modeling. *Assessment*, *12*, 320-329.

- Gignac, G. E. (2006a). A confirmatory examination of the factor structure of the Multidimensional Aptitude Battery (MAB): Contrasting oblique, higher-order, and nested factor models. *Educational and Psychological Measurement*, *66*, 136-145.
- Gignac, G. E. (2006b). The WAIS-III as a nested factors model: A useful alternative to the more conventional oblique and higher-order models. *Journal of Individual Differences*, *27*, 73-86.
- Gignac, G. E. (2008). Higher-order models versus direct hierarchical models: g as superordinate or breadth factor? *Psychology Science Quarterly*, *50*, 21-43.
- Gignac, G. E. (2014). Fluid intelligence shares closer to 60% of its variance with working memory capacity and is a better indicator of general intelligence. *Intelligence*, *47*, 122-133.
- Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, *55*, 57-68.
- Gottfredson, L. S. (1997). Why g Matters: The Complexity of Everyday life. *Intelligence*, *24*, 79-132.
- Grob, A., Meyer, C.S. & Hagmann-von Arx, P. (2009). *Intelligence and development scales (IDS)*. Bern: Huber.
- Gustafsson, J. E. & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research*, *28*, 407-434.
- Hagmann-von Arx, P., Meyer, C.S. & Grob, A. (2008). Assessing intellectual giftedness with the WISC-IV and the IDS. *Zeitschrift für Psychologie*, *216*, 172-179.
- Hakstian A. R. & Cattell, R. B. (1975). *The comprehensive Ability Battery*. Champaign: Institute for Personality and Ability Testing.
- Hanses, P. (2009). Stabilität von Hochbegabung. In D.H. Rost (Hrsg.), *Hochbegabte und hochleistende Jugendliche. Befunde aus dem Marburger Hochbegabtenprojekt* (2. Aufl., S. 93-159). Münster: Waxmann.
- Heller, K. A., Gaedike, A. K. & Weinländer, H. (1985). *Kognitiver Fähigkeitstest (KFT 4-13)* (2. Aufl.). Weinheim: Beltz.
- Heller, K. A., Kratzmeier, H. & Lengfelder, A. (1998). *Matrizen-Test-Manual Band 1 zu den Standard Progressive Matrices von J. C. Raven*. [Test manual of the *Standard Progressive Matrices* of J. C. Raven]. Göttingen, Germany: Beltz-Test.
- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Göttingen: Beltz-Test.

-
- Helmke, A. & Schrader, F.-W. (2010). Determinanten der Schulleistung. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 90-102). Weinheim: Beltz.
- Holzinger, K. J. & Swineford, F. (1937). The Bi-Factor Method. *Psychometrika*, 2, 41-54.
- Horn, J. L. & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology*, 57, 253–270.
- Hu, L. & Bentler, P. M. (1998). Fit Indices in Covariance Structure Modeling: Sensitivity to Underparameterized Model Misspecification. *Psychological Methods*, 3, 424-453.
- Hu, L. & Bentler, P. M. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus new Alternatives. *Structural Equation Modeling*, 6, 1-55.
- Jäger, (1982) Mehrmodale Klassifikation von Intelligenzleistungen. Experimentell kontrollierte Weiterentwicklung eines deskriptiven Intelligenzstrukturmodells. *Diagnostica*, 28, 195-226.
- Jäger, A. O., Süß, H.-M. & Beauducel, A. (1997) Berliner Intelligenzstruktur Test, Form 4. Göttingen: Hogrefe.
- Jensen, A. R. (1998). *The g-factor: The Science of Mental Ability*. Westport, CN: Praeger.
- Jensen, A. R. & Weng, L.-J. (1994). What is a good g? *Intelligence*, 18, 231–258.
- Johnson, W., Bouchard, T. J., Krueger, F., McGue, M. & Gottesman, I. I. (2004). Just one g: consistent results from three test batteries. *Intelligence*, 32, 95-107.
- Johnson, W., te Nijenhuis, J. & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence*, 36, 81-95.
- Kaufman, A. S. & Kaufman, N. L. (1993). *The Kaufman Adolescent and Adult Intelligence Test*. Circle Pines, MN: American Guidance Service.
- Kaufman, A. S. & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children – Second Edition, manual*. Circle Pines, MN: American Guidance Service.
- Kell, H. J., Lubinski, D., Benbow, C. P. & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science*, 24, 1831–1836.
- Kell, H. J. & Lang, J. W. B. (Eds.) (2019). The Great Debate: General Ability and Specific Abilities in the Prediction of Important Outcomes. [Special issue]. *Journal of Intelligence*.
- Kell, H. J. & Lang, J. W. B. (2018). The Great Debate: General Ability and Specific Abilities in the Prediction of Important Outcomes. [Special issue]. *Journal of Intelligence*, 6, 39.

- Kersting, M., Althoff, K. & Jäger, A. O. (2008). *WIT-2. Der Wilde-Intelligenztest. Verfahrenshinweise*. Göttingen: Hogrefe.
- KMK (2009). *Grundsatzposition der Länder zur begabungsgerechten Förderung*. (Beschluss der Kultusministerkonferenz vom 10.12.2009).
https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2009/2009_12_12-Begabungsgerechte-Foerderung.pdf [27.01.2020].
- KMK (2015). *Förderstrategie für leistungsstarke Schülerinnen und Schüler* (Beschluss der Kultusministerkonferenz vom 11.05.2015).
https://www.kmk.org/fileadmin/Dateien/pdf/350-KMK-TOP-011-Fu-Leistungsstarke_-_neu.pdf [27.01.2020].
- Koch, T., Holtmann, J., Bohn, J. & Eid, M. (2018). Explaining General and Specific Factors in Longitudinal, Multimethod, and Bifactor Models: Some Caveats and Recommendations. *Psychological Methods, 23*, 505-523.
- Kubinger, K. D. & Wurst, E. (2000). *Adaptives Intelligenz Diagnostikum - Version 2.1 (AID 2)*. Göttingen: Beltz.
- Liu, R., Qian, H., Luo, X. & Woo, A. (2018). Relative Diagnostic Profile: A Subscore Reporting Framework. *Educational and Psychological Measurement, 78*, 1072–1088.
- Lohman, D. F. (2009). Identifying academically talented students: Some general principles, two specific procedures. In L.V. Shavinina (Hrsg.), *International Handbook on giftedness* (S. 971-998). Amsterdam, NL: Elsevier.
- Lohman, D.F. & Hagen, E.P. (2001). *Cognitive abilities test (Form 6): Research Handbook*. Itasca, IL: Riverside.
- Lohman, D.F. & Korb, K.A. (2006). Gifted today but not tomorrow? Longitudinal changes in ITBS and CogAT scores during elementary school. *Journal for the Education of the Gifted, 29*, 451-484.
- Lohman, D.F., Korb, K.A. & Lakin, J.M. (2008). Identifying academically gifted English-language learners using nonverbal tests – A comparison of the Raven, NNAT, and CogAT. *Gifted Child Quarterly, 52*, 275-296.
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) " 'General intelligence,' objectively determined and measured." *Journal of Personality and Social Psychology, 86*, 96-111.
- Lubinski, D. (2009). Exceptional Cognitive Ability: The Phenotype. *Behavior Genetics, 39*, 350–358.

-
- Lubinski, D., Webb, R. M., Morelock, M. J. & Benbow, C. P. (2001). Top 1 in 10,000: A 10-Year Follow-Up of Profoundly Gifted. *Journal of Applied Psychology*, 86, 718-729.
- Lubinski, D. & Benbow, C. B. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science*, 1, 316–345.
- Lukesch, H., Kormann, A. & Mayrhofer, S. (2002). *PSB-R 4–6. Prüfungssystem für Schul- und Bildungsberatung für 4. bis 6. Klassen. Revidierte Fassung*. Göttingen: Hogrefe.
- Major, J. T., Johnson, W. & Bouchard, T. J. (2011). The dependability of the general factor of intelligence: Why small single-factor models do not adequately represent g. *Intelligence*, 39, 418-433.
- Mansolf, M. & Reise, P. (2017). When and why the second-order and bifactor models are distinguishable. *Intelligence*, 61, 120-129.
- Marx, H. (1992). Methodische und inhaltliche Argumente für und wider eine frühe Identifikation und Prädiktion von Lese- Rechtschreibschwierigkeiten. *Diagnostica*, 38, 249-268.
- Marx, H., Jansen, H. & Skowronek, H. (2000). Prognostische, differentielle und konkurrente Validität des Bielefelder Screenings zur Früherkennung von Lese-Rechtschreibschwierigkeiten (BISC). In: M. Hasselhorn, W. Schneider & H. Marx (Hrsg.) *Diagnostik von Lese-Rechtschreibschwierigkeiten. Test und Trends Bd. 1*. Göttingen: Hogrefe.
- Marx, H. & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In M. Hasselhorn & W. Schneider (Hrsg), *Frühprognose schulischer Kompetenzen*. Göttingen: Hogrefe.
- Maydeu-Olivares, A. & Coffman, D. L. (2006). Random Intercept Item Factor Analysis. *Psychological Methods*, 11, 344-362.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- McGrew, K. S. (2009). CHC theory and the Human Cognitive Abilities Project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37, 1–10.
- Meijer, R. R., Boevé, A. J. Tendeiro, J. N., Bosker, R. J. & Albers, C. J. (2017). The Use of Subscores in Higher Education: When is this useful?. *Frontiers in Psychology*, 8, 305
- Molenaar, D. (2016). On the distortion of model fit in comparing the bifactor model and the higher-order factor model. *Intelligence*, 57, 60-63.

-
- Morgan, G. B., Hodge, K. J., Wells, K. E. & Watkins, M. W. (2015). Are fit Indices Biased in Favor of Bi-factor Models in Cognitive Ability Research?: A Comparison of Fit in Correlated Factors, Higher-Order, and Bi-Factor Models via Monte Carlo Simulations. *Journal of Intelligence*, 3, 2-20.
- Murphy, K. (2017). What can we learn from “Not much more than g”? *Journal of Intelligence*, 5, 8.
- Muthén, L. K. & Muthén, B. O. (1998-2007). *Mplus user's guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.
- Murray, A. L. & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus the higher-order models of human cognitive ability structure. *Intelligence*, 41, 407-422.
- Naglieri, J.A. (1997). *Naglieri Nonverbal Ability Test: Multilevel technical manual*. San Antonio, TX: Harcourt Brace Educational Measurement.
- Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. V., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and Unknowns. *American Psychologist*, 51, 77-101.
- Pietschnig, J. & Voracek, M. (2015). One century of global IQ gains. A Formal meta-analysis of the Flynn effect (1909-2013). *Perspectives on Psychological Science*, 10, 282-306.
- Preckel, F. (2010). Intelligenztests in der Hochbegabungsdiagnostik. In: F. Preckel, W. Schneider & H. Holling (Hrsg.), *Diagnostik von Hochbegabung. Reihe: Jahrbuch der pädagogisch-psychologischen Diagnostik. Tests und Trends*. Göttingen: Hogrefe.
- Preckel, F. & Krampen, G. (2016). Ergebnisse einer szientometrischen Analyse von Publikationen zwischen 1980 und 2014. *Psychologische Rundschau*, 67, 1-14.
- Preckel, F. & Vock, M. (2013). *Hochbegabung. Ein Lehrbuch zu Grundlagen, Diagnose und Fördermöglichkeiten*. Göttingen: Hogrefe.
- Raven, J. C. (1941). Standardization of progressive matrices, 1938. *British Journal of Medical Psychology*, 19, 137-150.
- Raven, J.C (1962). *Advanced Progressive Matrices*. London, UK: Lewis.
- Ree, M. J. & Earles, J. A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, 44 321-332.
- Ree, M. J. & Earles, J. A. (1991). The stability of g across different methods of estimation. *Intelligence*, 15, 271-278.
- Reeve, C. L. & Blacksmith, N. (2009). Identifying g: A review of current factor analytic practices in the science of mental abilities. *Intelligence*, 37, 487–494.

-
- Reinert, G., Baltes, P. B. & Schmidt, L. R. (1965). Faktorenanalytische Untersuchungen zur Differenzierungshypothese der Intelligenz: die Leistungsdifferenzierungshypothese. *Psychologische Forschung*, 28, 246-300.
- Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, 47, 667-696.
- Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second order factor analysis. *Multivariate Behavioral Research*, 23, 51-67.
- Rost, D. H. (2009). Grundlagen, Fragestellung, Methode In: D. H. Rost (Hrsg.), *Hochbegabte und hochleistende Jugendliche* (S. 1-92). Münster: Waxmann.
- Rost, D. H. (2010). Stabilität von Hochbegabung. In: F. Preckel, W. Schneider & H. Holling (Hrsg.), *Diagnostik von Hochbegabung* (S. 233-265). Göttingen: Hogrefe.
- Rost, D. H. (2013). *Handbuch Intelligenz*. Weinheim: Beltz.
- Rost, D. H. & Buch, S. (2018). Hochbegabung. In D. H. Rost, J. R. Sparfeldt & S. R. Buch (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (5. Aufl., S. 226-242). Weinheim: Beltz.
- Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnick, F. & Spinath, F. (2015). Intelligence and school grades: A meta-analysis. *Intelligence*, 53, 118-137.
- Saß, S., Kampa, N. & Köller, O. (2017). The interplay of g and mathematical abilities in large-scale assessments across grades. *Intelligence*, 63, 33-44.
- Schlagheck, W. & Petermann, F. (2006). Hochbegabendiagnostik mit dem HAWIK-III und AID 2. *Kindheit und Entwicklung*, 15, 93-99.
- Schmidt-Atzert, L. & Amelang, M. (2012). *Psychologische Diagnostik* (5. Aufl.). Berlin: Springer.
- Schmiedek, F. & Li, S.-S. (2004). Toward an alternative representation for disentangling age-associated differences in general and specific cognitive abilities. *Psychology and Aging*, 19, 40-56.
- Schult, J. & Sparfeldt, J. R. (2016). Do non-g factors of cognitive ability tests align with specific academic achievements? A combined bifactor modeling approach. *Intelligence*, 59, 96-102.
- Schwarz, G. (1978). Estimating the dimensions of a model. *The Annals of Statistics*, 6, 461-464.
- Sinharay, S. (2010). How often do Subscores have added value? Results from Operational and simulated data. *Journal of Educational Measurement*, 47, 150-174.

- Sparfeldt, J. R., Buch, S. R., Rost, D. H. & Lehmann, G. (2008). Akkuratessse selbstberichteter Zensuren. *Psychologie in Erziehung und Unterricht*, 55, 68–75.
- Sparfeldt, J. R. & Rost, D. H. (2008). Leserbrief zu „Hochbegabtendiagnostik mit dem HAWIK-III und AID 2“ (Schlagheck & Petermann, 2006, *Kindheit und Entwicklung*, 15, 93-99). *Kindheit und Entwicklung*, 17, 68.
- Spearman, C. (1904). “General intelligence”, objectively determined and measured. *American Journal of Psychology*, 15, 201-292.
- Spearman, C. (1925) Some Issues in the Theory of “g” (including the Law of Diminishing Returns). *Nature*, 116, 436-439.
- Spearman, C. (1927). *The Abilities of Man: Their Nature and Measurement*. Macmillan: New York, USA, 1927.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.
- Stelzl, I. (1982). Der Ärger mit den Differenzen. In D. H. Rost (Hrsg.), *Fehler und Fallen der Statistik, Standardwerke aus Psychologie und Pädagogik* (S. 184-210). Münster: Waxmann.
- Sternberg, R. J. & Davidson, J. E. (Eds.) (2005). *Conceptions of giftedness (2nd ed.)*. New York, NY: Cambridge University Press.
- Süß, H.-M. (2001). Prädiktive Validität der Intelligenz im schulischen und außerschulischen Bereich. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 109–135). Lengerich: Pabst.
- Süß, H.-M., Kersting, M. & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen. *Diagnostica*, 37, 334-352.
- 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.
- Tent, L. & Birkel, P. (2010). Zensuren. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (4. Aufl., S. 949–958). Weinheim: Beltz.
- Tewes U., Rossmann, P. & Schallberger, U. (1999). *Hamburg-Wechsler-Intelligenztest für Kinder (3. Aufl.)*. Bern: Huber.
- Thorndike, R. L. (1987). Stability of factor loadings. *Personality and individual Differences*, 8, 585–586.

-
- Valerius, S. & Sparfeldt, J. R. (2014). Consistent *g*- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory Factor Analyses using three test batteries. *Intelligence*, *44*, 120–133.
- Valerius, S. & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell. *Zeitschrift für Pädagogische Psychologie*, *29*, 101-108.
- Vernon, P. A. (1989). The generality of *g*. *Personality and Individual Differences*, *10*, 803-804.
- Wai, J. & Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence*, *39*, 443--455.
- Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: The Psychology Cooperation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2003). *Wechsler Intelligence Scale for Children – Fourth Edition*. San Antonio, TX: Psychological Corporation.
- Wee, S. (2018). Aligning predictor-criterion bandwidths: specific abilities as predictors of specific performance. *Journal of Intelligence*, *6*, 40.
- Weiß, R. H. (2006). *CFT 20-R. Grundintelligenztest Skala 2*. Göttingen, Germany: Hogrefe.
- Woodcock, R. W., McGrew, K. S. & Mather, N. (2001). *Woodcock-Johnson Psychoeducational Battery – Third Edition*. Itasca, IL: Riverside.
- Yung, Y.-F., Thissen, D. & McLeod L. D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika*, *65*, 113-128.

ANHANG
A. Mplus Syntaxen

Im Folgenden werden die *Mplus* Syntaxen dargestellt, die den Analysen zu den ersten beiden Artikeln dieser Dissertation zugrunde lagen.

A 1 Mplus Syntax für das kombinierte Higher Order Modell der KFT-, BIS- und Fluid-Batterie (Valerius & Sparfeldt, 2014)

```

title: Konfirmatorische Faktorenanalyse (CFA) für kombiniertes Modell: KFT-HO, BIS-HO, Fluid
data:  file=alle_01-06-12_oes.dat;

variable:names= vpn klasse m d e f p g
          bd og lo an ch tg st ef wa tm xg zp dr zn sc
          cft1 cft2 cft3 cft4 spm
          zvt_a zvt_b zvt_c zvt_d zvt_ges
          v1 v2 q1 q2 n1 n2
          v1_1 v1_2 v2_1 v2_2 q1_1 q1_2 q2_1 q2_2 n1_1 n1_2 n2_1 n2_2;

usevariables=
! Subtests des BIS-V, BIS-N und BIS-F
AN BD CH DR EF LO OG SC ST TG TM WA XG ZN ZP

! Subtests des KFT-V, KFT-Q und KFT-N
V1_1 V1_2 V2_1 V2_2 ! V1 und V2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
Q1_1 Q1_2 Q2_1 Q2_2 ! Q1 und Q2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
N1_1 N1_2 N2_1 N2_2 ! N1 und N2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel

! Subtests des CFT 20-R und SPM; zusammen die "Fluid-Batterie"
CFT1 CFT2 CFT3 CFT4 SPM;

missing=all(999);      ! Fehlende Werte werden identifiziert durch den Wert 999
cluster=klasse;        ! Die Beobachtungen sind innerhalb der Klassen geclustert

analysis:  type=complex; ! Die Analyse berücksichtigt die Clusterung innerhalb der Klassen

model:     ! CFA für kombiniertes Higher Order Modell

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! BIS HO-Modell

! Verbaler Inhaltsfaktor: BIS-V
v by tm*          ! freigesetzt / Fixierung aufgehoben
ef wa tg st;

! Numerischer Inhaltsfaktor: BIS-N
n by zn*          ! freigesetzt / Fixierung aufgehoben
xg zp sc dr;

! Figuraler Inhaltsfaktor: BIS-F
f by an*          ! freigesetzt / Fixierung aufgehoben
lo bd og ch;

```

```

v@1;          ! fixiert auf 1
n@1;          ! fixiert auf 1
f@1;          ! fixiert auf 1

! Second Order g-Faktor für die BIS-Batterie: BIS-g
  bis by v* n f;
  bis@1;      ! fixiert auf 1

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! KFT HO-Modell

! Verbaler Faktor: KFT-V
  kftv by   v1_1-v1_2*      ! freigesetzt / Fixierung aufgehoben
           (lv1)           ! Parameter gleichgesetzt
           v2_1-v2_2
           (lv2);          ! Parameter gleichgesetzt
  kftv@1;    ! fixiert auf 1

! Quantitativer Faktor: KFT-Q
  kftq by   q1_1-q1_2*      ! freigesetzt / Fixierung aufgehoben
           (lq1)           ! Parameter gleichgesetzt
           q2_1-q2_2
           (lq2);          ! Parameter gleichgesetzt
  kftq@1;    ! fixiert auf 1

! Nonverbaler (figuraler) Faktor: KFT-N
  kftn by   n1_1-n1_2*      ! freigesetzt / Fixierung aufgehoben
           (ln1)           ! Parameter gleichgesetzt
           n2_1-n2_2
           (ln2);          ! Parameter gleichgesetzt
  kftn@1;    ! fixiert auf 1

! Second Order g-Faktor für den KFT: KFT-g
  kft_g by kftv*           ! freigesetzt / Fixierung aufgehoben
  kftq kftn;
  kft_g@1;                 ! fixiert auf 1

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! "Fluid-Batterie": Fluid-g
  fluid by cft1*           ! freigesetzt / Fixierung aufgehoben
  cft2 cft3 cft4 spm;
  fluid@1;                 ! fixiert auf 1

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

output: sampstat standardized tech1 tech2 residual;

```

A 2 Mplus Syntax für das kombinierte Nested-Factor Modell der KFT-, BIS- und Fluid-Batterie (Valerius & Sparfeldt, 2014)

```

title: Konfirmatorische Faktorenanalyse (CFA) für kombiniertes Nested Factor Modell: KFT-NF,
      BIS-NF, Fluid

data:  file=alle_01-06-12_oes.dat;

variable: names= vpn klasse m d e f p g
            bd og lo an ch tg st ef wa tm xg zp dr zn sc
            cft1 cft2 cft3 cft4 spm
            zvt_a zvt_b zvt_c zvt_d zvt_ges
            v1 v2 q1 q2 n1 n2
            v1_1 v1_2 v2_1 v2_2 q1_1 q1_2 q2_1 q2_2 n1_1 n1_2 n2_1 n2_2;

usevariables=
! Subtests des BIS-V, BIS-N und BIS-F
AN BD CH DR EF LO OG SC ST TG TM WA XG ZN ZP

! Subtests des KFT-V, KFT-Q und KFT-N
V1_1 V1_2 V2_1 V2_2 ! V1 & V2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
Q1_1 Q1_2 Q2_1 Q2_2 ! Q1 & Q2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
N1_1 N1_2 N2_1 N2_2 ! N1 & N2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel

! Subtests des CFT 20-R und SPM; zusammen die "Fluid-Batterie"
CFT1 CFT2 CFT3 CFT4 SPM;

missing=all(999); ! Fehlende Werte werden identifiziert durch den Wert 999
cluster=klasse;  ! Die Beobachtungen sind innerhalb der Klassen geclustert

analysis: type=complex;! Die Analyse berücksichtigt die Clusterung innerhalb der Klassen

model:  ! CFA für kombiniertes Nested Factor Modell

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! BIS NF-Modell

! Verbaler Inhaltsfaktor: BIS-V
v by tm*      ! freigesetzt / Fixierung aufgehoben
ef wa tg st;

! Numerischer Inhaltsfaktor: BIS-N
n by zn*      ! freigesetzt / Fixierung aufgehoben
xg zp sc dr;

! Figuraler Inhaltsfaktor: BIS-F
f by an*      ! freigesetzt / Fixierung aufgehoben
lo bd og ch;

v@1;          ! fixiert auf 1
n@1;          ! fixiert auf 1
f@1;          ! fixiert auf 1

v with n@0 f@0;
n with f@0;   ! orthogonale Inhaltsfaktoren BIS-V, BIS-N, BIS-F

! Orthogonaler Generalfaktor der BIS-Batterie: BIS-g
bis by AN*    ! freigesetzt / Fixierung aufgehoben
BD DR EF LO OG
SC ST TG TM WA
XG ZN ZP CH;
bis@1;        ! fixiert auf 1

bis with v@0 n@0 f@0 kftv@0 kftq@0 kftn@0;
! BIS-g orthogonal zu BIS-V, N, F sowie KFT-V, Q, N

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! KFT NF-Modell

```

```

! Verbaler Faktor: KFT-V
  kftv by  v1_1-v1_2*      ! freigesetzt / Fixierung aufgehoben
          (lv1)           ! Parameter gleichgesetzt
          v2_1-v2_2
          (lv2);          ! Parameter gleichgesetzt
          kftv@1;        ! fixiert auf 1

! Quantitativer Faktor: KFT-Q
  kftq by  q1_1-q1_2*      ! freigesetzt / Fixierung aufgehoben
          (lq1)           ! Parameter gleichgesetzt
          q2_1-q2_2
          (lq2);          ! Parameter gleichgesetzt
          kftq@1;        ! fixiert auf 1

! Nonverbaler (figuraler) Faktor: KFT-N
  kftn by  n1_1-n1_2*      ! freigesetzt / Fixierung aufgehoben
          (ln1)           ! Parameter gleichgesetzt
          n2_1-n2_2
          (ln2);          ! Parameter gleichgesetzt
          kftn@1;        ! fixiert auf 1

  kftv with kftq@0 kftn@0;
  kftq with kftn@0;      ! orthogonale KFT-Faktoren KFT-V, KFT-Q, KFT-N

! Orthogonaler Generalfaktor der KFT-Batterie: KFT-g
  kft_g by v1_1-v1_2*      ! freigesetzt / Fixierung aufgehoben
          v2_1-v2_2 q1_1-q1_2 q2_1-q2_2
          n1_1-n1_2 n2_1-n2_2;
          kft_g@1;        ! fixiert auf 1

  kft_g with kftv@0 kftq@0 kftn@0 v@0 n@0 f@0;
  ! KFT-g orthogonal zu KFT-V, Q, N sowie BIS-V, N, F

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! "Fluid-Batterie: Fluid-g

  fluid by cft1*           ! freigesetzt / Fixierung aufgehoben
          cft2 cft3 cft4 spm;
          fluid@1;        ! fixiert auf 1

  fluid with v@0 n@0 f@0 kftv@0 kftq@0 kftn@0;
  ! Fluid-g orthogonal zu BIS-V, N, F sowie KFT-V, Q, N

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

output: sampstat standardized tech1 tech2 residual;

```

A 3 Mplus Syntax für das batteriespezifische Generalfaktormodell der KFT-, BIS- und Fluid-Batterie (Valerius & Sparfeldt, 2014)

```
title: Konfirmatorische Faktorenanalyse (CFA) für batteriespezifisches Generalfaktormodell
      (GF-battery specific)

data:  file=alle_01-06-12_oes.dat;

variable: names= vpn klasse m d e f p g
          bd og lo an ch tg st ef wa tm xg zp dr zn sc
          cft1 cft2 cft3 cft4 spm
          zvt_a zvt_b zvt_c zvt_d zvt_ges
          v1 v2 q1 q2 n1 n2
          v1_1 v1_2 v2_1 v2_2 q1_1 q1_2 q2_1 q2_2 n1_1 n1_2 n2_1 n2_2;

usevariables=
! Subtests des BIS-V, BIS-N und BIS-F
AN BD CH DR EF LO OG SC ST TG TM WA XG ZN ZP

! Subtests des KFT-V, KFT-Q und KFT-N
V1_1 V1_2 V2_1 V2_2 ! V1 & V2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
Q1_1 Q1_2 Q2_1 Q2_2 ! Q1 & Q2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
N1_1 N1_2 N2_1 N2_2 ! N1 & N2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel

! Subtests des CFT 20-R und SPM; zusammen die "Fluid-Batterie"
CFT1 CFT2 CFT3 CFT4 SPM;

missing=all(999); ! Fehlende Werte werden identifiziert durch den Wert 999
cluster=klasse;  ! Die Beobachtungen sind innerhalb der Klassen geclustert

analysis: type=complex; ! Analyse berücksichtigt Klassencluster

model:  ! CFA für batteriespezifisches Generalfaktorenmodell (GF-battery specific)

! BIS-Generalfaktor-Modell
bis by AN*
BD CH DR EF LO OG SC
ST TG TM WA XG ZN ZP;
bis@1;

! KFT-Generalfaktor-Modell
kft by V1_1*
V1_2 V2_1 V2_2
Q1_1 Q1_2 Q2_1 Q2_2
N1_1 N1_2 N2_1 N2_2;
kft@1;

! Fluid-Generalfaktor-Modell
fluid by cft1*
cft2 cft3 cft4 spm;
fluid@1;

output: sampstat standardized residual;
```

A 4 Mplus Syntax für das Generalfaktormodell (Valerius & Sparfeldt, 2014)

```

title: Konfirmatorische Faktorenanalyse (CFA) für Generalfaktormodell (GF-general)

data: file=alle_01-06-12_oes.dat;

variable: names= vpn klasse m d e f p g
           bd og lo an ch tg st ef wa tm xg zp dr zn sc
           cft1 cft2 cft3 cft4 spm
           zvt_a zvt_b zvt_c zvt_d zvt_ges
           v1 v2 q1 q2 n1 n2
           v1_1 v1_2 v2_1 v2_2 q1_1 q1_2 q2_1 q2_2 n1_1 n1_2 n2_1 n2_2;

usevariables=
! Subtests des BIS-V, BIS-N und BIS-F
AN BD CH DR EF LO OG SC ST TG TM WA XG ZN ZP

! Subtests des KFT-V, KFT-Q und KFT-N
V1_1 V1_2 V2_1 V2_2 ! V1 & V2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
Q1_1 Q1_2 Q2_1 Q2_2 ! Q1 & Q2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
N1_1 N1_2 N2_1 N2_2 ! N1 & N2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel

! Subtests des CFT 20-R und SPM; zusammen die "Fluid-Batterie"
CFT1 CFT2 CFT3 CFT4 SPM;

missing=all(999); ! Fehlende Werte werden identifiziert durch den Wert 999
cluster=klasse; ! Die Beobachtungen sind innerhalb der Klassen geclustert

analysis: type=complex; ! Analyse berücksichtigt Klassencluster

model: ! CFA für Generalfaktorenmodell (GF-general)

       g by AN*                               ! freigesetzt / Fixierung aufgehoben
       BD CH DR EF LO OG SC ST
       TG TM WA XG ZN ZP
       V1_1 V1_2 V2_1 V2_2 Q1_1 Q1_2 Q2_1 Q2_2
       N1_1 N1_2 N2_1 N2_2
       cft1 cft2 cft3 cft4 spm;
       g@1;                                     ! fixiert auf 1

output: sampstat standardized residual;

```

A 5 Mplus Syntax für das kombinierte Bifactor (S-1) Modell der Zusatzanalyse zu Valerius & Sparfeldt (2014)

```

title: CFA für kombiniertes Bifactor (S-1) Modell: KFT-NF, BIS-NF, Fluid

data:  file=alle_01-06-12_oes.dat;

variable: names= vpn klasse m d e f p g
          bd og lo an ch tg st ef wa tm xg zp dr zn sc
          cft1 cft2 cft3 cft4 spm
          zvt_a zvt_b zvt_c zvt_d zvt_ges
          v1 v2 q1 q2 n1 n2
          v1_1 v1_2 v2_1 v2_2 q1_1 q1_2 q2_1 q2_2 n1_1 n1_2 n2_1 n2_2;

usevariables=
! Subtests des BIS-V, BIS-N und BIS-F
AN BD CH DR EF LO OG SC ST TG TM WA XG ZN ZP

! Subtests des KFT-V, KFT-Q und KFT-N
V1_1 V1_2 V2_1 V2_2 ! V1 & V2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
Q1_1 Q1_2 Q2_1 Q2_2 ! Q1 & Q2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel
N1_1 N1_2 N2_1 N2_2 ! N1 & N2 jew. gebildet aus 2 Subtesthälften "even & odd" parcel

! Subtests des CFT 20-R und SPM; zusammen die "Fluid-Batterie"
CFT1 CFT2 CFT3 CFT4 SPM;

missing=all(999); ! Fehlende Werte werden identifiziert durch den Wert 999
cluster=klasse;  ! Die Beobachtungen sind innerhalb der Klassen geclustert

analysis: type=complex;! Die Analyse berücksichtigt die Clusterung innerhalb der Klassen

model:  ! CFA für kombiniertes Nested Factor Modell

!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! BIS NF-Modell (S-1)

! Verbaler Inhaltsfaktor: BIS-V
v by tm*          ! freigesetzt / Fixierung aufgehoben
ef wa tg st;

! Numerischer Inhaltsfaktor: BIS-N
n by zn*          ! freigesetzt / Fixierung aufgehoben
xg zp sc dr;

! Figuraler Inhaltsfaktor: BIS-F
! f by an*        ! freigesetzt / Fixierung aufgehoben
! lo bd og ch;

v@1;              ! fixiert auf 1
n@1;              ! fixiert auf 1
! f@1;           ! fixiert auf 1

v with n@0;! f@0;
!n with f@0;     ! orthogonale Inhaltsfaktoren BIS-V, BIS-N, BIS-F

! Orthogonaler Generalfaktor der BIS-Batterie: BIS-g
bis by AN*        ! freigesetzt / Fixierung aufgehoben
BD DR EF LO OG
SC ST TG TM WA
XG ZN ZP CH;
bis@1;           ! fixiert auf 1

```

```

bis with v@0 n@0 !f@0
kftv@0 kftq@0;! kftn@0; ! BIS-g orthogonal zu BIS-V, N, F sowie KFT-V, Q, N
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! KFT NF-Modell (S-1)

! Verbaler Faktor: KFT-V
kftv by v1_1-v1_2* ! freigesetzt / Fixierung aufgehoben
      (lv1) ! Parameter gleichgesetzt
v2_1-v2_2
      (lv2); ! Parameter gleichgesetzt
kftv@1; ! fixiert auf 1

! Quantitativer Faktor: KFT-Q
kftq by q1_1-q1_2* ! freigesetzt / Fixierung aufgehoben
      (lq1) ! Parameter gleichgesetzt
q2_1-q2_2
      (lq2); ! Parameter gleichgesetzt
kftq@1; ! fixiert auf 1

! Nonverbaler (figuraler) Faktor: KFT-N
! kftn by n1_1-n1_2* ! freigesetzt / Fixierung aufgehoben
! (ln1) ! Parameter gleichgesetzt
! n2_1-n2_2
! (ln2); ! Parameter gleichgesetzt
! kftn@1; ! fixiert auf 1

kftv with kftq@0;! kftn@0;
! kftq with kftn@0; ! orthogonale KFT-Faktoren KFT-V, KFT-Q, KFT-N

! Orthogonaler Generalfaktor der KFT-Batterie: KFT-g
kft_g by v1_1-v1_2* ! freigesetzt / Fixierung aufgehoben
v2_1-v2_2 q1_1-q1_2 q2_1-q2_2
n1_1-n1_2 n2_1-n2_2;
kft_g@1; ! fixiert auf 1

kft_g with kftv@0 kftq@0 !kftn@0
v@0 n@0 ;!f@0;
! KFT-g orthogonal zu KFT-V, Q, N sowie BIS-V, N, F
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

! "Fluid-Batterie: Fluid-g

fluid by cft1* ! freigesetzt / Fixierung aufgehoben
cft2 cft3 cft4 spm;
fluid@1; ! fixiert auf 1

fluid with v@0 n@0! f@0
kftv@0 kftq@0;! kftn@0;
! Fluid-g orthogonal zu BIS-V, N, F sowie KFT-V, Q, N
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!

output: sampstat standardized tech1 tech2 residual;

```

A 6 Mplus Syntax für das Nested-Factor Modell allgemeiner und spezifischer Intelligenz- und Schulleistungsfaktoren (Valerius & Sparfeldt, 2015)

```

title: (CFA) für NF-Modell: g- und drei spezifische
       Intelligenzfaktoren (Verbal, Numerisch, Figural)
       korreliert mit allgemeinem und spezifischen Schulleistungsfaktoren (NF)

data: file=alle_01-06-12_oes.dat;

variable: names= vpn klasse m d e f p g
           bd og lo an ch tg st ef wa tm xg zp dr zn sc
           cft1 cft2 cft3 cft4 spm
           zvt_a zvt_b zvt_c zvt_d zvt_ges
           v1 v2 q1 q2 n1 n2
           v1_1 v1_2 v2_1 v2_2 q1_1 q1_2 q2_1 q2_2 n1_1 n1_2 n2_1 n2_2;

usevariables=
! Zensuren Mathe, Deutsch, Englisch, Physik, Geschichte, Französisch
m d e p g f

! BIS-Subtests
bd og lo an ch tg st ef wa tm xg zp dr zn sc

! KFT-Subtests
v1 v2 q1 q2 n1 n2;

missing=all(999);           ! Fehlende Werte werden identifiziert durch den Wert 999
cluster=klasse;             ! Die Beobachtungen sind innerhalb der Klassen geclustert
centering=groupmean(m e p g f); ! Zentrierung der Zensuren am Clustermittelwert

analysis: type=complex;     ! Die Analyse berücksichtigt Klassencluster

model: ! CFA für Nested Factor Modell: allgemeine und spezifische Intelligenz- und
       ! Schulleistungsfaktoren

       ! Spezifischer Intelligenzfaktor Verbal
       v by tm*               ! freigesetzt / Fixierung aufgehoben
       ef wa tg st v1 v2;
       v@1;                   ! fixiert auf 1

       ! Spezifischer Intelligenzfaktor Numerisch
       n by zn*               ! freigesetzt / Fixierung aufgehoben
       xg zp sc dr q1 q2;
       n@1;                   ! fixiert auf 1

       ! Spezifischer Intelligenzfaktor Figural
       fi by an*              ! freigesetzt / Fixierung aufgehoben
       lo bd og ch n1 n2;
       fi@1;                  ! fixiert auf 1

       v with n@0 fi@0;       ! Spezifische Intelligenzfaktoren orthogonal zueinander
       n with fi@0;

       ! Generalfaktor der Intelligenz: sg = Spearmans g
       sg by og*              ! freigesetzt / Fixierung aufgehoben
       an ch tg st ef wa tm xg
       dr zn sc v1 zp lo bd
       v2 q1 q2 n1 n2;
       sg@1;                  ! fixiert auf 1

       ! Allgemeiner Schulleistungsfaktor (sl)
       sl by d*               ! freigesetzt / Fixierung aufgehoben
       m e p g f;
       sl@1;                  ! fixiert auf 1

       ! Spezifisch verbaler Schulleistungsfaktor
       slv by d*              ! freigesetzt / Fixierung aufgehoben
       e f;
       slv@1;                 ! fixiert auf 1

       ! Spezifisch mathematisch/numerischer Schulleistungsfaktor
       sln by m*(1)           ! freigesetzt / Fixierung aufgehoben

```

```
p(1);                ! sowie Parameter gleichgesetzt
sln@1;              ! fixiert auf 1

! Orthogonalität der NF-Faktoren
slv with sln@0;     ! Spezifische SL-faktoren orthogonal zueinander
sl with slv@0 sln@0 ! Allg. SL orth. zu spez. verb. und num. SL-faktoren sowie
v@0 n@0 fi@0;       ! zu spez. verb., num. und fig. Intelligenzfaktoren

sg with v@0 n@0 fi@0 ! Int. Generalfak. orthogonal zu spez. Int.faktoren (V, N, F)
slv@0 sln@0;        ! sowie zu spez. verb. und math./nawi Schulleistungsfaktoren
```

```
output: sampstat standardized tech1 tech2 residual modindices;
```

A 7 Mplus Syntax für das Bifactor (S-1) Modell allgemeiner und spezifischer Intelligenzfaktoren korreliert mit allgemeinem und spezifischen NF-Schulleistungsfaktoren (Valerius & Sparfeldt, 2015)

```

title: (CFA) für Bifactor (S-1) Modell: g- und zwei spezifische
Intelligenzfaktoren (Verbal, Numerisch)
korreliert mit allgemeinem und spezifischen Schulleistungsfaktoren (NF)

data: file=alle_01-06-12_oes.dat;

variable: names= vpn klasse m d e f p g
bd og lo an ch tg st ef wa tm xg zp dr zn sc
cft1 cft2 cft3 cft4 spm
zvt_a zvt_b zvt_c zvt_d zvt_ges
v1 v2 q1 q2 n1 n2
v1_1 v1_2 v2_1 v2_2 q1_1 q1_2 q2_1 q2_2 n1_1 n1_2 n2_1 n2_2;

usevariables=
! Zensuren Mathe, Deutsch, Englisch, Physik, Geschichte, Französisch
m d e p g f

! BIS-Subtests
bd og lo an ch tg st ef wa tm xg zp dr zn sc

! KFT-Subtests
v1 v2 q1 q2 n1 n2;

missing=all(999); ! Fehlende Werte werden identifiziert durch den Wert 999
cluster=klasse; ! Die Beobachtungen sind innerhalb der Klassen geclustert
centering=groupmean(m e p g f); ! Zentrierung der Zensuren am Clustermittelwert

analysis: type=complex; ! Die Analyse berücks. die Clusterung innerhalb d. Klassen

model: ! CFA für Nested Factor Modell: allgemeine und spezifische Intelligenz und
!Schulleistungsfaktoren

! Spezifischer Intelligenzfaktor Verbal
v by tm* ! freigesetzt / Fixierung aufgehoben
ef wa tg st v1 v2;
v@1; ! fixiert auf 1

! Spezifischer Intelligenzfaktor Numerisch
n by zn* ! freigesetzt / Fixierung aufgehoben
xg zp sc dr q1 q2;
n@1; ! fixiert auf 1

v with n@0; ! Spezifische Intelligenzfaktoren orthogonal zueinander

! Generalfaktor der Intelligenz: sg = Spearmans g
sg by og* ! freigesetzt / Fixierung aufgehoben
an ch tg st ef wa tm xg
dr zn sc v1 zp lo bd
v2 q1 q2 n1 n2;
sg@1; ! fixiert auf 1

! Allgemeiner Schulleistungsfaktor (sl)
sl by d* ! freigesetzt / Fixierung aufgehoben
m e p g f;
sl@1; ! fixiert auf 1

! Spezifisch verbaler Schulleistungsfaktor
slv by d* ! freigesetzt / Fixierung aufgehoben
e f;
slv@1; ! fixiert auf 1

! Spezifisch mathematisch/numerischer Schulleistungsfaktor
sln by m*(1) ! freigesetzt / Fixierung aufgehoben
p(1); ! sowie Parameter gleichgesetzt
sln@1; ! fixiert auf 1

```

```
! Orthogonalität der NF-Faktoren
slv with sln@0;           ! Spezifische SL-faktoren orthogonal zueinander
sl with slv@0 sln@0      ! Allg. SL orth. zu spez. verb. und num. SL-faktoren sowie
v@0 n@0;                 ! zu spez. verb., num. Intelligenzfaktoren

sg with v@0 n@0          ! Int. Generalfak. orthogonal zu spez. Int.faktoren (V, N)
slv@0 sln@0;            ! sowie zu spez. verb. und math./nawi Schulleistungsfaktoren

output: sampstat standardized tech1 tech2 residual modindices;
```

B. Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Ferner habe ich die Stellen der Arbeit, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, unter Angabe der jeweiligen Quelle als Entlehnung kenntlich gemacht. Dies gilt ebenso für Tabellen, Abbildungen sowie bildliche und grafische Darstellungen, die nicht eigenhändig von mir angefertigt wurden.

Ort, Datum

Unterschrift

C. Publikationen

Der nachfolgende Abschnitt enthält die bei den jeweiligen Zeitschriftenverlagen veröffentlichten Artikel der ersten beiden Studien dieser Dissertation sowie das Manuskript der dritten Studie, welches sich aktuell im Begutachtungsprozess befindet.

C 1 Publikation Studie 1

Valerius, S. & Sparfeldt, J. R. (2014). Consistent g- as well as consistent verbal-, numerical- and figural-factors in nested factor models? Confirmatory Factor Analyses using three test batteries. *Intelligence*, 44, 120-133.

C 2 Publikation Studie 2

Valerius, S. & Sparfeldt, J. R. (2015). Zusammenhänge allgemeiner und spezifischer Intelligenzfaktoren mit allgemeinen und spezifischen Schulleistungen im Nested-Factor-Modell. *Zeitschrift für Pädagogische Psychologie*, 29, 101–108.

C 3 Manuskript Studie 3

Valerius, S., Lotz, C. & Sparfeldt, J. R. (eingereicht) „Hochbegabt“ oder „nicht hochbegabt“ – zur Übereinstimmung der Hochbegabungsidentifikation mit SPM, CFT 20-R und KFT-N. *Psychologie in Erziehung und Unterricht*.

Aus urheberrechtlichen Gründen werden die Zeitschriftenartikel in der Online-Publikation nicht abgebildet. Diese sind über die verlagseigenen Homepages verfügbar.