

Empirische Sonderpädagogik, 2015, Nr. 2, S. 99-116
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Befragung von Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf Lernen: Ergebnisse zur Messinvarianz

Lena Nusser¹, Claus H. Carstensen^{1,2} & Cordula Artelt²

¹ Leibniz-Institut für Bildungsverläufe e.V. (IfBi)

² Otto-Friedrich-Universität Bamberg

Zusammenfassung

Dieser Artikel betrachtet die messinvariante Erfassung bildungsrelevanter Konstrukte mit Hilfe schriftlicher Befragungen bei Schülerinnen und Schülern an Förderschulen und Hauptschulen in der 5. Jahrgangsstufe. Um optimale Administrationsbedingungen für Schülerinnen und Schüler mit sonderpädagogischem Förderbedarf Lernen zu erforschen, wurde ein experimentelles Design implementiert. Inwieweit angepasste Erhebungsinstrumente und unterschiedliche Beschulungsformen sowie anzunehmende Kompetenzunterschiede der Befragten eine messäquivalente Erfassung der Lesemotivation und des akademischen Selbstkonzeptes ermöglichen, wird durch Mehrgruppenvergleiche konfirmatorischer Faktorenanalysen untersucht. Die Ergebnisse deuten darauf hin, dass vergleichende Analysen zwischen Schülergruppen an Förderschulen und Hauptschulen für bestimmte Konstrukte und Faktoren sinnvoll interpretierbar sind.

Schlüsselwörter: Messinvarianz, Mehrgruppenvergleich, sonderpädagogischer Förderbedarf, Lesemotivation, akademisches Selbstkonzept

Questionnaires for Students with Special Educational Needs in the Area of Learning: Results from Multi-Group Analysis

Abstract

This article focuses on measurement invariance of the assessment of educationally relevant constructs via written questionnaires for students at special schools and at low track schools attending 5th grade. To examine optimal conditions of administration for students with special educational needs in the area of learning an experimental design was implemented. If accommodated questionnaires, different school enrollments as well as competence differences allow equivalent assessment of reading motivation and academic self-concepts will be investigated with multi-group comparison of confirmatory factor analysis. The results indicate that comparisons between groups of students at special schools and low track schools are meaningful for certain constructs.

Keywords: measurement invariance, multi group comparison, special educational needs, reading motivation, academic self-concept

Persönlichkeitseigenschaften, sozioökonomische oder ethnische sowie weitere personenbezogene Merkmale werden in der sozialwissenschaftlichen und psychologischen Forschung häufig mit Hilfe von Fragebögen erfasst (Scholl, 2003). Der Einsatz standardisierter Instrumente ist in diesem Kontext die häufigste Form, um den Anforderungen einer objektiven Erfassung gerecht zu werden. Ziel solcher Befragungen sind nicht zuletzt Gruppenvergleiche und differenzierte Analysen, um Disparitäten in Ausprägung und auch Veränderungen dieser Merkmale zu untersuchen. Voraussetzung hierfür ist die messäquivalente oder auch messinvariante Erfassung der zugrundeliegenden Konstrukte. Andernfalls würden verschiedene Ausprägungen eines Konstruktes, z.B. in Abhängigkeit von Geschlecht, nicht ein und dasselbe Konstrukt in beiden Gruppen in äquivalenter Art und Weise abbilden. Mittelwertvergleiche wären in einem solchen Fall unzulässig, da sie zu verzerrten Ergebnissen führen würden. Gerade aber das hohe Maß an Standardisierung zur objektiven Erfassung von Konstrukten kann bei bestimmten Personengruppen einer validen und messinvarianten Erfassung entgegenstehen (Lipski, 2000). Dies betrifft vor allem Kinder mit Entwicklungsverzögerungen (Kränzl-Nagl & Wilk, 2000), wie sie auch bei Schülerinnen und Schülern mit einer Lernbehinderung zu finden sind.

Zudem gibt es empirische Hinweise, dass das Leistungsniveau und das Leseverständnis der Befragten Effekte auf die Vergleichbarkeit der Angaben haben können, so dass Daten von Schülerinnen und Schülern unterschiedlicher Schultypen nicht zwingend äquivalent sein müssen (Byrne, Shavelson & Muthén, 1989; Steinmetz, Schmidt, Tina-Booh, Wiczorek & Schwartz, 2009). Möchte man jedoch vergleichende Analysen verschiedener Schülergruppen durchführen, kommt der Messinvarianz eine zentrale Bedeutung zu.

Dies ist insbesondere nicht trivial, wenn Daten von Schülerinnen und Schülern an Förderschulen mit anderen Schülergruppen

verglichen werden. Eine schriftliche Befragung von Förderschülerinnen und -schülern stellt eine besondere Herausforderung dar, da sie heterogene Kompetenzprofile und häufig Einschränkungen in der Lesefertigkeit, im sprachlichen Verständnis und in der Aufmerksamkeitsspanne aufweisen (Grünke, 2004). Es stellt sich deshalb die Frage, wie Fragebögen für Schülerinnen und Schüler mit kognitiven Beeinträchtigungen bzw. einem sonderpädagogischen Förderbedarf Lernen (SPF-L) gestaltet werden können, um zum Beispiel bildungsrelevante Faktoren messinvariant zu anderen Schülergruppen ohne Beeinträchtigung zu erheben und somit vergleichende Analysen sinnvoll interpretierbar sind.

Hierfür wurde für Schülerinnen und Schüler an Förderschulen mit dem Schwerpunkt Lernen ein experimentelles Design zur Gestaltung von schriftlichen Befragungen entwickelt und ein angepasster Fragebogen konzipiert, der die Besonderheiten dieser Gruppe berücksichtigen soll. Ob bildungsrelevante Konstrukte trotz dieser Anpassungen bei Schülerinnen und Schülern an Förderschulen im Vergleich zu jenen an Hauptschulen messäquivalent erhoben werden können, wird im Folgenden untersucht.

Schülerinnen und Schüler an Förderschulen Lernen

Schülerinnen und Schüler, bei denen ein sonderpädagogischer Förderbedarf diagnostiziert wird, werden in Deutschland nicht nur integrativ, sondern überwiegend im stark ausdifferenzierten Förderschulsystem in kleineren Lerngruppen unterrichtet. Die Zuweisung zu einer sonderpädagogischen Fördereinrichtung erfolgt basierend auf der Empfehlung der Ständigen Kultusministerkonferenz zu einer der acht spezifizierten Förderschwerpunkte, wie beispielsweise Lernen, Sprache, Sehen oder Hören (Ständige Konferenz der Kultusminister der Länder, 1994). Die größte Gruppe der Schülerinnen und Schüler stellt mit anteilig fast 40% jene mit einem Förderbedarf Lernen dar (Bil-

dungsbericht, 2014). Da insbesondere für den deutschsprachigen Raum eine einheitlich anzuwendende Definition für dieses Phänomen noch aussteht, existieren bisher keine konsistenten Standards zur Diagnostik des sonderpädagogischen Förderbedarfs Lernen (Bildungsbericht, 2014; Bos, Müller & Stubbe, 2010). Aus diesem Grund ist an Förderschulen mit dem Schwerpunkt Lernen eine äußerst heterogene Schülerschaft mit einer stark ausgeprägten Leistungsspanne zu finden, die sich hinsichtlich ihrer Kompetenzprofile stark unterscheiden können (Bos et al., 2010; Gebhardt, Oelkrug & Tretter, 2013).

In Annäherung an das Erscheinungsbild wird das Vorliegen eines SPF-L als überdauernde und weitreichende Einschränkung bei der Bewältigung schulischer Anforderungen beschrieben (Klauer & Lauth, 1997), die sich im Besonderen bei dem „Erwerb kognitiv-verbaler und abstrakter Inhalte (z. B. Lesen, Rechnen)“ (Grünke, 2004, S. 65) zeigt. Als weitere Charakteristika werden diesen Schülerinnen und Schülern schwächere sprachliche Fähigkeiten, geringe Lernumfänge als auch reduzierte Aufmerksamkeit zugeschrieben.

Bisher wurde diese Gruppe von Schülerinnen und Schülern selten bzw. lediglich mit geringen Stichprobenumfängen in large-scale-assessments integriert. Diese Tatsache bedingt sich auch dadurch, dass keine Standards existieren, wie diese Schülergruppe sinnvoll und vergleichbar in standardisierte Befragungen und Kompetenztestungen einbezogen werden kann (Hoermann, 2007). Um den Zugang zum Aufgabenmaterial für bestimmte Schülergruppen mit Beeinträchtigungen zu erleichtern, sowie konstrukt-irrelevante und behinderungsbezogene Einschränkungen zu verringern, werden häufig verschiedene Akkommodationen im Rahmen von Kompetenztests implementiert (Cormier, Altman, Shyyan & Thurlow, 2010). Hierzu gehören zum Beispiel die Darbietung der Aufgaben in Brailleschrift für Kinder und Jugendliche mit visuellen Einschränkungen oder aber eine Reduktion

der Aufgaben, mehr Testzeit oder auch Read-Aloud-Akkommodationen für Schülergruppen mit kognitiven Einschränkungen (Koretz & Barton, 2004). Insgesamt sind Forschungsergebnisse zu Akkommodationen und die Vergleichbarkeit von Testergebnissen von Schülerinnen und Schülern mit und ohne Behinderung sehr heterogen, so dass diskutiert wird, ob bestimmte Anpassungen die Instrumente möglicherweise so stark verändern, dass angepasste und nicht angepasste Aufgaben nicht mehr vergleichbar und messinvariant sind (Cormier et al., 2010).

Messinvarianz

Messinvarianz oder auch Messäquivalenz fragt danach, „whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute“ (Horn & McArdle, 1992, S. 117). Die Messinvarianzforschung findet vielseitig Beachtung in verschiedenen Disziplinen und Forschungskontexten.

Um zu prüfen, ob in verschiedenen Substichproben die zu untersuchenden Konstrukte messäquivalent erhoben werden und somit die strukturellen Beziehungen zwischen den Indikatoren und latenten Variablen invariant und vergleichbar sind, bietet sich die Mehrgruppenanalyse konfirmatorischer Faktorenmodelle als vielseitige und robuste Methode an (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). Dieses Vorgehen ermöglicht, dass die verschiedenen Modellparameter schrittweise hinsichtlich ihrer Messinvarianz untersucht werden (Chen, 2008; Sass, 2011).

Hierfür werden für die zu betrachtenden Gruppen die Faktorenmodelle simultan geschätzt und schrittweise Restriktionen in Form von gleichgesetzten Parametern über die Gruppen hinweg eingeführt. Starke faktorielle Invarianz herrscht dann, wenn die (a) konfigurale, (b) metrische und (c) skalare Invarianz gegeben ist (Byrne et al., 1989; Sass, 2011).

Zur Prüfung dieser Formen der Invarianz wird zunächst ein konfirmatorisches Modell, welches identische Messbeziehungen zwischen Indikatoren und Faktoren annimmt, für beide interessierenden Gruppen separat geschätzt. Diese Spezifikation dient in den nachfolgenden Analysen zur Messinvarianz mit simultanen Schätzungen als Baseline-Modell. In einem Modell zur (a) konfiguralen Invarianz wird die gleiche Anzahl an freien und fixierten Parametern in beiden Gruppen geschätzt. Die (b) metrische Invarianz nimmt gleiche Faktorladungen für die beobachteten Gruppen an; somit kann von einer identischen Beziehung zwischen den Indikatoren und ihren latenten Faktoren in allen Gruppen ausgegangen werden. Um die Annahme der (c) skalaren Invarianz zu erfüllen, werden zusätzlich die Intercepts der Indikatoren restringiert und über die Gruppen hinweg gleichgesetzt. Personen mit gleichen latenten Faktorwerten geben demnach ähnliche Antworten für die jeweiligen Items an (Sass, 2011). Auf dieser Stufe der Messäquivalenz sind Gruppenvergleiche der Faktormittelwerte sinnvoll interpretierbar (Brown, 2006; Dimitrov, 2010).

In der Praxis hat sich herausgestellt, dass die Hypothese der Messinvarianz für alle Parameter über verschiedene Gruppen hinweg sehr streng und selten realistisch ist. Aus diesem Grund haben Byrne et al. (1989) das Konzept der partiellen Invarianz eingeführt, das davon ausgeht, dass nicht alle Parameter invariant sind. Führen zusätzliche Modellrestriktionen zu einer signifikanten Verschlechterung des Modells, können auf Grund theoretischer Überlegungen und auch explorativer Vorgehensweise einzelne Indikatoren identifiziert werden, deren Parameter in den Gruppen frei geschätzt werden (Vandenberg & Lance, 2000). Um trotz eingeschränkter Messinvarianz sinnvolle Gruppenvergleiche für die gemessenen Konstrukte durchführen zu können, sollte jeder Faktor mit mindestens zwei Indikatoren, die metrische und skalare Invarianz aufweisen, repräsentiert sein (Byrne et al., 1989; Steenkamp & Baumgartner, 1998).

Fragestellung

Bezogen auf die dargelegten Herausforderungen bei der Befragung von Schülerinnen und Schülern mit SPF-L soll im Folgenden die messinvariante Erfassung der Konstrukte Lesemotivation und akademisches Selbstkonzept für die Gruppen der Schülerinnen und Schüler an Förderschulen und Hauptschulen untersucht werden. Grundsätzlich stellt sich die Frage, ob die Konstrukte in diesen beiden Gruppen die gleiche faktorielle Struktur aufweisen und somit die gleichen Indikatoren auf den entsprechenden latenten Faktoren laden. Ist diese Voraussetzung erfüllt, kann die Messinvarianz und somit die Vergleichbarkeit der Faktorladungen und der Intercepts für die beiden Schülergruppen untersucht werden.

Es ist anzunehmen, dass Schülerinnen und Schüler an Hauptschulen eher als die Gruppen an Förderschulen Lernen mit abstraktem sprachlichem Material arbeiten und gegebenenfalls vertrauter mit Test- und Befragungssituationen sind. Daher sollten sie vergleichsweise besser - sowohl dispositionell als auch situativ - in der Lage sein, sich mit dem Befragungsinstrument und den Items auseinanderzusetzen. Ihr Antwortverhalten würde mehr valide und konsistente Antworten liefern, welches die Faktorladungen und Intercepts beeinflusst.

Auf Grund reduzierter Aufmerksamkeitsspannen bei Schülerinnen und Schülern mit SPF-L könnten für diese Zielgruppe Positionseffekte erwartet werden, wenn Items nicht zu Beginn, sondern erst am Ende nach etwa 20- bis 25-minütiger Befragung präsentiert werden. Auf Grund gesunkener Konzentration und Bereitschaft zur Bearbeitung ist eine erhöhte Tendenz zu Response-Sets zu erwarten, die einer validen und messäquivalenten Erfassung entgegenstehen. Um diese Effekte zu untersuchen, erfolgte die Erfassung der Konstrukte bei dieser Schülergruppe mittels Fragebögen in zwei Varianten mit rotierter Itemreihenfolge. Deshalb wird auch hier die Messäquivalenz der Daten geprüft.

Methode

Stichprobe

Die zugrundeliegenden Daten stammen aus der Startkohorte 3 des Nationalen Bildungspanels¹ (NEPS; Blossfeld, Roßbach & von Maurice, 2011). Insgesamt nahmen in der 5. Jahrgangsstufe 57 Förderschulen mit dem Schwerpunkt Lernen aus dem ganzen Bundesgebiet teil; 587 Schülerinnen und Schüler beteiligten sich an der Erhebung, nachdem von ihren Erziehungsberechtigten eine Einverständniserklärung zur Teilnahme an der NEPS-Studie eingeholt worden war. Das durchschnittliche Alter der Schülergruppe betrug $M = 11.44$ Jahre ($SD = 0.65$), 44.1% der Teilnehmenden war weiblich. Als Referenz zur Stichprobe der Förderschülerinnen und -schüler dient die Schülergruppe an Hauptschulen. Insgesamt nahmen in dieser Schulform 745 Schülerinnen und Schüler aus 42 Institutionen teil. Das Durchschnittsalter lag bei $M = 11.30$ Jahre ($SD = 0.67$) und 45.5% der Schülergruppe war weiblich. Die Daten enthalten nicht für alle diese Schülerinnen und Schüler Informationen über einen vorhandenen sonderpädagogischen Förderbedarf. Auf Grund der Angaben des Statistischen Bundesamtes kann davon ausgegangen werden, dass nur eine geringe Anzahl von Personen mit SPF-L an Hauptschulen zu finden sind (Integrationsrate von 1.59% für das Schuljahr 2010/2011 [eigene Berechnung]; Statistisches Bundesamt, 2011).

Design

Die Erhebungen fanden an einem Schulvormittag bzw. an zwei Schulvormittagen im Klassenkontext statt. Die schriftliche Befragung der Schülerinnen und Schüler erfolgte

jeweils am Ende einer Testsitzung. Schülerinnen und Schüler an Hauptschulen erhielten einen umfangreichen Fragebogen, der neben der sozialen und ethnischen Herkunft auch weitere bildungsrelevante Aspekte erfragt (Frahm et al., 2011). Insgesamt dauerte die papierbasierte Befragung 40 Minuten. Nach einer kurzen Instruktion wurde der Fragebogen von den Schülerinnen und Schülern selbstständig ausgefüllt. Die Erhebung an Förderschulen unterlag einem experimentellen Design, um optimale Bedingungen für Befragungen und die Administrationsbedingungen zu erforschen (Heydrich, Weinert, Nusser, Artelt & Carstensen, 2013). So wurde das Instrument für diese Gruppe hinsichtlich (a) der Länge, (b) des Inhalts, (c) der Reihenfolge und (d) des Darbietungsmodus angepasst, um den spezifischen Bedürfnissen dieser Schülergruppe gerecht zu werden. Im Mittel dauerte die Befragung 36 am ersten bzw. 30 Minuten am zweiten Erhebungstag.

- a) Die Fragebogeninhalte wurden um etwa ein Drittel reduziert (von 185 auf 125 Items) und auf zwei Erhebungstage aufgeteilt. So konnten die inhaltlichen, zeitlichen und kognitiven Anforderungen an die Zielpersonen verringert werden, ohne auf zentrale Inhaltsbereiche verzichten zu müssen.
- b) Die Auswahl der Inhalte orientierte sich u. a. an der Bedeutsamkeit der Fragen für die besondere Zielgruppe. Beispielsweise wurden Fragen zum idealistischen und realistischen Schulabschluss entfernt, da die vorgegebenen Antwortmöglichkeiten nicht die Realität von Förderschülerinnen und -schüler widerspiegeln.
- c) Zur Identifizierung potentieller Positionseffekte durch abnehmende Aufmerksamkeit wurde das Instrument am zwei-

¹ Diese Arbeit nutzt Daten des Nationalen Bildungspanels (NEPS) Startkohorte 3 (Klasse 5), doi: 10.5157/NEPS:SC3:1.0.0. Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird NEPS vom Leibniz-Institut für Bildungsverläufe e.V. (IfBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.

ten Erhebungstag in zwei Varianten mit rotierter Itemreihenfolge dargeboten. Die Zuweisung der Instrumentenversionen 1 (vorwärts) und 2 (rückwärts) erfolgte zufällig auf Klassenebene.

- d) Um eine Strukturierungshilfe zu bieten und eine mögliche schwächere Lesefertigkeit als Hürde bei der validen Beantwortung zu mindern, wurden alle Items von geschulten Personen mit Hilfe eines standardisierten Skriptes vorgelesen.

Messung und Instrumente

Die Forschungsfragen zur Messinvarianz werden beispielhaft anhand verschiedener (Sub-)Skalen aus den Konstruktbereichen Lesemotivation und akademisches Selbstkonzept untersucht, deren theoretische Grundlage und Relevanz kurz vorgestellt werden.

Lesemotivation. Lesemotivation beschreibt die Absicht und das Bedürfnis einer Person, in einer Situation bestimmte Texte zu lesen (Schiefele, 1996). Die Lesemotivation wird als prädiktiv für die Lesekompetenz angesehen, da sich auch unter der Kontrolle kognitiver und sozialer Einflussvariablen signifikant positive Effekte zeigen (Artelt, Naumann & Schneider, 2010). Eine gering ausgeprägte Lesemotivation kann somit auch als eine relevante Ursache für eine niedrige Lesekompetenz verstanden werden. Zusammenhänge sind vor allem für den Faktor der intrinsischen Lesemotivation zu erwarten, der das Leseverhalten und die Häufigkeit von Leseaktivitäten und somit wiederum die Lesekompetenz positiv beeinflusst (McElvany, Kortenbruck & Becker, 2008; Möller & Schiefele, 2004).

Für die Erfassung der Lesemotivation wurden in die Instrumente des NEPS je drei Items für die Subskalen Leselust, Lesen aus Interesse und Selbstkonzept Lesen aus dem Fragebogen zur habituellen Lesemotivation (Möller & Bonerad, 2007) integriert (Frahm et al., 2011; s. Tabelle 1). Das administrierte Antwortformat weist eine vierstufige Li-

kert-Skala auf (1 = *stimme gar nicht zu*; 2 = *stimme eher nicht zu*; 3 = *stimme eher zu*; 4 = *stimme völlig zu*). Die Subskalen Leselust und Lesen aus Interesse weisen eine gute interne Konsistenz auf (Förderschule: $C\alpha = .839$ und $.760$; Hauptschule: $C\alpha = .900$ und $.846$), während die Werte für die Subskala Selbstkonzept Lesen nicht mehr akzeptable sind ($C\alpha = .524$ bzw. $.546$).

Akademisches Selbstkonzept. Das Selbstkonzept wird als die Selbstwahrnehmung einer Person aufgefasst, die durch Erfahrungen unterschiedlichster Art und deren Interpretation geformt wird (Shavelson, Hubner & Stanton, 1976). Für den schulischen Bereich ist besonders das akademische Selbstkonzept relevant, welches nach fachspezifischen Facetten differenziert erfasst werden kann (Byrne, 1996; Marsh, 1990). Es wird angenommen, dass das akademische Selbstkonzept als Fähigkeitskognition einer Person Zusammenhänge mit Leistungen, hier vor allem der schulischen Performanz, aufweist (Möller & Köller, 2004). In der NEPS-Schülerbefragung wurden drei Subskalen zum akademischen Selbstkonzept in Anlehnung an die PISA-Erhebung 2000 implementiert (Kunter et al., 2002), wobei zwischen dem verbalen, dem mathematischen und dem schulischen Selbstkonzept differenziert wird. Jede der drei Skalen wurde jeweils mit drei Items erfragt (Wohlkinger, Ditton, von Maurice, Haugwitz & Blossfeld, 2011; s. Tabelle 1). Das administrierte Antwortformat ist eine vierstufige Likert-Skala (1 = *trifft gar nicht zu*; 2 = *trifft eher nicht zu*; 3 = *trifft eher zu*; 4 = *trifft völlig zu*). Wie von Marsh (1986) vorgeschlagen, wird ein negativ-formuliertes Item aus den Analysen ausgeschlossen, da dieses zur Vermeidung einseitiger Antwortmuster dient und auf Grund unzureichender Validität nicht in die Skalenbildung einbezogen wird. Die Werte der internen Konsistenz der Subskalen betragen für die Stichprobe der Schülerinnen und Schüler mit SPF-L $C\alpha = .707$, $.813$ und $.806$ und liegen damit in einem

Tabelle 1: Items zur Erfassung der Lesemotivation und des akademischen Selbstkonzeptes

Itemformulierung	Förderschule			Hauptschule		
	N	M	SD	N	M	SD
Lesemotivation						
a) Es macht mir Spaß, Bücher zu lesen. (LL)	480	2.91	1.19	614	2.77	1.13
b) Ich finde Lesen interessant. (LL)	453	2.93	1.16	603	2.73	1.11
c) Wenn ich genügend Zeit hätte, würde ich noch mehr lesen. (LL)	463	2.58	1.24	605	2.53	1.17
d) Ich lese gern etwas über neue Dinge. (LI)	467	2.91	1.17	605	2.75	1.14
e) Ich bin überzeugt, dass ich beim Lesen eine Menge lernen kann. (LI)	467	3.15	1.10	605	2.91	1.11
f) Lesen ist wichtig, um Dinge richtig zu verstehen. (LI)	469	3.35	0.98	606	3.09	1.06
g) Ich habe manchmal Schwierigkeiten, einen Text wirklich gut zu verstehen. (r) (SK)	457	2.51	1.17	598	2.68	1.09
h) Ich kann Texte sehr gut und schnell verstehen. (SK)	465	2.89	1.13	599	2.78	1.01
i) Ich muss vieles erst mehrmals lesen, bevor ich es richtig verstanden habe. (r) (SK)	456	2.35	1.21	592	2.58	1.12
Akademisches Selbstkonzept						
a) Im Fach Deutsch bin ich ein hoffnungsloser Fall. (r) (V)	465	2.97	1.14	652	2.86	0.99
b) Im Fach Deutsch lerne ich schnell. (V)	475	3.02	1.06	661	2.83	0.91
c) Im Fach Deutsch bekomme ich gute Noten. (V)	458	2.97	1.07	649	2.72	0.86
d) Im Fach Mathematik bekomme ich gute Noten. (M)	469	3.14	1.06	658	2.99	0.93
e) Mathematik ist eines meiner besten Fächer. (M)	468	3.05	1.16	659	2.73	1.15
f) Ich war schon immer gut in Mathematik. (M)	466	2.89	1.13	649	2.59	1.09
g) In den meisten Schulfächern lerne ich schnell. (S)	469	3.06	1.01	657	3.06	0.81
h) In den meisten Schulfächern schneide ich in Klassenarbeiten gut ab. (S)	469	2.97	1.04	652	2.94	0.83
i) Ich bin in den meisten Schulfächern gut. (S)	473	3.27	0.94	660	3.08	0.81

Anmerkungen. LL = Leselust. LI = Lesen aus Interesse. SK = Selbstkonzept Lesen. V = Verbales Selbstkonzept. M = Mathematisches Selbstkonzept. S = Schulisches Selbstkonzept. (r) = rekodiert.

akzeptablen bis guten Bereich. An Hauptschulen liegen die Werte bei $Cr\alpha = .728$, $.861$ und $.820$.

Implementation in die Erhebungsinstrumente. Das Regelschulinstrument enthält die Items zu den drei Skalen des akademischen Selbstkonzeptes im mittleren Teil an den Positionen 84-92 der insgesamt 185 Items. Die Items zur Lesemotivation folgen an den Positionen 137-145. Alle Schülerinnen und

Schüler an Förderschulen erhielten die Fragen zu diesen Themenblöcken am zweiten Erhebungstag, an dem insgesamt 63 Items administriert wurden. Für die Vorwärts- und Rückwärts-Varianten der Instrumente wurden die Fragenblöcke modulweise rotiert, damit die inhaltlichen Zusammenhänge und logischen Reihenfolgen bestimmter Themengebiete erhalten blieben. In Instrument 1 wurden die Items zur Lesemotivation zu Beginn an den Positionen 9-17 prä-

sentiert, die Fragen zum akademischen Selbstkonzept folgten an den Positionen 18-26. Im rotierten Instrument 2 standen die Items zum akademischen Selbstkonzept an Position 28-36, während die Fragen zur Lesemotivation am Ende an den Positionen 55-63 folgten.

Methodisches Vorgehen

Zunächst wird die theoretisch postulierte Faktorstruktur in einer konfirmatorischen Faktorenanalyse (KFA) spezifiziert und hinsichtlich etwaiger Fehlspezifikationen mit Hilfe der Fitindizes geprüft. Nach der Bestätigung eines theoriegeleiteten Modells, dient dieses in den anschließenden Analysen des Mehrgruppenvergleiches als Baseline-Modell (Brown, 2006; Byrne & van de Vijver, 2010).

Das weitere Vorgehen folgt den im Theorieteil vorgestellten Analysenschritten zur Überprüfung der Messinvarianz. Bei diesem Bottom-Up-Ansatz werden zunächst alle Parameter frei geschätzt. Erst nach und nach werden Restriktionen in das Modell eingeführt, die Faktorladungen und Intercepts gleichsetzen. Da die Invarianz der Faktorladungen geprüft werden soll, werden diese Parameter zunächst alle frei geschätzt. Zu Identifikationszwecken des Modells wird die Varianz der latenten Faktoren auf 1 fixiert.

Die Beurteilung der Güte der Modelle erfolgt anhand der gängigen Fitindizes. Neben dem χ^2 -Wert, der stark abhängig von der Stichprobengröße ist (Cheung & Rensvold, 2002; Dimitrov, 2010), werden weitere Fitmaße herangezogen, die verschieden sensitiv gegenüber unterschiedlichen Aspekten der Modellpassung sind (Brown, 2006). Diese sind der *Root Mean Square Error of Approximation* (RMSEA), der *Standardized Root Mean Square Residual* (SRMR), der *Comparative Fit Index* (CFI) sowie der *Tucker Lewis Index* (TLI). Etabliert haben sich in der Forschungspraxis die Anwendungsregeln, dass RMSEA < .05, SRMR < .06 sowie TLI und CFI > .95 auf ei-

ne gute Modellgüte hindeuten (Hu & Bentler, 1999).

Für die Durchführung der KFA und Mehrgruppenvergleiche gilt die Voraussetzung der Normalverteilung der Variablen (Dimitrov, 2010), die im vorliegenden Fall nicht gegeben ist. Aus diesem Grund wird die robustere Schätzmethode *Restricted Maximum Likelihood* (MLR) gewählt. Fehlende Werte werden als *missing at random* behandelt. Somit werden im Sinne des *Full Information Maximum Likelihood*-Ansatzes auch unvollständige Datenreihen einbezogen (Sass, 2011). Auch wenn keine Variablen unterschiedlicher Ebenen simultan analysiert werden, wird die genestete Datenstruktur mit zwei Ebenen (Klassen, Schülerinnen und Schüler) in den Analysen berücksichtigt und eine Korrektur des Standardfehlers vorgenommen.

Die Beurteilung der Veränderungen der Passung bei schrittweise eingeführten Restriktionen und Gleichsetzungen der Parameter über die Substichproben hinweg erfolgt mittels des Satorra-Bentler- χ^2 -Differenztests (Satorra & Bentler, 2001). Die Invarianz der Parameter ist gegeben, wenn dieser nicht signifikant ausfällt.

Für alle Analysen wurde die Software Mplus Version 7 (Muthén & Muthén, 2012) genutzt.

Ergebnisse

Messinvarianz zu Skalen der Lesemotivation

Baseline-Modell. Wird mittels einer KFA die anzunehmende dreifaktorielle Struktur erzwungen, ergibt sich keine ausreichend gute Modellpassung für die beiden Stichproben. Item h lädt bei hohen Residuen auf allen drei Dimensionen. Dieses Item entstammt der Subskala Selbstkonzept Lesen, welche bereits durch eine unzureichende interne Konsistenz auffiel. Nach Ausschluss dieses Items sowie der Spezifizierung einer Korrelation der Fehlerterme zweier Items

der Subskala Lesen aus Interesse, erreichen die Modelle zufriedenstellende Fitindices für beide Schülergruppen (Förderschulen: $\chi^2(16) = 20.451$, $p = .201$, TLI = .992, CFI = .995, RMSEA = .024; SRMR = .023; Hauptschulen: $\chi^2(16) = 14.712$, $p = .546$, TLI = 1.001, CFI = 1.000, RMSEA = .000, SRMR = .016; s. Tabelle 2). Die Fehlerkovarianzen deuten darauf hin, dass in diesem Modell bestimmte Zusammenhänge nicht ausreichend berücksichtigt sind. Da die Daten beider Subgruppen für das spezifizierte Baseline-Modell jedoch eine hinreichende Passung aufweisen, wird mit der Prüfung der Messinvarianz fortgefahren.

Positionseffekte. In einem ersten Schritt wird geprüft, ob die beiden eingesetzten Versionen mit rotierten Itemreihenfolgen für die Stichprobe der Förderschülerinnen und -schüler zu einer messinvarianten Messung der Konstrukte geführt haben. Es könnte ver-

mutet werden, dass die Position der Items am Anfang bzw. am Ende des Fragebogens einen Effekt auf die Erfassung der Faktoren zur Lesemotivation hat. In einem Mehrgruppenvergleich wird daher die Messäquivalenz der Daten der beiden rotierten Instrumente untersucht.

Das aufgestellte dreifaktorielle Modell führt auch in den beiden Subgruppen zu einer sehr guten Passung (s. Tabelle 3), so dass es als Baseline-Modell für die Prüfung der Messinvarianz zugrunde gelegt werden kann (Instrument 1: $\chi^2(16) = 20.707$, $p = .190$, TLI = .982, CFI = .990, RMSEA = .033, SRMR = .029; Instrument 2: $\chi^2(16) = 19.195$, $p = .259$, TLI = .989, CFI = .993, RMSEA = .029, SRMR = .025). Das Modell der konfiguralen Messinvarianz für beide Instrumentengruppen, für die die gleiche Anzahl an Faktoren sowie das gleiche Muster an freien und fixierten Parametern spezifiziert ist, ergibt ebenso eine sehr

Tabelle 2: Faktorladungen, Faktorkorrelationen und Modellpassungen zur Lesemotivation

Förderschule (n = 493)							Hauptschule (n = 622)						
Faktorladungen							Faktorladungen						
	LL	LI	SK					LL	LI	SK			
Item a	.823***							.899***					
Item b	.877***							.927***					
Item c	.686***							.777***					
Item d		.718***							.752***				
Item e		.650***							.862***				
Item f		.657***							.745***				
Item g			.822***							.568***			
Item i			.587***							.994***			
Faktorkorrelationen							Faktorkorrelationen						
LL	1.00							LL	1.00				
LI	.903***	1.00						LI	.944***	1.00			
SK	-.150*	-.242**	1.00					SK	-.150**	-.233***	1.00		
Fitindices							Fitindices						
χ^2	df	p	TLI	CFI	RMSEA	SRMR	χ^2	df	p	TLI	CFI	RMSEA	SRMR
20.451	16	.201	.992	.995	.024	.023	14.712	16	.546	1.001	1.000	.000	.016

Anmerkungen. LL = Leselust. LI = Lesen aus Interesse. SK = Selbstkonzept Lesen.

*** $p < .001$; ** $p < .01$; * $p < .05$. Angaben in standardisierten Werten.

Tabelle 3: Indikatoren zur Messinvarianz für die Skalen zur Lesemotivation in den verschiedenen Instrumenten für Förderschülerinnen und -schüler

Modell	χ^2	df	p	TLI	CFI	RMSEA	SRMR	$\Delta\chi^2(\Delta df)$	p
Instrument 1 (n = 263)	20.707	16	.190	.982	.990	.033	.029		
Instrument 2 (n = 230)	19.195	16	.259	.989	.993	.029	.025		
Konfigurale Invarianz	40.316	32	.149	.985	.991	.032	.027		
Metrische Invarianz	46.962	40	.209	.990	.993	.027	.044	5.312 (8)	.724
Skalare Invarianz	49.580	45	.296	.994	.995	.020	.044	2.172 (5)	.825

gute Passung ($\chi^2(32) = 40.316$, $p = .149$, TLI = .985, CFI = .991, RMSEA = .032, SRMR = .027). Wird die zusätzliche Restriktion der gleichen Faktorladungen zur Prüfung der metrischen Invarianz in beiden Gruppen hinzugefügt, führt dies nicht zu einer signifikanten Verschlechterung des Modellfits ($\Delta\chi^2(8) = 5.312$, $p = .724$). Die Annahme gleicher Intercepts in beiden Gruppen zur Prüfung der skalaren Invarianz bringt ebenfalls keine bedeutsame Verschlechterung des χ^2 -Wertes ($\Delta\chi^2(5) = 2.172$, $p = .825$) oder der Fitindices mit sich.

Für das Konstrukt der Lesemotivation ist somit sowohl konfigurale, metrische als auch skalare Messinvarianz gegeben. Somit kann davon ausgegangen werden, dass das Konstrukt der Lesemotivation mit drei latenten Faktoren äquivalent mit den beiden rotierten Befragungsinstrumenten bei Schülerinnen und Schülern mit SPF-L erhoben wird.

Vergleich mit der Hauptschule. Als weitere Forschungsfrage ist zu untersuchen, ob die Annahme der starken faktoriellen Invarianz auch für beide Gruppen der Schülerinnen und Schüler an Förderschulen und an Hauptschulen gelten kann. Die eingesetzten Instrumente unterscheiden sich sowohl hinsichtlich der Länge als auch im Darbietungsmodus. In einem analogen Vorgehen wird ein Mehrgruppenvergleich für die Gruppen der Schülerschaft an Förderschulen und Hauptschulen vorgenommen. Basierend auf dem bestätigten Baseline-Modell werden konfigurale, metrische und skalare Messinvarianz geprüft (s. Tabelle 4).

Die simultane Schätzung des Modells bei nicht fixierten Parametern führt zu einer guten Passung ($\chi^2(32) = 35.054$, $p = .325$, TLI = .998, CFI = .999, RMSEA = .013, SRMR = .019). Die Restriktion gleicher Faktorladungen in beiden Gruppen führt zu einer signifikanten Steigerung des χ^2 -Wertes ($\Delta\chi^2(8) = 23.292$, $p = .003$). Erlaubt man dem Modell, die Ladung eines Indikators auf

Tabelle 4: Indikatoren zur Messinvarianz für die Skalen zur Lesemotivation bei Schülerinnen und Schülern an Hauptschulen und Förderschulen

Modell (n = 1115)	χ^2	df	p	TLI	CFI	RMSEA	SRMR	$\Delta\chi^2(\Delta df)$	p
Konfigurale Invarianz	35.054	32	.325	.998	.999	.013	.019		
Metrische Invarianz	53.441	40	.076	.993	.995	.025	.051	23.292 (8)	.003
Partielle metrische Invarianz	45.388	39	.223	.997	.998	.017	.037	11.673 (7)	.112
Skalare Invarianz*	52.720	44	.173	.996	.997	.019	.040	7.719 (5)	.172

* Modell basiert auf partieller metrischer Invarianz.

den latenten Faktor Leseinteresse (Item e) in beiden Gruppen frei zu schätzen, wird der Anstieg des χ^2 -Wertes ($\Delta\chi^2(7) = 11.673$, $p = .112$) nicht signifikant. Dies ist ein Fall von partieller metrischer Invarianz. Da jedoch für den Faktor Leseinteresse zwei Items messäquivalente Faktorladungen aufweisen, kann dennoch von einer invarianten Erfassung des Faktors ausgegangen werden (Brown, 2006; Byrne et al., 1989). Die zusätzliche Annahme gleicher Intercepts aller acht Indikatoren kann auf Grund des χ^2 -Differenz-Tests ($\Delta\chi^2(5) = 7.719$, $p = .172$) beibehalten werden.

Trotz der partiellen metrischen Invarianz kann somit auf Grundlage der Ergebnisse von einer starken Messinvarianz ausgegangen werden, die zeigt, dass die Erfassung an Förderschulen und Hauptschulen für das Konstrukt der Lesemotivation messinvariant erfolgt. Ist starke Messäquivalenz gegeben, können Differenzen zwischen den Faktormittelwerten beider Gruppen sinnvoll interpretiert werden, da sie sich auf existierende Gruppenunterschiede beziehen und nicht auf nicht invariante Messungen zurückzuführen sind.

Messinvarianz zu Skalen des akademischen Selbstkonzeptes

Baseline-Modell. Ein konfirmatorisches Modell mit der theoretisch zugrundeliegenden Struktur ergibt zunächst lediglich eine akzeptable Modellpassung (Förderschule: $\chi^2(17) = 45.990$, $p < .001$, TLI = .942, CFI = .965, RMSEA = .059; SRMR = .038; Hauptschule: $\chi^2(17) = 136.536$, $p < .001$, TLI = .878, CFI = .926, RMSEA = .102, SRMR = .054). Da das Item d der Subskala mathematisches Selbstkonzept zusätzlich auf dem Faktor des verbalen Selbstkonzeptes lädt und somit nicht dem theoretisch postulierten Ladungsmuster entspricht, wird dieses Item aus den weiteren Analysen ausgeschlossen. Das nachfolgende spezifizierte Modell führt zu einer hinreichenden Modellpassung für beide Stichproben (Förderschule: $\chi^2(11) = 20.989$, $p = .034$,

TLI = .974, CFI = .986, RMSEA = .043, SRMR = .024; Hauptschule: $\chi^2(11) = 38.677$, $p < .001$, TLI = .954, CFI = .976, RMSEA = .061, SRMR = .027; s. Tabelle 5).

Positionseffekte. Betrachtet man für die Stichprobe der Schülerinnen und Schüler mit SPF-L an Förderschulen die Messinvarianz der beiden Instrumentenvarianten, zeigen sich folgende Modellpassungen im Mehrgruppenvergleich (s. Tabelle 6).

Das zuvor aufgestellte Baseline-Modell weist auch in den beiden Subgruppen der beiden Förderschulinstrumente eine gute Passung auf (Instrument 1: $\chi^2(11) = 20.700$, $p = .037$, TLI = .953, CFI = .976, RMSEA = .058, SRMR = .034; Instrument 2: $\chi^2(11) = 10.015$, $p = .529$, TLI = 1.006, CFI = 1.000, RMSEA = .000, SRMR = .023). Die weitere Prüfung der konfiguralen, metrischen und skalaren Messinvarianz zeigt, dass zusätzliche Restriktionen keine signifikante Verschlechterung der Passung oder des χ^2 -Wertes ergeben, so dass auch für das Konstrukt des akademischen Selbstkonzeptes von einer invarianten Messung mit den beiden rotierten Instrumentenversionen ausgegangen werden kann.

Vergleich mit der Hauptschule. Der Vergleich mit der Stichprobe an Hauptschulen zeigt jedoch, dass bei Gleichsetzung der Parameter die Modellpassung teilweise signifikant abnimmt (s. Tabelle 7).

Die Gleichsetzung der Faktorladungen aller Indikatoren führt zu einem signifikanten χ^2 -Differenz-Test ($\Delta\chi^2(7) = 17.688$, $p = .013$). Für die Faktoren des verbalen und des schulischen Selbstkonzeptes muss jeweils die Ladung eines Indikators (Item c und h) in beiden Gruppen frei geschätzt werden, um keine signifikante Modellverschlechterung zu erreichen ($\Delta\chi^2(5) = 9.830$, $p = .080$). Die Gleichsetzung der Intercepts aller Variablen führt ebenso zu einer signifikanten Verschlechterung des χ^2 -Wertes ($\Delta\chi^2(4) = 14.555$, $p < .001$). Ein Intercept des schulischen Selbstkonzeptes (Item i) muss über beide Gruppen frei geschätzt wer-

den, um keinen signifikanten χ^2 -Differenz-Test zu erhalten ($\Delta\chi^2(3) = 1.459, p = .692$). Dieses Item weist somit keinen invarianten Intercept für beide Schülergruppen auf.

Für das Konstrukt des akademischen Selbstkonzeptes gibt es Fälle partieller metrischer und partieller skalarer Messinvarianz. Bezüglich der metrischen Messinvari-

anz sind zwei Ladungen der insgesamt sieben Items als nicht invariant einzustufen. Zusätzlich weist ein weiterer Indikator einen nicht messäquivalenten Intercept auf. Somit ist die Voraussetzung nach Byrne et al. (1989) zweier messinvarianter Indikatoren pro latentem Faktor nicht gegeben.

Tabelle 5: Faktorladungen, Faktorkorrelationen und Modellpassungen zum akademischen Selbstkonzept

Förderschule (n = 490)							Hauptschule (n = 674)						
Faktorladungen							Faktorladungen						
	V	M	S				V	M	S				
Item b	0.718***						.745***						
Item c	0.774***						.765***						
Item e		0.820***						.842***					
Item f		0.775***						.849***					
Item g			0.770***							.786***			
Item h			0.754***							.756***			
Item i			0.784***							.792***			
Faktorkorrelationen							Faktorkorrelationen						
V	1.00						V	1.00					
M	0.353***	1.00					M	.078	1.00				
S	0.807***	0.575***	1.00				S	.704***	.423***	1.00			
Fitindices							Fitindices						
χ^2	df	p	TLI	CFI	RMSEA	SRMR	χ^2	df	p	TLI	CFI	RMSEA	SRMR
20.989	11	.034	.974	.986	.043	.024	38.677	11	<.001	.954	.976	.061	.027

Anmerkungen. V = Verbales Selbstkonzept. M = Mathematisches Selbstkonzept. S = Schulisches Selbstkonzept.

*** $p < .001$; ** $p < .01$; * $p < .05$. Angaben in standardisierten Werten.

Tabelle 6: Indikatoren zur Messinvarianz für die Skalen des akademischen Selbstkonzepts in den verschiedenen Instrumenten für Förderschülerinnen und -schüler

Modell	χ^2	df	p	TLI	CFI	RMSEA	SRMR	$\Delta\chi^2(\Delta df)$	p
Instrument 1 (n = 261)	20.700	11	.037	.953	.976	.058	.034		
Instrument 2 (n = 229)	10.015	11	.529	1.006	1.000	.000	.023		
Konfigurale Invarianz	32.353	22	.072	.973	.986	.044	.030		
Metrische Invarianz	42.171	29	.054	.974	.982	.043	.057	9.747 (7)	0.203
Skalare Invarianz	48.635	33	.039	.973	.979	.044	.063	6.777 (4)	0.148

Diskussion

In diesem Artikel wurde die Faktorstruktur der Konstrukte Lesemotivation und des akademischen Selbstkonzeptes betrachtet sowie die Homogenität dieser bei einem rotierten Erhebungsinstrument und bei verschiedenen Schülergruppen untersucht. Zur Prüfung der Messinvarianz wurden konfirmatorische Mehrgruppenvergleiche mit den Restriktionen gleicher Parameter (Faktorstruktur, Faktorladungen, Intercepts) angewendet.

Modellergebnisse

Für die drei Subskalen der Lesemotivation konnten durch Ausschluss eines Items und die Spezifikation einer Fehlertermkorrelation hinreichende Modellpassungen erreicht werden. Das ausgeschlossene Item h der Subskala Selbstkonzept Lesen könnte möglicherweise auf Grund der anderen Polung in der Formulierung nicht zur Faktorstruktur passen bzw. bei einer etwas oberflächlichen Bearbeitung der Fragen zu nicht validen Antworten führen. Die spezifizierte Residuenkorrelation zwischen den Items e und f lässt vermuten, dass weitere Zusammenhänge (noch) nicht adäquat im Modell berücksichtigt sind. Wenn man den Inhalt dieser beiden Items betrachtet, zielen sie augenscheinlich auf das Thema (schulisches) Lernen, wohingegen das Item d derselben Subskala die Freude an der Tätigkeit

Lesen betont. Somit kann ein systematischer Messfehler vermutet werden.

Auch für die Skala des akademischen Selbstkonzeptes wurde ein Item aus den Analysen ausgeschlossen, um einen akzeptablen Modellfit zu erhalten. Die KFA zeigt, dass das nachfolgend ausgeschlossene Item d des mathematischen Selbstkonzeptes zusätzlich eine signifikante Faktorladung für die verbale Komponente aufweist. Dieses Befundmuster einer Kreuzladung ist in beiden Stichproben zu finden und führt dazu, dass dieses Item nicht in der intendierten Form zu interpretieren ist. Möglicherweise führt die einheitliche Formulierung der aufeinanderfolgenden Items (c: „Im Fach Deutsch bekomme ich gute Noten“ und d: „Im Fach Mathe bekomme ich gute Noten“) zu nicht nach Schulfächern differenzierten Angaben.

Positionseffekte

Die Modelltests konnten zeigen, dass die rotierten Instrumente für die Schülerinnen und Schüler mit SPF-L an Förderschulen zu einer messinvarianten Erfassung der Subskalen zur Lesemotivation und zum akademischen Selbstkonzept führten. Die Annahme, dass eine abnehmende Aufmerksamkeit im Verlauf der Befragung zu weniger validen Antworten führen könnte, hat sich nicht bestätigt. Der administrierte Umfang der Befragung stellt sich als angemessen für diese Schülergruppe dar. Zudem ist festzustellen,

Tabelle 7: Indikatoren zur Messinvarianz für die Skalen des akademischen Selbstkonzeptes bei Schülerinnen und Schülern an Hauptschulen und Förderschulen

Modell (n = 1164)	χ^2	df	p	TLI	CFI	RMSEA	SRMR	$\Delta\chi^2(\Delta df)$	p-value
Konfigurale Invarianz	59.313	22	<.001	.962	.980	.054	.026		
Metrische Invarianz	77.216	29	<.001	.962	.974	.053	.081	17.688 (7)	.013
Partielle Metrische Invarianz	69.792	27	<.001	.964	.977	.052	.061	9.830 (5)	.080
Skalare Invarianz*	83.516	31	<.001	.962	.972	.054	.062	14.555 (4)	<.001
Partielle Skalare Invarianz*	72.391	30	<.001	.968	.977	.049	.061	1.455 (3)	.692

* = Modell basiert auf partieller metrischer Invarianz.

dass es nicht zu vorzeitigem Abbrechen der Befragung seitens der Schülerinnen und Schüler kam. Auf Grund der vorstrukturierten Bearbeitung mit Hilfe des Vorleseleitfadens wurden alle Teilnehmenden durch das gesamte Instrument geführt. Eine eigenständige Bearbeitung des Fragebogens hätte möglicherweise auf Grund der hohen Leselast einen erhöhten Anteil fehlender Angaben hervorrufen können (Gresch, Strietholt, Kanders & Solga, 2014).

Vergleiche mit der Hauptschule

Die Ergebnisse zur Messinvarianz im Vergleich mit der Gruppe der Hauptschülerinnen und -schüler sind weniger eindeutig zu beurteilen. Betreffend der Skalen zur Lesemotivation kann trotz partieller Messinvarianz von einer starken faktoriellen Invarianz ausgegangen werden.

Die Ladung des Item e auf den Faktor Lesen aus Interesse ist über die Schülergruppen hinweg nicht invariant. Für die Hauptschulgruppe stellt dieses Item den Indikator mit der stärksten Ladung dar, während es für die Schülergruppe an Förderschulen den schwächsten Indikator darstellt (s. Tabelle 2). Dieses Item fragt nach der Überzeugung, durch die Tätigkeit Lesen viel lernen zu können. Diese Formulierung scheint von beiden Gruppen konzeptuell unterschiedlich aufgefasst zu werden, so dass es zu nicht äquivalenten Faktorladungen kommt (Chen, 2008). Da die beiden anderen Indikatoren für den Faktor Lesen aus Interesse invariant sind, könnte man für vergleichende Analysen in Erwägung ziehen, dieses Item auszuschließen (Cheung & Rensvold, 1999).

Das Modell zum akademischen Selbstkonzept weist bei den Schülerinnen und Schüler mit SPF-L eine bessere Passung auf als in der Hauptschulstichprobe. Insgesamt fließen nur sieben Indikatoren in das Modell ein. Daher ist das Ergebnis zweier nicht äquivalenter Faktorladungen (Item c und h) und einem nicht äquivalenten Intercept (Item i) nicht trivial für vergleichende Analy-

sen. Bei Ausschluss der drei genannten nicht messinvarianten Indikatoren würden die Faktoren des verbalen und des schulischen Selbstkonzeptes lediglich jeweils mit einem Item repräsentiert. Unauffälliger ist diesbezüglich die Facette des mathematischen Selbstkonzeptes, das zwei Items mit messinvarianten Faktorladungen und Mittelwerten aufweist. Somit ist der hier vorgestellte Konstruktbereich in seiner Gesamtheit nicht für schulformübergreifende Analysen geeignet.

Zwei Items weisen invariante Faktorladungen auf. Diese beiden sind sich inhaltlich ähnlich, als dass sie jeweils nach Noten im Fach Deutsch bzw. in Klassenarbeiten fragen. Die Problematik dieser Fragestellung liegt darin begründet, dass die Bewertung an Förderschulen anders erfolgt als an Hauptschulen: Häufig gibt es an Förderschulen individuelle Berichterstattungen über Lernfortschritte anstelle von Noten. So geben 21.5 bzw. 23.9 % der Schülerinnen und Schüler an Förderschulen an, dass sie in ihrem letzten Zeugnis keine Note für das Fach Mathe bzw. Deutsch erhalten haben. In Hauptschulen machen nur 2.2 bzw. 4.0% der Zielpersonen diese Angabe.

Zusätzlich zeigten die Mehrgruppenvergleiche, dass der Intercept des Items i („Ich bin in den meisten Schulfächern gut.“) nicht messinvariant in den beiden Schülergruppen ist. Die Schülerinnen und Schüler an Förderschulen weisen hier einen höheren Mittelwert auf als die Schülergruppe an Hauptschulen. Diese Tatsache allein erklärt nicht die Nicht-Äquivalenz des Parameters. Grund für die mangelnde Messinvarianz könnte jedoch sein, dass Schülerinnen und Schüler, die einen ähnlichen Faktormittelwert für die Skala des schulischen Selbstkonzeptes aufweisen, in den beiden Stichproben nicht mit dem gleichen Ausmaß an Zustimmung auf dieses Item reagieren (Chen, 2008).

Wird das akademische Selbstkonzept dieser verschiedenen Schülerpopulationen betrachtet und verglichen, sind Referenzgruppeneffekte zu bedenken. In Förder-

schulen ist eine Art „Schonraum“ geschaffen worden, in dem auch besonderer Wert auf soziale Entwicklung sowie schulische und persönliche Erfolgserlebnisse gelegt wird. Da das akademische Selbstkonzept von schulischen Erfahrungen, Erfolgs- und Misserfolgserlebnissen, sozialen Vergleichen und der Interpretation dieser geformt wird, spiegeln sich pädagogische und didaktische Maßnahmen des differenzierten Unterrichts und der segregierten Schulsettings darin wider (Bos et al., 2010). Auf Grund der unterschiedlichen Rahmenbedingungen könnte sich das akademische Selbstkonzept bei diesen Schülergruppen nicht nur in Bezug auf die mittlere Ausprägung anders darstellen, sondern sich auch bezüglich der relativen Beiträge der Einzelitems (d. h. konzeptuell) unterscheiden.

Ausblick

Die Summe der Anpassungen der Administrationsbedingung bei der vorliegenden Studie hat für die Gruppe der Schülerinnen und Schüler mit SPF-L zu aussichtsreichen Ergebnissen geführt. Daher lohnt es sich, diese Akkommodationen bei schriftlichen Befragungen dieser und auch weiterer Schülergruppen ausführlicher zu untersuchen sowie weiterführende Untersuchungen zu spezifischen Effekten der Anpassungen, wie z. B. dem Vorlesen, durchzuführen. Die dargestellten Ergebnisse sind beispielhafte Resultate, die sich nicht zwingend auf weitere Konstrukte und Skalen übertragen lassen. Hierzu wären weitere Studien wünschenswert.

In der Sonderpädagogik als auch in der Bildungsforschung bleiben die messinvarianten und damit vergleichbaren Messungen von bildungsrelevanten Konstrukten und Kompetenzen weiterhin ein wichtiges Forschungsfeld. Zentrale Fragen bezüglich der Entwicklung von Schülerinnen und Schülern an Förderschulen im Vergleich zu Schülerinnen und Schülern ohne SPF in anderen Schulformen oder auch im Vergleich zu so-

genannten Integrationskindern im allgemeinen Schulsystem sind bisher noch unbeantwortet und bedürfen vertiefter Analysen sowie weiterer Forschung. Die experimentellen Machbarkeitsstudien im Rahmen des NEPS und die damit generierten Daten können hierzu einen wichtigen Beitrag leisten.

Literaturverzeichnis

- Artelt, C., Naumann, J. & Schneider, W. (2010). Lesemotivation und Lernstrategien. In E. Klieme, C. Artelt, J. Hartig, N. Jude, O. Köller, M. Prenzel, W. Schneider & P. Stanat (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 73-112). Münster: Waxmann.
- Bildungsbericht (2014). *Bildung in Deutschland 2014. Ein indikatorengestützter Bericht mit einer Analyse zur Bildung von Menschen mit Behinderungen*. Bielefeld: Bertelsmann Verlag.
- Blossfeld, H.-P., Roßbach, H.-G. & von Maurice, J. (Hrsg.) (2011). Education as a life-long Process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft, Sonderheft* 14.
- Bos, W., Müller, S. & Stubbe, T. C. (2010). Abgehängte Bildungsinstitutionen: Hauptschulen und Förderschulen. In G. Quenzel & K. Hurrelmann (Hrsg.), *Bildungsverlierer. Neue Ungleichheiten* (S. 375-397). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Byrne, B. M. (1996). Academic self-concept: Its structure, measurement, and relation to academic achievement. In B. A. Bracken (Hrsg.), *Handbook of self-concept: developmental, social, and clinical considerations* (S. 287–316). New York: Wiley.
- Byrne, B. M., Shavelson, R. J. & Muthén, B. O. (1989). Testing for the equivalence of factor covariance and mean structures. The issue of partial measurement invari-

- ance. *Psychological Bulletin*, 105(3), 456–466.
- Byrne, B. M. & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10(2), 107–132.
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95(5), 1005–1018.
- Cheung, G. W. & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, 25, 1–27.
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233–255.
- Cormier, D. C., Altman, J., Shyyan, V. & Thurlow, M. L. (2010). *A summary of the research on the effects of test accommodations: 2007–2008* (Technical Report 56). Minneapolis: University of Minnesota, National Center on Educational Outcomes.
- Dimitrov, D. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121–149.
- Frahm, S., Goy, M., Kowalski, K., Sixt, M., Strietholt, R., Blatt, I., Bos, W. & Kandera, M. (2011): Transition and development from lower secondary to upper secondary school. In H.-P. Blossfeld, H.-G. Roßbach & J. Maurice (Hrsg.), *The National Educational Panel Study: need, main features, and research potential* (S. 217–232). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gebhardt, M., Oelkrug, K. & Tretter, T. (2013). Das mathematische Leistungsspektrum bei Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf in der Sekundarstufe. Ein explorativer Querschnitt der fünften bis neunten Klassenstufe in Münchner Förderschulen. *Empirische Sonderpädagogik*, 5(2), 130–143.
- Gresch, C., Strietholt, R., Kandera, M. & Solga, H. (2014). Reading-aloud versus self-administrated student questionnaires: An experiment on data quality. Manuscript submitted for publication.
- Grünke, M. (2004). Lernbehinderung. In G. W. Lauth, M. Grünke & J. C. Brunstein (Hrsg.), *Interventionen bei Lernstörungen. Förderung, Training und Therapie in der Praxis* (S. 65–77). Göttingen: Hogrefe.
- Heydrich, J., Weinert, S., Nusser, L., Artelt, C. & Carstensen, C. H. (2013). Including students with special educational needs into large-scale assessments of competencies. Challenges and approaches within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online*, 5(2), 217–240.
- Hoermann, B. (2007). *Die Unsichtbaren in PISA, TIMSS & Co. Kinder mit Lernbehinderungen in nationalen und internationalen Schulleistungsstudien*. Verfügbar unter: http://bildungswissenschaft.univie.ac.at/fe2/fileadmin/user_upload/inst_bildungswissenschaft/Diplomarbeit_Hoermann.pdf.
- Horn, J. L. & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18, 117–144.
- Hu, L.-T. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Klauer, K. J. & Lauth, G. W. (1997). Lernbehinderungen und Leistungsschwierigkeiten bei Schülern. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie, Themenbereich D, Serie I, Pädagogische Psychologie) (S. 701–738). Göttingen: Hogrefe.
- Koretz, D. & Barton, K. (2004). Assessing students with disabilities: Issues and evi-

- dence. *Educational Assessment*, 9(1/2), 29-60.
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.-J. & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Kränzl-Nagl, R. & Wilk, L. (2000). Möglichkeiten und Grenzen standardisierter Befragung unter besonderer Berücksichtigung der Faktoren soziale und personale Wünschbarkeit. In F. Heinzel (Hrsg.), *Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive* (S. 59-75). Weinheim: Juventa.
- Lipski, J. (2000). Zur Verlässlichkeit der Angaben von Kindern bei standardisierten Befragungen. In F. Heinzel (Hrsg.), *Methoden der Kindheitsforschung. Ein Überblick über Forschungszugänge zur kindlichen Perspektive* (S. 77-86). Weinheim: Juventa.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.
- Marsh, H. W. (1990). The Structure of academic self-concept: The Marsh/Shavelson Model. *Journal of Educational Psychology*, 82(4), 623-636.
- McElvany, N., Kortenbruck, M. & Becker, M. (2008). Lesekompetenz und Lesemotivation. Entwicklung und Mediation des Zusammenhangs durch Leseverhalten. *Zeitschrift für Pädagogische Psychologie*, 22(3-4), 207-219.
- Möller, J. & Bonerad, E.-M. (2007). Fragebogen zur habituellen Lesemotivation. *Psychologie in Erziehung und Unterricht*, 54, 259-267.
- Möller, J. & Köller, O. (2004). Die Genese akademischer Selbstkonzepte: Effekte dimensionaler und sozialer Vergleiche. *Psychologische Rundschau*, 55(1), 19-27.
- Möller, J. & Schiefele, U. (2004). Motivationale Grundlagen der Lesekompetenz. In U. Schiefele, C. Artelt, W. Schneider & P. Stanat (Hrsg.), *Struktur, Entwicklung und Förderung von Lesekompetenz. Vertiefende Analysen im Rahmen von PISA 2000* (S. 101-124). Wiesbaden: Verlag für Sozialwissenschaften.
- Muthén, L. K. & Muthén, B. O. (2012). Mplus Version 7 [Computersoftware]. Los Angeles, CA.
- Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, 29(4), 347-363.
- Satorra, A. & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514.
- Schiefele, U. (1996). *Motivation und Lernen mit Texten*. Göttingen: Hogrefe.
- Scholl, A. (2003). *Die Befragung. Sozialwissenschaftliche Methode und kommunikationswissenschaftliche Anwendung*. Konstanz: UVK (UTB).
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Selfconcept: Validation of construct interpretations. *Review of Educational Research*, 46, 407-444.
- Statistisches Bundesamt (2011). Bildung und Kultur. Allgemeinbildende Schule. Schuljahr 2010/2011. Fachserie 11. Wiesbaden.
- Ständige Konferenz der Kultusminister der Länder (1994). Empfehlungen zur sonderpädagogischen Förderung in den Schulen in der Bundesrepublik Deutschland. Verfügbar unter: http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1994/1994_05_06-Empfehlung-sonderpaed-Foerderung.pdf
- Steenkamp, J.-B. E. M. & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78-107.
- Steinmetz, H., Schmidt, P., Tina-Booh, A., Wiczorek, S. & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between

educational groups in human values measurement. *Quality & Quantity*, 43(4), 559–616.

Vandenberg R. J. & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–69.

Wohlkinger, F., Ditton, H., von Maurice, J., Haugwitz, M. & Blossfeld, H.-P. (2011). Motivational concepts and personality aspects across the life course. In H.-P. Blossfeld, H.-G. Roßbach & J. Maurice (Hrsg.), *The National Educational Panel Study: need, main features, and research potential* (S. 155–168). Wiesbaden: VS Verlag für Sozialwissenschaften.

Lena Nusser

Leibniz-Institut für Bildungsverläufe e.V.
Wilhelmsplatz 3
96047 Bamberg
lenu.nusser@lifbi.de

Erstmalig eingereicht: 03.11.2014

Überarbeitung eingereicht: 02.02.2015

Angenommen: 07.02.2015