

Empirische Sonderpädagogik, 2014, Nr. 2, S. 99-113

Entwicklung und erste Validierung einer Kurzversion des „Fragebogens zur Erfassung von Dimensionen der Integration von Schülern (FDI 4-6)“ von Haerberlin, Moser, Bless und Klaghofer

Martin Venetz¹, Carmen Zurbriggen² & Michael Eckhart³

¹ Hochschule für Heilpädagogik Zürich

² Universität Freiburg (Schweiz)

³ Pädagogische Hochschule Bern

Zusammenfassung

Der „Fragebogen zur Erfassung der Dimensionen der Integration von Schülern“ (FDI 4-6; Haerberlin, Moser, Bless & Klaghofer, 1989) ist ein Instrument zur Selbsteinschätzung des emotionalen, sozialen und leistungsmotivationalen Integriertseins. Im vorliegenden Beitrag wird anhand einer Untersuchungsstichprobe ($N = 744$) eine 12 Items umfassende Kurzversion entwickelt und anhand der Originalstichprobe ($N = 1620$) überprüft. Konfirmatorische Faktorenanalysen bestätigen die postulierte dreifaktorielle Struktur. Multiple Gruppenvergleiche bzgl. Schulleistungen, Verhaltenskompetenzen und sprachlicher Herkunft belegen die Messäquivalenz der Kurzform. Die internen Konsistenzen der Kurzskalen fallen zufriedenstellend aus ($.80 \leq \alpha \leq .91$). Die Skalen weisen hypothesenkonforme Zusammenhänge mit dem Wohlbefinden im Schulunterricht, Verhaltenskompetenzen (aus Sicht der Lehrperson), der Lern- und Leistungsmotivation, dem akademischen Selbstkonzept sowie den Schulleistungen auf. Mit dem Kurzfragebogen KFDI 4-6 liegt ein reliables, valides und sehr ökonomisches Instrument zur Messung schulischen Integriertseins von Schülerinnen und Schülern vor.

Schlüsselwörter: Inklusion, schulische Integration, Kinder, Fragebogen, Gütekriterien

Development and preliminary validation of a short questionnaire based on the FDI 4-6 by Haerberlin, Moser, Bless, and Klaghofer measuring the educational integration of pupils

Abstract

The „Fragebogen zur Erfassung der Dimensionen der Integration von Schülern“ (FDI 4-6; Haerberlin, Moser, Bless & Klaghofer, 1989) is a self-report questionnaire measuring educational integration of pupils from an emotional, social, and motivational perspective. A short form comprising 12 items was developed on the basis of a sample of 744 pupils, and replicated using the sample of the original study ($N = 1620$). Confirmatory factor analyses support the postulated three-dimensional structure. Multiple group comparisons with respect to academic achievement, behavioural competencies, and linguistic background confirm the measurement equivalence of the short form. Internal consistencies of the three scales are satisfactory ($.80 \leq \alpha \leq .91$). As hypothesized, the scales are correlated with wellbeing in the classroom, behavioural competencies (teacher ratings), goal orientations, academic self-concept, and academic achievement. This short form KFDI 4-6 represents a reliable, valid, and highly economical instrument measuring the educational integration of pupils.

Key words: Inclusion, educational integration, children, questionnaire, psychometric properties

Bereits Ende der 1980er-Jahre wies die Forschergruppe um Haerberlin darauf hin, dass schulische Integration weit mehr als eine schulorganisatorische Maßnahme bedeutet (Haerberlin, Bless, Moser & Klaghofer, 1990). Maßnahmen wie die Aufhebung von Sonderklassen oder die Institutionalisierung von Gesamtschulen ermöglichen zwar den Zugang zum regulären Bildungssystem, sie bewirken hingegen nicht per se jenen Zustand, bei dem sich eine Schülerin bzw. ein Schüler in einem sozialen System – der Schule und der Schulklasse – eingebettet oder eben integriert fühlt.

Im Rahmen des INTSEP-Forschungsprogrammes des Heilpädagogischen Instituts der Universität Freiburg (Schweiz) werden seit beinahe 30 Jahren Fragen zu Wirkungen der schulischen Integration vs. Separation nachgegangen. Um in Erfahrung zu bringen, in welchem Ausmaß das Miteinander der Schülerinnen und Schüler sowie das gemeinsame Lernen gelingt, inwiefern sich diese in ihrer Schule wohl, akzeptiert und den kognitiven Anforderungen gewachsen fühlen, hat die Forschergruppe den „Fragebogen zur Erfassung der Dimensionen der Integration von Schülern (FDI 4-6)“ entwickelt (Haerberlin, Moser, Bless & Klaghofer, 1989). Das subjektive Integriertsein umfasst demnach drei grundlegende pädagogische Zielvorstellungen: emotionales Wohlbefinden in der Schule, positive Beziehungen zu Mitschülerinnen und Mitschülern in der Schulklasse sowie Vertrauen in die eigene schulische Leistungsfähigkeit.

Emotional-soziale Aspekte der schulischen Entwicklung sind vor allem – aber nicht nur – im sonderpädagogischen Bereich von zentraler Bedeutung. Daher werden sie in Studien zu Effekten von schulischer Integration bzw. Inklusion auf Schülerinnen und Schüler mit leichtem bis moderatem sonderpädagogischen Förderbedarf neben dem leistungsbezogenen Fortschritt häufig untersucht (s. Forschungsüberblicke von Bless & Mohr, 2007; Ruijs & Peetsma, 2009).

Beschreibung, Entwicklung und Gütekriterien des FDI 4-6

Der FDI 4-6 zielt auf die emotionale (z.B. „Ich gehe gerne in die Schule“), soziale (z.B. „Ich habe sehr viele Freunde in meiner Klasse“) und leistungsmotivationale (z.B. „Ich kann auch sehr schwere Aufgaben lösen“) Dimension der Integration von Schülerinnen und Schülern der vierten bis sechsten Primarstufe. Geht es um sozial-emotionale Aspekte der Integration, bietet sich das Erheben von Selbsteinschätzungen von Schülerinnen und Schülern an (für weitere methodische Zugänge s. Eckhart, 2011). Kinder und Jugendliche werden in diesem Sinne als Expertinnen und Experten betrachtet.

Für die Konstruktion des FDI wurde von Haerberlin und Mitarbeitenden zu den drei theoretisch begründeten Dimensionen in einem ersten Schritt eine grosse Anzahl von Items formuliert. Um eine Überforderung bezüglich Sprachverständnis und Differenzierungsvermögen zu vermeiden, wurde zum einen besonderen Wert auf altersadäquate Formulierungen gelegt und zum anderen wurden vierstufige Likert-Skalen (*stimmt gar nicht – stimmt weniger – stimmt ziemlich – stimmt genau*) verwendet. Nach einer Validierung durch Fachpersonen wurden die Items in einer Testphase datenbasiert ($N = 125$) anhand von Trennschärfe- und Schwierigkeitsanalysen auf 45 reduziert und im Rahmen einer Pilotstudie ($N = 387$) erprobt (Moser, Bless & Haerberlin, 1989). Der Fragebogen wurde schließlich einer für die Deutschschweiz repräsentativen Stichprobe von 1279 Schülerinnen und Schülern der 4. und 5. Primarklasse (Grundstufe) vorgelegt. Nach einem Jahr erfolgte eine zweite Messung mit derselben Stichprobe. Die Überprüfung der faktoriellen Struktur erfolgte mittels exploratorischer Faktorenanalyse. Die internen Konsistenzen der Skalen (Cronbachs α) lagen für Messzeitpunkt 1 zwischen .89 und .93, für Messzeitpunkt 2 zwischen .90 und .94; die Retest-Reliabilitäten zwischen .57 und .73 (Haerberlin et al., 1989).

Weiterentwicklung des FDI

Der FDI wurde in den letzten zwei Jahrzehnten wiederholt in der Integrationsforschung eingesetzt (z.B. Bless, Schüpbach & Bonvin, 2004; Gebhardt, Schwab, Krammer & Gasteiger Klicpera, 2012; Kronig, Haeberlin & Eckhart, 2000; Sauer, Ide & Borchert, 2007; Venetz & Tarnutzer, 2011) und gilt als bewährtes Verfahren zur Messung des schulischen Integriertseins von Kindern und Jugendlichen. Allerdings lässt sich das Verfahren in seiner ursprünglichen Form nicht mehr ohne weiteres anwenden: Erstens ist der FDI mit 45 Items und einer Bearbeitungszeit von 20 bis 30 Minuten zeitlich recht aufwändig, was Schülerinnen und Schülern mit sonderpädagogischem Förderbedarf nicht ohne weiteres zugemutet werden kann. Deshalb scheint – zwecks ökonomischer Erfassung – eine kürzere Version des FDI angebracht. Dies würde auch der neuerdings zu beobachtenden Entwicklung im Bereich der Testkonstruktion entsprechen, nach der Skalen von vier bis sechs Items aufgrund ihrer erhöhten Homogenität den Vorrang gegenüber längeren Versionen gegeben wird (Schweizer, 2011). Zweitens ist eine testtheoretische Überprüfung notwendig. So stellt sich allgemein die Frage der Konstruktvalidität. Fraglich ist auch die Benennung der Dimension *Leistungsmotivationale Integration*: Sowohl die damaligen theoretischen Überlegungen als auch bisherige Befunde lassen vermuten, dass es sich eher um das Konstrukt akademisches Selbstkonzept als Leistungsmotivation handelt. Um die Testwerte von Kindern mit gegenüber Kindern ohne Schulleistungsschwächen auch tatsächlich vergleichen zu können, ist drittens eine Überprüfung der Messinvarianz des FDI unerlässlich. Viertens bedarf es einer sprachlichen Anpassung des Verfahrens: Angezeigt sind eine gendergerechte Formulierung sowie weitere sprachliche Aktualisierungen. So beinhalten einzelne Items relativ komplexe Formulierungen, was gerade für Schülerinnen und Schüler mit sprachlichen Schwächen entsprechende Hürden verursacht.

Zielsetzungen und Hypothesen

Zentrales Anliegen dieses Beitrags ist die Entwicklung eines Kurzfragebogens, mit dem die drei Dimensionen schulischer Integration auf Basis des FDI reliabel, valide und ökonomisch erfasst werden können. Die Anforderungen können dabei wie folgt präzisiert werden:

Erstens sollen die internen Konsistenzen der entwickelten Kurzskalen mindestens .70 betragen. Nach Kline (2011) können Reliabilitätskoeffizienten um .90 als „excellent“, solche um .80 als „good“ und solche um .70 noch als „adequate“ betrachtet werden.

Zweitens soll die theoretisch postulierte faktorielle Struktur, nämlich drei moderat interkorrelierende Faktoren, mit den Items der Kurzversion gut abgebildet werden können.

Drittens soll die Messinvarianz bei sonderpädagogisch relevanten Schülergruppen gewährleistet werden, sodass auffällige Mittelwertsunterschiede zwischen diesen Gruppen sinnvoll interpretiert werden können.

Viertens soll die Kurzversion konstruktvalid sein. Plausibilitätsüberlegungen legen dabei folgende Hypothesen nahe:

1. Es wird erwartet, dass alle FDI-Kurzskalen positiv und substantiell mit dem Wohlbefinden im Schulunterricht korreliert sind, die stärkste Beziehung jedoch zwischen der Skala *Emotionale Integration* (EI) und dem Wohlbefinden besteht.
2. Die Skala *Soziale Integration* (SI) korreliert substantiell negativ mit der Einschätzung der Lehrperson zu Peer-Problemen.
3. Generell wird erwartet, dass die Skala *Leistungsmotivationale Integration* (LI) sowohl mit dem akademischen Selbstkonzept, der Lern- und Leistungsmotivation sowie den Schulleistungen positiv korreliert. Im Besonderen soll folgende Hypothese geprüft werden: Mit der Skala LI wird primär nicht die Lern- und Leistungsmotivation, sondern vielmehr das akademische Selbstkonzept erfasst.

Methoden

Stichproben und Durchführung

Für die Entwicklung einer Kurzversion wurde auf eine Stichprobe (im Folgenden Untersuchungsstichprobe genannt) zurückgegriffen, die im Rahmen einer Studie zur schulischen Integration von Kindern mit sonderpädagogischem Förderbedarf gewonnen wurde (Venetz, Tarnutzer, Zurbriggen & Sempert, 2012). Sie umfasst 782 Schülerinnen und Schüler (davon 46.3% weiblich) aus 40 Regelklassen (89.6%) und 16 Klassen mit besonderem Lehrplan (10.4%) der vierten (1.7%), fünften (8.6%) und sechsten (89.7%) Primarschulstufe. Das Alter beträgt im Mittel 12.2 Jahre ($SD = 0.87$ Jahre). Für Validierungszwecke konnten die Daten des ersten Messzeitpunktes aus der Studie von Haeblerlin et al. (1989) verwendet werden. Die Originalstichprobe umfasst 1773 Schülerinnen und Schüler (davon 47.0% weiblich) aus 74 Regelklassen (81.4%) und 42 Schulklassen mit besonderem Lehrplan (18.6%) der vierten (52.7%), fünften (43.0%) und sechsten (4.3%) Primarschulstufe. Das Alter beträgt im Mittel 10.7 Jahre ($SD = 0.84$ Jahre). In beiden Stichproben fand die schriftliche Befragung im Klassenverbund statt.

Entwicklung einer Kurzversion

Für die Entwicklung einer Kurzversion wurde ein dreistufiges Vorgehen gewählt: Der erste Schritt beinhaltete die Kürzung der 45 Items umfassenden Originalversion um fünf Items pro Dimension auf 30 total Items. Als Ausschlusskriterium diente die Höhe der im Testmanual (Haeblerlin et al., 1989) berichteten Faktorladungen: Von jeder Dimension wurden jene fünf Items ausgeschlossen, die auf dem theoretisch postulierten Faktor die jeweils tiefsten Ladungen aufwiesen. Die nun auf 30 Items reduzierte und sprachlich leicht angepasste Version wurde den Schülerinnen und Schülern der Untersuchungsstichprobe vorgelegt.

In einem zweiten Schritt wurde die Anzahl Items pro Dimension anhand der Analyseergebnisse in der Untersuchungsstichprobe sukzessive weiterverringert: In für jede Dimension separaten Analysen wurde jeweils dasjenige Item ausgeschlossen, das mit möglichst geringfügigen Verlusten der internen Konsistenz verbunden war, wobei für den Ausschlussentscheid auch inhaltliche Überlegungen eine Rolle spielten. Als Maß der internen Konsistenz diente der Koeffizient ω (McDonald, 1999). Gegenüber dem herkömmlichen Cronbachs α hat dieser den Vorteil, dass die Reliabilitätsschätzungen adäquater sind und zudem mit Vertrauensintervallen versehen werden können (Kelley & Cheng, 2012; Raykov, 2002, 2007). Der Koeffizient gibt das Ausmaß an, in dem eine latente Variable die von den Items geteilte Varianz reflektiert. Zur Validierung der Ergebnisse wurde in einem dritten Schritt geprüft, ob sich die Befunde mit den Daten der Originalstichprobe von Haeblerlin et al. (1989) replizieren lassen.

Überprüfung der Kurzversion

Die psychometrischen Eigenschaften der entwickelten Kurzversion wurden unter dem Aspekt der faktoriellen Validität, der Messinvarianz sowie der Konstruktvalidität untersucht.

Die *faktorielle Validität* wurde sowohl in der Untersuchungs- als auch in der Originalstichprobe mit konfirmatorischen Faktorenanalysen überprüft. Zur Beurteilung der Modellgüte wurden neben dem χ^2 -Modelltest auch Fit-Indizes herangezogen, namentlich der Comparative Fit Index (CFI), der Tucker-Lewis Index (TLI), das Standardized Root Mean Square Residual (SRMR) sowie der Root Mean Square Error of Approximation (RMSEA). CFI- und TLI-Werte um .95 oder höher, SRMR- bzw. RMSEA-Werte um .08 bzw. .06 oder tiefer sprechen für einen guten Fit zwischen dem postulierten Modell und den Daten (Hu & Bentler, 1999).

Zur Klärung der Frage, ob die Kurzversion für verschiedene Subgruppen äquivalent ist,

wurde sie in der Untersuchungsstichprobe mittels multiplen Gruppenvergleichen auf konfigurale (gleiche Faktorenstruktur), metrische (zusätzlich gleiche Faktorenladungen) und skalare (zusätzlich gleiche Itemmittelwerte) *Messinvarianz* getestet (Brown, 2006). Da es sich dabei um genestete Modelle handelt, wurde jeweils mit dem χ^2 -Differenztest geprüft, ob das metrische bzw. skalare Messmodell signifikant weniger gut auf die Daten passt als das konfigurale Modell. Allerdings halten Cheung und Rensvold (2002) den Test für zu restriktiv und kommen aufgrund von Simulationsstudien zum Schluss, dass erst CFI-Differenzen grösser .01 als bedeutsam zu interpretieren sind.

Insgesamt wurden in separaten Analysen drei Mehrgruppenvergleiche durchgeführt: Erstens wurden Schülerinnen und Schüler, die aufgrund curriculärer Schulleistungstests in Mathematik oder Deutsch zum ersten Quartil gehören, mit den übrigen Schülerinnen und Schülern verglichen. Zweitens wurden Schülerinnen und Schüler mit und ohne Verhaltensauffälligkeiten verglichen. Der Gruppe mit Auffälligkeiten wurden Schülerinnen und Schüler zugeordnet, die auf Basis von Normwerten für die deutschsprachige Lehrversion des SDQ (Goodman, 2001; Klasen, Woerner, Rothenberger & Goodman, 2003) als „auffällig“ oder „grenzwertig auffällig“ gelten. Im Rahmen des dritten Mehrgruppenvergleichs wurden Schülerinnen und Schüler mit Deutsch als Erstsprache solchen mit Deutsch als Zweitsprache gegenübergestellt.

Zur Abklärung der *Konstruktvalidität* der FDI-Kurzversion sind einfache und multiple Korrelationskoeffizienten berechnet worden. Dabei wurden die Skalensummenwerte mit Angaben aus drei Informationsquellen in Beziehung gesetzt: (1) mit Einschätzungen der Lehrperson (L-Daten) zu Verhaltenskompetenzen ihrer Schülerinnen und Schüler; (2) mit Ergebnissen aus curricularen Schulleistungstests in Deutsch und Mathematik (T-Daten); (3) mit Selbsteinschätzungen der Schülerinnen und Schüler (Q-Daten) zum Befinden im Unterricht, zur Lern- und Leistungsmotiva-

tion sowie zum akademischen Selbstkonzept.

Messinstrumente zur Validierung

Zur Erfassung der Verhaltenskompetenzen wurde die Lehrversion des SDQ (Goodman, 2001) eingesetzt, die sich aus den fünf Subskalen *Emotionale Probleme*, *Hyperaktivität*, *Verhaltensprobleme*, *Peer-Probleme* und *Prosoziales Verhalten* zusammensetzt. Das akademische Selbstkonzept wurde mit der Skala *Absolutes Selbstkonzept* aus den Skalen zur Erfassung des schulischen Selbstkonzepts (SESKO; Schöne, Dickhäuser, Spinath & Stiensmeier-Pelster, 2002) erhoben, die motivationalen Zielorientierungen mit den Skalen *Lernziel-*, *Annäherungs-Leistungs-* sowie *Vermeidungs-Leistungs-Zielorientierung* aus den Skalen zur Erfassung der Lern- und Leistungsmotivation (SELLMO; Spinath, Stiensmeier-Pelster, Schöne & Dickhäuser, 2002). Das Befinden im Unterricht wurde, basierend auf dem Modell der *Positiven* (PA) und *Negativen Aktivierung* (NA) (Yik, Russell & Feldman Barrett, 1999), mit den PANAVA-Kurzskalen (Schallberger, 2005) und methodisch zudem mit zwei unterschiedlichen Ansätzen erfasst (s. dazu Kahnemann, 1999): zum einen mit einem konventionellen Fragebogen, zum anderen mit Hilfe der Experience Sampling Method (ESM; Hektner, Schmidt & Csikszentmihalyi, 2007). Die 14 im Verlaufe einer Schulwoche mittels ESM *in situ* gewonnenen Daten zum aktuellen Befinden wurden zu einem Gesamtwert aggregiert.

Statistikprogramme und vorbereitende Analysen

Die Modelle mit latenten Variablen wurden in Mplus (Version 7.0 und 7.11; Muthén & Muthén, 1998-2012) spezifiziert, die anderen Berechnungen in SPSS (Version 20 und 21) durchgeführt. Vor den Hauptanalysen wurden die Daten im Hinblick auf Ausreißer, fehlende Werte und Verteilungsform untersucht. Mittels Mahalanobis-Distanz wurden in der Untersuchungsstichprobe 38 (4.9%), in der

Originalstichprobe 153 (8.6%) Schülerinnen und Schüler als multivariate Ausreißer identifiziert und aus den Analysen ausgeschlossen. Da die fehlenden Werte in den FDI-Items keine besonderen Muster aufwiesen, wurde im Rahmen konfirmatorischer Faktorenanalysen die Funktion Full Information Maximum Likelihood (FIML) verwendet. Weil die FDI-Items zudem von einer (multivariaten) Normalverteilung abweichen (Mardia-Test), wurde ein robuster Schätzer gewählt (MLR; Muthén & Muthén, 1998-2012).

Ergebnisse

Fragebogenkürzung

Die Analyseergebnisse zur Fragebogenkürzung sind Tabelle 1 zu entnehmen (aus Platzgründen ist darin nicht jeder einzelne Kürzungsschritt dargestellt). Betrachtet man zunächst die Ergebnisse der Untersuchungsstichprobe, kann als Hauptbefund festgehalten werden, dass die Dimensionen schulischen Integriertseins auch mit vier Items (bzw. Indikatoren) pro Faktor noch reliabel genug erfasst werden können: Ihre interne Konsistenz kann mit ω -Werten von .80 (LI), .84 (SI) und .91 (EI) als gut bis sehr gut bezeichnet werden. Im Vergleich zu den Modellen mit zehn Items pro Faktor beträgt der Reliabilitätsverlust ($\Delta\omega$) nach Kürzung auf vier Items pro Faktor zwischen .01 (EI) und .08 (LI). Nennenswert ist ferner die hier nicht weiter diskutierte Beobachtung, dass die Modellgüte mit abnehmender Anzahl Items pro Faktor zunimmt.

Der zweite zentrale Befund ist, dass sich diese Ergebnisse mit den Daten der Originalstichprobe weitgehend replizieren lassen (s. dazu den unteren Teil von Tab. 1): Auch hier haben die Kürzungen in etwa dieselben Auswirkungen auf die internen Konsistenzen einerseits und die Fitmaße andererseits. Deskriptive Modellvergleiche zeigen zudem, dass die Reduktion von 15 (Vollversion) auf vier Items pro Dimension zu Reliabilitätsverlusten von minimal .04 (EI) und maximal .12 (LI) führt.

Faktorielle Validität

Gesamthaft betrachtet sprechen die Ergebnisse konfirmatorischer Faktorenanalysen für das postulierte Modell mit drei interkorrelierenden Faktoren erster Ordnung (s. Tab. 2), und zwar in der Untersuchungs- wie auch in der Originalstichprobe. Obwohl die signifikanten χ^2 -Tests auf keinen exakten Modellfit hinweisen, sprechen der RMSEA ($\leq .06$) sowie das SRMR ($\leq .08$) für eine gute Modellpassung. Der CFI und der TLI erreichen dann den geforderten Grenzwert ($\approx .95$), wenn im Modell sechs oder weniger Indikatoren pro Faktor berücksichtigt werden. Mit Blick auf die Fitmaße vermag von allen Modellen mit drei interkorrelierenden Faktoren jenes mit vier Indikatoren pro Faktor die implizierte Struktur am besten zu reproduzieren.

Auch der inferenzstatistische Vergleich des Dreifaktorenmodells mit vier Indikatoren pro Faktor und dem Einfaktormodell mit 12 Indikatoren mittels χ^2 -Differenztest (Untersuchungsstichprobe: $\Delta\text{MLR}\chi^2 = 653.63$, $df = 3$, $p < .001$; Originalstichprobe: $\Delta\text{MLR}\chi^2 = 2889.96$, $df = 3$, $p < .001$) und die entsprechenden Fitmaße lassen auf die faktorielle Validität der dreifaktoriellen Struktur schließen.

Vor dem Hintergrund dieser Ergebnisse und der Zielsetzung, die Dimensionen schulischen Integriertseins ökonomisch, aber dennoch messgenau zu erfassen, wird für die nachfolgenden Analysen zur Messinvarianz und zur Konstruktvalidität die Kurzversion mit vier Items pro Faktor verwendet.

Messinvarianz

Tabelle 3 gibt einen Überblick über die Fitmaße der berechneten Modelle sowie die Ergebnisse der χ^2 -Differenztests. Da die Fitmaße der in separaten Modellen verglichenen Gruppen auf einen guten Fit hinweisen (alle $\text{CFI} \geq .95$, $\text{SRMR} \leq .08$, $\text{RMSEA} \leq .06$), kann von konfigurationaler Messinvarianz ausgegangen werden.

Für die Überprüfung der metrischen Invarianz wurde ein Basismodell spezifiziert, in

Tabelle 1: Fit-Indizes und Reliabilitäten (ω) der FDI-Dimensionen in der Untersuchungs- und der Originalstichprobe

	Items	MLR χ^2	df	CFI	SRMR	RMSEA	$\lambda_{\min}/\lambda_{\max}$	ω (VI _{90%})
Untersuchungsstichprobe (N = 744)								
EI	10	293.31***	35	.927	.044	.100	.58/.87	.92 (.91-.93)
	8	135.98***	20	.959	.037	.088	.56/.88	.90 (.90-.91)
	6	51.13***	7	.979	.025	.079	.66/.89	.90 (.89-.91)
	4	3.06	2	.999	.005	.027	.82/.90	.91 (.90-.92)
SI	10	126.85***	35	.951	.037	.059	.39/.80	.86 (.84-.88)
	8	60.99***	20	.973	.028	.052	.60/.80	.87 (.85-.89)
	6	26.79**	9	.983	.022	.052	.61/.81	.86 (.84-.88)
	4	4.54	2	.995	.012	.041	.67/.80	.84 (.82-.86)
LI	10	257.49***	35	.903	.047	.092	.55/.75	.88 (.87-.90)
	8	159.56***	20	.924	.043	.097	.63/.76	.88 (.87-.89)
	6	62.95***	9	.956	.032	.090	.68/.72	.85 (.84-.87)
	4	9.28**	2	.989	.017	.070	.63/.77	.80 (.78-.82)
Originalstichprobe (N = 1620)								
EI	15	1266.07***	90	.905	.044	.090	.56/.84	.94 (.93-.94)
	10	691.21***	35	.923	.043	.108	.65/.85	.93 (.92-.93)
	8	255.68***	20	.963	.032	.085	.60/.77	.92 (.91-.92)
	6	50.07***	9	.991	.014	.053	.69/.88	.91 (.90-.91)
	4	1.99	2	1.000	.003	.000	.81/.88	.90 (.90-.91)
SI	15	1282.71***	90	.838	.060	.090	.37/.76	.89 (.88-.89)
	10	431.33***	35	.918	.044	.084	.56/.77	.87 (.86-.88)
	8	129.17***	20	.971	.027	.058	.55/.79	.85 (.84-.86)
	6	65.83***	9	.980	.023	.062	.56/.80	.85 (.83-.86)
	4	21.48***	2	.988	.018	.078	.57/.81	.81 (.79-.83)
LI	15	1000.86***	90	.911	.043	.079	.62/.78	.93 (.93-.94)
	10	466.53***	35	.932	.038	.087	.66/.80	.91 (.91-.92)
	8	369.33***	20	.929	.039	.104	.65/.80	.90 (.89-.90)
	6	145.36***	9	.955	.031	.097	.68/.74	.86 (.86-.87)
	4	4.23	2	.999	.007	.027	.64/.76	.81 (.80-.82)

Anmerkungen. EI: Emotionale Integration; SI: Soziale Integration; LI: Leistungsmotivationale Integration. MLR χ^2 : Robuste Maximum-Likelihood-Schätzung nach Yuan-Bentler. CFI: Comparative Fit Index; SRMR: Standardized Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; λ : standardisierte Faktorladungen; ω : Reliabilität (VI_{90%}: 90-Prozent-Vertrauensintervall).

** : p < .01; *** : p < .001.

Tabelle 2: Ergebnisse konfirmatorischer Faktorenanalysen verschiedener FDI-Versionen in der Untersuchungs- und Originalstichprobe

Modell	MLR χ^2	df	p	CFI	TLI	SRMR	RMSEA (VI _{90%})	CFit
Untersuchungssstichprobe (N = 744)								
30 Items (3 Faktoren)	1365.61	402	< .001	.895	.887	.054	.057 (.053-.060)	< .001
24 Items (3 Faktoren)	803.37	249	< .001	.925	.917	.048	.055 (.050-.059)	.033
18 Items (3 Faktoren)	411.03	132	< .001	.948	.940	.047	.053 (.048-.059)	.170
12 Items (3 Faktoren)	127.26	51	< .001	.977	.970	.036	.045 (.035-.055)	.799
12 Items (1 Faktor)	1509.13	54	< .001	.555	.457	.170	.190 (.182-.199)	< .001
Originalstichprobe (N = 1620)								
45 Items (3 Faktoren)	5476.83	942	< .001	.866	.859	.052	.055 (.053-.056)	< .001
30 Items (3 Faktoren)	2484.98	402	< .001	.908	.901	.043	.057 (.054-.059)	< .001
24 Items (3 Faktoren)	1272.86	249	< .001	.941	.935	.035	.050 (.048-.053)	.404
18 Items (3 Faktoren)	525.66	132	< .001	.968	.963	.031	.043 (.039-.047)	.999
12 Items (3 Faktoren)	130.99	51	< .001	.989	.986	.021	.031 (.025-.038)	> .999
12 Items (1 Faktor)	3651.47	54	< .001	.525	.419	.174	.203 (.198-.208)	< .001

Anmerkungen. MLR χ^2 : Robuste Maximum-Likelihood-Schätzung nach Yuan-Bentler. CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; SRMR: Standardized Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; VI_{90%}: 90-Prozent-Vertrauensintervall; CFit: p-Wert des Close-Fit-Tests (Wahrscheinlichkeit RMSEA \leq .05).

Tabelle 3: Ergebnisse zur Messinvarianzprüfung der FDI-Kurzskalen: Gruppenvergleiche bezüglich Schulleistung, Verhaltenskompetenzen und sprachlicher Herkunft in der Untersuchungsstichprobe

	MLR χ^2	df	CFI	TLI	SRMR	RMSEA (90% VI)	CFit	Δ MLR χ^2	Δ df	p
Schulleistung: Prozentrang > 25 (n = 431) vs. Prozentrang \leq 25 (n = 245)										
Gruppe PR > 25	133.17***	51	.961	.950	.043	.061 (.049-.074)	.072			
Gruppe PR \leq 25	66.35	51	.986	.981	.039	.035 (.000-.057)	.857			
Basismodell	201.00***	102	.969	.960	.042	.054 (.043-.064)	.283			
Metrisches Modell	209.32***	111	.969	.964	.049	.051 (.040-.062)	.413	8.49	9	.486
Skalares Modell	228.12***	120	.966	.963	.050	.052 (.041-.062)	.383	26.82	18	.082
Verhaltenskompetenzen: ohne (n = 573) vs. mit Verhaltensauffälligkeiten (n = 171)										
Gruppe Ohne	87.90**	51	.985	.980	.039	.036 (.022-.048)	.974			
Gruppe Mit	77.99**	51	.966	.957	.049	.056 (.028-.079)	.334			
Basismodell	166.57***	102	.980	.975	.041	.041 (.030-.052)	.900			
Metrisches Modell	176.10***	111	.980	.977	.047	.040 (.028-.050)	.941	9.15	9	.424
Skalares Modell	207.58***	120	.974	.971	.050	.044 (.034-.054)	.819	41.81	18	.001
Sprachliche Herkunft: Deutsch als Erstsprache (n = 432) vs. Deutsch als Zweitsprache (n = 308)										
Gruppe DaE	101.57***	51	.976	.968	.038	.048 (.034-.061)	.581			
Gruppe DaZ	80.42**	51	.977	.970	.043	.043 (.024-.061)	.719			
Basismodell	182.10***	102	.976	.969	.040	.046 (.035-.057)	.715			
Metrisches Modell	190.88***	111	.976	.972	.048	.044 (.033-.055)	.816	8.76	9	.460
Skalares Modell	211.12***	120	.973	.970	.050	.045 (.035-.055)	.772	28.86	18	.050

Anmerkungen. MLR χ^2 : Robuste Maximum-Likelihood-Schätzung nach Yuan-Bentler. CFI: Comparative Fit Index; TLI: Tucker-Lewis Index; SRMR: Standardized Root Mean Square Residual; RMSEA: Root Mean Square Error of Approximation; VI_{90%}: 90-Prozent-Vertrauensintervall; CFit: p -Wert des Close-Fit-Tests (Wahrscheinlichkeit RMSEA \leq .05). Δ MLR χ^2 : korrigierte MLR χ^2 -Differenz zum Basismodell.

** : $p < .01$, *** : $p < .001$.

dem alle Parameter jeweils beider Gruppen frei geschätzt wurden. Werden nun die Faktorenladungen jeweils in beiden Gruppen restringiert (Metrisches Modell in Tab. 3), so fällt die Modellgüte im Vergleich zum Basismodell in allen drei Gruppenvergleichen nicht signifikant schlechter aus (Schulleistung: $\Delta\text{MLR}\chi^2 = 8.49$, $df = 9$, $p = .486$; Verhaltenskompetenzen: $\Delta\text{MLR}\chi^2 = 9.15$, $df = 9$, $p = .424$; sprachliche Herkunft: $\Delta\text{MLR}\chi^2 = 8.76$, $df = 9$, $p = .460$), sodass auch die metrische Invarianz als gegeben betrachtet werden kann.

Für die Überprüfung skalarer Invarianz werden im Modell zusätzlich die entsprechenden Item-Mittelwerte in beiden Gruppen gleichgesetzt (Skalares Modell in Tab. 3). Der Fit dieses Modells unterscheidet sich in zwei Gruppenvergleichen nicht signifikant von dem des Basismodells (Schulleistung: $\Delta\text{MLR}\chi^2 = 26.82$, $df = 18$, $p = .082$; sprachliche Herkunft: $\Delta\text{MLR}\chi^2 = 28.86$, $df = 18$, $p = .050$). Im Gruppenvergleich bezüglich Verhaltenskompetenzen unterscheidet sich die Modellgüte des skalaren Modells zwar signifikant ($\Delta\text{MLR}\chi^2 = 41.81$, $df = 18$, $p = .001$) von jenem des Basismodells, allerdings dürfte dieser Unterschied praktisch kaum bedeutsam sein, da die Differenz des CFI mit .006 kleiner ist als der von Cheung und Rensvold (2002) vorgeschlagene Grenzwert von .01.

Zusammenfassend liefern die Befunde klare Hinweise dafür, dass mit der FDI-Kurzversion in den verglichenen Subgruppen das schulische Integriertsein äquivalent gemessen wird.

Konstruktvalidität

Für diese Analysen wurden die entsprechenden Items zu Skalensummenwerten zusammengefasst. Tabelle 4 informiert über einige statistische Kennwerte der FDI-Kurzskalen sowie der -Items. Fünf Beobachtungen seien dazu erwähnt: 1) Alle Trennschärfkoeffizienten der Items (r_{it}) liegen zwischen .58 und .84. 2) Die α - und ω -Koeffizienten (s. dazu Tab. 1) sind gleich, 3) die Mittelwerte aller

Skalen liegen im positiven Bereich (d.h. über dem theoretischen Mittelwert von 10), 4) die Skala SI weist eine deutlich rechtssteile Verteilung auf, und 5) die Skalenwerte sind gering bis moderat miteinander korreliert ($.17 \leq r \leq .29$).

Eine Übersicht über die Zusammenhänge zwischen den FDI-Kurzskalen und den Instrumenten zur Validierung bietet Tabelle 5. Die Zusammenhänge zwischen FDI- und Befindensskalen entsprechen den in Hypothese 1 formulierten Erwartungen: So bestehen zwischen den Kriteriumsvariablen SI bzw. LI und den Prädiktorvariablen PA und NA (PA erfasst das Ausmaß positiver, NA das Ausmaß negativer Affekte) signifikante Zusammenhänge kleiner bis mittlerer Größe – erfasst mittels bilanzierenden Fragebogenurteilen (Habituell: $R = .29$ bzw. $.27$, beide $p < .001$, $f^2 = .09$ bzw. $.08$) oder mittels aggregierten Werten zum aktuellen Befinden (Aktuell: beide $R = .24$, $p < .001$, $f^2 = .06$). Der Zusammenhang zwischen der Skala EI und der affektiven Grundstimmung ist mit $R = .55$ ($p < .001$, $f^2 = .44$) bzw. $.42$ ($p < .001$, $f^2 = .22$) klar höher.

Auch Hypothese 2 (eine substantielle, negative Beziehung zwischen den Skalen SI und *Peer-Probleme* (PP)) bewährt sich: Die Selbstberichte zum sozialen Integriertsein und das Ausmaß an Problemen mit Gleichaltrigen aus Sicht der Lehrperson sind hoch negativ korreliert ($r = -.45$, $p < .001$). Im Vergleich dazu sind die Beziehungen zu den übrigen SDQ-Skalen deutlich tiefer, in ihrer Richtung aber dennoch erwartungskonform ($.07 \leq |r| \leq .18$).

Die Skala LI ist – konform mit Hypothese 3 – mit den Skalen zur Lern- und Leistungsmotivation ($R = .43$, $p < .001$, $f^2 = .23$), der Selbstkonzeptskala ($r = .72$, $p < .001$) sowie den Schulleistungen ($r = .46$ bzw. $.40$; gesamthaft $R = .47$, alle $p < .001$) substantiell positiv korreliert. Eine hierarchische Regressionsanalyse zeigt zudem, dass die Skalen zur Lern- und Leistungsmotivation neben dem akademischen Selbstkonzept – obwohl statistisch signifikant ($\Delta F(3, 708) = 10.54$, $p < .001$) – kaum zusätzliche Varianz (ΔR^2

Tabelle 4: Interne Konsistenz (Cronbachs α) der KFDI-Skalen, (part-whole-korrigierte) Trennschärfekoeffizienten (r_{it}) der KFDI-Items sowie Verteilungskennwerte

Skala Item-Nummer und Item-Text	α	r_{it}	M	SD	Schiefe	Exzess
Emotionale Integration (EI)	.91		12.11	3.08	-0.71	0.01
1. Ich gehe gerne in die Schule.		.84	3.04	0.83	-0.57	-0.24
4. Ich habe keine Lust, in die Schule zu gehen. (-)		.78	3.26	0.89	-0.83	-0.15
7. Mir gefällt es in der Schule.		.81	3.11	0.84	-0.73	-0.05
10. Die Schule macht Spass.		.78	2.80	0.90	-0.41	-0.55
Soziale Integration (SI)	.84		14.08	2.39	-1.51	2.16
2. Ich habe sehr viele Freundinnen oder Freunde in meiner Klasse.		.68	3.43	0.79	-1.26	0.82
5. Ich komme mit meinen Mitschülerinnen und Mitschülern sehr gut aus.		.71	3.56	0.68	-1.43	1.48
8. In meiner Klasse fühle ich mich allein. (-)		.62	3.68	0.68	-2.17	-1.19
11. Mit meinen Mitschülerinnen und Mitschülern verbringe ich mich sehr gut.		.69	3.43	0.75	-1.19	0.87
Leistungsmotivationale Integration (LI)	.80		12.03	2.42	-0.35	0.00
3. Ich lerne schnell.		.65	2.89	0.81	-0.34	-0.40
6. Ich kann auch sehr schwierige Aufgaben lösen.		.61	2.84	0.78	-0.21	-0.42
9. Ich bin ein guter Schüler/eine gute Schülerin.		.63	3.06	0.71	-0.45	0.09
12. In der Schule ist mir vieles zu schwierig. (-)		.58	3.23	0.75	-0.71	0.02

Anmerkungen. Untersuchungsstichprobe (N = 744). Die mit einem (-) gekennzeichneten Items sind umgepolt. Kodierung: 1 = stimmt gar nicht, 2 = stimmt eher nicht, 3 = stimmt eher, 4 = stimmt genau.

Die Skala EI korreliert (r) mit .28 mit der Skala SI und mit .29 mit der Skala LI, die Skalen SI und LI sind mit .17 korreliert (alle $p < .05$).

= .021) in der Skala LI aufzuklären vermögen. Das bedeutet, dass die Dimension LI im Wesentlichen das Konstrukt akademisches Selbstkonzept abbildet.

Diskussion

Ziel dieser Untersuchung war die Entwicklung und teststatistische Überprüfung einer Kurzversion des FDI von Haerberlin et al. (1989). Mit dem „Kurzfragebogen zur Erfas-

sung von Dimensionen der Integration (KFDI 4-6)“ steht ein Instrument zur Verfügung, das die drei Dimensionen schulischen Integriertseins sehr ökonomisch, hinreichend reliabel und – wie erste Analysen zeigen – konstruktvalide zu messen vermag.

Mit insgesamt 12 Items (vier Items pro Dimension) ist der KFDI gegenüber der 45 Items umfassenden Originalversion wesentlich kürzer. Die Bearbeitung des Kurzfragebogens dürfte weniger als fünf Minuten in Anspruch nehmen. Die Items sind sprachlich

Tabelle 5: Einfache (*r*) und multiple (*R*) Korrelationen der KFDI-Skalen mit Validierungsskalen

FDI-Skalen	Selbstbericht (Q-Daten)						Fremdbericht (L-Daten)						Test (T-Daten)				
	Befinden Habituell			Aktuell			Selbst- konzept			Verhaltenskompetenzen						Schulleistung	
	PA	NA	PA	NA	PA	NA	SK	LZO	ALZ	VLZ	EP	VP	HY	PP	PRO	Ma	De
Emotionale Integration	<i>R</i>	.55		.42				.43				.22					.08
	<i>r</i>	.55	-.37	.42	-.30		.28	.42	.17	-.04	-.05	-.16	-.14	-.10	.21	-.06	-.08
Soziale Integration	<i>R</i>	.29		.24				.23				.46					.07
	<i>r</i>	.26	-.25	.19	-.24		.14	.19	.03	-.11	-.16	-.11	-.07	-.45	.18	-.03	-.07
Leistungsmotivationale Integration	<i>R</i>	.27		.24				.43				.37					.47
	<i>r</i>	.25	-.23	.17	-.24		.72	.30	.15	-.22	-.29	-.15	-.28	-.07	.05	.46	.40

Anmerkungen. Untersuchungsstichprobe ($649 \leq n \leq 744$). Die grau hinterlegten Zellen beziehen sich auf die zentralen Hypothesen. Kursiv gesetzte Korrelationskoeffizienten sind signifikant ($p < .05$). Multiple Korrelationen der jeweiligen FDI-Skala mit allen Validierungsskalen eines Konstrukts. PA: Positive Aktivierung, NA: Negative Aktivierung. SK: akademisches Selbstkonzept. LZO: Lernzielorientierung, ALZ: Annäherungs-Leistungsziel-Orientierung, VLZ: Vermeidungs-Leistungsziel-Orientierung, EP: Emotionale Probleme, VP: Verhaltensprobleme, HY: Hyperaktivität, PP: Peer-Probleme, PRO: Prosoziales Verhalten. De: Deutsch, Ma: Mathematik.

einfach formuliert und in der Regel auch für Kinder und Jugendliche mit geringem bis moderatem sonderpädagogischen Förderbedarf oder mit Deutsch als Zweitsprache gut verständlich. Analysen zur Messäquivalenz legen zudem den Schluss nahe, dass gruppenspezifische Eigenschaften (sprachliche Herkunft, Schulleistung und Verhaltenskompetenzen) das Antwortverhalten kaum verzerren und sich demzufolge (auf Skalenwerten basierende) Mittelwertsdifferenzen als tatsächliche Unterschiede auf Konstruktebene interpretieren lassen. Auch die internen Konsistenzen der Skalen sind trotz ihrer Kürze mit Werten zwischen .80 und .90 als gut bis sehr gut zu beurteilen.

Die konfirmatorischen Faktorenanalysen sprechen klar für die Validität der postulierten Faktorenstruktur des KFDI von drei gering bis moderat interkorrelierenden Faktoren erster Ordnung. Für den vielleicht etwas überraschenden Befund, nämlich dass die Anpassungsgüte des Modells mit abnehmender Anzahl Items pro Faktor zunimmt, sind wohl mehrere Gründe verantwortlich: Die besseren CFI-/TLI-Werte mit wenigen Items pro Faktor dürften mit Befunden aus Simulationsstudien (s. z.B. Marsh, Tau, Balla & Grayson, 1998; Kenny & McCoach, 2003) zusammenhängen, wonach sich der CFI und der TLI mit zunehmender Anzahl Items pro Faktor im Allgemeinen verschlechtern. Allerdings zeigen Kenny und McCoach (2003) auch, dass sich der RMSEA mit zunehmender Anzahl Items pro Faktor tendenziell verbessert, was in der vorliegenden Studie gerade nicht der Fall ist. Dies wiederum dürfte auf die Auswahl der Items zurückzuführen sein: Die für die Kurzversion ausgewählten Items sind demnach besonders gute Repräsentanten der erfassten Dimensionen, d.h. sie lassen sich charakterisieren durch hohe Faktorladungen, unbedeutenden Faktornebenladungen sowie geringen und unkorrelierten Messfehlern. Mit zunehmender Anzahl Items pro Faktor – so die „bad apple hypothesis“ (Kenny & McCoach, 2003, S. 349) – steigt die Wahrscheinlichkeit, unter all den guten auch schlechte Items zu haben, was sich entsprechend nega-

tiv auf die Anpassungsgüte des Modells auswirkt.

Die Beziehungen zwischen den FDI-Kurzskalen und den Skalen, die zur Bestimmung der Konstruktvalidität berechnet wurden, fallen gesamthaft betrachtet erwartungskonform aus. Zudem konnten die vermuteten gerichteten Zusammenhänge zwischen ausgewählten Skalen allesamt bestätigt werden. Der sprachliche Gehalt der Items der Skala LI sowie ihre Beziehung zur akademischen Selbstkonzept-Skala einerseits und zu den Skalen der Lern- und Leistungsmotivation andererseits legen eine Umbenennung der Dimension nahe: Die Bezeichnung *Kompetenz- bzw. Fähigkeitsbezogene Integration* oder – den Bezeichnungen *Emotionale* und *Soziale Integration* entsprechend – *Kognitive Integration* wäre wohl treffender.

Als Fazit bleibt festzuhalten, dass der KFDI in testologischer Hinsicht gute bis sehr gute Eigenschaften aufweist und vor allem für Untersuchungen empfohlen werden kann, in welchen das schulische Integriertsein von Kindern und Jugendlichen relativ global erfasst werden soll. Die sprachliche Einfachheit macht ihn gerade auch für sonder- und inklusionspädagogische Forschungsfragen hoch attraktiv. Die Kürze des Instruments erlaubt es zudem, den KFDI in Studien einzusetzen, in welchen Selbstberichte schulischen Integriertseins zwar nicht im Zentrum stehen, aber dennoch mitberücksichtigt werden wollen. Auch in der Praxis bietet sich der Einsatz des KFDI mit Einschränkungen an. Als Screeningverfahren kann er Lehrpersonen sowie Fachkräften der Sonder- bzw. Heilpädagogik wertvolle Anhaltspunkte zur Früherkennung mangelnden Integriertseins liefern und zur Evaluation von Interventionen oder Förderprogrammen genutzt werden. Für eine differenzierte Einzelfallanalyse sollte hingegen weiterhin die Originalversion eingesetzt werden.

Damit der KFDI in den erwähnten Anwendungsbereichen möglichst gewinnbringend genutzt werden kann, sind weitere Entwicklungsschritte notwendig: Erstens ist zu prüfen, ob sich die Verteilungskennwerte der

Skala SI mit schärfer formulierten Items verbessern lassen. Für den Einsatz des KFDI in Längsschnittuntersuchungen sind zweitens Analysen zur zeitlichen Stabilität sowie zur Messinvarianz über die Zeit erforderlich. Drittens sollten weitere Untersuchungen zur Validität durchgeführt werden, und zwar insbesondere auch zur kriterialen und prognostischen Validität. Schließlich sollten viertens Normen zur Verfügung gestellt werden.

Literaturverzeichnis

- Bless, G., Schüpbach, M. & Bonvin, P. (2004). *Klassenwiederholung – Determinanten, Wirkungen und Konsequenzen*. Bern: Haupt.
- Bless, G. & Mohr, K. (2007). Die Effekte von Sonderunterricht und gemeinsamem Unterricht auf die Entwicklung von Kindern mit Lernbehinderungen. In J. Walter & F. B. Wember (Hrsg.), *Sonderpädagogik des Lernens. Handbuch Sonderpädagogik* (S. 375-383). Göttingen: Hogrefe.
- Brown (2006). *Confirmatory factor analysis for applied research*. New York: The Guilford Press.
- 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.
- Eckhart, M. (2011). Soziale Integrationsprozesse in Schulklassen. Methodische Annäherung und Entwicklung eines Computerprogramms zur Analyse sozialer Prozesse in Schulklassen (Sozio). In J. Steppacher & A. Lanfranchi (Hrsg.), *Integration gelingt. Gute Praxis wahrnehmen, Neues entwickeln* (S.136-147). Bad Heilbrunn: Klinkhardt.
- Gebhardt, M., Schwab, S., Krammer, M. & Gasteiger Klicpera, B. (2012). Achievement and integration of students with and without special educational needs (SEN) in the fifth grade. *Journal of Special Education and Rehabilitation*, 13(3-4), 7-19.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of American Academy of Child and Adolescent Psychiatry*, 40(11), 1337-1345.
- Haeberlin, U., Moser, U., Bless, G. & Klaghofer, R. (1989). *Integration in die Schulklasse. Fragebogen zur Erfassung von Dimensionen der Integration von Schülern FDI 4-6*. Bern: Haupt.
- Haeberlin, U., Bless, G., Moser, U. & Klaghofer, R. (1990). *Die Integration von Lernbehinderten. Versuche, Theorien, Forschungen, Enttäuschungen, Hoffnungen*. Bern: Haupt.
- Hektner, J. M., Schmidt, J. A. & Csikszentmihalyi, M. (2007). *Experience Sampling Method. Measuring the quality of everyday life*. Thousand Oaks: Sage.
- Hu, L. & 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.
- Kahnemann, D. (1999). Objective happiness. In D. Kahnemann, E. Diener & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 3-25). New York: Russell Sage Foundation.
- Kelley, K. & Cheng, Y. (2012). Estimation of and confidence interval formation for reliability coefficients of homogeneous measurement instruments. *Methodology*, 8(2), 39-50.
- Kenny, D. A. & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling*, 10(3), 333-351.
- Klasen, H., Woerner, W., Rothenberger, A. & Goodman, R. (2003). Die deutsche Fassung des Strengths and Difficulties Questionnaire (SDQ-Deu) – Übersicht und Bewertung erster Validierungs- und Normierungsbe-funde. *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 52(7), 491-502.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.
- Kronig, W., Haeberlin, U. & Eckhart, M. (2000). *Immigrantenkinder und schulische Selektion*. Bern: Haupt.
- Marsh, H. W., Hau, K.-T., Balla, J. R. & Grayson, D. (1998). Is more ever too much? The

- number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research*, 33(2), 181-220.
- McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Erlbaum.
- Moser, U., Bless, G. & Haerberlin, U. (1989). Fragebogen zur Erfassung von Dimensionen der Integration von Schülern (FDI 4-6). *Psychologie in Erziehung und Unterricht*, 36(1), 17-26.
- Muthén, L. K. & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Raykov, T. (2002). Analytical estimation of standard error and confidence interval for scale reliability. *Multivariate Behavioral Research*, 37(1), 89-103.
- Raykov, T. (2007). Reliability if deleted, not 'alpha if deleted': Evaluation of scale reliability following component deletion. *British Journal of Mathematical and Statistical Psychology*, 60, 201-216.
- Ruijs, N. M. & Peetsma, T. T. D. (2009). Effects of inclusion on students with and without special educational needs reviewed. *Educational Research Review*, 4, 67-79.
- Sauer, S., Ide, S. & Borchert, J. (2007). Zum Selbstkonzept von Schülerinnen und Schülern an Förderschulen und integrativer Beschulung: Eine Vergleichsuntersuchung. *Heilpädagogische Forschung*, 33(3), 135-142.
- Schöne, C., Dickhäuser, O., Spinath, B. & Stiensmeier-Pelster, J. (2002). Das Fähigkeitsselfkonzept und seine Erfassung. In J. Stiensmeier-Pelster & F. Rheinberg (Hrsg.), *Diagnostik von Motivation und Selbstkonzept* (S. 3-14). Göttingen: Hogrefe.
- Schallberger, U. (2005). *Kurzskalen zur Erfassung der Positiven Aktivierung, Negativen Aktivierung und Valenz in Experience Sampling Studien (PANAVA-KS)*. Zürich: Psychologisches Institut der Universität Zürich.
- Schweizer, K. (2011). Some thoughts concerning the recent shift from measures with many items to measures with few items (Editorial). *European Journal of Psychological Assessment*, 27(2), 71-72.
- Spinath, B., Stiensmeier-Pelster, J., Schöne, C. & Dickhäuser, O. (2002). *Die Skalen zur Erfassung der Lern- und Leistungsmotivation (SELLMO)*. Göttingen: Hogrefe.
- Venetz, M. & Tarnutzer, R. (2011). Schulisches Integriertsein und Befinden im Unterricht. In J. Steppacher & A. Lanfranchi (Hrsg.), *Integration gelingt. Gute Praxis wahrnehmen, Neues entwickeln* (S.103-118). Bad Heilbrunn: Klinkhardt.
- Venetz, M., Tarnutzer, R., Zurbriggen, C. & Sempert, W. (2012). *Emotionales Erleben im Unterricht und schulbezogene Selbstbilder. Vergleichende Analysen von Lernenden in integrativen und separativen Schulformen*. Bern: SZH/CSPS.
- Yik, M. S. M., Russell, J. A. & Feldman Barrett, L. (1999). Structure of self-reported current affect: Integration and beyond. *Journal of Personality and Social Psychology*, 77(3), 600-619.

Dank

Ein besonderer Dank geht an Prof. Dr. Gérard Bless von der Universität Freiburg (Schweiz), der uns die Daten der Originalstichprobe für die Validierungsanalysen zur Verfügung gestellt hat.

Dr. Martin Venetz

Interkantonale Hochschule für
Heilpädagogik
Schaffhauserstr. 239
Postfach 5850
CH-8050 Zürich
martin.venetz@hfh.ch