

Bezugsgruppeneffekte auf das Fähigkeitsselbstkonzept Deutsch vor und nach dem Übertritt in die Sekundarstufe I

Markus P. Neuenschwander¹, Stephan Rösselet²,
Edith Niederbacher¹, Benno Rottermann³

¹ Pädagogische Hochschule FHNW, Solothurn

² Institut für Externe Schulevaluation auf der Sekundarstufe II IFES, Zürich

³ Pädagogische Hochschule FHNW, Brugg-Windisch

Zusammenfassung: Das Fähigkeitsselbstkonzept (FSK) von Schülerinnen und Schülern wird von ihren Leistungen, aber auch von sozialen Vergleichen mit Bezugsgruppen beeinflusst. Diese Vergleiche führen dazu, dass Kinder mit gleichen Leistungen in leistungsschwächeren Klassen ein höheres FSK aufweisen als in leistungsstärkeren Klassen (Big-Fish-Little-Pond Effekt). Über die Wirkung von früheren Klassenleistungen auf das FSK beim Verbleib in der Primarstufe im Vergleich zum Wechsel beim Übertritt in die Sekundarstufe I ist bisher wenig bekannt. Wir analysierten daher Vergleichsprozesse mittels Strukturgleichungsmodellen am Beispiel des Faches Deutsch anhand einer Längsschnittstichprobe mit zwei Gruppen: Eine Gruppe ($N=682$) verblieb in der Primarstufe, die andere Gruppe ($N=394$) wechselte im gleichen Alter nach dem 5. Schuljahr in die gegliederte Sekundarstufe I. Während beim Verbleib in der Primarstufe die individuelle Testleistung positiv und die Klassenleistung des 5. Schuljahres das FSK Deutsch im 6. Schuljahr negativ beeinflusste, fiel dieser Effekt beim Wechsel in die Sekundarstufe I weg. Die Bedeutung von sozialen Vergleichen auf das FSK in verschiedenen Stadien des Übergangs in die Sekundarstufe I wird diskutiert.

Schlüsselbegriffe: Fähigkeitsselbstkonzept, Bezugsgruppeneffekt, Leistungen, Transition, Längsschnitt

Effects of Reference Groups on German Academic Self-Concept Before and After the Transition to Lower Secondary School

Summary: Students' academic self-concept (ASC) depends on their individual achievement and the social comparison with reference groups. These comparisons lead to higher ASC for students with equal achievements in lower achieving classes compared to those in higher achieving classes (Big-Fish-Little-Pond Effect). Little is known about the effects of prior class achievement on ASC when students remain in primary school compared to same aged students who move on to lower secondary education. Therefore, we analyzed social comparisons using structural equation models for the case of the subject German by a longitudinal sample with two groups. The first group of $N=682$ students stayed in primary education, while the second group of $N=394$ same aged students transitioned to ability-grouped lower secondary education after 5th grade. While individual achievement showed positive and class achievement in 5th grade showed negative impact on students' ASC in 6th grade for students staying in primary education, this effect was not found for students transitioning to lower secondary education. The significance of social comparisons on ASC in varying stages of the transition to lower secondary education is discussed.

Keywords: Academic self-concept, frame of reference, achievement, transition, longitudinal data

Das schulische Fähigkeitsselbstkonzept (FSK) als mentales Modell einer Person über ihre schulischen Fähigkeiten und Eigenschaften (vgl. Moschner & Dickhäuser, 2006) ist ein wichtiger

Indikator für das Selbstbild von Kindern und Jugendlichen. Es hängt mit den schulischen Leistungen zusammen, wobei zwischen dem FSK und den Leistungen keine eindeutige Kau-

salitätsrichtung, sondern eine wechselseitige Beeinflussung identifiziert werden konnte (Marsh & O'Mara, 2008). Das FSK stellt überdies einen bedeutenden Einflussfaktor für den Verlauf des schulischen Interesses, die Entwicklung der Bildungsaspirationen und die Formation von Bildungsentscheidungen dar (Nagy, Trautwein, Köller, Baumert & Garrett, 2006; Trautwein, Lüdtke, Marsh, Köller & Baumert, 2006). Das FSK ist den Schulfächern entsprechend in verschiedene fachspezifische Facetten aufgeteilt (Markus & Wurf, 1987). Es wird durch soziale Vergleiche der eigenen Leistungen mit jenen von Bezugspersonen (externale Vergleiche) sowie temporale Vergleiche mit früheren Leistungen und dimensionale Vergleiche mit Leistungen in anderen Fächern (internale Vergleiche) beeinflusst (vgl. Köller, Trautwein, Lüdtke & Baumert, 2006). Im Folgenden wird gezeigt, wie sich das Fähigkeitsselbstkonzept Deutsch aufgrund von Vergleichen mit den schulischen Leistungen Gleichaltriger beim Übergang in die Sekundarstufe I und beim Verbleib in der Primarschule unabhängig vom Alter verändert.

Der Einfluss von Bezugsgruppen

Personen streben bei der Beurteilung ihrer Fähigkeiten nach sozialen Vergleichen mit anderen (Festinger, 1954). Für die Bewertung der schulischen Leistung bilden schulische Bezugspersonen wichtige soziale Referenzen von Schülerinnen und Schülern (vgl. Marsh, 2005). Dabei stellt die Schulklasse die wohl wichtigste Bezugsgruppe (*pond*) dar. Weil sich Schulklassen in ihrer durchschnittlichen Leistung voneinander unterscheiden, ist das FSK der Schülerinnen und Schüler (*fishes*) bei gleicher Leistung je nach Schulklasse unterschiedlich ausgeprägt (Big-Fish-Little-Pond Effekt; Marsh, 2005). Ein Kind mit einer bestimmten Leistung hat aufgrund seiner besseren Leistungsposition ein höheres FSK in einer leistungsschwächeren Klasse als in einer leistungsstärkeren Klasse, d. h. das FSK korreliert mit den Klassenleistungen negativ (Kontrasteffekt; vgl. Chmielewski, Dumont & Trautwein, 2013). Diese Effekte der Klassenleis-

tungen auf das FSK wurden wiederholt belegt und konnten für verschiedene Fächer und in unterschiedlichen Kulturen bestätigt werden (Pan-Human Theory; Marsh, 2005).

Stabilität des FSK und Veränderung bei Übergängen

Zahlreiche Studien belegten die hohe Stabilität des FSK von Schülerinnen und Schülern über die Zeit (vgl. zum Überblick Neuenschwander, 2005). Die Gründe hierfür liegen einerseits in der hohen Stabilität der Leistungsposition in einer Klasse (Helmke & Weinert, 1997). Andererseits bleibt auch die schulische Bezugsgruppe über die Zeit stabil, solange keine Übertritte stattfinden. Weil während der Primarstufe kaum Wechsel stattfinden, bleibt das FSK in dieser Phase stabil (Cole et al., 2001).

Der Übertritt in die Sekundarstufe I ist von Veränderungen der Schülerinnen und Schüler in zahlreichen Dimensionen begleitet: zum Beispiel die Strukturierung von Lernprozessen, Leistungsanforderungen, die Zusammensetzung und Organisation der Schulklasse und anderes (Sirsch, 2000). Zudem befinden sich Schülerinnen und Schüler in diesem Alter in einer Entwicklungsphase, die mit weitreichenden körperlichen und psychosozialen Veränderungen einhergehen (Barber & Olsen, 2004). Alle diese Veränderungen beeinflussen ihr FSK. Frühere Forschung zeigte jedoch exemplarisch für 13-Jährige, dass sich das FSK nicht pubertätsbedingt verändert (Arens, Seeshing Yeung, Craven, Watermann & Hasselhorn, 2013). Vielmehr verliert es an Stabilität über die Zeit, wenn ein Bezugsgruppenwechsel wie beim Schulübertritt in die Sekundarstufe I stattfindet, weil dieser zu Veränderungen der Leistungspositionen innerhalb der Klasse führt, die sich auf das FSK auswirken (Cole et al., 2001). Für diese Veränderungen des FSK sind also primär leistungsbezogene Merkmale der Klassenzusammensetzung (Komposition) und nicht entwicklungsbedingte Faktoren verantwortlich (Arens et al., 2013; Watermann, Klingebiel & Kurtz, 2010).

Die vorliegende Studie

Obwohl die Effekte der Bezugsgruppe auf das FSK von Schülerinnen und Schülern gut belegt sind, wurden Effekte des Bezugsgruppenwechsels beim Übertritt von der Primarstufe in die leistungssegregierte Sekundarstufe I bisher weniger untersucht (Schwarzer, Lange & Jerusalem, 1982; Wouters, De Fraine, Colpin, Van Damme & Verschueren, 2012). Dabei sind Übergänge für das Verstehen der Veränderung des FSK besonders interessant. Sie ermöglichen zu untersuchen, wie frühere Bezugsgruppen im Verhältnis zur aktuellen Bezugsgruppe das FSK beeinflussen. Insbesondere könnten Bezugsgruppeneffekte je nach Phase in einem Schulübergang unterschiedliche Wirkungen entfalten: Möglicherweise wirken sich Vergleichsprozesse vor einem Schulübergang weniger schnell auf das FSK aus als nach dem Schulübergang (Arens & Watermann, 2015). Schulübergänge, insbesondere der Übergang in die gegliederte Sekundarstufe I, erfordern eine Neupositionierung in der Bezugsgruppe, sodass frühere Vergleiche mit der Bezugsgruppe zugunsten von aktuellen Vergleichen an Bedeutung verlieren (Becker & Neumann, 2016).

In der vorliegenden Studie werden deshalb die Folgen sozialer Vergleiche für das FSK in Deutsch von Schülerinnen und Schülern mit und ohne Übertritt von der Primarstufe in die Sekundarstufe I untersucht. Im Zentrum steht die Frage, wie frühere und aktuelle individuelle Leistungen und Klassenleistungen das FSK beeinflussen, wenn die Kinder in der gleichen Klasse verbleiben oder aber in die Sekundarstufe I wechseln. Damit soll geklärt werden, ob Vergleiche mit der früheren Bezugsgruppe bei einer stabilen bzw. sich wechselnden Bezugsgruppe die gleiche Bedeutung besitzen. Um allfällige Alterseffekte zu kontrollieren, werden die längsschnittlichen Effekte in zwei Gruppen, die im gleichen Alter sind, berechnet. Unsere Analyse beschränkt sich auf das FSK Deutsch, weil das Fach Deutsch bei der Selektion in die Sekundarstufe I eine zentrale Bedeutung spielt (Neuen-

schwander, 2014). Wir vergleichen zwei Gruppen einer Stichprobe: Kinder der ersten Gruppe (Gruppe ohne Übertritt) verbleiben im fünften und sechsten Schuljahr in der Primarstufe (Übertritt nach dem sechsten Schuljahr). Kinder der zweiten Gruppe (Gruppe mit Übertritt) wechseln nach dem fünften Schuljahr in die gegliederte Sekundarstufe I. In den vier Schweizer Kantonen, die wir in unsere Studie einbezogen haben, umfasst die Sekundarstufe I drei Schulniveaus, wobei die Quoten der Schülerinnen und Schüler in den einzelnen Schulniveaus zwischen den Kantonen deutlich variieren (Neuenschwander, 2009). Die Klassentestleistungen in Schulniveaus mit hohen Anforderungen sind eher hoch, während die Klassentestleistungen in Schulniveaus mit tiefen Anforderungen eher tief sind. Wir postulieren folgende Hypothesen (vgl. Abb. 1):

Hypothese 1: Wir erwarten aufgrund der dargestellten Befunde zu den Bezugsgruppeneffekten für beide Gruppen (mit/ohne Übertritt) und zu beiden Messzeitpunkten positive Zusammenhänge der individuellen Deutschleistung (Gruppe mit Übertritt: $\gamma_{11} > 0$, $\gamma_{23} > 0$, Gruppe ohne Übertritt: $\gamma_{11} > 0$, $\gamma_{23} > 0$) und negative Zusammenhänge der Deutschleistungen der aktuellen Klasse mit dem FSK Deutsch der Schülerinnen und Schüler (Gruppe mit Übertritt: $\gamma_{12} < 0$, $\gamma_{24} < 0$, Gruppe ohne Übertritt: $\gamma_{12} < 0$, $\gamma_{24} < 0$, vgl. Abb. 1).

Vergleichsprozesse erhalten besonders viel Aufmerksamkeit, wenn Kinder nach dem Übertritt in eine leistungssegregierte Klasse eintreten, weil sie ihre Leistungsposition in der neuen Bezugsgruppe neu bestimmen müssen. Die Ergebnisse dieser aktuellen Vergleichsprozesse überlagern die Ergebnisse der früheren Vergleichsprozesse im Schuljahr vor dem Übertritt (Watermann et al., 2010). Daher verlieren die frühere Bezugsgruppe und die früheren individuellen Leistungen ihren Einfluss auf das FSK (Schwarzer et al., 1982). Die neue Bezugsgruppe stellt hingegen eine wichtige neue Informationsquelle dar und bildet die zentrale soziale Referenz für das FSK.

Hypothese 2: Zusätzlich zu den Zusammenhängen gemäß Hypothese 1 erwarten wir für die Gruppe mit Übertritt keinen Zusammenhang zwischen den individuellen Leistungen und den Klassenleistungen im Fach Deutsch im fünften Schuljahr und dem FSK Deutsch im sechsten Schuljahr ($\gamma_{21} = 0, \gamma_{22} = 0$, Abb. 1). Für die Gruppe ohne Übertritt werden diese längsschnittlichen Zusammenhänge hingegen angenommen.

Hypothese 3: Zusätzlich zu den Zusammenhängen gemäß Hypothese 1 erwarten wir für die Gruppe ohne Übertritt, dass die Klassenleistungen im Fach Deutsch der fünften Klasse mit dem FSK Deutsch der sechsten Klasse negativ ($\gamma_{22} < 0$) und die individuellen Leistungen positiv ($\gamma_{21} > 0$) zusammenhängen (Abb. 1). Für die

Schülerinnen und Schüler bleibt die Bezugsgruppe in der gleichen Zeit stabil. Vergleichsprozesse im fünften Schuljahr führen zu einer Leistungsposition in der Klasse, die das FSK im sechsten Schuljahr vorhersagt.

Um den Effekt des Schulüberganges explizit zu formulieren, postulieren wir die bereits in den Hypothesen 2 und 3 dargestellten Zusammenhänge zwischen den Leistungen und Klassenleistungen im fünften Schuljahr und dem FSK im sechsten Schuljahr zusätzlich in Form einer Moderationshypothese: In der Gruppe mit Übertritt sagen die individuellen Leistungen (γ_{21} , M1) und die Klassentestleistungen (γ_{22} , M2) im fünften Schuljahr das FSK im sechsten Schuljahr weniger stark vorher als in der Gruppe ohne Übertritt (Abb. 1).

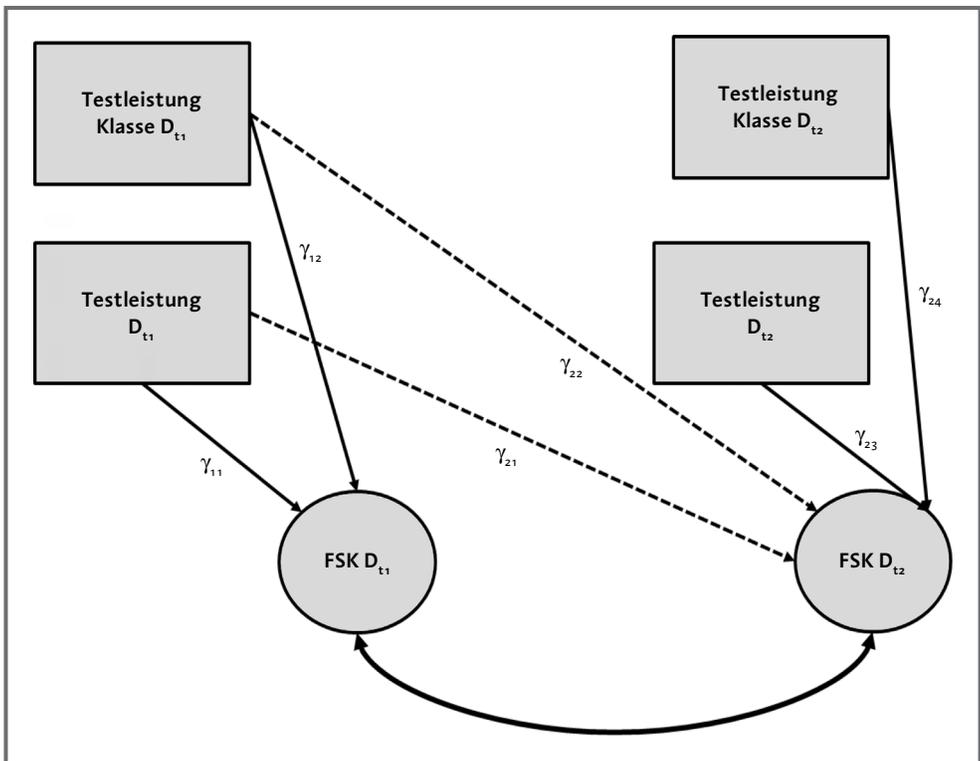


Abb. 1: Darstellung der postulierten Zusammenhänge

Anmerkungen: Gestrichelte Linien: Pfade werden nur für die Gruppe ohne Übertritt postuliert. FSK = Fähigkeitsselbstkonzept, D = Deutsch, t₁ = erster Messzeitpunkt, t₂ = zweiter Messzeitpunkt.

Tab. 1: Stichprobendesign und Stichprobengrößen (N)

	fünfte Klasse	sechste Klasse	N_{weiblich}	$N_{\text{männlich}}$
Gruppe ohne Übertritt (Kantone AG, BL)	Primarstufe	Primarstufe	202	192
Gruppe mit Übertritt (Kantone BE, LU)	Primarstufe	Sekundarstufe I	343	339

Anmerkungen: AG = Aargau, BL = Basel-Landschaft, BE = Bern, LU = Luzern.

Methode

Die Hypothesenprüfung erfolgt mit Daten der Längsschnittstudie *Wirkungen der Selektion* (WiSel¹). Für die Studie wurden Schulen aus den Schweizer Kantonen Aargau, Basel-Landschaft, Bern und Luzern zufällig ausgewählt und zur Teilnahme angefragt. Vor den Erhebungen wurde von den Kantonen, den Schulleitungen, den Lehrpersonen, den Eltern und den Kindern das ausdrückliche Einverständnis zur Teilnahme an der Studie eingeholt. Das Schuljahr beginnt in allen einbezogenen Kantonen Mitte August. Die erste Erhebungswelle (W1) fand zwischen Oktober 2011 und März 2012 im fünften Schuljahr statt, d. h. 10 bis 24 Wochen nach Beginn des Schuljahres (Teilnahme am Deutschtest: $N = 1514$), die zweite Erhebungswelle (W2) im sechsten Schuljahr zwischen Oktober 2012 und Dezember 2012, d. h. 10 bis 16 Wochen nach Beginn des Schuljahres (Teilnahme am Deutschtest: $N = 1816$). Die Längsschnittstichprobe bestand aus 1076 Schülerinnen und Schülern, die sowohl zum ersten als auch zum zweiten Messzeitpunkt den Deutschtest absolvierten.

Die ausgewählten Kantone unterschieden sich in Bezug auf den Übertrittszeitpunkt in die Sekundarstufe I. In den Kantonen Aargau und Basel-Landschaft fand der Übertritt nach dem fünften Schuljahr statt. Die Schülerinnen und Schüler in diesen Kantonen hatten bei der zweiten Erhebungswelle den Übertritt in die Sekundarstufe I bereits vollzogen (Gruppe mit Übertritt). In den Kantonen Bern und Luzern fand der Übertritt erst nach dem sechsten Schuljahr statt. Diese Schülerinnen und Schüler befanden sich bei der zweiten Erhebung noch immer in der Primarstufe, d. h. sie wechselten die Schule nicht (Gruppe ohne Übertritt). Je nach Kanton beginnt das Selektionsverfahren 12 bis 18 Monate vor dem Übertritt. Auf der Basis von Noten, Testleistungen, der Beurteilung von überfachlichen Kompetenzen und Empfehlungen von Lehrpersonen wird ein Kind ca. 3–6 Monate vor dem Übertritt einem Schulniveau der Sekundarstufe I zugewiesen.

Stichprobe

Die Gruppe mit Übertritt bestand aus $N_{\text{MO}} = 394$ Schülerinnen und Schülern (51,3 % weiblich, 74,7 % mit schweizerischer Staatsbürgerschaft oder Doppelstaatsbürgerschaft, durchschnittliches Alter 10.8 Jahre, vgl. Tab. 1). In der Gruppe ohne Übertritt befanden sich $N_{\text{OO}} = 682$ Schülerinnen und Schüler (50,3 % weiblich, 78,8 % mit schweizerischer Staatsbürgerschaft oder Doppelstaatsbürgerschaft, durchschnittliches Alter 10.7 Jahre).

Um systematische Rücklaufverzerrungen im Längsschnitt zu überprüfen, wurden die Schülerinnen und Schüler, die den Leistungstest nur in der ersten Welle ausgefüllt hatten (Querschnitt QS), hinsichtlich ihres FSK Deutsch und den Deutschleistungen zum ersten Messzeitpunkt mit denjenigen Schülerinnen und Schülern, die die Leistungstests zu beiden Messzeitpunkten ausgefüllt hatten (Längsschnitt LS), verglichen. Die beiden Gruppen unterschieden sich nicht in Bezug auf das FSK Deutsch ($M_{\text{QS}} = 4.59$, $SD_{\text{QS}} = 0.88$, $N_{\text{QS}} = 363$; $M_{\text{LS}} = 4.63$, $SD_{\text{LS}} = 0.82$, $N_{\text{LS}} = 896$; $t(1257) = -.86$, $p > .05$, Cohens $d = 0.01$). In Bezug auf die Deutschleistungen zum ersten Messzeitpunkt gab es einen signifikanten Unterschied zwischen den Gruppen ($M_{\text{QS}} = -0.05$, $SD_{\text{QS}} = 1.06$, $N_{\text{QS}} = 415$; $M_{\text{LS}} = 0.07$, $SD_{\text{LS}} = 1.03$, $N_{\text{LS}} = 1076$; $t(1489) = -1.98$, $p < .05$), die Effektstärke des Unterschieds war jedoch gering (Cohens $d = 0.01$).

Erhebungsinstrumente und verwendete Konzepte

FSK Deutsch

Das FSK der Schülerinnen und Schüler im Fach Deutsch wurde mit vier Items (z. B. „Ich bin im Fach Deutsch gut“ oder „Ich kann die Aufgaben gut lösen, die ich im Deutschunterricht bearbeiten muss“)

¹ Wir danken dem Schweizerischen Nationalfonds für die Finanzierung der Studie (Projektnummer 100013_134594/1).

im Schülerfragebogen erhoben (Eccles et al., 1989, Übersetzung ins Deutsche mit Rückübersetzung) und auf einer Antwortskala von *stimmt überhaupt nicht* (1) bis *stimmt voll und ganz* (6) beantwortet (FSK zu t1: Gruppe mit Übertritt: $\alpha = .90$, $M = 4.69$, $SD = 0.83$, $N = 363$, Anteil Missing 7,9%; Gruppe ohne Übertritt: $\alpha = .88$, $M = 4.59$, $SD = 0.81$, $N = 533$, Anteil Missing 21,8%; FSK zu t2: Gruppe mit Übertritt: $\alpha = .91$, $M = 4.59$, $SD = 0.83$, $N = 385$, Anteil Missing 2,3%; Gruppe ohne Übertritt: $\alpha = .90$, $M = 4.57$, $SD = 0.88$, $N = 665$, Anteil Missing 2,5%; Neuschwander, Rösselet et al., 2013; Neuschwander, Rösselet, Scheffler & Rottermann, 2014).

Leistungstest Deutsch

Die Leistungstests im Fach Deutsch bestanden aus einer Auswahl an lehrplanvaliden Aufgaben von Moser, Buff, Angelone und Hollenweger (2011; Dokumentation inkl. Reliabilitäts- und Validationsinformationen in Neuschwander, Rottermann & Rösselet, 2013). Der Deutschtest der ersten Welle umfasste sechs Aufgaben (Wortschatz, Anfangsbausteine/Vorsilben, Textverständnis und Lesen, Grammatik I und II, Orthografie) mit insgesamt 83 Items. Die einzelnen Items wurden dichotom (0 = falsch, 1 = richtig) kodiert. Die Schülerinnen und Schüler der Gesamtstichprobe lösten im Durchschnitt 48.5 der 83 Items korrekt. Die Reliabilität des Tests wurde mittels Split-half-Reliabilität erste Testhälfte vs. zweite Testhälfte ($r_{SH} = .80$) und gerade vs. ungerade Items ($r_{SH} = .92$) überprüft.

Der Deutschtest der zweiten Welle umfasste zehn Aufgaben (Wortschatz I und II, Wortarten, Adjektive, Textverständnis I, II und II, Grammatik, Anfangsbausteine/Vorsilben, Orthografie) mit insgesamt 85 Items. Eine Teilmenge der Items der ersten Welle wurde in einem Ankeritem-Design wiederholt vorgelegt. Die Schülerinnen und Schüler lösten im Durchschnitt 50.3 der 85 Items (Split-half-Reliabilität erste Testhälfte vs. zweite Testhälfte $r_{SH} = .82$, gerade vs. ungerade Items $r_{SH} = .95$).

Die Items konnten anschließend mit einer Linking-Methode, adaptiert von Haberman (2009), zu einem Testwert für jede Welle verrechnet werden. Mit dieser Linking-Methode wurde die Leistung mithilfe der Item Response Theorie (IRT; Yen & Fitzpatrick, 2006) auf einer Metrik abgebildet (warm likelihood estimates), auf der direkt die Veränderung der Kompetenzwerte beschrieben werden kann. Die Leistung der Schülerinnen und Schüler wurde auf

der Klassenebene durch Mittelwertbildung zum jeweiligen Messzeitpunkt aufgrund aller verfügbaren Werte in der Klasse aggregiert. Dieses Vorgehen ermöglicht die längsschnittliche Schätzung von Parametern bei gleichen und wechselnden Gruppenzusammensetzungen. Die Klassen der Gruppe ohne Übertritt waren zu beiden Messzeitpunkten bis auf die Drop-Outs und Drop-Ins identisch. Die Klassen der Gruppe mit Übertritt setzten sich nach dem Übertritt neu zusammen.

Durchführung

Die *Schülerbefragungen* und die *Leistungstests* wurden unter Anleitung von instruierten Projektmitarbeitenden im Klassenverband während der Unterrichtszeit durchgeführt. Das Ausfüllen der Schülerfragebogen beanspruchte inklusive Instruktion zwei Lektionen zu 45 Minuten. Der Leistungstest im Fach Deutsch dauerte eine Lektion (10 Minuten Einführung, 35 Minuten Testzeit). Er wurde 2–3 Wochen nach der Schülerbefragung durchgeführt.

Auswertungsmethode

Die Auswertungen der Daten und die Überprüfung der Hypothesen erfolgten schrittweise mit *Mplus 7.11* (Muthén & Muthén, 1998–2012). Im ersten Schritt wurden Korrelationen zwischen den verschiedenen Variablen berechnet. Anschließend wurden die Faktorenstruktur des FSK (Messinvarianzen) sowie die Strukturgleichungsmodelle mit den in den Hypothesen propagierten Pfaden getestet. Bei den Berechnungen wurde der gegenüber Verzerrungen aufgrund nichtnormalverteilter Daten robustere MLR-Schätzer verwendet. Für die durchgeführten Modellvergleiche mittels χ^2 -Differenzen wurde die Satorra-Bentler-Korrektur der χ^2 -Werte (TRD) angewendet. Zur Beurteilung des Modellfits wurden verschiedene Maße herangezogen: Comparative Fit Indices (CFI) über .95, Root Mean Square Errors of Approximation (RMSEA) unter .06 und Standardized Root Mean Square Residuals (SRMR) unter .08 weisen auf einen guten Fit der Modelle mit den erhobenen Daten hin (Hu & Bentler, 1999). Die Analysen zu den Mess- und Strukturmodellen basierten nur auf Daten derjenigen Personen, die zu beiden Messzeitpunkten den Leistungstest ausgefüllt hatten. Die Analysen zum Messmodell und zum Strukturmodell konnten so mit der jeweils gleichen Stichprobe

durchgeführt werden. Eine Überprüfung der Messmodelle ohne Ausschluss der Schülerinnen und Schüler, die nicht an beiden Tests teilgenommen hatten, führte zu vergleichbaren Ergebnissen. Fehlende Werte wurden mit der FIML-Prozedur, die in *Mplus* implementiert ist, behandelt, was gemäß Graham (2009) zu robusten Schätzwerten führt. Um die Mehrebenenstruktur der Daten (Schülerinnen und Schüler in Klassen) zu berücksichtigen, wurden die Auswertungen wie bei Becker und Neumann (2016) mit dem Analyseverfahren in *Mplus* Type = Complex durchgeführt. Die wechselnde Klassenzugehörigkeit bei der Gruppe mit Übertritt wurde in den Analysen berücksichtigt.

Ergebnisse

Korrelationen

Eine erste Übersicht über die Zusammenhänge des FSK Deutsch mit den Leistungen und Klassenleistungen ist für die Gruppe mit Übertritt (MÜ) und die Gruppe ohne Übertritt (OÜ) der Tabelle 2 zu entnehmen. Mittelwerte, Standardabweichungen und Korrelationen wurden mit den oben beschriebenen Einstellungen mit *Mplus* berechnet. Die Mittelwerte der FSK-Variablen wurden latent geschätzt, *Mplus* gibt keine Standardabweichungen dieser latenten Werte aus. Die Ladungen der einzelnen Items wurden frei geschätzt und weder zwischen den beiden Messzeitpunkten noch zwischen den Gruppen gleichgesetzt, im Unterschied zur Modellprüfung in Tabelle 2.

Die in Tabelle 2 dargestellten deskriptiven Ergebnisse belegen die Stabilität des FSK Deutsch ($r_{FSKt1*FSKt2} = .66$, respektive $.55$) und der Deutschleistungen ($r_{LDt1*LDt2} = .67$, respektive $.65$) über die Zeit. Der ebenfalls positive Zusammenhang der Klassenleistung fiel für die Gruppe mit Übertritt, in welcher die Klassen nicht identisch waren, tiefer aus ($r_{KLDt1*KLDt2_MÜ} = .30$) als bei der Gruppe ohne Übertritt ($r_{KLDt1*KLDt2_OÜ} = .67$).

Die positiven Zusammenhänge zwischen Deutschleistung und FSK Deutsch fielen zum ersten Messzeitpunkt in beiden Gruppen ähnlich hoch aus ($r_{LDt1*SKDt1_MÜ} = .43/r_{LDt1*SKDt1_OÜ} = .35$). Zum zweiten Messzeitpunkt war der Zusammenhang von individueller Deutschleistung und FSK Deutsch für die Gruppe mit Übertritt geringer ($r_{LDt2*SKDt2_MÜ} = .09/r_{LDt2*SKDt2_OÜ} = .42$). Dies ist ein Hinweis darauf, dass die Deutschleistung direkt nach dem Übertritt nicht zur Bildung des FSK Deutsch herangezogen wird. Der Zusammenhang von Klassenleistung und FSK war in der fünften Klasse bei der Gruppe ohne Übertritt signifikant positiv ($r_{KLDt1*SKDt1_MÜ} = .07/r_{KLDt1*SKDt1_OÜ} = .13$) und in der sechsten Klasse nicht signifikant ($r_{KLDt2*SKDt2_MÜ} = -.08/r_{KLDt2*SKDt2_OÜ} = .03$).

Messinvarianz

Mit der Testung der Messinvarianz wird sichergestellt, dass ein Messmodell für verschiedene Gruppen und Messzeitpunkte gleich gebildet

Tab. 2: Mittelwerte und Korrelationen

	Ohne Übertritt	FSK D _{t1}	FSK D _{t2}	L D _{t1}	L D _{t2}	KL D _{t1}	KL D _{t2}
Mit Übertritt	M	4.51	4.56	-.05	.28	-.11	.25
FSK D _{t1}	4.66	–	.66**	.43**	.37**	.13**	.08
FSK D _{t2}	4.57	.55**	–	.38**	.42**	.04	.08
L D _{t1}	.28	.35**	.06	–	.67**	.39**	.28*
L D _{t2}	.28	.31**	.09	.65**	–	.32**	.45**
KL D _{t1}	.24	.07	.01	.50**	.31**	–	.67**
KL D _{t2}	.28	.38**	.03	.60**	.67**	.30**	–

Anmerkungen: FSK = Fähigkeitsselbstkonzept, L = Leistung, KL = Klassenleistung, D = Deutsch, _{t1} = erster Messzeitpunkt, _{t2} = zweiter Messzeitpunkt; $N_{MÜ} = 396$, $N_{OÜ} = 682$; Werte oberhalb Diagonale: Gruppe ohne Übertritt (OÜ), Werte unterhalb Diagonale: Gruppe mit Übertritt (MÜ).

* $p < .05$, ** $p < .01$.

werden kann. Die Messinvarianz wird durch den Vergleich von unterschiedlich restriktiven Modellen nachgewiesen. Dem Step-Up-Ansatz (Brown, 2006; Christ & Schlüter, 2012) folgend, testeten wir in einem ersten Schritt ein Basismodell zur Überprüfung der konfiguralen Messinvarianz (Modell 1). Dieses Modell 1 bildet ein Messmodell mit einem Faktor mit vier Indikatoren (Items FSK Deutsch) zu den zwei Messzeitpunkten in den zwei Gruppen. Im Basismodell wurden die Korrelationen zwischen den Faktoren sowie die Ladungen zwischen Indikatoren und latenter Variable frei geschätzt. Die Residualvarianzen der jeweils gleichen Indikatoren wurden über die beiden Messzeitpunkte korreliert (Autokorrelation). Es wurden keine weiteren Korrelationen zwischen den Indikatoren zugelassen.

Die anschließenden Invarianzanalysen wurden jeweils mit beiden Gruppen (mit/ohne Übertritt) gleichzeitig durchgeführt. In einem zweiten Schritt wurde das Vorliegen metrischer Messinvarianz überprüft (Christ & Schlüter, 2012). Dazu wurden die Ladungen der Indikatoren auf den latenten Faktoren für beide Messzeitpunkte und für beide Gruppen gleichgesetzt. Der Vergleich des Modells 2 mit Modell 1 zeigt, ob die einzelnen Items zu den verschiedenen Messzeitpunkten für den latenten Faktor die gleiche Bedeutung haben.

Der Modell-Fit für das Modell 1 (Konfigurale Messinvarianz) war sehr gut ($N=1071$; $\chi^2=41.64$; $df=30$; $p>.05$; CFI=1.00; RMSEA=.03; SRMR=.02). Das metrisch messinvariante Messmodell (Modell 2) wies einen signifikant schlechteren Fit als Modell 1 auf ($N=1071$; $\chi^2=125.91$, $df=42$; $p<.01$; CFI=.97; RMSEA=.06; SRMR=.23; TRD=89.3; $\Delta df=12$; $p<.01$). Nach Sichtung der Modifikationsindices wurden die Ladungen von Item 1 für die Gruppe mit Übertritt zum ersten und zweiten Messzeitpunkt freigesetzt. Dieses partiell metrisch invariante Modell 3 wies keinen signifikant schlechteren Fit als Modell 1 auf ($N=1071$; $\chi^2=58.90$; $df=40$; $p<.05$; CFI=.99; RMSEA=.03; SRMR=.07; TRD=17.54; $\Delta df=10$; $p>.05$).

Hypothesentestung

Zur Prüfung der Hypothesen wurden fünf Modelle gerechnet. In Tabelle 3 sind die nicht standardisierten Pfadkoeffizienten der Analysen dargestellt.

In Modell 1 wurden nur die querschnittlichen Pfade der Leistungen und der Klassenleistungen im Fach Deutsch auf das FSK Deutsch der jeweiligen Messzeitpunkte für beide Gruppen (mit/ohne Übertritt) frei geschätzt (γ_{11} , γ_{12} , γ_{23} , γ_{24}). Die Korrelationen der Leistungen und Klassentestleistungen zwischen den beiden Messzeitpunkten wurden frei geschätzt. Die Pfade von den Leistungen und Klassenleistungen im Fach Deutsch der fünften Klasse auf das FSK Deutsch in der sechsten Klasse wurden 0 gesetzt. Die Ergebnisse zu Modell 1 belegten für die Gruppe *mit Übertritt* positive querschnittliche Zusammenhänge der Leistungen ($\gamma_{11}=.41$, $\gamma_{23}=.37$) und negative Zusammenhänge der Klassenleistungen im Fach Deutsch ($\gamma_{12}=-.27$, $\gamma_{24}=-.17$) mit dem FSK Deutsch. D. h. je besser die individuellen Leistungen sind, desto höher ist das FSK. Je besser die Klassenleistungen sind, desto tiefer ist das FSK. Die Korrelation des FSK zwischen den Messzeitpunkten war signifikant ($r=.55$), ebenfalls korrelierten die Leistungen und Klassentestleistungen signifikant zwischen den Messzeitpunkten. Bei der Gruppe *ohne Übertritt* waren die positiven querschnittlichen Zusammenhänge von Deutschleistungen und FSK Deutsch ebenfalls vorhanden ($\gamma_{11}=.36$, $\gamma_{23}=.37$), diejenigen von den Klassenleistungen mit dem FSK Deutsch erwiesen sich jedoch als gering und nicht signifikant ($\gamma_{12}=.03$, $\gamma_{24}=-.17$). Die Korrelation des FSK zwischen den Messzeitpunkten war signifikant ($r=.61$), ebenfalls korrelierten die Leistungen und Klassentestleistungen signifikant zwischen den Messzeitpunkten.

Die in Hypothese 1 formulierten querschnittlichen Zusammenhänge zu den Bezugsgruppeneffekten wurden mit Ausnahme des Effekts von der Klassenleistung auf das FSK im fünften Schuljahr in der Gruppe ohne Übertritt gefunden (Hypothese 1 zum Teil bestätigt).

Tab. 3: Modelltestung mittels Strukturgleichungen (unstandardisierte Pfade)

Modell	Gruppe	UVs	FSK D _{t₁}				FSK D _{t₂}				
			b	SE	Est./SE	p	b	SE	Est./SE	p	
M ₁	Mit Übertritt	LD _{t₁}	0.41	0.07	6.26	<.01	0.15	0.06	2.76	<.01	
		KLD _{t₁}	-0.27	0.13	-2.07	<.05					
M ₁	Ohne Übertritt	LD _{t₂}	.12	0.36	0.05	6.95	<.01	0.37	0.04	9.44	<.01
		KLD _{t₂}									
M ₂	Mit Übertritt	LD _{t₁}	0.45	0.07	6.06	<.01	0.10	0.07	1.54	.12	
		KLD _{t₁}	-0.29	0.14	-2.09	<.05	-0.04	0.11	-0.36	.72	
M ₂	Ohne Übertritt	LD _{t₂}	.14	0.47	0.05	0.28	<.01	0.29	0.06	4.80	<.01
		KLD _{t₂}									
M ₃	Mit Übertritt	LD _{t₁}	0.49	0.07	7.04	<.01	0.22	0.05	4.76	<.01	
		KLD _{t₁}	-0.37	0.14	-2.70	<.01	-0.21	0.09	-2.28	<.05	
M ₃	Ohne Übertritt	LD _{t₂}	.16	0.44	0.05	8.72	<.01	0.22	0.05	4.76	<.01
		KLD _{t₂}									
M ₄	Mit Übertritt	LD _{t₁}	0.48	0.07	6.99	<.01	0.22	0.05	4.89	<.01	
		KLD _{t₁}	-0.32	0.14	-2.39	<.05	-0.12	0.11	-1.07	.28	
M ₄	Ohne Übertritt	LD _{t₂}	.16	0.44	0.05	8.79	<.01	0.22	0.05	4.89	<.01
		KLD _{t₂}									
M ₅	Mit Übertritt	LD _{t₁}	0.47	0.07	6.36	<.01	0.15	0.07	2.18	<.05	
		KLD _{t₁}	-0.37	0.14	-2.67	<.01	-0.20	0.10	-2.05	<.05	
M ₅	Ohne Übertritt	LD _{t₂}	.15	0.45	0.05	8.84	<.01	0.26	0.06	4.21	<.01
		KLD _{t₂}									

Anmerkungen: FSK = Fähigkeitsselbstkonzept, D = Deutsch, L = Leistung, KL = Klassenleistung, _{t₁} = erster Messzeitpunkt, _{t₂} = zweiter Messzeitpunkt; N_{MÜ} = 396, N_{OU} = 682.

In Modell 2 wurden zusätzlich zu den Korrelationen in Modell 1 die Zusammenhänge der Leistungen und Klassenleistungen im Fach Deutsch der fünften Klasse mit dem FSK Deutsch der sechsten Klasse (γ_{21} und γ_{22}) für beide Gruppen (mit/ohne Übertritt) frei geschätzt. Dieses Modell diente insbesondere der Testung, ob die Klassenleistungen der fünften Klasse mit dem FSK Deutsch der sechsten Klasse in beiden Gruppen zusammenhängen. Gemäß Hypothese 2 hingen die Leistungen und Klassenleistungen im Fach Deutsch in der fünften Klasse und das FSK Deutsch in der sechsten Klasse bei der Gruppe *mit Übertritt* nicht signifikant zusammen ($\gamma_{21} = .10$, $\gamma_{22} = -.04$). Die Korrelation des FSK zwischen den Messzeitpunkten war signifikant ($r = .55$), ebenfalls korrelierten die Leistungen und Klassentestleistungen signifikant zwischen den Messzeitpunkten. Bei der Gruppe *ohne Übertritt* erwiesen sich die Zusammenhänge von Leistungen ($\gamma_{21} = .29$) und Klassenleistungen ($\gamma_{22} = -.42$) der fünften Klasse mit dem FSK Deutsch der sechsten Klasse gemäß Hypothese 3 als signifikant. Die Klassenleistungen im fünften Schuljahr unterdrückten den im Modell 1 gefundenen Effekt der Klassenleistungen im sechsten Schuljahr auf das FSK im sechsten Schuljahr. Die Klassenleistungen im sechsten Schuljahr haben nach dem Übertritt unter Kontrolle derjenigen des fünften Schuljahres keinen signifikanten Effekt mehr auf das FSK im sechsten Schuljahr. Die Korrelation des FSK zwischen den Messzeitpunkten war signifikant ($r = .62$), ebenfalls korrelierten die Leistungen und Klassentestleistungen signifikant zwischen den Messzeitpunkten.

Im nächsten Schritt wurden schließlich mittels χ^2 -Differenztest (Satorra-Bentler Scaled Chi Square Test) die Moderationshypothesen (M1, M2) getestet, ob sich die längsschnittlichen Pfade von den Leistungen im Fach Deutsch (γ_{21}) und den Klassenleistungen (γ_{22}) auf das FSK Deutsch zwischen den beiden Gruppen (mit/ohne Übertritt) signifikant voneinander unterscheiden.

In Modell 3 wurden die längsschnittlichen Pfade (γ_{21} und γ_{22}) von den Leistungen und Klassenleistungen im Fach Deutsch auf das FSK Deutsch für die beiden Gruppen (mit/ohne Übertritt) gleichgesetzt. Dieses Modell diente

der Testung, ob sich die Effekte der Deutschleistungen und -klassenleistungen der fünften Klasse auf das FSK Deutsch in der sechsten Klasse zwischen den beiden Gruppen (mit/ohne Übertritt) voneinander unterscheiden.

In Modell 4 wurde nur der längsschnittliche Pfad von den Deutschleistungen im fünften Schuljahr auf das FSK Deutsch im sechsten Schuljahr (γ_{21} , M1) für die beiden Gruppen (mit/ohne Übertritt) gleichgesetzt. Dieses Modell diente der Testung, ob sich allein die Effekte der Leistungen der fünften Klasse auf das FSK Deutsch in der sechsten Klasse zwischen den beiden Gruppen (mit/ohne Übertritt) voneinander unterscheiden.

In beiden Gruppen (mit/ohne Übertritt) waren die Korrelationen des FSK zwischen den Messzeitpunkten signifikant. Ebenfalls korrelierten die Leistungen und Klassentestleistungen signifikant zwischen den Messzeitpunkten.

In Modell 5 wurde nur der längsschnittliche Einfluss der Klassenleistungen im Fach Deutsch auf FSK Deutsch (γ_{22} , M2) für die beiden Gruppen (mit/ohne Übertritt) gleichgesetzt. Dieses Modell diente der Testung, ob sich alleine die Zusammenhänge der Klassenleistungen der fünften Klasse auf das FSK Deutsch in der sechsten Klasse zwischen den beiden Gruppen (mit/ohne Übertritt) voneinander unterscheiden. In beiden Gruppen (mit/ohne Übertritt) waren die Korrelationen des FSK zwischen den Messzeitpunkten signifikant. Ebenfalls korrelierten die Leistungen und Klassentestleistungen signifikant zwischen den Messzeitpunkten. Gemäß Tabelle 4 zeigten sich aufgrund der signifikanten χ^2 -Differenzwerte (rechte Spalte) signifikante Unterschiede zwischen den Zusammenhängen der Leistungen und Klassenleistungen Deutsch der fünften Klasse mit dem FSK Deutsch in der sechsten Klasse in den beiden Gruppen. Wenn allerdings beide Pfade gleichzeitig geprüft wurden, unterschied sich das Modell nicht signifikant von Modell 2. In den Modellen 3 bis 5 wurden die Kennwerte jeweils mit dem Modell 2 verglichen. Diese Ergebnisse bestätigen die in den Moderationshypothesen M1 und M2 formulierten Erwartungen zu den unterschiedlichen Vergleichsprozessen mit und ohne Übertritt in die Sekundarstufe I.

Tab. 4: Gütekriterien der Strukturgleichungsmodelle

Modell	RMSEA	CFI	SRMR	TLI	χ^2, df	p	$\Delta\chi^2$ korr.
M1	.053	.965	.083	.957	242.0, 96	<.01	
M2	.050	.971	.084	.962	215.7, 92	<.01	
M3	.050	.970	.077	.962	221.2, 94	<.01	5.52 (n.s.)
M4	.050	.970	.077	.961	219.5, 93	<.01	3.94 ($p < .05$)
M5	.050	.970	.082	.961	219.9, 93	<.01	4.50 ($p < .05$)

Diskussion

Erstmals wurden Bezugsgruppeneffekte auf das FSK Deutsch beim Übertritt in die gegliederte Sekundarstufe I im Vergleich zum Verbleib von Gleichaltrigen in der Primarstufe analysiert. Die Ergebnisse illustrieren in Weiterführung zu Becker und Neumann (2016) den Big-Fish-Little-Pond Effekt und die komplexen Vergleichsprozesse beim Übergang in die Sekundarstufe I, bei denen neben den eigenen Leistungen die aktuellen Leistungen der Bezugsgruppe (Marsh, 2005) für das FSK bestimmend sind. Wenn Kinder hingegen in der gleichen Bezugsgruppe bleiben, sind die früheren individuellen Leistungen und Klassentestleistungen einflussreicher, der Effekt der aktuellen Bezugsgruppe fällt weg. Die Ergebnisse zeigen am Beispiel des FSK Deutsch, welche Auswirkungen der Übertritt in die Sekundarstufe I auf Vergleichsprozesse mit der Bezugsgruppe hat und wie sich Vergleichsprozesse in verschiedenen Situationen der Schulkarriere unterschiedlich stark und nachhaltig auf das FSK auswirken.

Für die Situation im Jahr vor dem Übertritt in die Sekundarstufe I bestätigen die Ergebnisse den Big-Fish-Little-Pond Effekt für das Fach Deutsch weitgehend: Schulische Leistungen hängen positiv und Klassenleistungen negativ mit dem FSK zusammen. Sind die Schülerinnen und Schüler in ein Leistungsniveau der Sekundarstufe I übergetreten, müssen sie sich neu orientieren. Dabei nimmt die Bedeutung aktueller sozialer Vergleiche und damit der aktuellen Klassenleistung für das FSK zu. Vorausgegangene Leistungen und Klassenleistungen sind aufgrund des Bezugsgruppenwechsels nicht

relevant. Möglicherweise werden Bezugsgruppeneffekte in der Sekundarstufe I durch die Gliederung in Schulniveaus verstärkt, weil sich die Klassentestleistungen je nach Schulniveau deutlich unterscheiden (Aust, Waterman & Grube, 2010). Je nach Schulniveau vergleichen sich die Schülerinnen und Schüler nach dem Übertritt mit einer deutlich leistungsstärkeren, mit einer etwa gleich leistungsstarken und einer deutlich weniger leistungsstarken Klasse als in der Primarstufe.

Anders ist die Situation beim Verbleib in der Primarstufe. Kinder im fünften Schuljahr, bei denen kein Schulübertritt unmittelbar bevorsteht, fokussieren in geringem Maß auf soziale Vergleiche mit den Gleichaltrigen in ihrer Klasse. Daher hängt in dieser spezifischen Situation das FSK abweichend zu Hypothese 1 nicht von den Klassenleistungen ab. Diese Kinder treten mit dem Wechsel in das sechsten Schuljahr in ein Selektionsverfahren, das in eine Zuweisungsentscheidung in ein Niveau der Sekundarstufe I mündet. Während dieses Selektionsverfahrens beginnen sich die Kinder intensiv untereinander zu vergleichen. Ihr FSK wird durch die Leistungsposition in der Klasse bestimmt, die sie allmählich bis zum fünften Schuljahr gewonnen haben, welche den Effekt der aktuellen Leistungsposition in der Klasse unterdrückt. Daher sind in Übereinstimmung mit den Moderationshypothesen die früheren individuellen und Klassenleistungen für das FSK beim Verbleib in der gleichen Klasse bestimmender als beim Übergang in die Sekundarstufe I. Die Ergebnisse belegen, dass bei einem Schulwechsel Vergleichsprozesse hohe Aufmerksamkeit erhalten und sich sofort auf das FSK auswirken. In anderen Situationen

wirken sich aber diese Vergleichsprozesse mit Verzögerung aus (etwa beim Verbleib in der gleichen Primarschulklasse). Für die Bezugsgruppentheorie resultiert, dass Vergleichsprozesse nicht nur durch Persönlichkeitszüge moderiert werden (Jonkmann, Becker, Marsh, Lüdtke & Trautwein, 2012), sondern auch durch Schulübertritte. Sie wirken sich vor allem dann auf das FSK aus, wenn sich Kinder in einer neuen Gruppe bezüglich ihrer Leistung positionieren müssen.

Der Vergleich von zwei gleichaltrigen Gruppen in unterschiedlichen Situationen zeigt, übereinstimmend mit Arens et al. (2013), dass nicht altersbedingte Veränderungen bzw. die eintretende Pubertät zu Veränderungen des FSK von Schülerinnen und Schülern führen, sondern der Bezugsgruppenwechsel beim Übertritt in die Sekundarstufe I. Der gefundene negative Zusammenhang der Klassenleistung mit dem FSK nach dem Übertritt kommt auch dadurch zustande, dass die Schülerinnen und Schüler mit dem Übertritt in die Sekundarstufe I in ein Schulniveau übertreten, das zu einer Homogenisierung der Leistungsvarianz beiträgt. Die Zuweisung zu einem höheren (bzw. tieferen) Schulniveau führt zu einer neuen Leistungsposition in der neuen Klasse und ist demnach analog zur Veränderung der Noten (Neuenschwander, 2016) mit einer Abnahme (bzw. Zunahme) des FSK verbunden, weil Schülerinnen und Schüler sich auch mit anderen Klassen, die sich in einem anderen Leistungsniveau befinden, vergleichen.

Einschränkend muss auf die methodische Herausforderung hingewiesen werden, Bezugsgruppeneffekte bei wechselnden Klassenzugehörigkeiten pfadanalytisch zu überprüfen, weshalb die Testleistungen manifest und nicht latent erfasst werden mussten. Immerhin konnten längsschnittlich und unter Kontrolle der Mehrebenenstruktur und des Alters individuelle und Klasseneffekte auf das FSK beim Verbleib in einer Gruppe im Vergleich zum Wechsel einer Gruppe getestet werden, sodass die Nachhaltigkeit von Bezugsgruppeneffekten überprüft werden konnte. Vor einer einge-

henden Interpretation dieser Befunde ist eine Replikation erforderlich. Zudem wurde das FSK drei Wochen vor dem Leistungstest erfasst. Es ist unklar, ob dieser zeitliche Abstand die Ergebnisse beeinflusst. Unklar ist auch, ob die Befunde gleich sind, wenn die Jugendlichen in eine Gesamtschule ohne Leistungsniveaus übertreten. Weil nur zwei Messzeitpunkte berücksichtigt worden sind, sind keine Aussagen über die Langzeitwirkung der Bezugsgruppeneffekte möglich. Zukünftig soll untersucht werden, ob die beschriebenen Bezugsgruppeneffekte in allen Fächern (zum Beispiel Mathematik, Naturwissenschaften, Geschichte, Musik u. a.) gleichermaßen auftreten und ob zwischen den Vergleichsprozessen und gewissen Fachinhalten Interaktionen existieren.

Implikationen für die Praxis

Soziale Vergleiche sind für das FSK entscheidend. Eine Neugruppierung von Schulklassen, insbesondere bei Neueinteilungen in leistungssegregierte Gruppen, ist daher für das FSK zentral. Entsprechend führt die Zuweisung in eine Schulform mit hohen Anforderungen zu einer relevanten Abnahme des FSK (Möller, Retelsdorf, Köller & Marsh, 2011). Gegebenenfalls kann dies bei Kindern mit tiefem FSK im Selektionsverfahren berücksichtigt werden. Überdies sollten Lehrpersonen und Eltern bei Klassenwechseln (z. B. wegen Umzug) bzw. bei Änderungen in der Leistungsposition in einer Gruppe Auswirkungen auf das FSK beobachten. Dem FSK schädigenden sozialen Vergleichen können Lehrpersonen mit einer verstärkten individuellen bzw. lernzielorientierten Bezugsnorm (Holodynski, 2007) und einer FSK-dienlichen Attribuierung von schulischen Erfolgen und Misserfolgen (Weiner, 2009) entgegenwirken. Ebenfalls können selbstwertdienliche individuelle Rückmeldungen etwa in Verbindung mit Leistungstests eingesetzt werden (Übersicht z. B. in Lenske, 2016). Diese Optionen können insbesondere vor und nach Schulübergängen geprüft werden, um ungünstigen Veränderungen des FSK entgegenzusteuern.

Literatur

- Arens, A. K., Seeshing Yeung, A., Craven, R. G., Watermann, R. & Hasselhorn, M. (2013). Does the timing of transition matter? Comparison of German students' self-perceptions before and after transition to secondary school. *International Journal of Educational Research*, 57, 1–11. <https://dx.doi.org/10.1016/j.ijer.2012.11.001>
- Arens, A. K. & Watermann, R. (2015). How an early transition to high-ability secondary schools affects students' academic self-concept: Contrast effects, assimilation effects, and differential stability. *Learning and Individual Differences*, 37, 64–71. <https://dx.doi.org/10.1016/j.lindif.2014.11.007>
- Aust, K., Waterman, R. & Grube, D. (2010). Konsequenzen von Leistungsgruppierungen für die Entwicklungsverläufe des allgemeinen und fachspezifischen Fähigkeitsselfkonzepts nach dem Übergang in die Sekundarstufe. *Zeitschrift für Erziehungswissenschaft*, 12, 328–351. https://dx.doi.org/10.1007/978-3-531-92216-4_14
- Barber, B. K. & Olsen, J. A. (2004). Assessing the transitions to middle and high school. *Journal of Adolescent Research*, 19, 3–30. <https://dx.doi.org/10.1177/0743558403258113>
- Becker, M. & Neumann, M. (2016). Context-related changes in academic self-concept development: On the long-term persistence of big-fish-little-pond effects. *Learning and Instruction*, 45, 31–39. <https://dx.doi.org/10.1016/j.learninstruc.2016.06.003>
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford. <https://dx.doi.org/10.1080/00036810600603377>
- Chmielewski, A. K., Dumont, H. & Trautwein, U. (2013). Tracking effects depend on tracking type: An international comparison of students' mathematics self-concept. *American Educational Research Journal*, 50, 925–957. <https://dx.doi.org/10.3102/0002831213489843>
- Christ, O. & Schlüter, E. (2012). *Strukturgleichungsmodelle mit Mplus: Eine praktische Einführung*. München: Oldenbourg Verlag. <https://dx.doi.org/10.1524/9783486714807>
- Cole, D. A., Maxwell, S. E., Martin, J. M., Peeke, L. G., Seroczynski, A. D., Tram, J. M., ... & Maschman, T. (2001). The development of multiple domains of child and adolescent self-concept: A cohort sequential longitudinal design. *Child Development*, 72, 1723–1746. <https://dx.doi.org/10.1111/1467-8624.00375>
- Eccles, J. S., Wigfield, A., Flanagan, C., Miller, C., Reuman, D. & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality*, 57, 283–310. <https://dx.doi.org/10.1111/j.1467-6494.1989.tb00484.x>
- Festinger, L. (1954). A theory of social comparison process. *Human Relations*, 7, 117–140. <https://dx.doi.org/10.1177/001872675400700202>
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549–576. <https://dx.doi.org/10.1146/annurev.psych.58.110405.085530>
- Haberman, S. J. (2009). *Linking parameter estimates derived from an item response model through separate calibrations*. ETS Research Report No. RR-09-40. Princeton, NJ: ETS.
- Helmke, A. & Weinert, F. (1997). Bedingungsfaktoren schulischer Leistungen. In F. E. Weinert (Hrsg.), *Psychologie des Unterrichts und der Schule* (Enzyklopädie der Psychologie, Bd. 3, S. 71–176). Göttingen: Hogrefe.
- Holodynski, M. (2007). Entwicklung der Leistungsmotivation. In M. Hasselhorn & W. Schneider (Hrsg.), *Handbuch der Entwicklungspsychologie* (S. 299–311). Göttingen: Hogrefe.
- 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, 1–55. <https://dx.doi.org/10.1080/10705519909540118>
- Jonkmann, K., Becker, M., Marsh, H. W., Lüdtke, O. & Trautwein, U. (2012). Personality traits moderate the Big-Fish-Little-Pond Effect of academic self-concept. *Learning and Individual Differences*, 22, 736–746. <https://dx.doi.org/10.1016/j.lindif.2012.07.020>
- Köller, O., Trautwein, U., Lüdtke, O. & Baumert, J. (2006). Zum Zusammenspiel von schulischer Leistung, Selbstkonzept und Interesse in der gymnasialen Oberstufe. *Zeitschrift für Pädagogische Psychologie*, 20, 27–39. <https://dx.doi.org/10.1024/1010-0652.20.12.27>
- Lenke, G. (Hrsg.). (2016). *Schülerfeedback in der Grundschule*. Münster: Waxmann.
- Markus, H. & Wurf, E. (1987). The dynamic self-concept: A social psychological perspective. *Annual Review of Psychology*, 38, 299–337. <https://dx.doi.org/10.1146/annurev.ps.38.020187.001503>
- Marsh, H. W. (2005). Big-Fish-Little-Pond Effect on academic self-concept. *Zeitschrift für Pädagogische Psychologie*, 19, 119–127. <https://dx.doi.org/10.1024/1010-0652.19.3.119>
- Marsh, H. W. & O'Mara, A. (2008). Reciprocal effects between academic self-concept, self-esteem, achievement, and attainment over seven adolescent years: Unidimensional and multidimensional perspectives of self-concept. *Personality and Social Psychology Bulletin*, 34, 542–552. <https://dx.doi.org/10.1177/0146167207312313>
- Möller, J., Retelsdorf, J., Köller, O. & Marsh, H. W. (2011). The reciprocal I/E model: An integration of models of relations between academic achievement and self-concepts. *American Educational Research Journal*, 48, 1315–1346. <https://dx.doi.org/10.3102/0002831211419649>
- Moschner, B. & Dickhäuser, O. (2006). Selbstkonzept. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 685–692). Weinheim: Beltz.
- Moser, U., Buff, A., Angelone, D. & Hollenweger, J. (2011). *Nach sechs Jahren Primarschule. Deutsch, Mathematik und motivational-emotionales Befinden am Ende der 6. Klasse*. Zürich: Bildungsdirektion Kanton Zürich.
- Muthén, L. K. & Muthén, B. O. (1998–2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Nagy, G., Trautwein, U., Köller, O., Baumert, J. & Garrett, J. (2006). Gender and course selection in upper secondary education: Effects of academic self-concept and intrinsic value. *Educational Research and Evaluation*, 12, 323–345. <https://dx.doi.org/10.1080/13803610600765687>
- Neuenschwander, M. P. (2005). *Unterrichtssystem und Unterrichtsqualität. Konturen einer Unterrichtstheorie für die Sekundarstufe und ihre empirische Bewährung*. Bern: Haupt.

14 Markus P. Neuenschwander, Stephan Rösselet, Edith Niederbacher, Benno Rottermann

- Neuenschwander, M. P. (2009). Systematisch benachteiligt? Ergebnisse einer Studie zu Bildungssystem und -beteiligung. *Pädagogische Führung*, 20(3), 36–39.
- Neuenschwander, M. P. (2014). Selektionsentscheidungen beim Übergang in die Sekundarstufe I und in den Arbeitsmarkt im Vergleich. In M. P. Neuenschwander (Hrsg.), *Selektion in Schule und Arbeitsmarkt* (S. 63–98). Zürich: Rüegger.
- Neuenschwander, M. P., Rösselet, S., Rottermann, B., Singer, A., Scheffler, L. & Ziehli, M. (2013). *Wirkungen der Selektion WiSel. Dokumentation des Fragebogens für Schülerinnen und Schüler der Welle 1 2011/2012 (Forschungsbericht)*. Solothurn: Pädagogische Hochschule Nordwestschweiz, Zentrum Lernen und Sozialisation.
- Neuenschwander, M. P., Rösselet, S., Scheffler, L. & Rottermann, B. (2014). *Wirkungen der Selektion WiSel. Dokumentation des Fragebogens für Schülerinnen und Schüler der Welle 2 2012/2013 (Forschungsbericht)*. Solothurn: Pädagogische Hochschule Nordwestschweiz, Zentrum Lernen und Sozialisation.
- Neuenschwander, M. P., Rottermann, B. & Rösselet, S. (2013). *Wirkungen der Selektion WiSel. Dokumentation der Leistungstests Mathematik und Deutsch der Welle 1 2011/2012 (Forschungsbericht)*. Solothurn: Pädagogische Hochschule Nordwestschweiz, Institut Forschung und Entwicklung.
- Neuenschwander, M. P. (2016). Bildungsungleichheit am Beispiel der Leistungsentwicklung in Deutsch und Mathematik beim Übergang in die Sekundarstufe I. In B. Ziegler (Hrsg.), *(Un-)Gleichheiten in der Demokratie* (S. 95–118). Zürich: Schulthess-Verlag.
- Schwarzer, R., Lange, B. & Jerusalem, M. (1982). Selbstkonzeptentwicklung nach einem Bezugsgruppenwechsel. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 14, 125–140.
- Sirsch, U. (2000). *Probleme beim Schulwechsel*. Münster: Waxmann.
- Trautwein, U., Lüdtke, O., Marsh, H. W., Köller, O. & Baumert, J. (2006). Tracking, grading, and student motivation: Using group composition and status to predict self-concept and interest in ninth-grade mathematics. *Journal of Educational Psychology*, 98, 788–806. <https://dx.doi.org/10.1037/0022-0663.98.4.788>
- Watermann, R., Klingebiel, F. & Kurtz, T. (2010). Die motivationale Bewältigung des Grundschulübergangs aus Schüler- und Elternsicht. In K. Maaz, J. Baumert, C. Gresch & N. McElvany (Hrsg.), *Der Übergang von der Grundschule in die weiterführende Schule. Leistungsgerechtigkeit und regionale, soziale und ethnisch-kulturelle Disparitäten* (S. 355–383). Bonn: Bundesministerium für Bildung und Forschung.
- Weiner, B. (2009). *Motivationspsychologie*. Weinheim: Beltz.
- Wouters, S., De Fraine, B., Colpin, H., Van Damme, J. & Verschueren, K. (2012). The effect of track changes on the development of academic self-concept in high school: A dynamic test of the big-fish–little-pond effect. *Journal of Educational Psychology*, 104, 793–805. <https://dx.doi.org/10.1037/a0027732>
- Yen, W. M. & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational Measurement* (pp. 111–154). Westport: Praeger Publisher.

Prof. Dr. Markus P. Neuenschwander
Mag. phil. Edith Niederbacher
Pädagogische Hochschule FHNW
Institut Forschung und Entwicklung
Zentrum Lernen und Sozialisation
Obere Sternengasse 7
CH-4502 Solothurn
E-Mail: markus.neuenschwander@fhnw.ch
edith.niederbacher@fhnw.ch

Lic. phil. Stephan Rösselet
Institut für Externe Schulevaluation
auf der Sekundarstufe II IFES
Stampfenbachstr. 117
CH-8006 Zürich
E-Mail: roesselet@ifes-ipes.ch

Dr. phil. Benno Rottermann
Pädagogische Hochschule FHNW
Institut Primarstufe
Bahnhofstr. 6
CH-5210 Brugg-Windisch
E-Mail: benno.rottermann@fhnw.ch