Schulische Segregation: Technische Dokumentation und Detailergebnisse

Onlinedokument zum Kapitel
„Auf die Mitschüler/innen kommt es an? Schulische Segregation – Effekte der Schul- und Klassenzusammensetzung in der Primarstufe und der Sekundarstufe I“
im Band 2 des Nationalen Bildungsberichts Österreich 2015

Horst Biedermann, Christoph Weber, Barbara Herzog-Punzenberger & Arvid Nagel

In diesem Dokument werden die methodische Vorgehensweise sowie die Detailergebnisse in Bezug auf das Kapitel „Schulische Segregation“ im Band 2 des Nationalen Bildungsberichts (NBB) 2015 (vgl. Biedermann, Weber, Herzog-Punzenberger & Nagel, 2016) dargestellt. Vorab soll darauf verwiesen werden, dass alle Analysen mit Daten der Überprüfung der Bildungsstandards in Mathematik, 4. Schulstufe (BIST-Ü-M4; vgl. BIFIE, 2015a) und 8. Schulstufe (BIST-Ü-M8; vgl. BIFIE, 2015b) durchgeführt wurden. Die Analysen basieren auf jeweils 10 plausiblen Werten für die Mathematikleistung (vgl. dazu Trendtel, 2015).

A. Verwendete Methoden

Die Messung der sozialen und ethnisch-kulturellen Segregation

In Abschnitt 5.1.1 des NBB-Beitrags (vgl. Biedermann et al., 2016, S. 146 ff.) wird die soziale und ethnisch-kulturelle Segregation in den österreichischen Bezirken zwischen Schulen (interschulische Segregation) beschrieben. Daneben wird auch auf die Segregation innerhalb von Schulen (intraschulische Segregation) und zwischen Schultypen (AHS vs. HS/NMS) eingegangen.

Zur Beschreibung der interschulischen Segregation in den einzelnen Bezirken wird der Dissimilaritätsindex (DI) von Duncan und Duncan (1955) herangezogen.

Der DI berechnet sich für einen Bezirk wie folgt:

$$DI = \frac{1}{2}\sum_{j}\left|\frac{n_j^A}{N^A} - \frac{n_j^B}{N^B}\right|$$

n_j^Abzw. n_j^B ist hierbei die Anzahl der Schüler/innen der Gruppe *A* bzw. *B* in der Schule *j*, N^A bzw. N^B ist die Anzahl der Schüler/innen der Gruppe *A* bzw. *B* im Bezirk. Ein Vorteil des DI, der einer der meist verwendeten Segregationsindices ist (vgl. u.a. Altrichter, Bacher, Beham, Nagy & Wetzelhütter, 2011; Jenkins, Micklewright & Schnepf, 2008; Leckie, Pillinger, Jones & Goldstein, 2012), ist seine geradlinige Interpretation. Ein Wert von DI = 0,3 bedeutet etwa, dass 30% der Schüler/innen der Gruppe A (oder B) die Schule wechseln müssten, um eine Gleichverteilung (keine Segregation) zu erreichen, wobei die Annahme gilt, dass die Schüler/innen nicht durch andere Schüler/innen ersetzt werden.

Ein Nachteil des DI – wie auch von anderen deskriptiven Segregationsindizes – ist der Umstand, dass der DI die Abweichung von einer Gleichverteilung quantifiziert und nicht die Abweichung von einer zufälligen Aufteilung von Schüler/innen mit bestimmten Merkmalen auf Schulen (siehe u.a. Allen, Burgess, Davidson & Windmeijer, 2015; Mazza & Punzo, 2015).[1] Eine Folge daraus ist, dass der DI überschätzt wird, wobei der Bias besonders bei kleinen Schulen und auch kleinen Gruppengrößen höher ausfällt (vgl. Allen, Burgess & Windmeijer, 2009). Würde man etwa 100 Schüler/innen mit Migrationshintergrund und 100 Schüler/innen ohne Migrationshintergrund zufällig auf zwei Schulen zu je 100 Schüler/innen aufteilen, würde man annähernd eine Gleichverteilung erreichen. Werden die

[1] Des Weiteren basieren deskriptive Segregationsindizes auf beobachteten Anteilen, die auch die Stichproben-Variation beinhalten.

DOI: http://dx.doi.org/10.17888/nbb2015-2-4-1

Schüler/innen jedoch zufällig auf 20 Schulen (mit je 10 Schüler/innen) aufgeteilt, so werden sich die Schulen zum Teil deutlich in ihrem Migrant/innenanteil unterscheiden, was vom DI als Segregation ausgewiesen wird, tatsächlich aber das Resultat einer Zufallszuteilung ist. Ein solcher Bias würde Bezirksvergleiche zum Teil sehr erschweren, da für Bezirke mit kleinen Schulen (z.B. Bezirke im Burgenland; vgl. dazu Freunberger, Robitzsch & Pham, 2014) die tatsächliche Segregation stärker überschätzt wird, als in Bezirken mit größeren Schulen (u.a. Wien).

Um diesem positiven Bias entgegen zu wirken, wurde die von Leckie et al. (2012) vorgeschlagene mehrstufige Methode verwendet. Dabei wird in einem ersten Schritt das Ausmaß der Segregation auf Basis eines Mehrebenen-Binomial-Response-Modells geschätzt. Im Fall von zwei Ebenen ist das Modell wie folgt zu beschreiben (vgl. ebd., S. 9):

$$y_j \sim Binomial(n_j, \pi_j)$$
$$logit(\pi_j) = \beta_0 + u_j$$
$$u_j \sim N(0, \sigma_u^2)$$

y_j ist dabei der beobachtete Anteil von Schüler/innen der Gruppe A (z.B. Migrant/innen) in Schule j, n_j ist die gesamte Anzahl der Schüler/innen der Schule j und π_j ist der unbekannte, zugrundeliegende Anteil der Gruppe A in Schule j. π_j wird durch eine Logit-Linkfunktion mit einem linearen Prädiktor $\beta_0 + u_j$ in Verbindung gesetzt, wobei β_0 die Konstante und u_j das Residuum auf Schulebene ist. Die Varianz des Residuums σ_u^2 ist eine direkte Maßzahl der Segregation, da sie die Variation des Schüler/innenanteils der Gruppe A (auf einer Logit-Skala) quantifiziert.

Ein entsprechendes Modell wird für jeden Bezirk i geschätzt. Auf Basis der Koeffizienten β_{0i} und σ_{ui}^2 wird im nächsten Schritt der DI für die einzelnen Bezirke simuliert. Die Simulationsmethode basiert auf fünf Schritten, die M mal wiederholt werden. Die fünf Schritte sind (vgl. ebd., S. 14):

1.) Simuliere einen Wert für jede der j Random-Effekte auf Schulebene: $u_j^m \sim N(0, \sigma_u^2)$
2.) Berechne die Werte für π_j^m entsprechend:
$$\pi_j^m = \frac{1}{1 + e^{-(\beta_0 + u_j^m)}}$$
3.) Berechne die Anzahl der Schüler/innen der Gruppe A und B:
$$n_j^{m,A} = \pi_j^m n_j;\ n_j^{m,B} = n_j - n_j^{m,A}$$
4.) Aggregiere die Anzahl der Schüler/innen über die J Schulen:
$$N^{m,A} = \sum_j n_j^{m,A};\ N^{m,B} = \sum_j n_j^{m,B}$$
5.) Berechne DI:
$$DI^m = \frac{1}{2}\sum_j \left|\frac{n_j^{m,A}}{N^{m,A}} - \frac{n_j^{m,B}}{N^{m,B}}\right|$$

Der Mittelwert vom DI über die m Iterationen ist Punktschätzer für den DI. Die berichteten Ergebnisse basieren auf $M = 5000$ Iterationen. Erwartungsgemäß zeigen die Ergebnisse, dass die DI-Werte auf Basis der Simulationsmethode speziell in Bezirken mit kleinen Gruppengrößen (Anteil der Migrant/innen) bzw. in Bezirken mit kleinen Schulen geringer ausfallen als DI-Werte auf Basis der beobachteten Häufigkeiten.

Berechnung der sozialen Segregation

Die Berechnung vom DI setzt dichotome Variablen voraus. Der Sozialstatus ist jedoch eine z-standardisierte Variable (vgl. Freunberger et al., 2014). Diese wurde zur Berechnung vom DI – analog zur Vorgehensweise von Jenkins et al. (2008) – am Median dichotomisiert.

Durch Schultypen erklärte Segregation

Zur Berechnung des Ausmaßes der sozialen und ethnisch-kulturellen Segregation, die durch Schultypen (AHS vs. HS/NMS) erklärten werden kann, wurde entsprechend Leckie et al. (2012) das oben beschriebene Modell um den Prädiktor Schultyp erweitert, wobei die Berechnung nicht auf Bezirksebene, sondern auf Bundesebene erfolgt.

$$y_j \sim Binomial(n_j, \pi_j)$$
$$logit(\pi_j) = \beta_0 + \beta_1 Schultyp + u_j$$
$$u_j \sim N(0, \sigma_u^2)$$

Die auf Schulebene erklärte Varianz (R^2) gibt nun an, welcher Anteil der zwischenschulischen Segregation (σ_u^2) durch die Schultypen (0 = AHS, 1 = HS/NMS) erklärt werden kann.

Intraschulische Segregation

Zur Berechnung der intraschulischen Segregation wurde das Zweiebenen-Binomial-Response-Modell auf eine dritte Ebene erweitert (Schüler/in, Klasse, Schule):

$$y_{ij} \sim Binomial(n_{ij}, \pi_{ij})$$
$$logit(\pi_{ij}) = \beta_0 + v_j + u_{ij}$$
$$v_j \sim N(0, \sigma_v^2)$$
$$u_{ij} \sim N(0, \sigma_u^2)$$

Dabei ist y_{ij} der Anteil der Schüler/innen der Gruppe A in der Klasse *i* der Schule *j*. Die Varianz σ_v^2 bildet die zwischenschulische Segregation ab, während σ_u^2 das Ausmaß der intraschulischen Segregation quantifiziert. Der relative Anteil der intra- bzw. interschulischen Segregation wurde wie folgt berechnet:

Gesamtsegregation = $\sigma_u^2 + \sigma_v^2$

Prozentsatz der intraschulischen Segregation an der Gesamtsegregation = $\frac{\sigma_u^2}{(\sigma_u^2 + \sigma_v^2)} * 100$

Prozentsatz der intraschulischen Segregation an der Gesamtsegregation = $\frac{\sigma_v^2}{(\sigma_u^2 + \sigma_v^2)} * 100$

Diese Analysen wurden ebenfalls nur auf Bundesebene und nicht differenziert für die einzelnen Bezirke durchgeführt.

Modellschätzung

Die Modelle zur Berechnung der interschulischen Segregation und zur Frage der durch Schultypen erklärten interschulischen Segregation wurden in M*plus* 7 (vgl. Muthén & Muthén, 1998–2012) unter Verwendung einer robusten Maximum Likelihood Schätzung (MLR) durchgeführt.[2] Für die Analyse der intra- und interschulischen Segregation wurde eine Bayes-Schätzung mit Standardeinstellung (2 unabhängige Markov Chain Monte Carlo (MCMC) Ketten; Gelman-Rubin-Konvergenzkriterium = 0,05; uninformative Priors; siehe dazu Asparouhov & Muthén, 2010) verwendet, da für Dreiebenenmodelle mit kategorialen unabhängigen Variablen keine Maximum Likelihood (ML) Schätzung in M*plus* verfügbar ist. Für die Bayes Schätzung können grundsätzlich keine MI-Daten (TYPE = IMPUTATION in M*plus*) verwendet werden.[3] Daher wurde die Schätzung für alle 10

[2] Es soll noch kurz darauf hingewiesen werden, dass für die Berechnung der interschulischen Segregation zum Teil in den Bezirken eine geringe Anzahl an Schulen vorliegt (z.T. auch kleiner 5). Simulationsstudien (u.a. Bryan & Jenkins, 2015; Stegmueller, 2013) zeigen, dass bei geringerem N auf Ebene 2 (L2) unter der Verwendung einer ML-Schätzung die Modellparameter verzerrt geschätzt werden. Stegmueller (2013) hält jedoch fest, dass Modelle, die nur einen Random Intercept beinhalten noch akzeptable Ergebnisse liefern, solange N auf L2 größer 15 ist. Wobei jedoch ein kleineres N in erster Linie zu deutlich unterschätzten Konfidenzintervallen führt, die in der aktuellen Anwendung nicht von Interesse sind (σ_u^2 wird als deskriptive Maßzahl verwendet). Folglich ist mit keinen substanziellen Verzerrungen durch die ML-Schätzung zu rechnen.

[3] Bei der ML-Schätzung (bzw. allgemein beim klassischen „Frequentist"-Ansatz) wird der Standardfehler eines Parameters zur Bestimmung dessen Signifikanz bzw. zur Berechnung des Konfidenzintervalls entsprechend Rubin (1976) berechnet. Eine

Datensätze getrennt durchgeführt. Als Punktschätzer der Koeffizienten ($\beta_{0,}, \sigma_u^2$ und σ_v^2) wurde – wie bei der Analyse von MI-Daten üblich (vgl. u.a. Enders, 2010) – der jeweilige Mittelwert aus den 10 Analysen verwendet. Da die gemittelten Koeffizienten lediglich zur Deskription verwendet werden, ist die Berechnung der „Signifikanz" (das Credibility Intervall für die Bayes Schätzung) hinfällig (bzw. auch nicht möglich, siehe Fußnote 2).

Die Analyse von Kompositionseffekten

Die Analyse der Kompositionseffekte basiert auf einem Dreiebenenregressionsmodell (Ebene 1/L1 = Schüler/innen; Ebene 2/L2 = Klassen und Ebene 3/L3 = Schulen) der Form:

$$Mathe_{ijk} = \beta_0 + \beta_1(X_{ijk}) + \beta_2(X_{.jk}) + \beta_3(X_{..k}) + e_{ijk} + u_{0jk} + v_{00k}$$

X_{ijk} ist ein L1-Merkmal (z.B. Sozialstatus einer Schülerin/eines Schülers), $X_{.jk}$ das auf L2 aggregierte L1-Merkmal (z.B. der durchschnittliche Sozialstatus einer Klasse) und $X_{..k}$ ist das auf L3 aggregierte L1-Merkmal (z.B. der durchschnittliche Sozialstatus einer Schule). Die L1- und L2-Variablen wurden am Gesamtmittelwert (grandmean) zentriert (vgl. Enders & Tofighi, 2007). Folglich sind β_2 der Schul- und β_3 der Klassenkompositionseffekt. Bei der Analyse wurde schrittweise vorgegangen. In einem ersten Schritt wurden die Effekte der Kompositionsmerkmale (Sozialstatus, Geschlecht, Migrationshintergrund und Familiensprache) getrennt analysiert (d.h. das jeweilige Merkmal wurde auf L1- und aggregiert auf L2 und L3 berücksichtigt). In einem zweiten Schritt wurden die Effekte der Kompositionsmerkmale simultan betrachtet. Da die Anteile der Migrant/innen sowie Schüler/innen mit nichtdeutscher Familiensprache auf L2 und L3 sehr hoch korreliert sind, wurde bei der simultanen Betrachtung nur der Anteil der Schüler/innen mit nichtdeutscher Familiensprache berücksichtigt. Im dritten Schritt wurden zusätzlich noch weitere Kovariaten kontrolliert (Alter auf L1, Privatschule, Urbanisierungsgrad, Kleinstschule[4] auf L3; siehe dazu Variablenübersicht in Tabelle 4.2 im NBB-Beitrag: Biedermann et al., 2016, S. 145).

Bei den Analysen wurde der Multilevel Manifest Covariate (MMC) Ansatz (siehe dazu u.a. Lüdtke, Marsh, Robitzsch, Trautwein, Asparouhov & Muthén, 2008) verwendet, d.h. die Kompositionsmerkmale (z.B. durchschnittlicher Sozialstatus der Klasse) wurden als Mittelwert über alle Schüler/innen der jeweiligen Klasse bzw. Schule berechnet. Die Reliabilität von L2-/L3-Variablen, die auf diese Weise aus L1-Daten aggregiert werden, kann eingeschränkt sein, wenn Daten nur für eine Stichprobe einer Klasse/Schule vorliegen. Eine verzerrte Schätzung der Effekte kann die Folge sein (vgl. Lüdtke et al., 2008). Eine Alternative stellt der Multilevel Latent Covariate (MLC) Ansatz dar, der u.a. von Lüdtke et al. (2008) bei einem geringen Sampling Ratio (% der Schüler/innen je Klasse/Schule) für formative Merkmale (z.B. Anteil Migrant/innen, durchschnittlicher Sozialstatus, …) empfohlen wird. Dabei wird, anders als beim klassischen Zugang, bei dem L1-Variablen manifest aggregiert (MMC- Ansatz; d.h. Klassenmittelwert von L1-Messung als L2-Merkmal) werden, das L2-Merkmal latent geschätzt, wodurch der Stichprobenfehler korrigiert wird. Die Verwendung des MLC-Ansatzes ist jedoch bei Dreiebenenmodellen nicht möglich, da der MLC-Ansatz eine implizite Groupmean Zentrierung durchführt und bei Dreiebenenmodellen die Berechnung von Kompositionseffekten bei Groupmean Zentrierung nicht möglich ist (vgl. Brincks, 2012).[5] Der Vergleich des MMC- und MLC-Ansatzes zur Berechnung von Schulkompositionseffekten im Rahmen von Zweiebenenmodellen zeigt jedoch keine wesentlichen Unterschiede der Ansätze. Des Weiteren ist darauf zu verweisen, dass auf L2 eine Vollerhebung vorliegt (d.h. Sampling Ratio auf L2 geht gegen 100%), was beim MLC-Ansatz entsprechend der Simulationsergebnisse von Lüdtke et al. (2008) zur Überschätzung der Kontexteffekte führen würde. Jedoch ist möglicherweise die Reliabilität des Schulmittelwerts dadurch eingeschränkt, da nur für die 4. bzw. 8. Schulstufe Daten vorliegen (grob gerechnet ergibt sich auf Schulebene ein Sampling Ratio von ca. 25% = 1 von 4 möglichen

entsprechende Prozedur für die Berechnung des Credibility Intervalls (das Bayes-Äquivalent zum Konfidenzintervall) für MI-Daten ist nicht verfügbar.

[4] Anzahl der getesteten Schüler/innen < 10.

[5] Im Zweiebenenmodell ist der Kompositionseffekt bei Grandmean Zentrierung gleich dem Effekt auf L2. Bei Groupmean Zentrierung ist der Kompositionseffekt gleich der Differenz von L2 und L1 Effekt (siehe dazu u.a. Enders & Tofighi, 2007).

Schulstufen). Hier zeigen jedoch die Vergleiche auf Basis von Zweiebenenmodellen, dass mit keinen wesentlichen Unterschieden zwischen den Verfahren zu rechnen ist. Alle Analysen zu den Kompositionseffekten wurden mit M*plus* unter Verwendung der MLR-Schätzung durchgeführt.

Berechnung von Effektstärken

Zur Beurteilung der Stärke der Kompositionseffekte wurden Effektstärken nach Marsh et al. (2012) berechnet:

$$ES = 2b(SD_{PRED})/(SD_{UV})$$

wobei b der unstandardisierte Kontexteffekt ist, SD_{Pred} ist die Standardabweichung des Prädiktors und SD_{UV} ist die L1-Standardabweichung der abhängigen Variable. ES ist vergleichbar mit Cohens *d* (vgl. u.a. Cohen, 1992) und wird in Anlehnung an Eder, Altrichter, Hofmann und Weber (2015) wie folgt interpretiert: Effektstärken ab 0,15 als schwach, ab 0,35 als mittel und ab 0,55 als stark bezeichnet.

B. Detailergebnisse

Ergebnisse zu den Kompositionseffekten (Abschnitt 5.1.2 des NBB-Beitrags: Biedermann et al., 2016, S. 153 ff.)

In Tabelle 1 sind die Ergebnisse der Mehrebenenmodelle für die 4. und 8. Schulstufe dargestellt. In der ersten Spalte (M1) sind jeweils die „bivariaten“ Effekte abgebildet. Dabei wurden die Kompositionseffekte für die einzelnen Merkmale getrennt in den Blick genommen (M1 fasst die Ergebnisse von 5 unterschiedlichen Modellen zusammen). Bei M2 wurden die Kompositionseffekte simultan berücksichtigt und bei M3 wurden zusätzlich Kontrollvariablen (siehe oben) berücksichtigt. Da die Familiensprache und der Migrationshintergrund auf L2 und L3 sehr hoch korreliert sind, wurde bei M2 und M3 jeweils nur die Sprache berücksichtigt. Die anschließende Tabelle 2 gibt die Ergebnisse für die nach HS/NMS vs. AHS differenzierten Analysen wieder (Ergänzung zu Tabelle 4.3 im NBB-Beitrag: Biedermann et al., 2016, S. 155). Die abschließende Tabelle 3 gibt schließlich noch die Intraklassenkorrelation (ICC) für die verwendeten Variablen wieder.

Tabelle 1: Kompositionseffekte für BIST-Ü-M4 und BIST-Ü-M8 (Ergänzung zu Abbildung 4.8 im NBB-Beitrag: Biedermann et al., 2016, S. 154)

	BIST-Ü-M4						BIST-Ü-M8					
	M1		M2		M3		M1		M2		M3	
	b	ES	b	ES	b	ES	b	ES	b	ES	b	ES
Schüler/innen												
Sozialstatus	52,8***	0,88	51,8***	0,87	48,9	0,82	15,3***	0,51	14,6***	0,48	14,1***	0,47
Männlich	15,0***	0,16	15,3***	0,17	17,3	0,19	12,0***	0,20	12,6***	0,21	13,0***	0,22
Migrations-hintergrund	–36,5***	–0,30					–16,8***	–0,21				
Nicht-Deutsche Familiensprache	–34,9***	–0,26	–10,9***	–0,08	–8,5	–0,06	–17,9***	–0,22	–11,9***	–0,15	–10,2***	–0,13
Klasse												
Durchschnittlicher Sozialstatus	21,8***	0,19	22,3***	0,20	22,0***	0,19	99,3***	1,99	92,0***	1,85	89,2***	1,79
%Männlich	–9,4	–0,03	–6,3	–0,02	–6,1	–0,02	–4,5	–0,03	14,4***	0,09	13,8***	0,09
%Migrations-hintergrund	–62,1***	–0,26					–186,1***	–1,39				
%Nicht-Deutsche Familiensprache	–57,1***	–0,20	5,0	0,02	4,2	0,02	–185,2***	–1,36	–57,1***	–0,42	–53,1***	–0,39
Schule												
Durchschnittlicher Sozialstatus	–9,8**	–0,07	–22,1***	–0,17	–11,7**	–0,09	–25,1***	–0,42	–29,8***	–0,50	–55,4***	–0,93
%Männlich	–2,4	–0,01	–2,6	–0,01	–3,4	–0,01	–67,9***	–0,27	–13,6	–0,05	–15,2	–0,06
%Migrations-hintergrund	–13,8*	–0,05					56,7***	0,39				
%Nicht-Deutsche Familiensprache	–44,3***	–0,13	–89,3***	–0,27	–56,6***	–0,17	44,4***	0,30	–16,6*	–0,11	–7,6	–0,05
R^2												
Schüler/innen	0.007 – 0.196		0,206		0,22		0,010 – 0,063		0,079		0,095	
Klasse	0.002 – 0.38		0,139		0,138		0,000 – 0,702		0,728		0,727	
Schule	0.000 – 0.058		0,252		0,243		0,027 – 0,193		0,28		0,638	

Anmerkungen: *** $p < 0{,}001$; ** $p < 0{,}01$; * $p < 0{,}05$. Unstandardisierte Koeffizienten (b) und Effektstärken (ES). Die Effekte der Kontrollvariablen sind nicht dargestellt.

Tabelle 2: BIST-Ü-M8 Getrennte Analysen nach Schultypen (Ergänzung zu Tabelle 4.3 im NBB-Beitrag: Biedermann et al., 2016, S. 155)

	AHS		HS		HS und Leistungsgruppen	
	b	ES	b	ES	b	ES
Schüler/innen						
Sozialstatus	17,596***	0,493	12,034***	0,360	9,631***	0,288
Männlich	13,867***	0,213	12,582***	0,220	11,827***	0,207
Nicht-Deutsche Alltagssprache	–12,581***	–0,128	–9,103***	–0,125	–8,11***	–0,111
Klasse						
Durchschnittlicher Sozialstatus	18,465***	0,220	105,427***	1,241	18,495***	0,243
%Männlich	22,719***	0,148	1,642	0,007	–6,025	–0,033
%Nicht-Deutsche Alltagssprache	–14,135	–0,066	–57,096***	–0,297	–9,218*	–0,068
Schule						
Durchschnittlicher Sozialstatus	0,921	0,010	–64,422***	–0,669	0,135	0,028
%Männlich	–0,698	–0,003	–17,672	–0,071	–7,769	0,000
%Nicht-Deutsche Alltagssprache	–49,216*	–0,224	–0,207	–0,002	–25,969**	–0,173
R^2						
Schüler/innen	0,096		0,073		0,302	
Klasse	0,177		0,632		0,745	
Schule	0,196		0,757		0,299	

Anmerkungen: *** p < 0,001; ** p < 0,01; * p < 0,05. Unstandardisierte Koeffizienten (b) und Effektstärken (ES). Die Effekte der Kontrollvariablen sind nicht dargestellt.

Tabelle 3: ICC für die verwendeten Variablen

Gesamt		Mathematik	Sozialstatus	Familiensprache	Migrationshintergrund	Geschlecht
BIST-Ü-M8	Klasse	0,322	0,067	0,035	0,034	0,052
	Schule	0,276	0,241	0,251	0,256	0,019
BIST-Ü-M8-HS	Klasse	0,508	0,110	0,043	0,042	0,019
	Schule	0,064	0,055	0,259	0,267	0,019
BIST-Ü-M8-AHS	Klasse	0,091	0,043	0,022	0,021	0,117
	Schule	0,164	0,107	0,156	0,184	0,019
BIST-Ü-M4	Klasse	0,075	0,066	0,026	0,032	0,000
	Schule	0,102	0,126	0,124	0,160	0,001

Anmerkung: Die ICC wurden wie folgt berechnet: $ICC_{Schule} = \sigma_v^2 \,/\, (\sigma_v^2 + \sigma_u^2 + \sigma_e^2)$ und $ICC_{Klasse} = \sigma_u^2 \,/\, (\sigma_v^2 + \sigma_u^2 + \sigma_e^2)$, wobei σ_v^2 die Varianz auf Schulebene, σ_u^2 die Varianz auf Klassenebene und σ_e^2 die Varianz auf Schüler/innenebene ist (vgl. u.a. Hox, 2010).

Literatur:

Allen, R., Burgess, S. M. & Windmeijer, F. A. (2009). *More reliable inference for segregation indices.* Centre for Market and Public Organisation.

Allen, R., Burgess, S., Davidson, R. & Windmeijer, F. (2015). More reliable inference for the dissimilarity index of segregation. *The Econometrics Journal*, *18* (1), 40–66.

Altrichter, H., Bacher, J., Beham, M., Nagy, G. & Wetzelhütter, D. (2011). The effects of a free school choice policy on parents' school choice behaviour. *Studies in Educational Evaluation*, *37* (4), 230–238.

Asparouhov, T. & Muthén, B. (2010). *Bayesian Analysis Using Mplus: Technical Implementation.* Mplus Technical Report.

Biedermann, H., Weber, C., Herzog-Punzenberger, B. & Nagel, A. (2016). Auf die Mitschüler/innen kommt es an? Schulische Segregation – Effekte der Schul- und Klassenzusammensetzung in der Primarstufe und der Sekundarstufe I. In M. Bruneforth, F. Eder, K. Krainer, C. Schreiner, A. Seel & C. Spiel (Hrsg.), *Nationaler Bildungsbericht Österreich 2015, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 133–174). Graz: Leykam. DOI: http://dx.doi.org/10.17888/nbb2015-2-4

BIFIE (2015a). *Standardüberprüfung 2013 Mathematik, 4. Schulstufe (BIST-Ü-M413).* Unveröffentlichter Datensatz. Salzburg: BIFIE.

BIFIE (2015b). *Standardüberprüfung 2012 Mathematik, 8. Schulstufe (BIST-Ü-M812).* Unveröffentlichter Datensatz. Salzburg: BIFIE.

Brincks, A. (2012). *The Implications of Centering in a Three-Level Multilevel Model.* Open Access Dissertation. Paper 743.

Bryan, M. L. & Jenkins, S. P. (2016). Multilevel modelling of country effects: a cautionary tale. *European Sociological Review*, 32 (1), 3–22.

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112* (1), 155–159.

Duncan, O. D. & Duncan, B. (1955). A methodological analysis of segregation indexes. *American Sociological Review*, 20 (2), 210–217.

Eder F., Altrichter, H., Hofmann, F. & Weber, C. (Hrsg.). (2015). *Evaluation der Neuen Mittelschule (NMS). Befunde aus den Anfangskohorten. Forschungsbericht.* Graz: Leykam. Zugriff am 27.10.2015 unter https://www.bifie.at/system/files/dl/eval_forschungsbericht.pdf

Enders, C. K. & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological methods, 12* (2), 121.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Press.

Freunberger, R., Robitzsch, A. & Pham, G. (2014). *Hintergrundvariablen und spezielle Analysen in der BIST-Ü-M4 2013. Technischer Bericht.* Verfügbar am 08.09.2015 unter https://www.bifie.at/node/2765

Hox, J. J. (2010). *Multilevel Analysis. Techniques and Applications.* New York. Routledge.

Jenkins, S. P., Micklewright, J. & Schnepf, S. V. (2008). Social segregation in secondary schools: how does England compare with other countries? *Oxford Review of Education*, 34 (1), 21–37.

Leckie, G., Pillinger, R., Jones, K. & Goldstein, H. (2012). Multilevel modeling of social segregation. *Journal of Educational and Behavioral Statistics, 37* (1), 3–30.

Lüdtke, O., Marsh, H. W., Robitzsch, A., Trautwein, U., Asparouhov, T. & Muthén, B. (2008). The multilevel latent covariate model: a new, more reliable approach to group-level effects in contextual studies. *Psychological methods*, *13* (3), 203–229.

Marsh, H. W., Lüdtke, O., Nagengast, B., Trautwein, U., Morin, A. J., Abduljabbar, A. S. & Köller, O. (2012). Classroom climate and contextual effects: Conceptual and methodological issues in the evaluation of group-level effects. *Educational Psychologist*, *47* (2), 106–124.

Mazza, A. & Punzo, A. (2014). On the Upward Bias of the Dissimilarity Index and Its Corrections. *Sociological Methods & Research*. DOI: http://dx.doi.org/10.1177/0049124114543242

Muthén, L. K. and Muthén, B. O. (1998–2012). *Mplus User's Guide. Seventh Edition.* Los Angeles, CA: Muthén & Muthén.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63 (3), 581–592.

Stegmueller, D. (2013). How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science*, *57* (3), 748–761.

Trendtel, M. (2015). Skalierung der Leistungsdaten und Linking zur Baseline-Erhebung. Technische Dokumentation. BIST-Ü Mathematik, 4 Schulstufe, 2013. Salzburg: BIFIE. Verfügbar am 08.04.2016 unter https://www.bifie.at/node/2782