bifie | Bildung

*standards

Hintergrundvariablen und spezielle Analysen

Technische Dokumentation – BIST-Ü Mathematik, 4. Schulstufe, 2013

Roman Freunberger
Alexander Robitzsch
Giang Pham

Bundesinstitut bifie

Bundesinstitut bifie

Bundesinstitut für Bildungsforschung, Innovation & Entwicklung
des österreichischen Schulwesens
Alpenstraße 121 / 5020 Salzburg
www.bifie.at

Hintergrundvariablen und spezielle Analysen
Technische Dokumentation – BIST-Ü Mathematik, 4. Schulstufe, 2013*
BIFIE I Department Bildungsstandards & Internationale Assessments (BISTA),
Salzburg 2014

* Der ursprüngliche Titel der Publikation lautete: *Hintergrundvariablen und spezielle Analysen in der BIST-Ü-M4 2013. Technische Dokumentation.*

Inhaltsverzeichnis

1 Hintergrundvariablen

In der Bildungsstandardüberprüfung von 2013 für Mathematik auf der 4. Schulstufe (BIST-UE-M4 2013) kamen insgesamt vier Kontextfragebögen zum Einsatz: für Schüler/innen (SFB = Schülerfragebogen), Eltern (EFB = Elternfragebogen), Lehrer/innen (LFB = Lehrerfragebogen) und die Schulleitung (SLFB = Schulleiterfragebogen). Die Variablen, welche der Analyse von gruppenspezifischen Effekten sowie sozialen Vergleichen (s. Technische Dokumentation zum fairen Vergleich) und in weiterer Folge der Ergebnisrückmeldung[1] dienten, wurden aus einzelnen oder mehreren Items aus den Kontextfragebögen abgeleitet und werden im Nachfolgenden als Hintergrundvariablen bezeichnet.

Die erste Ebene der Ergebnisrückmeldung umfasste die Schüler/innen (Individualrückmeldung), die Unterrichtsgruppen und die Schulen. Eine Zusammenfassung der Schulergebnisse erfolgte an die Schulaufsicht. Die zweite Ebene involvierte die Landesberichte und den Bundesbericht, diese Ebene wird als Systemebene bezeichnet. Die konzeptionelle Hierarchie der Ergebnisrückmeldung – definiert durch die jeweilige Rezipientengruppe – bedingt weitestgehend auch die Analysestrategien sowohl für die Leistungsschätzer[2] als auch für abgeleitete Variablen (s. Kapitel 5).

Einige Fragebogen-Items kamen in der BIST-UE-M4 2013 sowohl im SFB als auch im EFB vor. Man kann hier entsprechend aus zwei Quellen die benötigte Information ableiten. Dies wurde zum Beispiel bei der Bildung des Sozialstatus gemacht (s. Kapitel 2). Man beachte hier, dass sowohl die Büchervariable (*Anzahl Bücher zu Hause*) als auch die Variablen zum HISEI (Highest International Socio-Economic Index of occupational status; Ganzeboom, De Graaf & Treiman, 1992) aus beiden Fragebögen (SFB und EFB) in die Berechnung mitaufgenommen wurden. Dieses Vorgehen war aus diversen Gründen nicht immer möglich (z. B. zu hoher Missing-Anteil, Konsistenz zu bisherigen Studien etc.), in diesen Fällen wurden für die jeweilige abgeleitete Variable die Daten aus dem SFB verwendet. Im SLFB sind Informationen zum Hintergrund auf Schulebene von Relevanz, diese wurden hauptsächlich für das Hintergrundmodell der Plausible-Value-Imputation verwendet (s. dazu die entsprechende Technische Dokumentation) und nicht direkt für die Ergebnisrückmeldung.

Hintergrundvariablen wurden entweder durch einfache Aggregationsschritte oder durch komplexere Analysen gebildet, welche in den nachfolgenden Kapiteln erläutert werden. Der Sozialstatus wird durch eine lineare Transformation aus spezifischen Variablen gebildet (s. Kapitel 2) und für die Skalen *Lernfreude* und *Selbstkonzept* werden jeweils Skalenwerte gebildet (s. Kapitel 4). Komplexere Analyseschritte sind für die Berechnung des Leistungsunterschieds zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund unter Kontrolle des Sozialstatus (s. Kapitel 3) sowie für die Bildung des Index der sozialen Benachteiligung (B, s. Kapitel 5) notwendig gewesen.

---

[1] Es sei an dieser Stelle auch auf die relevanten Berichte zu den Ergebnisrückmeldungen hingewiesen, diese sind Online im Downloadbereich unter https://www.bifie.at/node/64/ abrufbar.

[2] Auf Systemebene wird auf die Gesamtpopulation inferiert und daher werden hier Statistiken basierend auf den Plausible Values (PVs; Mislevy, Beaton, Kaplan & Sheehan, 1992) rückgemeldet. Auf Schul- und Individualebene sollten möglichst genaue Punktschätzer rückgemeldet werden, hierzu werden Weighted-Likelihood-Estimates (WLEs; Warm, 1989) anstelle der PVs berechnet.

2 Bestimmung des Sozialstatus

Für die Berechnung des Sozialstatus (S) wurden Variablen verwendet, die den sozioökonomischen Hintergrund abbilden (Tabelle 1; siehe auch OECD, 2005, 2009, 2012). Die z-standardisierten Werte der Variablen `HISEI`, `BUCH` (Anzahl der Bücher zu Hause) und `ELTAUSB` (elterliche Ausbildung) gingen in die Bestimmungsgleichung (s. Gleichung 3) mit ein. Der HISEI entspricht dem höchsten ISEI von entweder Vater (FISEI) oder Mutter (MISEI). Der ISEI bezieht Beruf, Ausbildung und Gehalt mit ein und wird aus den Berufsklassifikationen nach ISCO-08 (2008 International Standard Classification of Occupations; Ganzeboom et al., 1992) abgeleitet (d. h., das Gehalt wird nicht direkt in Fragebögen erfasst). Die ISCO-Kodierung erfolgt am BIFIE, daraus wird der HISEI mithilfe von Transfertabellen ermittelt. Für die Büchervariable und den HISEI wurden Werte aus dem EFB und dem SFB herangezogen, für die Variable der elterlichen Ausbildung wurden die Angaben aus dem EFB verwendet. Der Sozialstatus ist die gewichtete Summe aus den besagten Variablen, zusätzlich können durch die z-Standardisierung die Werte als Abweichungen von seinem Mittelwert interpretiert werden.

Tabelle 1: *Hintergrundvariablen für die Berechnung des Sozialstatus.*

Variable	Label	Werte
`ELTAUSB`	Höchste Ausbildung Eltern (4-stufig)	1 = max. Pflichtschule; 2 = Berufsschule, Meisterausbildung, Gesundheitspflege; 3 = Matura; 4 = Universität, FH, pädag. Akad.
`BUCH`	Anzahl Bücher zu Hause	1 = 0–10; 2 = 11–25; 3 = 26–100; 4 = 101–200; 5 = mehr als 200
`HISEI`	Highest International Socio-Economic Index of occupational status	Min. = 11.56, Max. = 88.96

Für den fairen Vergleich wurden fehlende Daten mittels multipler Imputation geschätzt (s. Technische Dokumentation zum fairen Vergleich), daraus resultierten 50 imputierte Datensätze. Für jede Imputation wurde für die oben genannten Variablen eine z-Transformation durchgeführt. Der Sozialstatus berechnet sich pro Imputation nach

$$\begin{aligned} S = {} & \frac{1}{6}\left(z(HISEI_{\text{SFB}}) + z(HISEI_{\text{EFB}}) + z(BUCH_{\text{SFB}}) + z(BUCH_{\text{EFB}})\right) \\ & + \frac{1}{3} z(ELTAUSB) \end{aligned} \tag{1}$$

Nachfolgend wurden pro Imputation von der Variable Sozialstatus S gewichtete Mittelwerte und Standardabweichungen berechnet. In Analogie zum Vorgehen bei der Berechnung von Statistiken aus multiplen Imputationen wurden anschließend ein Mittelwert (MW) und eine Standardabweichung (SD) über die 50 Imputationen berechnet. Mit den Werten von S aus jeder Imputation und mit den gepoolten Werten MW und SD wurde eine z-Transformation durchgeführt:

$$z(S) = \frac{S - MW(S)}{SD(S)} \tag{2}$$

Man erhält somit für jede der 50 Imputationen einen Wert für $z(S)$.

Für die Systemberichte (i. e. Bundes- und Landesberichte) wurde der Sozialstatus auf Basis der imputierten Daten aus der Plausible-Value-Ziehung (10 Datensätze) berechnet. Hierzu wurde für jeden der 10 Imputationsdatensätze der Sozialstatus anhand der Gleichung 3 bestimmt. Die Koeffizienten und der Interzept dieser Bestimmungsgleichung ergaben sich aus einer linearen Regression mit $z(S)$ als abhängige Variable (AV) und den nicht z-standardisierten Variablen `BUCH` (EFB und SFB), `HISEI` (EFB und SFB) und `ELTAUSB` (EFB) als unabhängige Variablen (UVs)[3]. Setzt man die sich ergebenden Werte ein, erhält man für M4 folgende Bestimmungsgleichung:

$$\begin{aligned} S = -3.318 + 0.348 \cdot & \left(\frac{BUCH_{\mathrm{SFB}} + BUCH_{\mathrm{EFB}}}{2}\right) \\ + 0.020 \cdot & \left(\frac{HISEI_{\mathrm{SFB}} + HISEI_{\mathrm{EFB}}}{2}\right) \\ + 0.445 \cdot & \, ELTAUSB \end{aligned} \tag{3}$$

Durch dieses Vorgehen ergibt sich für die 10 PV-Datensätze jeweils eine z-transformierte Variable des Sozialstatus, die mit der Sozialstatus-Variable der Rückmeldedaten auf Gruppen- und Schulebene (s. fairer Vergleich) vergleichbar ist.

3 Leistungsunterschiede nach Migrationsstatus

Für die BIST-UE-M4 2013 wurden drei Variablen zum Migrationshintergrund definiert, zwei dichotome Variablen und eine trichotome Variable. Der Migrationshintergrund eines Kindes wird nach OECD-Kriterien durch das Herkunftsland/Geburtsland der Eltern und Kinder ermittelt. Wenn mindestens ein Elternteil im Inland geboren wurde, weist das Kind keinen Migrationshintergrund auf. Die trichotome Variable wurde gebildet, um den Migrationsstatus, basierend auf der Definition nach OECD (2012), in drei Kategorien abzubilden (Kinder mit mindestens einem Elternteil aus Deutschland bilden allerdings eine Ausnahme, sie werden zur Gruppe der Schülerinnen und Schüler ohne Migrationshintergrund gezählt): (i) Inland (mind. 1 Elternteil in Österreich oder Deutschland geboren), (ii) Migrant 2. Generation (Eltern im Ausland geboren, aber Schüler/in in Österreich oder Deutschland geboren) und (iii) Migrant 1. Generation (Eltern und Kind im Ausland geboren).

Die dichotome Variable unterscheidet nur zwischen Schülerinnen und Schülern mit bzw. ohne Migrationshintergrund. Auch hier wurde Deutschland, abweichend zur OECD-Definition, als Ausnahme behandelt (BIFIE-Definition). Wurde mindestens ein Elternteil in Österreich oder Deutschland geboren, definierte dies die Zugehörigkeit zur Gruppe der Schüler/innen ohne Migrationshintergrund. Für die Rückmeldung auf Systemebene wurde – im Sinne der Vollständigkeit – auch der Anteil an Schülerinnen und Schülern mit Migrationsstatus laut OECD-Definition (i. e. inkl. Kinder deutscher Herkunft) berichtet. Für alle leistungsspezifischen Vergleiche wurde ausschließlich der Migrationsstatus nach BIFIE-Definition herangezogen.

[3] Die Regression wurde für den ersten Imputationsdatensatz zur Berechnung des fairen Vergleichs gerechnet, nicht pro Imputation gesondert.

Für den Bundesergebnisbericht und die Landesergebnisberichte wurde eine Grafik erstellt, in der die Gruppe der Schüler/innen mit Migrationshintergrund und die Gruppe der Schüler/innen ohne Migrationshintergrund hinsichtlich ihrer Mathematikleistung verglichen wurden. Zusätzlich wurde in einer gesonderten Grafik der Leistungsunterschied um den Sozialstatus korrigiert dargestellt. Die Frage nach der Leistungsdifferenz zwischen den Gruppen unter Kontrolle des Sozialstatus kann z. B. mittels einer linearen multiplen Regression analysiert werden. Man muss allerdings davon ausgehen, dass sich in den Kovariaten (z. B. Sozialstatus) Verteilungsunterschiede in den beiden Gruppen ergeben können (s. Abbildung 1). Das nachfolgende Verfahren verweist auf aktuelle Konzepte aus dem Bereich der Ökonometrie zum Thema Kausaler Inferenz (vgl. Gangl, 2010; Imbens & Wooldridge, 2009; Lüdtke, Robitzsch, Köller & Winkelmann, 2010; Morgan & Winship, 2007; Winship & Morgan, 1999) und versucht diese Heterogenität auszugleichen.

Bedient man sich der Terminologie des *Potential-Outcome*-Ansatzes, so kann man die Annahme treffen, dass der Migrationsstatus einen Treatmentindikator bzw. die Treatmentvariable[4] D darstellt. Schüler/innen können sich nun in zwei Zuständen befinden, sie weisen entweder einen Migrationshintergrund auf oder nicht. Die Problematik an der Erfassung des kausalen Effekts (des Migrationshintergrunds auf die Leistung) ist, dass die Individuen nicht zufällig der einen oder anderen Gruppe zugeordnet werden können, sondern eine *natürliche* Zuteilung stattfindet.

Die Treatmentgruppe wäre demnach die Gruppe der Schüler/innen mit Migrationshintergrund ($D = 1$, n_1 = Anzahl Schüler/innen mit Migrationshintergrund), die Kontrollgruppe wäre die Gruppe der Schüler/innen ohne Migrationshintergrund ($D = 0$, n_0 = Anzahl Schüler/innen ohne Migrationshintergrund). Mit Y wird die Mathematikleistung der Individuen bezeichnet. Im Sinne des kontrafaktischen Modells würde man nun die Frage stellen, welchen Leistungsunterschied Schüler/innen mit im Vergleich zu Schülerinnen und Schülern ohne Migrationshintergrund mit gleichem Sozialstatus aufweisen. Der kausale Effekt zwischen Treatment und Kontrollgruppe wird als Average Treatment Effect (ATE) bezeichnet und ist definiert durch

$$ATE = E[\delta] = E[Y|S = s, D = 1] - E[Y|S = s, D = 0], \tag{4}$$

dieser wäre unter Annahme einer randomisierten Zuweisung der Einheiten zum Treatment unverzerrt. Der ATE kann als erwarteter Wert von δ definiert werden und zum Beispiel durch eine einfache Mittelwertdifferenz geschätzt werden. Die beobachtete Mittelwertdifferenz zwischen den beiden Schülergruppen beträgt für Österreich 64 Punkte. Da der Migrationsstatus mit dem Sozialstatus konfundiert ist, muss man annehmen, dass der Mittelwertunterschied *verzerrt* ist. Für eine unverzerrte Schätzung der Leistungsdifferenz wurde daher ein Matchingverfahren angewendet, bei dem man im Wesentlichen Schüler/innen aus beiden Gruppen mit gleichem Sozialstatus vergleicht. Der kausale Effekt auf Individualebene wäre demnach definiert durch

$$\delta_i = y_{i1} - y_{i0}. \tag{5}$$

[4]Man könnte allerdings auch den Begriff *Bedingung* verwenden.

Die individuellen potenziellen Ergebnisse y_{i1} (Treatmentbedingung) und y_{i0} (Kontrollbedingung) können nicht für jedes Individuum beobachtet werden, da eine Schülerin oder ein Schüler nicht beiden Gruppen gleichzeitig angehören kann. Die nichtbeobachtbaren Ergebnisse, die potenziell möglich wären, werden aus einem statistischen Modell geschätzt. In der BIST-UE-M4 2013 wurde in einem Regressionsmodell unter Berücksichtigung der konfundierenden Variable *Sozialstatus* die Mathematikleistung für Schüler/innen mit und ohne Migrationshintergrund separat vorhergesagt.

Die Grundidee für dieses Vorgehen ist, die beiden Gruppen unter Berücksichtigung einer gleichen Verteilung des Sozialstatus zu vergleichen. Dies kann auf die Annahme zurückgeführt werden, dass zwischen Treatment D_i und potenziellen Outcomes $(y_{i0},\ y_{i1})$ keine strikte Unabhängigkeit vorliegt. Erst nach Einführung von Kovariaten kann man von einem *unkonfundierten* Verhältnis sprechen (Kenny, 1979; Schafer & Kang, 2008)[5]. In Anlehnung an das Vorgehen zur Berechnung von Treatmenteffekten durch eine ANCOVA mit nichtparallelen Regressionsgeraden wird ein Regressionsmodell mit einer konfundierenden Variable (Sozialstatus, S) für beide Treatmentgruppen separat wie folgt berechnet: Für die Gruppe der Schüler/innen mit Migrationshintergrund durch die Gleichung

$$\begin{aligned} \mu_1(Y|S=s) = \mu_1(s) &= E(Y|S=s, D=1) \\ &= \beta_{10} + \beta_{11}s + \beta_{12}s^2 + \beta_{13}s^3 + \beta_{14}s\mathbf{1}[s \geq 0], \end{aligned} \tag{6}$$

und für die Gruppe der Schüler/innen ohne Migrationshintergrund durch die Gleichung

$$\begin{aligned} \mu_0(Y|S=s) = \mu_0(s) &= E(Y|S=s, D=0) \\ &= \beta_{00} + \beta_{01}s + \beta_{02}s^2 + \beta_{03}s^3 + \beta_{04}s\mathbf{1}[s \geq 0] \end{aligned} \tag{7}$$

Für Schüler/innen mit Migrationshintergrund gilt $y_{i1} = y_i$, das heißt, dass die potenziellen Outcomes für die Gruppe der Schüler/innen mit Migrationshintergrund den beobachteten Leistungswerten entsprechen und es gilt weiters $y_{i0} = \mu_0(s_i)$, die Werte für y_{i0} werden für diese Schülergruppe aus dem Modell vorhergesagt. Für Schüler/innen ohne Migrationshintergrund gilt entsprechend $y_{i0} = y_i$ und $y_{i1} = \mu_1(s_i)$. Der ATE berechnet sich dann nach

$$\begin{aligned} ATE &= \frac{1}{n}\sum_i (y_{i1} - y_{i0}) \\ &= \frac{1}{n}\left(\sum_{i,D_i=1} (y_i - \hat{\mu}_0(s_i)) + \sum_{i,D_i=0} (\hat{\mu}_1(s_i) - y_i)\right) \end{aligned} \tag{8}$$

Vielfach wird auch der ATT (Average Treatment Effect on the Treated) berichtet. Der ATT wäre von Relevanz, wenn man auf die Population aller Schüler/innen mit Migrationshintergrund inferieren möchte und die Frage ist, was der vorhergesagte kausale Effekt wäre, wenn Schüler/innen aus der Treatmentgruppe (mit Migrationshintergrund) auf den hypothetischen Status *Schüler/in ohne Migrationshintergrund* wechseln würden ($D_i = 1$ zu $D_i = 0$). Der ATT berechnet sich nach

[5]In der Missing-Data-Terminologie würden die unbeobachteten potenziellen Outcomes somit als „missing at random“ (MAR) definiert werden können.

$$ATT = \frac{1}{n_1} \sum_{i,D_i=1} (y_i - \hat{\mu}_0(s_i)) \tag{9}$$

Tabelle 2: *Beispieldaten zur Berechnung des ATE und ATT.*

D_i	s_i	y_i	y_{i0}	y_{i1}	$y_{i1} - y_{i0}$
0	–1.620	551	551	**446**	–105
0	–0.764	483	483	**475**	–7
0	–0.605	502	502	**481**	–20
0	–0.271	439	439	**495**	55
0	1.093	574	574	**554**	–19
0	1.109	443	443	**555**	112
0	1.452	586	586	**570**	–15
0	1.564	701	701	**575**	–126
1	–1.496	410	**463**	410	–52
1	–1.251	385	**479**	385	–94
1	–0.865	489	**503**	489	–14
1	1.104	443	**587**	443	–144
1	1.213	580	**591**	580	–11

Anmerkungen: Fett geschriebene Werte sind nicht beobachtbar und werden durch das Regressionsmodell ermittelt.

Abbildung 1 zeigt unter anderem, dass bei einem mittleren Sozialstatus Schüler/innen ohne Migrationshintergrund eine höhere Mathematikleistung zeigen als Schüler/innen mit Migrationshintergrund. Der Leistungsunterschied, berechnet durch eine einfache Mittelwertdifferenz, sinkt nach Kontrolle des Sozialstatus von 64 Punkten auf 34 Punkte (ATE)[6]. Die konfundierende Variable Sozialstatus kann somit nur einen Teil des Gruppenunterschieds aufklären.

Beispiel zur Berechnung des ATT und ATE

Tabelle 2 zeigt beispielhaft fiktive Daten von 13 Personen für $D_i = 1$ und $D_i = 0$, man beachte hier, dass fett geschriebene Werte nicht beobachtbar sind und daher durch das Regressionsmodell vorhergesagt werden. Zur Berechnung des ATE werden die Differenzwerte von $y_{i1} - y_{i0}$ gemittelt, was in diesem Beispiel eine mittlere Differenz von 34 Punkten ergibt, berechnet nach $ATE = \frac{1}{n} \sum_i (y_{i1} - y_{i0}) = \frac{1}{13}(-442) = -34$. Zur Berechnung des ATT benötigt man den Mittelwert über y_{i0}, dieser beträgt $\hat{\mu}_0(s_i) = 531$. Die mittlere Differenz $y_i - \hat{\mu}_0(s_i)$ über die Gruppe $D_i = 1$ entspricht dem ATT, in diesem Beispiel: $ATT = \frac{1}{n_1} \sum_{i,D_i=1} (y_i - \hat{\mu}_0(s_i)) = \frac{1}{5} \sum_{i,D_i=1} (y_i - 525) = \frac{1}{5}(-316) = -63$.

Der Zusammenhang mit Regressionsmodellen

Abbilung 1 zeigt, dass sich die vorhergesagten Mathematikleistungen für Schüler/innen mit Migrationshintergrund und ohne Migrationshintergrund bei gleichem Sozialstatus unterscheiden. Diese Mathematikleistungen wurden mithilfe der Funktion $\mu_1(s)$ bzw. $\mu_0(s)$ mit einer nichtlinearen Regression bestimmt. Dabei fällt auf, dass die Differenz $\mu_1(s) - \mu_0(s)$ von s abhängig ist, also der „Treatmenteffekt“ Migrationshintergrund mit dem Sozialstatus variiert. Der ATE bestimmt eine gewichtete Differenz mit dem Ansatz $ATT =$

[6] Der ATT wäre für Österreich bei 28 Punkten.

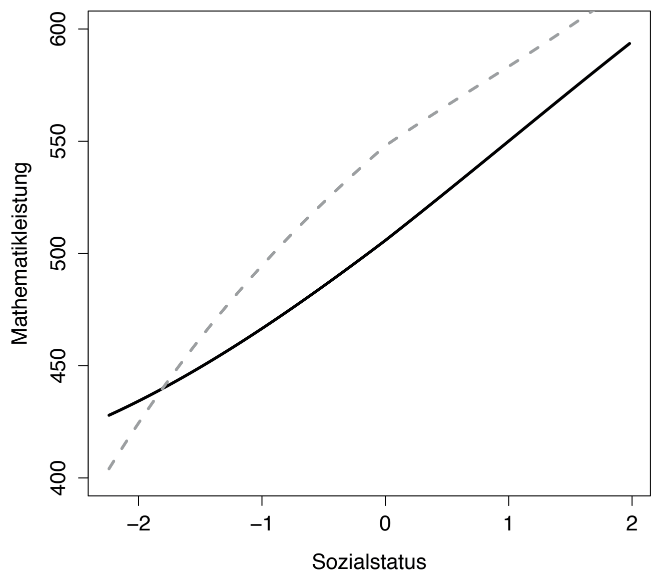

Abbildung 1: *Durch das Regressionsmodell vorhergesagte Mathematikleistung. Die Distanz zwischen den Linien zeigt den Leistungsunterschied zwischen Schülerinnen und Schülern mit (durchgezogene Linie) und ohne Migrationshintergrund (gestrichelte Linie), in Abhängigkeit vom Sozialstatus.*

$\int[\mu_1(s) - \mu_0(s)]w(s)ds$, wobei die Gewichtungsfunktion w der Dichtefunktion des Sozialstatus in der gesamten Population entspricht. Der Unterschied in Mathematikleistungen zwischen der Gruppe der Schüler/innen mit und ohne Migrationshintergrund wird demzufolge im Allgemeinen von der Wahl der Gewichtung w abhängen. Der ATT lässt sich schreiben als $ATE = \int[\mu_1(s) - \mu_0(s)]w_1(s)ds$, wobei w_1 die Dichtefunktion des Sozialstatus in der Gruppe der Schüler/innen mit Migrationshintergrund bezeichnet.

Häufig verwendet man als Berechnungsmethode für den Unterschied zwischen Schüler/innen mit und ohne Migrationshintergrund unter Kontrolle des Sozialstatus (*bedingter Effekt*) eine lineare Regression $Y = E(Y|S = s, D = d) = \beta_0 + aD + \beta_1 s$. Der bedingte Effekt ist dann der Regressionskoeffizient a. Wenn die lineare Regression gilt, so folgt $\mu_1(s) = \beta_0 + a + \beta_1 s$ sowie $\mu_0(s) = \beta_0 + \beta_1 s$. In diesem Fall ist der Unterschied zwischen Schülerinnen und Schülern mit Migrationshintergrund und Schülerinnen und Schülern ohne Migrationshintergrund für jeden Sozialstatus konstant und es gilt $ATE = ATT = a$. Die Umkehrung gilt jedoch im Allgemeinen nicht: Wenn die Größe des Unterschieds vom Sozialstatus abhängig ist, so wird man im Allgemeinen mit der linearen Regression nicht den durchschnittlichen Effekt (ATE), sondern einen eher artifiziellen Effekt schätzen, welcher eine Gewichtungsfunktion $\tilde{w}$ datengetrieben wählt, $\int[\mu_1(s) - \mu_0(s)]\tilde{w}(s)ds$, und *nicht* die Sozialstatus-Verteilung in der Population widerspiegelt (siehe Morgan and Winship [2007], Ch. 5).

4 Die Skalen *Selbstkonzept* und *Lernfreude*

In Tabelle 3 sind die Variablen aus dem SFB dargestellt, die für die Skalen *Lernfreude* (LF) und *Selbstkonzept* (SK) verwendet wurden. Die Beantwortungskategorien waren jeweils vierstufig: 1 = *stimme völlig zu*, 2 = *stimme eher zu*, 3 = *stimme eher nicht zu*, 4 = *stimme überhaupt nicht zu*. Zur Berechnung eines Skalenwerts war es notwendig, die Polung von bestimmten Fragen umzukehren (s. Spalte *Pol.* in Tabelle 3). Ein hoher Skalenwert kann demnach mit einer hohen Ausprägung auf der jeweiligen Skala interpretiert werden. Die Skalenwerte ergeben sich aus den Mittelwerten der jeweils vier Indikatoren, diese werden wiederum in vier Kategorien eingestuft und als prozentuelle Anteile rückgemeldet.

Tabelle 3: *Variablen für die Skalen Selbstkonzept (SK) und Lernfreude (LF).*

Item	Itemformulierung	Skala	Pol.	MW	SD	λ
LF01	Ich hätte in der Schule gern mehr Mathematik.	LF	–	2.46	1.08	0.75
LF02	Ich lerne gern Mathematik.	LF	–	2.83	1.03	0.82
LF03	Mathematik ist langweilig.	LF	+	3.16	1.01	0.69
LF04	Ich mag Mathematik.	LF	–	3.11	1.02	0.89
SK01	Normalerweise bin ich gut in Mathematik.	SK	–	3.35	0.78	0.80
SK02	Mathematik fällt mir schwerer als vielen meiner Mitschülerinnen und Mitschüler.	SK	+	3.23	0.97	0.67
SK03	Ich bin einfach nicht gut in Mathematik.	SK	+	3.29	0.98	0.65
SK04	Ich lerne schnell in Mathematik.	SK	–	3.11	0.90	0.73

Anmerkungen: SK = Selbstkonzept, LF = Lernfreude, Pol. = Polung der Itemformulierung, MW = Mittelwert, SD = Standardabweichung, λ = Faktorladung auf die jeweilige Skala.

Zur Bestimmung der internen Konsistenz von den Skalenwerten wurde Cronbachs α (Cortina, 1993; Cronbach, 1951) und, basierend auf einer einfaktoriellen konfirmatorischen Faktorenanalyse, ω bestimmt(McDonald, 1999; Zinbarg, Yovel, Revelle & McDonald, 2006)[7]. Beide Skalen weisen eine zufriedenstellend hohe Reliabilität auf (LF: $\alpha = .864$, $\omega = .867$; SK: $\alpha = .803$, $\omega = .804$). Durch eine einfaktorielle Faktorenanalyse mit Hauptachsenmethode wurden nachfolgend die Eigenvektoren ermittelt. Das Verhältnis der Eigenwerte vom ersten Faktor zum zweiten Faktor beträgt bei der Skala Lernfreude 51.58, bei der Skala Selbstkonzept 17.13. Zusammenfassend lassen sich die Ergebnisse im positiven Sinne für die Skalenkonstruktion interpretieren.

5 Der Index der sozialen Benachteiligung

Der *Index der sozialen Benachteiligung* (B) wurde in Anlehnung an Bruneforth, Weber und Bacher (2012) berechnet und soll die soziale Zusammensetzung, insbesondere belastende Kompositionseffekte einer Schule reflektieren. Dieser Sozialindex setzt sich aus vier Kontextvariablen zusammen. Für jede/n Schüler/in werden Dummyvariablen (d) generiert, die mit einer dichotomen Kodierung (0/1) angeben, ob für die jeweilige Person die nachfolgende Definition (pro Variable) zutreffend ist:

- d_{1ij} = Schüler/in hat Migrationshintergrund

[7] Nachfolgende Analysen basieren auf gemittelten Statistiken, gerechnet für jede der 10 PV-Imputationen.

- d_{2ij} = Schüler/in hat andere Muttersprache als Deutsch
- d_{3ij} = Eltern von Schüler/in haben maximal Pflichtschulabschluss
- d_{4ij} = Schüler/in aus unterem Quintil des HISEI

Der Index berechnet sich auf Individualebene für Schüler/in i in Schule j durch

$$B_{ij} = 100 + 100 \cdot \left(\frac{1}{4}(d_{1ij} + d_{2ij} + d_{3ij} + d_{4ij})\right) \tag{10}$$

Aus den vier erhaltenen dummykodierten Variablen wird pro Schüler/in der Index mit 100 multipliziert und eine Konstante von 100 addiert. Pro Person können sich somit Werte von 100, 125, 150, 175, 200 ergeben, je nachdem ob 0, 1, 2, 3 oder 4 der Merkmale zutreffen. Nachfolgend werden die Schülerwerte auf Schulebene durch die eine Aggregationsvorschrift zusammengefasst:

$$\bar{B}_{\bullet j} = \frac{1}{n_j}\sum_{i=1}^{n_j} B_{ij} \tag{11}$$

Zusätzlich wird auf Schulebene eine Kategorisierungsvariable eingeführt, welche die Werte von $\bar{B}_{\bullet j}$ in vier Kategorien einteilt:

1. 100 – 115: geringe Belastung
2. 116 – 125: mittlere Belastung
3. 126 – 135: hohe Belastung
4. $>$ 135: sehr hohe Belastung

Die durch dieses Vorgehen erhaltenen Werte können als Rohwerte oder beobachtete Werte des Index der sozialen Benachteiligung definiert werden. Tabelle 4 zeigt die Korrelationen zwischen den Dummyvariablen, Belastungsindex und dem Leistungsschätzer der Mathematikkompetenz (WLE), auf Schüler- ($n = 73655$) und Schulebene ($n = 3048$) gemeinsam mit den Mittelwerten der jeweiligen Variablen[8]. Zur Berechnung der Korrelationen auf Schulebene wurden die individuellen Werte der Schüler/innen aggregiert (gemittelt). Man kann in den Ergebnissen beobachten, dass die Korrelationen auf Schulebene tendenziell höher ausfallen als auf Individualebene. Die Variablen zur Bestimmung von B korrelieren durchgehend negativ mit der Mathematikleistung, eine höhere Belastung geht somit mit einer geringeren Leistung einher.

Für die Rückmeldung des Index der sozialen Benachteiligung wurden auf Bundes- und Landesebene PVs (Plausible Values; Mislevy, 1991; Mislevy et al., 1992), auf Ebene der Schulaufsicht EAPs (expected a posteriori; Raudenbush & Bryk, 2002) berechnet. Der Index B entspricht dem Mittelwert über einen Score individueller Belastung, der sich aus den vier Dummyvariablen ergibt. Der Schulmittelwert für die Mathematikleistung wird in den Schulberichten als Mittelwert der individuellen WLEs berechnet, die Rückmeldung auf Bundes- und Landesebene basiert auf den PVs. Ein vergleichbares Vorgehen wurde für die Rückmeldung des Index der sozialen Benachteiligung gewählt und soll nachfolgend beschrieben werden.

[8]Durch die Dichotomisierung können die Mittelwerte der Dummyvariablen als Anteile interpretiert werden.

Tabelle 4: *Belastungsindex: Korrelationsmatrix.*

	d_{1ij}	d_{2ij}	d_{3ij}	d_{4ij}	B	WLE
d_{1ij}	**.190**	.692	.296	.252	.818	–.241
d_{2ij}	.922	**.174**	.261	.209	.788	–.240
d_{3ij}	.559	.548	**.078**	.226	.562	–.212
d_{4ij}	.225	.201	.307	**.199**	.625	–.197
B	.909	.895	.715	.550	**116**	–.315
WLE	–.387	–.362	–.369	–.262	–.441	**533**

Anmerkungen: Werte auf der Hauptdiagonale sind Mittelwerte der einzelnen Variablen, Werte oberhalb der Hauptdiagonale Korrelationen basierend auf individuellen Schülerwerten. Werte unterhalb der Hauptdiagonale sind Korrelationen basierend auf aggregierten Schulwerten. WLE (Warm's weighted likelihood estimate) = Personenschätzer der Mathematikleistung in M4.

Ein spezieller Umstand, der vor allem auf der Primarstufe zum Tragen kommt und bei Analysen berücksichtigt werden sollte, sind die stark heterogenen Schulgrößen. Auf der vierten Schulstufe gibt es zahlreiche Kleinschulen (Schulen $<$ 10 Schüler/innen; Tabelle 5), die sich zusätzlich in den Bundesländern unterschiedlich verteilen. Eine mögliche Einschränkung der Reliabilität von Indizes basierend auf Schulmittelwerten kann diesem Umstand geschuldet sein. Lüdtke et al. (2008) zeigten z. B., dass die Schätzung von Kontexteffekten basierend auf aggregierten Variablen auf Gruppenebene (e. g. Sozialindex) verzerrt ist, in Abhängigkeit von der Anzahl an Level-1-Einheiten (e. g. Schüler/innen in Schulen). Diesen Effekt gilt es auszugleichen. Dazu wird ein Nullmodell spezifiziert nach

$$\begin{aligned} B_{ij} &= \mu_0 + u_j + \epsilon_{ij} = \mu_j + \epsilon_{ij} \\ Var(u_j) &= \tau^2, Var(\epsilon_{ij}) = \sigma^2, \end{aligned} \tag{12}$$

hier wird der Wert B einer Person i in Schule j vorhergesagt durch den Schulmittelwert μ_j und die Abweichung ϵ_{ij} des Schülerwerts vom Schulmittelwert (vgl. auch Glas, Scheerens & Thomas, 2003; Goldstein, 1997; Lüdtke et al., 2008; Lüdtke, Marsh, Robitzsch & Trautwein, 2011; Rumberger & Palardy, 2004). Die beiden Varianzen τ^2 und σ^2 entsprechen der Varianz zwischen den bzw. innerhalb der Schulen und werden nachfolgend für die Bestimmung des ICC (Intraclass Correlation Coefficient) verwendet.

Der ICC gibt den Anteil der Gesamtvarianz an, welcher durch Unterschiede auf Gruppenebene (hier Schulen) erklärt wird (Raudenbush & Bryk, 2002). Der ICC berechnet sich nach

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2} \tag{13}$$

Betrachtet man die Werte des ICC in Tabelle 5 pro Bundesland, so wird ersichtlich, dass diese sehr heterogen sind. Im Burgenland hätte man z. B. einen ICC von .055, was bedeuten würde, dass 5.5 % der Unterschiede im Index der sozialen Benachteiligung auf Unterschiede zwischen Schulen zurückführbar wären, in Wien wären es 28.4 %. Zusätzlich kann der ICC(2) berechnet werden, hierzu wird σ^2 in Gleichung 13 durch die (angenommene) Anzahl an Schülerinnen und Schülern pro Schule dividiert (σ^2/n). Man erhält damit

Tabelle 5: *Schulgrößen pro Bundesland.*

BL	Schulen	Schulen < 10	τ^2	σ^2	ICC	ICC(2,20)	ICC(2,60)
Bgld	178	80 (45 %)	22.4	388.5	.055	.536	.776
K	239	50 (21 %)	24.6	316.3	.072	.609	.823
NÖ	609	100 (16 %)	52.3	454.8	.103	.697	.873
OÖ	559	93 (17 %)	105.8	494.1	.176	.811	.928
Sbg	184	40 (22 %)	78.4	492.7	.137	.761	.905
Stmk	479	126 (26 %)	87.9	378.6	.188	.823	.933
T	377	156 (41 %)	47.0	532.5	.081	.638	.841
V	163	61 (37 %)	53.0	642.3	.076	.623	.832
W	260	2 (1 %)	261.5	660.3	.284	.888	.960

Anmerkungen: Schulen < 10 = Anzahl an Schulen mit weniger als 10 Schülerinnen und Schülern, τ^2 = Varianz zwischen Schulen, σ^2 = Varianz innerhalb von Schulen, ICC = Intraclass Correlation Coefficient, ICC(2,20) = ICC(2) bei einer Schulgröße von 20 Schülerinnen und Schülern, ICC(2,60) = ICC(2) bei einer Schulgröße von 60 Schülerinnen und Schülern.

eine Aussage darüber, inwieweit sich die Reliabilität unter Annahme einer bestimmten Anzahl von Level-1-Einheiten (Schüler/innen) verändern würde. Die Werte in Tabelle 5 zeigen einen Anstieg der Reliabilität bei einer möglichen Gruppengröße von 20 oder 60 Schülerinnen und Schülern pro Schule.

Tabelle 6: *Vergleich der Schätzer für den Index der sozialen Benachteiligung.*

	$\bar{B}_{\bullet j}$		$\bar{B}_{j,PV}$		$\bar{B}_{j,EAP}$	
BL	MW	SD	MW	SD	MW	SD
Bgld	110.8	20.3	110.9	4.4	110.9	3.8
K	109.7	18.6	109.8	5.1	109.8	4.6
NÖ	112.3	22.7	112.2	7.1	112.3	6.7
OÖ	114.8	25.0	114.8	10.9	114.8	10.8
Sbg	113.9	24.2	114.0	9.3	113.9	8.9
Stmk	112.3	22.1	112.3	9.9	112.3	9.8
T	114.3	24.4	114.3	7.3	114.3	6.7
V	116.8	26.5	116.9	7.4	116.8	6.6
W	128.2	30.3	128.2	15.8	128.2	15.7

Anmerkungen: BL = Bundesland, $\bar{B}_{\bullet j}$ = beobachteter Wert, $\bar{B}_{j,PV}$ = Plausible Values, $\bar{B}_{j,EAP}$ = EAP, MW = Mittelwert, SD = Standardabweichung.

Die Rückmeldung des Index der sozialen Benachteiligung passiert auf Systemebene in Form von Anteilen von Schülerinnen und Schülern in Schulen, verteilt auf die vier Kategorien. Es wird ersichtlich, wie viel Prozent der Schüler/innen Schulen mit geringer, mittlerer, hoher und sehr hoher Belastung besuchen. Auf Ebene der Schulaufsicht wird die Index-Kategorie pro Schule (mit mind. 10 Schülerinnen und Schülern) rückgemeldet. Hinsichtlich der hohen Anzahl an Kleinschulen (cf. Tabelle 5) ist mit Einschränkungen der Reliabilität zu rechnen, da in manchen Schulen der Belastungsindex aufgrund weniger Schüler/innen erfasst wird. Die Reliabilität der beobachteten Mittelwerte bestimmt sich nach

$$Rel(\bar{B}_{\bullet j}) = \lambda_j = ICC(2) = \frac{\tau^2}{\tau^2 + \sigma^2/n_j} \tag{14}$$

Der EAP-Schätzer für die Rückmeldung auf Ebene der Schulaufsicht wird definiert durch

$$B_{j,EAP} = \lambda_j \cdot \bar{B}_{\bullet j} + (1 - \lambda_j) \cdot \hat{\mu}_0 \tag{15}$$

Es wird ersichtlich, dass Werte mit geringer Reliabilität stärker in Richtung des Mittelwerts gezogen werden. Die Schätzer, die für die Rückmeldung auf Bundes- und Landesebene verwendet wurden (PVs), werden aus der Verteilung $PV \sim N(B_{j,EAP}, \sigma^2_{j,PV})$ gezogen mit $\sigma^2_{j,PV} = \frac{1}{\frac{1}{\tau^2} + \frac{n_j}{\sigma^2}}$.

Die unterschiedlichen Schätzer werden in Tabelle 6 im Vergleich dargestellt. Entsprechend den Eigenschaften der verwendeten Schätzer (e. g., Mislevy, 1991; Rost, 2004; Von Davier, Gonzalez & Mislevy, 2009) ist hervorzuheben, dass sich diese vorwiegend hinsichtlich der geschätzten Standardabweichung unterscheiden, die Mittelwerte hingegen relativ ähnlich zueinander sind. Für die Systemebene, mit Inferenz auf die Gesamtpopulation, würde die PV-Schätzung die beste Approximation der Verteilung liefern, auf Ebene der Schulaufsicht sind die EAPs als möglichst genaue Punktschätzer eine geeignete Variante für die Rückmeldung der Daten.

Literatur

Bruneforth, M., Weber, C. & Bacher, J. (2012). Chancengleichheit und garantiertes Bildungsminimum in Österreich. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012. Band 2. Fokussierte Analysen bildungspolitischer Schwerpunktthemen* (S. 187–226). Graz: Leykam.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334.

Gangl, M. (2010). Causal inference in sociological research. *Annual Review of Sociology, 36*(1), 21–47.

Ganzeboom, H., De Graaf, P. & Treiman, D. (1992). A standard international socioeconomic index of occupational status. *Social Science Research, 21*(1), 1–56.

Glas, C., Scheerens, J. & Thomas, S. (2003). *Educational evaluation, assessment and monitoring: A systematic approach.* Taylor & Francis.

Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement, 8*(4), 369–395.

Imbens, G. W. & Wooldridge, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature, 47*(1), 5–86.

Kenny, D. A. (1979). *Correlation and causality.* John Wiley & Sons Inc.

Lüdtke, O., Marsh, H. W., Robitzsch, A. & Trautwein, U. (2011). A 2 × 2 taxonomy of multilevel latent contextual models: Accuracy bias trade offs in full and partial error correction models. *Psychological Methods, 16*(4), 444–467.

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.

Lüdtke, O., Robitzsch, A., Köller, O. & Winkelmann, H. (2010). Kausale Effekte in der Empirischen Bildungsforschung. Ein Vergleich verschiedener Ansätze zur Schätzung des Effekts des Einschulungsalters. In W. Bos, E. Klieme & O. Köller (Hrsg.), *Schulische Lernangelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (S. 257–284). Münster: Waxmann.

McDonald, R. P. (1999). *Test theory.* Mahwah NJ: Erlbaum.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, *56*, 177–196.

Mislevy, R. J., Beaton, A. E., Kaplan, B. & Sheehan, K. M. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement*, *29*, 133–161.

Morgan, S. L. & Winship, C. (2007). *Counterfactuals and causal inference.* Cambridge: Cambridge University Press.

OECD. (2005). *PISA 2003: Technical report.* Paris: OECD Publishing.

OECD. (2009). *PISA 2006: Technical report.* Paris: OECD Publishing.

OECD. (2012). *PISA 2009: Technical report.* Paris: OECD Publishing.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical linear models.* Thousand Oaks: Sage.

Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion.* Bern: Huber.

Rumberger, R. W. & Palardy, G. J. (2004). Multilevel models for school effectiveness research. In D. Kaplan (Hrsg.), *Handbook of quantitative methodology for the social sciences* (S. 235–258). Thousand Oaks, CA: Sage Publications.

Schafer, J. L. & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*(4), 279–313.

Von Davier, M., Gonzalez, E. & Mislevy, R. (2009). What are plausible values and why are they useful. *IERI monograph series*, *2*, 9–36.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*, 427–450.

Winship, C. & Morgan, S. L. (1999). The estimation of causal effects from observational data. *Annual Review of Sociology*, *25*(1), 659–706.

Zinbarg, R. E., Yovel, I., Revelle, W. & McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: A comparison of estimators for ω_h. *Applied Psychological Measurement*, *30*(2), 121–144.