

bifie | Bildung

 standards

Hintergrundvariablen und spezielle Analysen

Technische Dokumentation – BIST-Ü Englisch,
8. Schulstufe, 2013

*Giang Pham
Roman Freunberger
Alexander Robitzsch*

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 Englisch, 8. Schulstufe, 2013

BIFIE | Department Bildungsstandards & Internationale Assessments (BISTA),
Salzburg 2014

Der Text sowie die Aufgabenbeispiele dürfen für Zwecke des Unterrichts in österreichischen Schulen sowie von den Pädagogischen Hochschulen und Universitäten im Bereich der Lehreraus-, Lehrerfort- und Lehrerweiterbildung in dem für die jeweilige Lehrveranstaltung erforderlichen Umfang von der Homepage (www.bifie.at) heruntergeladen, kopiert und verbreitet werden. Ebenso ist die Vervielfältigung der Texte und Aufgabenbeispiele auf einem anderen Träger als Papier (z. B. im Rahmen von Power-Point-Präsentationen) für Zwecke des Unterrichts gestattet.

Inhaltsverzeichnis

3 1 Hintergrundvariablen

4 2 Variablenbildung

4 2.1 Sozialstatus

5 2.2 Selbstkonzept und Lernfreude

7 2.3 Index der sozialen Benachteiligung

10 3 Leistungsunterschiede nach Migrationsstatus unter Kontrolle des Sozialstatus

15 Literaturverzeichnis



1 Hintergrundvariablen

In der Bildungsstandardüberprüfung von 2013 für Englisch auf der 8. Schulstufe (BIST-Ü E8 2013) kam neben dem Schulleiterfragebogen zur Erhebung der schulbezogenen Kontextdaten auch der Schülerfragebogen (SFB) zum Einsatz, um die schülerbezogenen Merkmale zu erheben. Die Variablen, welche der Analyse von gruppenspezifischen Effekten sowie sozialen Vergleichen (vgl. Kapitel Fairer Vergleich, Pham & Robitzsch, in press) und in weiterer Folge der ErgebnISRückmeldung¹ dienten, wurden aus einzelnen oder mehreren Items aus den Kontextfragebögen abgeleitet und werden im Nachfolgenden als Hintergrundvariablen bezeichnet.

Bei der Rückmeldung der BIST-Ü wurde zwischen zwei Ebenen differenziert. Die erste Ebene der ErgebnISRückmeldung ist die Praxisebene. Diese Ebene umfasste die Individualrückmeldungen (an die Schüler/innen), die Gruppenrückmeldungen (an die Lehrpersonen) und die Schulrückmeldung (an die Schulleiter/innen). Eine Zusammenfassung der Schulergebnisse erfolgte des Weiteren an die Schulaufsicht. Auf der zweiten Ebene fanden sich die Landesergebnisberichte und der Bundesergebnisbericht, diese Ebene wird als Systemebene bezeichnet. Die konzeptionelle Hierarchie der ErgebnISRückmeldung – definiert durch die jeweilige Rezipientengruppe – bedingt weitestgehend auch die Analysestrategien sowohl für die Leistungsschätzer² als auch für abgeleitete Variablen (s. Abschnitt 2.3).

Die rückgemeldeten Hintergrundvariablen sind in Tabelle 1 aufgelistet. Hintergrundvariablen wurden entweder durch einfache Aggregationsschritte oder durch komplexere Analysen gebildet, welche mit einem Stern (*) in Tabelle 1 gekennzeichnet sind, und in den nachfolgenden Kapiteln erläutert werden. Der Sozialstatus wird durch eine lineare Transformation aus spezifischen Variablen gebildet (s. Abschnitt 2.1). Für die Skalen *Lernfreude* und *Selbstkonzept* werden jeweils Skalenwerte gebildet (s. Abschnitt 2.2). Komplexere Analyseschritte sind für die Bildung des Index der sozialen Benachteiligung (*B*, s. Abschnitt 2.3) sowie für die Berechnung des Leistungsunterschieds zwischen Schülerinnen und Schülern mit und ohne Migrationshintergrund unter Kontrolle des Sozialstatus (s. Abschnitt 3) notwendig gewesen.

¹Es sei an dieser Stelle auch auf den Link zu den ErgebnISRückmeldungen hingewiesen, diese sind online im Downloadbereich unter <https://www.bifie.at/node/64/> abrufbar.

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

Tabelle 1: Die rückgemeldeten Hintergrundvariablen in der BIST-Ü E8 2013

Variable	Label	Werte
female	Geschlecht (weiblich)	0=männlich; 1=weiblich
migrant	Migrationsstatus	0=einheimisch; 1=mit Migrationshintergrund ³
ndtspr	Erstsprache (nicht deutsch)	0=deutsch; 1=eine andere Sprache
eltausb	höchste elterliche Ausbildung (4-stufig)	1=max. Pflichtschule; 2=Berufsschule, Meisterausbildung, Gesundheitspflege; 3=Matura; 4=Universität, FH, pädag. Akad.
HISEI	Highest International Socio-Economic Index of occupational status	Min.=11.56; Max.=88.96
buch	Anzahl Bücher zu Hause	1=0–10; 2=11–25; 3=26–100; 4=101–200; 5= mehr als 200
zsozstatus*	Sozialstatus	siehe Kapitel 2.1
bela_cat*	Index der sozialen Benachteiligung auf Schulebene	siehe Kapitel 2.3
SK_cat*	Skalenwert Selbstkonzept	siehe Kapitel 2.2
LF_cat*	Skalenwert Lernfreude	siehe Kapitel 2.2
wohl1	Schulfreude ⁴	Likertskala 1–5: 1=sehr gern; 5=sehr ungern
wohl2	Zufriedenheit mit der Klasse ⁵	Likertskala 1–5: 1=sehr zufrieden; 5=sehr unzufrieden
testmot	Bemühen im Vergleich zu einer Schularbeit ⁶	1=viel weniger; 2=weniger; 3=genauso; 4=mehr; 5=viel mehr

2 Variablenbildung

2.1 Sozialstatus

Die Bestimmungsgleichung für den Sozialstatus setzt sich aus den z-standardisierten Werten (mit z gekennzeichnet) der drei Variablen $HISEI$ (z_{HISEI}), $buch$ (z_{buch}) und $eltausb$ ($z_{eltausb}$) zusammen. Der HISEI (*Highest International Socio-Economic Index of occupational status*, Ganzeboom, De Graaf & Treiman, 1992) entspricht dem höchsten ISEI der Eltern, bezogen auf den Index der Mutter ($MISEI$) oder des Vaters ($FISEI$). Der ISEI bezieht den Beruf, die Ausbildung und das Gehalt mit ein und basiert auf den Berufsklassifikationen nach der ISCO-08 (*2008 International Standard Classification of Occupations*, Ganzeboom & Treiman, 2010).

Zunächst wurde der Mittelwert ($sozstatus$) der drei oben genannten z-Variablen berechnet. Als Datenbasis für die Berechnung der Variable $sozstatus$ wurden die imputierten Datensätze verwendet. Eingesetzt wurden 50 für den fairen Vergleich imputierte Datensätze

³Für nähere Erläuterungen lesen Sie bitte die genaue Definition im Glossar zur Rückmeldung (BIFIE, 2014).

⁴Die zugehörige Frage im Fragebogen lautete: „Wie gern gehst du in die Schule?“

⁵Die zugehörige Frage im Fragebogen lautete: „Wie zufrieden bist du mit deiner Klasse (deinen Mitschülerinnen und Mitschülern) insgesamt?“

⁶Die zugehörige Frage im Fragebogen lautete: „Wie sehr hast du dich im Vergleich zu einer Schularbeit bei diesem Test bemüht?“

(siehe Pham & Robitzsch, in press) für die Schul- und Gruppenrückmeldungen (Praxisebene), weil diese Variable ein zentraler Prädiktor im Fairen-Vergleichs-Modell ist, und die für diesen Zweck imputierten Werte die optimalen Schätzer für jede einzelne Schule bzw. Unterrichtsgruppe sind. Für die Systemberichte (Systemebene) dienten die Werte aus den 10 gezogenen Plausible Values (PVs, Mislevy, 1991; Mislevy et al., 1992) als Grundlage, da diese die besten Schätzungen für die Populationsverteilungen liefern. Pro imputiertem Datensatz wurden die drei Stammvariablen z-standardisiert und die Variable *sozstatus* gebildet, wie in Gleichung (1) dargestellt:

$$sozstatus = \frac{1}{3}(z_{HISEI} + z_{buch} + z_{eltausb}) \quad (1)$$

Daraus resultierten 50 *sozstatus*-Werte für die Analysen auf Praxisebene und 10 Werte für die Analysen auf Systemebene. Danach wurden der gepoolte Mittelwert ($\bar{\mu}_p$) und die Standardabweichung (δ_p) aller 50 bzw. 10 vorhandenen Werte der Variable *sozstatus* berechnet. Auf Basis dieser Werte konnte eine z-Transformation der Variable *sozstatus* in allen imputierten Datensätze vorgenommen werden, wie in Gleichung (2) dargestellt:

$$z_{sozstatus} = \frac{sozstatus - \bar{\mu}_p}{\delta_p} \quad (2)$$

Letztendlich wurden die 50 bzw. 10 Werte von $z_{sozstatus}$ gemittelt. Das Ergebnis entsprach dem zurückgemeldeten Sozialstatus ($z_{sozstatus}$) in den Rückmeldungen:

$$z_{sozstatus} = \frac{1}{I} \sum_1^I z_{sozstatus}, \quad \text{mit } I = \begin{cases} 50, & \text{Analyse für Praxisebene} \\ 10, & \text{Analyse für Systemebene} \end{cases} \quad (3)$$

Durch die z-Standardisierung können die $z_{sozstatus}$ -Werte als normierte Abweichungen vom Mittelwert der Population interpretiert werden. Die Verwendung unterschiedlicher Datensätze für die Berechnungen auf Praxis- und Systemebene galt auch für alle anderen Analysen.

2.2 Selbstkonzept und Lernfreude

Die Skalen akademisches Selbstkonzept und fachspezifische Lernfreude stammen aus dem Schülerfragebogen, der in TIMSS 2007 (Trends in International Mathematics and Sciences Study, Martin & Preuschoff, 2009) eingesetzt wurde. Die deutsche Übersetzung der Items wurde aus dem österreichischen Schülerfragebogen von TIMSS 2007 übernommen (IEA, 2007) und für das Fach Englisch angepasst. Die Beantwortungskategorien sind 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 der positiv formulierten Items so umzukehren, dass ein hoher Skalenwert mit einer hohen Ausprägung auf der jeweiligen Skala korrekt interpretiert werden kann. Die Formulierungen, Polungen und die deskriptive Statistik der Items sind in Tabelle 2 dargestellt.

Wie in Tabelle 2 sichtbar wird, gab es bei jedem Item einen kleinen Anteil (< 5%) an fehlenden Angaben. Vor der Bildung der Skalenwerte für die Rückmeldungen auf der Praxisebene wurde daher eine einfache Imputation durchgeführt, um fehlende Werte von Schülerinnen und Schülern mit mindestens einer Angabe bei der jeweiligen Skala durch Schätzwerte zu ersetzen. Zur Schätzung eines fehlenden Werts \hat{x}_{pi} von Person p und Item i wurden der

Tabelle 2: Itembeschreibung der Skalen Lernfreude und Selbstkonzept

Item	Itemformulierung	Skala	Pol.	MW	SD	N_miss	p_miss	r_{it}	λ
LF01	Ich hätte in der Schule gern mehr Englisch.	LF	+	2.2	0.9	1861	2.4 %	0.84	0.72
LF02	Ich lerne gern Englisch.	LF	+	2.5	0.9	2005	2.6 %	0.82	0.78
LF03	Englisch ist langweilig.	LF	-	3.0	1.0	2380	3.1 %	0.84	0.74
LF04	Ich mag Englisch.	LF	+	2.9	0.9	1756	2.3 %	0.79	0.88
SK01	Normalerweise bin ich gut in Englisch.	SK	+	3.1	0.8	1489	1.9 %	0.79	0.81
SK02	Englisch fällt mir schwerer als vielen meiner Mitschülerinnen und Mitschüler.	SK	-	3.1	0.9	2145	2.8 %	0.82	0.71
SK03	Ich bin einfach nicht gut in Englisch.	SK	-	3.2	0.9	2096	2.7 %	0.80	0.78
SK04	Ich lerne schnell in Englisch.	SK	+	2.8	0.9	2176	2.8 %	0.81	0.75

Anmerkungen: SK = Selbstkonzept, LF = Lernfreude, Pol. = Polung der Itemformulierung, MW = Mittelwert, SD = Standardabweichung, N_miss = Anzahl fehlender Angaben, p_miss = Anteil fehlender Angaben, r_{it} = Trennschärfe des jeweiligen Items, λ = Faktorladung auf die jeweilige Skala.

Skalenmittelwert der Person μ_p , der Itemmittelwert μ_i über alle Personen hinweg und der Gesamtmittelwert μ über alle Personen und die vier Items der jeweiligen Skala herangezogen, wie in Gleichung (4) dargestellt:

$$\hat{x}_{pi} = \begin{cases} 1, & \text{wenn } \hat{x} < 1 \\ 4, & \text{wenn } \hat{x} > 4 \\ \hat{x} & \text{wenn } 1 \leq \hat{x} \leq 4 \end{cases} \quad (4)$$

Die Skalenwerte (Selbstkonzept: *SK*, Lernfreude: *LF*) ergeben sich aus den Mittelwerten der jeweils vier Indikatoren. Zur Bestimmung der internen Konsistenz der Skalen wurde Cronbachs α (Cortina, 1993; Cronbach, 1951) berechnet. Beide Skalen weisen eine zufriedenstellend hohe Reliabilität auf: $\alpha_{LF} = 0.86$, $\alpha_{SK} = 0.85$. Die Zugehörigkeit der Items zu den Skalen wird durch die hohe Trennschärfe r_{it} (siehe Tabelle 2) bestätigt, die durch die Korrelation zwischen dem Wert des jeweiligen Items mit dem Summenwert der drei übrigen Items berechnet wird.

Zusätzlich wurde die Dimensionalität und Reliabilität der Skalen mithilfe einer konfirmatorischen Faktorenanalyse (KFA) untersucht. Durch die einfaktorielle Faktorenanalyse mit Hauptachsenmethode wurden nachfolgend die Eigenvektoren ermittelt. Bei jeder Skala gab es nur einen Faktor mit einem Eigenwert größer als 1. Das Verhältnis der Eigenwerte vom ersten Faktor zum zweiten Faktor betrug bei der Skala Lernfreude 51.10, bei der Skala Selbstkonzept 30.38. Das einfaktorielle KFA-Modell für beide Skalen lieferte zufriedenstellende Anpassung. Die standardisierten Residualdiskrepanzwurzeln (*SRMR*) lagen unter dem Cut-off-Wert von 0.08 (vgl. Hu & Bentler, 1999): $SRMR_{LF} = 0.04$, $SRMR_{SK} = 0.07$. Alle Items wiesen eine Faktorladung über 0.70 auf (siehe Tabelle 2). Die Reliabilitäten der Faktorenwerte ω (McDonald, 1999; Zinbarg, Yovel, Revelle & McDonald, 2006) sind

$\omega_{LF} = 0.86$, $\omega_{SK} = 0.85$. Durch diese Ergebnisse wurden die Eindimensionalität der Skalen und die Reliabilität der Skalenwerte bestätigt.

Für die Rückmeldungen auf der Praxisebene wurden die Skalenwerte, die aus dem oben beschriebenen Vorgehen resultierten, in vier Kategorien eingestuft: niedrig (bis 1.5), eher niedrig (bis 2.5), eher hoch (bis 3.5) und hoch (> 3.5). Für die Ergebnisberichte auf Systemebene wurden die gepoolten Ergebnisse unter Verwendung der 10 PVs mithilfe einer analogen Vorgehensweise berechnet (vgl. Abschnitt 2.1).

2.3 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
- d_{2ij} = Schüler/in hat andere Muttersprache als Deutsch
- d_{3ij} = Eltern von Schüler/in haben maximal Pflichtschulabschluss
- d_{4ij} = Schüler/in stammt aus unterem Quintil des HISEI

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

$$B_{ij} = 100 + 100 \cdot \left(\frac{1}{4} (d_{1ij} + d_{2ij} + d_{3ij} + d_{4ij}) \right) \quad (5)$$

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

$$\bar{B}_{\bullet j} = \frac{1}{n_j} \sum_{i=1}^{n_j} B_{ij} \quad (6)$$

aggregiert und an die Schülerdaten angehängt. Zusätzlich wird auf Schulebene eine Kategorisierungsvariable eingeführt, die die gerundeten 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 beobachtete Werte des Index der sozialen Benachteiligung interpretiert werden. Tabelle 3 zeigt die Korrelationen zwischen den Dummyvariablen, Belastungsindex und dem Leistungsschätzer der Englischkompetenz⁷ auf Schüler- ($n = 76728$) und Schulebene ($n = 1410$) gemeinsam mit den Mittelwerten der jeweiligen Variablen. Zur Berechnung der Korrelationen auf Schulebene wurden die individuellen Werte aggregiert (gemittelt). Durch die Dichotomisierung können die Mittelwerte der Dummyvariablen als Prozentanteile interpretiert werden.

Tabelle 3: Index der sozialen Benachteiligung: Korrelationsmatrix, BIST-Ü E8 2013

	d_{1ij}	d_{2ij}	d_{3ij}	d_{4ij}	B	TWLE
d_{1ij}	.183	.771	.214	.242	.813	-.148
d_{2ij}	.967	.171	.217	.225	.806	-.176
d_{3ij}	.610	.631	.114	.205	.544	-.217
d_{4ij}	.378	.393	.620	.191	.616	-.229
B	.939	.947	.791	.633	116	-.274
TWLE	-.339	-.370	-.607	-.623	-.520	518

Anmerkungen: Werte auf der Hauptdiagonale: Mittelwerte der einzelnen Variablen, Werte oberhalb der Hauptdiagonale: Korrelationen basierend auf individuellen Schülerwerten. Werte unterhalb der Hauptdiagonale: Korrelationen basierend auf aggregierten Schulwerten. TWLE (Transformed Warm's Weighted Likelihood Estimate) = auf BIST-Metrik transformierter Personenschätzer der Schülerleistungen.

Für die Rückmeldung des Index der sozialen Benachteiligung auf allen Ebenen wurden nicht die beobachteten Werte, sondern auf Systemebene wie üblich die PVs (Plausible Values; Mislevy, 1991; Mislevy et al., 1992), und auf Ebene der Schulaufsicht die EAPs (Expected A Posteriori, Raudenbush & Bryk, 2002) berechnet. Die Verwendung von EAPs für die Rückmeldung an die Schulaufsicht lässt sich damit begründen, dass die Reliabilität von Indizes, die auf Schulmittelwerten basieren, stark von der Schülerzahl abhängt (vgl. Lüdtke et al., 2008). Diesen Effekt gilt es auszugleichen. Dazu wird ein Nullmodell spezifiziert nach

$$\begin{aligned}
 B_{ij} &= \gamma_{00} + \gamma_{01} \cdot z_j + U_j + R_{ij} = \mu_B(j) + U_j + R_{ij}, \\
 z_i &= \begin{cases} 0, & \text{falls Schule } j \text{ eine APS}^8 \text{ ist,} \\ 1, & \text{falls Schule } j \text{ eine AHS}^8 \text{ ist,} \end{cases} \\
 \mu_B(j) &= \gamma_{00} + \gamma_{01} \cdot z_j, \\
 \text{Var}(U_j) &= \tau^2, \text{Var}(R_{ij}) = \sigma^2
 \end{aligned} \tag{7}$$

Hier ist γ_{00} der Mittelwert des Index aller Schüler/innen der allgemeinbildenden Pflichtschulen (APS⁸); γ_{01} die Abweichung des Index-Mittelwerts aller Schüler/innen der allgemeinbildenden höheren Schulen (AHS) vom Mittelwert der APS-Schülerinnen und Schüler. Der Wert B_{ij} eines Schülers i in Schule j wird vorhergesagt durch den Gesamtmittelwert $\mu_B(j)$ der entsprechenden Schulform der Schule j , die Abweichung U_j des Schulmittelwerts vom $\mu_B(j)$, und die Abweichung R_{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

⁷Der Englisch-Gesamtwert ist der Mittelwert der Leistungsschätzer (WLE = Warm's Mean Weighted Likelihood Estimate; Warm, 1989) der drei Kompetenzbereiche Hören, Lesen und Schreiben.

⁸Schulen, die bereits als NMS geführt werden, zählen zur Gruppe der APS. Dies gilt demnach auch für AHS, die NMS-Klassen führen.

τ^2 und σ^2 entsprechen der Varianz zwischen bzw. innerhalb der Schulen und werden nachfolgend für die Bestimmung des ICC (Intraclass Correlation Coefficient) verwendet.

Tabelle 4: Reliabilität des Index und Anzahl getesteter Schüler/innen nach Bundesland

BL	Anzahl Schulen	Durchschnittliche Schülerzahl	τ^2	σ^2	ICC	ICC(2,10)	ICC(2,50)	γ_{01}
Bgld	49	50	24.0	441.9	.05	.35	.73	-6.4
K	85	60	42.0	417.1	.09	.50	.83	-7.8
NÖ	300	48	46.2	411.3	.10	.53	.85	-9.1
OÖ	277	50	144.2	482.5	.23	.75	.94	-10.2
Sbg	92	58	113.5	483.4	.19	.70	.92	-9.4
Stmk	208	51	116.6	397.8	.23	.75	.94	-8.2
T	129	53	50.9	500.8	.09	.50	.84	-9.1
V	67	58	65.6	593.1	.10	.53	.85	-11.8
W	203	69	219.8	664.0	.25	.77	.94	-23.3

Anmerkungen: τ^2 = Varianz zwischen Schulen, σ^2 = Varianz innerhalb Schulen, ICC = Intraclass Correlation Coefficient, ICC(2, 10) = ICC(2) bei einer festgelegten Schulgröße von 10 Schülerinnen und Schülern, ICC(2, 50) = ICC(2) bei einer festgelegten Schulgröße von 50 Schülerinnen und Schülern, γ_{01} = Mittelwertdifferenz zwischen AHS und APS.

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

$$ICC = \frac{\tau^2}{\tau^2 + \sigma^2} \quad (8)$$

Betrachtet man die Werte des ICC in Tabelle 4 pro Bundesland, so wird ersichtlich, dass diese sehr heterogen sind. Im Burgenland hätte man z. B. einen ICC von .05, was bedeuten würde, dass 5% der Unterschiede im Index der sozialen Benachteiligung auf Unterschiede zwischen Schulen zurückführbar wären, in Wien wären es 25%. Zusätzlich kann der ICC(2) (Raudenbush & Bryk, 2002) als Maß der Reliabilität berechnet werden, hierzu wird σ^2 in Gleichung (8) durch die Anzahl an Schülerinnen und Schülern pro Schule (z. B. Schule j mit n_j getesteten Schülerinnen und Schülern) dividiert:

$$Rel(\bar{B}_{\bullet j}) = \lambda_j = ICC(2) = \frac{\tau^2}{\tau^2 + \sigma^2/n_j} \quad (9)$$

Man erhält damit eine Aussage darüber, inwieweit sich die Reliabilität unter Annahme einer bestimmten Anzahl von Level-1-Einheiten (Schüler/innen) verändern würde. Die Werte in Tabelle 4 zeigen einen Anstieg der Reliabilität bei einer festgelegten möglichen Anzahl von 10 ($ICC(2, 10)$) oder 50 ($ICC(2, 50)$) Schülerinnen und Schülern.

Lüdtke et al. (2008) zeigten z. B., dass die Schätzung von Kontexteffekten basierend auf aggregierten Variablen auf Gruppenlevel (z. B. Sozialindex) je nach Anzahl an Level-1-Einheiten (z. B. Anzahl von Schülerinnen und Schülern) verzerrt sein kann. Wie in der Tabelle 4 ersichtlich ist, ist die Reliabilität des beobachteten Werts des Index in einer Schule mit 10 Schülerinnen und Schülern in vielen Bundesländern (mit kleinem ICC) kleiner als 0.7.

Auf Ebene der Schulaufsicht wurden die Indizes pro Schule rückgemeldet. Da an manchen Schulen der Belastungsindex nur anhand der Angaben von wenigen Schülerinnen und Schülern erfasst wurde, ist dort bei den beobachteten Schulwerten mit einer Einschränkung der Reliabilität zu rechnen. Aus diesem Grund wurde für die Rückmeldung an die Schulaufsicht der EAP von $\bar{B}_{\bullet j}$ ($B_{j,EAP}$, Raudenbush & Bryk, 2002) berechnet. Der $B_{j,EAP}$ jeder Schule wird definiert durch:

$$B_{j,EAP} = \lambda_j \cdot \bar{B}_{\bullet j} + (1 - \lambda_j) \cdot \mu_B(j) \quad (10)$$

Es zeigt sich, dass bei dieser Berechnung Schulwerte ($\bar{B}_{\bullet j}$) mit geringer Reliabilität (λ_j) stärker in Richtung des schulformspezifischen Mittelwerts des jeweiligen Bundeslands $\mu_B(j)$ tendieren. Für die Schulen mit weniger als 10 Schülerinnen und Schülern wurden die Werte durch *NA* (nicht vorhanden) ersetzt. Diese Vorgehensweise wurde analog zur Analyse in der BIST-Ü M4 2013 durchgeführt (siehe Freunberger, Robitzsch & Pham, in press). In der Primarstufe betraf dieses Ersetzen 708 Schulen, in der Sekundarstufe I betraf es nur 14 Schulen.

Auf Systemebene wurde der Index der sozialen Benachteiligung in Form von Schülerverteilungen rückgemeldet, d. h. wie verteilen sich die Schüler/innen der getesteten Schulstufe prozentuell auf die vier Kategorien. Dafür wurden die PVs berechnet, die sich ergeben durch

$$B_{j,PV} = \frac{\sigma^2/n_j + \tau^2}{\tau^2 \cdot \sigma^2/n_j} \quad \text{mit} \quad \sigma_{j,PV}^2 = \frac{1}{\frac{1}{\tau^2} + \frac{n_j}{\sigma^2}} \quad (11)$$

Die PVs des Index einer Schule j wurden aus einer Verteilung mit $B_{j,PV} \sim N(B_{j,EAP}, \sigma_{j,PV}^2)$ gezogen. Die verschiedenen Schätzer werden in Tabelle 5 verglichen.

Tabelle 5: Vergleich der Schätzer für den Index der sozialen Benachteiligung

	$\bar{B}_{\bullet j}$		$B_{j,EAP}$		$B_{j,PV}$	
	MW	SD	MW	SD	MW	SD
Bgld	113.3	6.4	113.2	4.7	112.9	4.9
K	113.4	7.8	113.5	6.5	112.9	6.4
NÖ	113.4	8.2	113.4	6.9	113.0	6.9
OÖ	116.2	13.1	116.2	12.0	115.7	11.8
Sbg	115.3	12.0	115.3	10.8	114.9	10.7
Stmk	115.0	11.6	115.0	10.8	114.2	10.3
T	114.7	8.6	114.9	7.1	114.2	7.2
V	116.9	9.8	116.9	8.5	115.8	8.4
W	131.9	19.1	131.9	18.3	130.8	17.7

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

3 Leistungsunterschiede nach Migrationsstatus unter Kontrolle des Sozialstatus

Für die BIST-Ü E8 2013 wurden zwei Variablen zum Migrationshintergrund definiert. Der Migrationshintergrund eines Kindes wird nach OECD-Kriterien (OECD, 2012) durch das

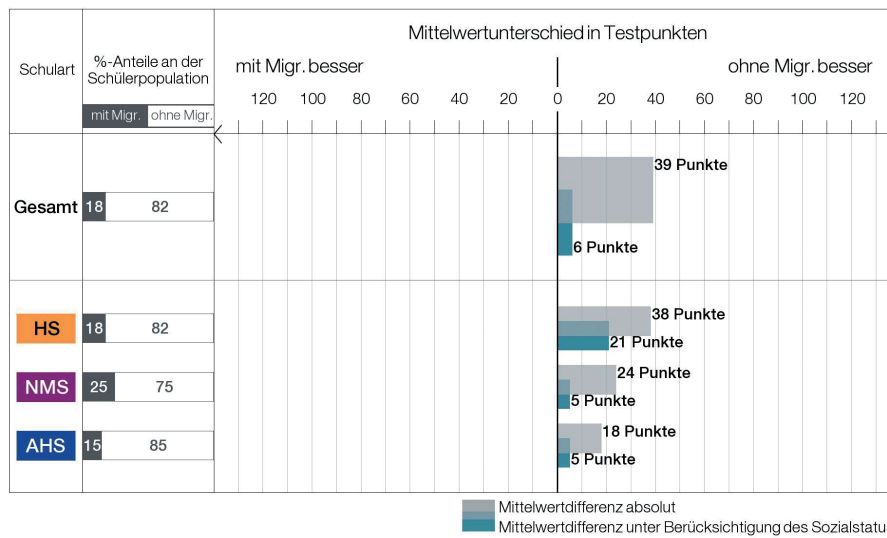


Abbildung 1: Unterschiede in der Englischkompetenz nach Migrationshintergrund unter Berücksichtigung des Sozialstatus (Bundesergebnisbericht)

Herkunftsland/Geburtsland der Eltern und des Kindes ermittelt. Wenn mindestens ein Elternteil im Inland geboren wurde, weist das Kind keinen Migrationshintergrund auf. Nach BIFIE-Definition wurde Deutschland, abweichend von der OECD-Definition, als Ausnahme behandelt. 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 (d. h. inkl. Kindern deutscher Herkunft) berichtet. Für alle leistungsspezifischen Vergleiche wurde ausschließlich der Migrationsstatus nach BIFIE-Definition herangezogen.

Für den Bundes- und Landesergebnisbericht wurde eine Grafik erstellt, in der die Gruppe der Schüler/innen mit und ohne Migrationshintergrund (*migrant*) hinsichtlich ihrer Englischleistung verglichen wurden. Zusätzlich wurde in einer gesonderten Grafik der Leistungsunterschied um den Sozialstatus (*zsozstatus*, vgl. Abschnitt 2.1) korrigiert dargestellt (siehe Abbildung 1). Mit diesem Vergleich wird aufgezeigt, wie groß die verbleibende Mittelwertdifferenz zwischen Kindern mit und ohne Migrationshintergrund wäre, wenn beide Schülergruppen im Durchschnitt den gleichen Sozialstatus hätten. Die zugrundeliegende Analyseverfahren wird im Folgenden beschrieben.

Für diesen Vergleich bestehen zumindest zwei gängige Methoden. Die erste Methode ist die multivariate Regressionsanalyse mit den beiden Variablen *migrant* und *zsozstatus* als Prädiktoren und der Schülerleistung als abhängiger Variable. Bei dieser Methode muss jedoch die Annahme getroffen werden, dass der Zusammenhang zwischen Sozialstatus und Schülerleistung in beiden Schülergruppen (mit und ohne Migrationshintergrund) äquivalent ist. Aufgrund unterschiedlicher Verteilung des Sozialstatus innerhalb dieser beiden Schülergruppen (s. Abbildung 2) erscheint jedoch die Plausibilität dieser Annahme sehr fraglich.

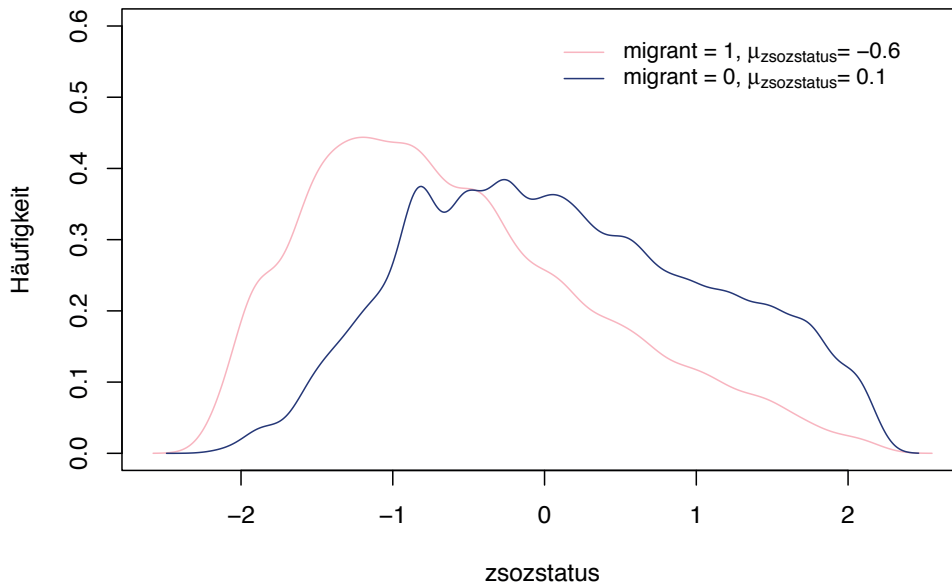


Abbildung 2: Verteilung des Sozialstatus nach Migrationshintergrund

Die alternative Methode beruht auf dem Potential-Outcome-Ansatz, der in der Forschung zur kausalen Inferenz häufig verwendet wurde (Gangl, 2010; Lütke, Robitzsch, Köller & Winkelmann, 2010; Imbens & Wooldridge, 2009; Morgan & Winship, 2007; Winship & Morgan, 1999). Mit dieser Methode wird die Annahme eines gleichen Effekts des Sozialstatus auf die Schülerleistung bei verschiedenen Schülergruppen nicht vorausgesetzt.

Bedient man sich der Terminologie des *Potential-Outcome*-Ansatzes, so kann angenommen werden, dass der Migrationsstatus einen Treatmentindikator bzw. die Treatmentvariable⁹ D darstellt. Schüler/innen können sich nun in zwei Zuständen befinden, sie weisen entweder einen Migrationshintergrund auf oder nicht. Die Problematik bei 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 individuelle Schülerleistung bezeichnet. Im Sinne des kontrafaktischen Modells würde man nun die Frage stellen, welchen Leistungsunterschied Schüler/innen mit Migrationshintergrund im Vergleich zu Schülerinnen und Schülern ohne Migrationshintergrund bei gleichem Sozialstatus aufweisen. Der Effekt zwischen Treatment und Kontrollgruppe wird als Average Treatment Effect (ATE) bezeichnet und ist definiert durch

$$ATE = E[\delta] = E[Y|S = s, D = 1] - E[Y|S = s, D = 0] \quad (12)$$

⁹Man könnte allerdings auch den Begriff *Bedingung* verwenden.

Dieser wäre unter Annahme einer randomisierten Zuweisung der Einheiten zum Treatment unverzerrt. Der ATE ist als erwarteter Wert von δ definiert und kann zum Beispiel durch eine einfache Mittelwertdifferenz geschätzt werden. Die beobachtete Mittelwertdifferenz zwischen den beiden Schülergruppen beträgt für Österreich 38 Punkte. Da der Migrationsstatus mit dem Sozialstatus konfundiert ist, muss davon ausgegangen werden, dass der beobachtete 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 jeweils identischem Sozialstatus vergleicht. Der kausale Effekt auf Individualebene wäre demnach definiert durch

$$\delta_i = y_{i1} - y_{i0} \quad (13)$$

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-Ü E8 2013 wurde in einem (nichtlinearen) Regressionsmodell unter Berücksichtigung der konfundierenden Variable **zsozstatus** die Englischleistung 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)¹⁰. In Anlehnung an das Vorgehen zur Berechnung von Treatmenteffekten durch eine ANCOVA mit nichtparallelen Regressionsgeraden wird ein Regressionsmodell mit einer konfundierenden Variable (**zsozstatus**, S) für beide Treatmentgruppen separat wie folgt berechnet: Für die Gruppe der Schüler/innen mit Migrationshintergrund durch die Gleichung

$$\begin{aligned} \mu_1(Y|S = s) &= \mu_1(s) = E(Y|S = s, D = 1) \\ &= \beta_{10} + \beta_{11}s + \beta_{12}s^2 + \beta_{13}s^3 + \beta_{14}s1[s > 0], \end{aligned} \quad (14)$$

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

$$\begin{aligned} \mu_0(Y|S = s) &= \mu_0(s) = E(Y|S = s, D = 0) \\ &= \beta_{00} + \beta_{01}s + \beta_{02}s^2 + \beta_{03}s^3 + \beta_{04}s1[s > 0] \end{aligned} \quad (15)$$

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

¹⁰In der Missing-Data-Terminologie würden die unbeobachteten potenziellen Outcomes somit als „missing at random“ (MAR) definiert werden können.

nach

$$\begin{aligned}
 ATE &= \frac{1}{n} \sum_i (y_{i1} - y_{i0}) \\
 &= \frac{1}{n} \left(\sum_{i, D_i=1} (y_i - \mu_0(s_i)) + \sum_{i, D_i=0} (\mu_1(s_i) - y_i) \right)
 \end{aligned} \tag{16}$$

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 Sozialstatus der *Schüler/in ohne Migrationshintergrund* wechseln würden ($D_i = 1$ zu $D_i = 0$). Der ATT berechnet sich nach

$$ATT = \frac{1}{n_1} \sum_{i, D_i=1} (y_i - \mu_0(s_i)) \tag{17}$$

Tabelle 6: 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.

Tabelle 6 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 $\mu_0(s_i) = 531$. Die mittlere Differenz $y_i - \mu_0(s_i)$ über die Gruppe $D_i = 1$ entspricht dem ATT, in diesem Beispiel: $ATT = \frac{1}{n_1} \sum_{i, D_i=1} (y_i - \mu_0(s_i)) = \frac{1}{5} \sum_{i, D_i=1} (y_i - 525) = \frac{1}{5}(-316) = -63$.

Abbildung 3 zeigt die Ergebnisse nach dem Potential-Outcome-Ansatz der BIST-Ü E8 2013. Der Leistungsunterschied, berechnet durch eine einfache Mittelwertdifferenz der Schülergruppen, beträgt 38 Punkte. Dieser Unterschied sinkt nach Kontrolle des Sozialstatus auf 6 Punkte (ATE). Der ATT beträgt nur 1.1 Punkte für Österreich. Die kon-

