

„Fairer Vergleich“

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



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

„Fairer Vergleich“

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 Ziel der Analyse

3 2 Stichproben

4 3 Kovariaten im Modell

4 3.1 Kontextvariablen

5 3.2 Englischleistungen

6 4 Umgang mit fehlenden Werten

6 5 Regressionsanalyse

7 5.1 Schätzmethode

9 5.2 Modellspezifikation

10 5.3 Ergebnisse

12 6 Bestimmung des Erwartungsbereichs

15 7 Zusammenfassung und Ausblick

16 Literaturverzeichnis



1 Ziel der Analyse

Zahlreiche Untersuchungen haben den erheblichen Einfluss von Kontextfaktoren auf die Schülerleistungen nachgewiesen. Auf Basis von mehr als 9000 Metaanalysen hat Hattie (2009) eine Synthese über den Effekt verschiedener Faktoren erarbeitet. Beispielsweise verfügt die Schulgröße über eine durchschnittliche Effektgröße von $d = 0.43$ (S. 74) oder der sozioökonomische Status der Schüler/innen hat im Durchschnitt einen Effekt von $d = 0.57$ (S. 61) auf die Schülerleistung. Es ist daher nötig, die Kontextfaktoren zu berücksichtigen, wenn ein Leistungsvergleich zwischen verschiedenen Schülergruppen von Interesse ist (*Fairer Vergleich; FV*), insbesondere weil die Lehrkräfte keinerlei Einfluss darauf haben. In Deutschland ist das Konzept des FVs seit Mitte des letzten Jahrzehnts durch die Vergleichsarbeiten (Isaac & Hosenfeld, 2008) verbreitet, die auf Systemebene als eine wesentliche Säule der Gesamtstrategie des Bildungsmonitorings dienen (KMK, 2006). In der Praxis gilt ein FV als die „*notwendige Voraussetzung zur validen Einschätzung der Wirkung von Unterricht*“ (Fiege, Reuther & Nachtigall, 2011, S. 136) und wird demzufolge als ein wichtiges Instrument zur Unterrichtsentwicklung angesehen (vgl. Helmke, Hosenfeld & Schrader, 2004). In Österreich führt das BIFIE seit 2012 im Auftrag des BMBF regelmäßig die Bildungsstandardüberprüfung (BIST-Ü) am Ende der 4. und der 8. Schulstufe durch, die eine wichtige Grundlage für die Qualitätsentwicklung an Schulen darstellt. In den BIST-Ü-Ergebnisrückmeldungen wird neben den an Bildungsstandards orientierten Kriterien auch ein *Erwartungsbereich (EB)* pro Schule/Gruppe bereitgestellt. Dieser wird auf Basis der Schul- oder Gruppenwerte mit ähnlichen strukturellen Rahmenbedingungen berechnet. Die relative Position des erreichten Werts im Vergleich zu diesem EB gibt einen Hinweis darauf, ob die eigenen Ergebnisse besser bzw. schwächer sind, als aufgrund der Standortbedingungen der Schule/Gruppe sowie der demografischen und sozioökonomischen Merkmale der Schüler/innen statistisch zu erwarten wäre. In diesem Kapitel wird der Prozess der FV-Analyse in der BIST-Ü im Fach Englisch auf der achten Schulstufe (E8) geschildert.

2 Stichproben

An der BIST-Ü E8 2013 haben 76.728 Schüler/innen aus 1410 Schulen mit 4769 Unterrichtsgruppen (UG) teilgenommen. Mit Berücksichtigung der Reliabilität der Schätzung wurden jedoch für extrem kleine Schulen und Gruppen (mit weniger als fünf Schülerinnen und Schülern) keine Erwartungsbereiche bestimmt. Aus diesem Grund sind insgesamt 1402 Schulen in die Berechnungen eingeflossen. Es gab 4699 Unterrichtsgruppen mit mindestens fünf Schülerinnen und Schülern, die in die Analyse auf Gruppenebene einbezogen wurden¹.

¹Eine Unterrichtsgruppe enthielt im Test keine Teilnehmer/innen im Testbereich Hören, weshalb die Stichprobe bei der FV-Gruppenanalyse in diesem Bereich nur 4698 Gruppen umfasste.

3 Kovariaten im Modell

3.1 Kontextvariablen

Zunächst werden die relevanten Kontextvariablen erläutert, die im Rahmen der BIST-Ü E8 2013 erhoben und zur Berechnung des Erwartungsbereichs herangezogen wurden. Tabelle 1 gibt einen Überblick über die verfügbaren Kontextvariablen (Kovariaten), die Leistungsunterschiede zwischen Schulen vorhersagen können.

Tabelle 1: Aufstellung der Kovariaten

Merkmalsname	Variable	Wert	Daten- -quelle	Anmerkungen
<i>Standortbezogene Merkmale</i>				
- Gemeindegröße	gemgroesse	1: > 500.000 2: bis 500.000 3: bis 50.000 4: bis 10.000 5: bis 3000	PLATO	Die Dummy-Variablen werden zusätzlich als Kovariaten im Modell mitaufgenommen.
- Urbanisierungsgrad	urbgrad	A: dicht B: durchschnittlich C: dünn besiedelt	PLATO	Die Dummy-Variablen werden zusätzlich als Kovariaten im Modell mitaufgenommen.
- Entfernung zur nächstgelegenen AHS	entfahs	1: weniger als 5 km 2: 5 bis 14 km 3: 15 bis 30 km 4: mehr als 30 km	SLFB	
<i>Schulbezogene Merkmale</i>				
- Schulart (AHS/APS)	ahs.fv	0: APS 1: AHS	PLATO	NMS und AHS mit NMS-Klassen zählen zur APS-Gruppe.
- Schulerhalter (öffentlich, privat)	privat	0: öffentlich 1: privat	PLATO	
- Schulgröße	log.schulgroesse	log-Wert	Tracking-daten	SSt8 Schülerzahl, log(N.schuelerK18).
<i>Aggregierte Schülermerkmale der Schule</i>				
- Anteil der Mädchen	female_idschool	Anteil (%)	Tracking-daten	
- Anteil der Schüler/innen mit/ohne Migrationshintergrund	migrant_idschool	Anteil (%)	Tracking-daten	
- Anteil der Schüler/innen, deren Erstsprache nicht Deutsch ist	ndtspr_idschool	Anteil (%)	SFB	

Tabelle 1: Aufstellung der Kovariaten

Merkmal	Variable	Wert	Daten- quelle	Anmerkungen
- Sozialstatus der Schüler/innen (Ausbildung und beruflicher Status der Eltern, Anzahl der Bücher zu Hause)	zsozstatus _idschool	z-Wert	SFB	siehe Kapitel Hintergrundvariablen (Pham, Freunberger & Robitzsch, 2015).
- Anteil der von der Testung ausgeschlossenen Schüler/innen	P.ausSCHULE	Anteil (%)	Tracking- daten	Verwendung bei der Analyse auf Schulebene.
- Anzahl der von der Testung ausgeschlossenen Schüler/innen	N.ausg	0–13	Tracking- daten	Verwendung bei der Analyse auf Gruppenebene.

Die Kovariaten umfassten die Schulart, die Gemeindegröße, den Urbanisierungsgrad, die Entfernung zur nächstgelegenen AHS, die Art des Schulträgers und die aggregierten Merkmale der Schülerschaft. Analog zur Vorgehensweise in der BIST-Ü im Fach Mathematik der achten Schulstufe (M8) im Jahr 2012 (Robitzsch, in Vorbereitung) wurde die logarithmierte Jahrgangsstärke anstatt der absoluten Schülerzahl der Schulen verwendet. Auf Schulebene wurde die Anzahl ausgeschlossener Schüler/innen in Prozentanteile transformiert, um den Schulgrößen-Effekt zu bereinigen. Diese Transformationen wurden auf Gruppenebene nicht durchgeführt, da die Größe der verschiedenen Unterrichtsgruppen weniger stark schwankte. Bis auf diesen Unterschied war die Kovariatenliste für das Analysemodell auf Gruppenebene identisch (mit Gruppen als Aggregationsebene).

3.2 Englischleistungen

Die Vollerhebung in der BIST-Ü E8 2013 erfolgte in drei Kompetenzbereichen Hören, Lesen und Schreiben. Die Skalierungsergebnisse² in diesen Bereichen dienten als Grundlage für die bereichsweisen FV-Analysen (abhängige Variablen). Der Englisch-Gesamtwert war der Mittelwert dieser drei Bereichswerte und wurde unmittelbar nach jedem Imputationsvorgang (siehe Abschnitt 4) gebildet. Pro Kompetenzbereich sowie für die Englisch-Gesamtleistung wurde für jede Schule/Gruppe ein Erwartungsbereich bestimmt und zurückgemeldet (siehe Abschnitt 6).

²Die Ergebnisse wurden auf einen Wertebereich von 200 bis 800 gestutzt.

4 Umgang mit fehlenden Werten

Bei der BIST-Ü E8 2013 waren fehlende Werte (Missings) vorhanden. Der Missinganteil der relevanten Variablen war unterschiedlich hoch (er variierte zwischen 0% und 11.5%). Missings lagen bei den individuellen Leistungsdaten und Kovariaten vor, dagegen waren die Kovariaten auf Schul- und Gruppenebene vollständig. Die Missinganteile der individuellen Variablen sind in Tabelle 2 angegeben.

Tabelle 2: *Missinganteile der individuellen Kovariaten*

Variable	Labels	Mittelwert	Missings	% Missings
female	Geschlecht	0.50	20	0.0%
migrant	Migrationshintergrund	0.18	661	0.9%
ndtspr	Andere Muttersprache als Deutsch	0.17	1724	2.2%
eltausb	Elternausbildung	2.60	8821	11.5%
HISEI	Highest International Socio-Economic Index of occupational status (auf Basis ISCO-08)	49.02	3700	4.8%
buch	Anzahl Bücher zu Hause	3.53	1244	1.6%
E8LTWLE	Skalierungsergebnis Hören	536.45	32	0.0%
E8RTWLE	Skalierungsergebnis Lesen	524.28	28	0.0%
E8WTWLE	Skalierungsergebnis Schreiben	495.19	127	0.2%

Um durch Missings verursachte Probleme wie eine reduzierte Stichprobengröße oder verzerrte Parameterschätzungen zu vermeiden (siehe Little & Rubin, 2002), wurden die fehlenden Daten unter Einbeziehung vorhandener Informationen multipel imputiert. Dafür wurde eine Mehrebenenimputation mithilfe des R-Pakets „pan“ (Zhao & Schafer, 2013) eingesetzt. Dabei wurden sowohl alle Variablen auf Gruppen- und Schulebene als auch die Individualvariablen (s. Tabelle 1) in einem Random-Intercept-Modell eingefügt. Die Schule wurde dabei als Cluster-Variable verwendet. Es wurde angenommen, dass die Daten zufällig fehlten (*missing at random*, siehe Rubin, 1976, vgl. Lüdtke und Robitzsch, 2010), d. h. das Auftreten von fehlenden Werten einer Variable ist durch die Ausprägung anderer beobachteter Variablen im Imputationsmodell, nicht jedoch durch die Ausprägung auf fehlenden Variablen bedingt. Durch die Durchführung der multiplen Imputation, die zu 50 imputierten Datensätzen führte, wurde die Unsicherheit bei der Ersetzung berücksichtigt. Darüber hinaus wurde die geschachtelte Datenstruktur in Betracht gezogen, da Schüler/innen einer Schule ähnlichere Werte auf einer Variablen aufweisen können als Schüler/innen verschiedener Schulen. Die Imputationen fanden separat nach Stratum (Schulform³ x Stadt/Land) statt, damit die stratumspezifische Zusammenhangsstruktur zwischen Variablen berücksichtigt werden konnte.

5 Regressionsanalyse

Nach der Datenimputation wurden die Individualdaten auf Analyseebene (Schul- bzw. Gruppenebene) aggregiert. Daraus entstanden auf jeder Analyseebene 50 aggregierte Da-

³Schulen, die bereits als NMS geführt werden, zählen zur Gruppe der APS. Dies gilt demnach auch für AHS, die NMS-Klassen führen.

tensätze, auf deren Basis die Regressionsanalysen spezifiziert und durchgeführt wurden. Außerdem erfolgten die Analysen getrennt nach Stratum. Anschließend wurden die einzelnen Ergebnisse nach den Kombinationsformeln von Rubin zusammengefasst (Lüdtke & Robitzsch, 2010).

5.1 Schätzmethode

Bei der FV-Analyse wurde eine große Anzahl von Prädiktoren verwendet, die z. T. sehr hoch miteinander korrelierten. Beispielsweise betrug die Korrelation zwischen dem Anteil der Migrantinnen und Migranten und dem Anteil von Schülerinnen und Schülern mit anderer Muttersprache als Deutsch $r = .97$. Die Korrelationsmatrix der Kovariaten auf Schulebene wird in Tabelle 3 dargestellt.

Variable	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]
gemgroesse [1]	1.00									
urbgrad [2]	.81	1.00								
entfahs [3]	.64	.63	1.00							
privat [4]	-.23	-.19	-.16	1.00						
log.schulgroesse [5]	-.42	-.35	-.26	-.03	1.00					
P.ausgSCHULE [6]	-.16	-.13	-.18	-.12	-.09	1.00				
female_idschool [7]	-.02	-.01	.02	.12	.09	-.07	1.00			
migrant_idschool [8]	-.66	-.59	-.47	-.03	.13	.42	-.05	1.00		
ndtspr_idschool [9]	-.63	-.57	-.46	-.04	.11	.42	-.06	.97	1.00	
zsozstatus_idschool ⁴ [10]	-.19	-.19	-.06	.38	.35	-.39	.16	-.37	-.40	1.00

Tabelle 3: Korrelationsmatrix der Kovariaten auf Schulebene

Mit der großen Anzahl der Variablen und dem starken Zusammenhang zwischen den Kovariaten gingen Probleme einher, die die Vorhersagekraft des Hintergrundmodells in Frage stellen können. Angesichts dessen haben wir das *Lasso-Verfahren* (Lasso) (Tibshirani, 1996, Tutz, 2012, S. 149–154) als Regularisierungsmethode gewählt, um eine optimale Prädiktion des erwarteten Werts zu gewinnen. Mit dem Lasso wird die absolute Größe einiger Regressionskoeffizienten vermindert und andere Regressionskoeffizienten werden auf null gesetzt. Ziel war es, nur die bedeutsamsten Prädiktoren von den miteinander korrelierten Variablen im Modell zu behalten. Dadurch konnten stabilere Vorhersagen für die erwarteten Werte gewonnen werden. Unsere Analysen fanden in R (R Core Team, 2013) unter Verwendung des glmnet-Pakets (Friedman, Hastie & Tibshirani, 2010) statt.

Bei der multivariaten Regressionsanalyse $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$ mit der Methode der kleinsten Quadrate (OLS) wird das Kriterium der kleinsten Fehlerquadratsumme verwendet, um die Regressionsparameter $\mathbf{b} = (b_1, \dots, b_p)$ zu bestimmen. Beim Lasso-Verfahren wird die Summe der Fehlerquadrate $\|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2$ und einem Strafterm (penalty) $Pen(\mathbf{b}) = \sum_{l=1}^p |b_l|$ (p ist die Anzahl der Prädiktoren) minimiert:

$$L_{\text{lasso}}(\mathbf{b}) = \|\mathbf{y} - \mathbf{X}\mathbf{b}\|^2 + \lambda \cdot Pen(\mathbf{b})$$

(vgl. Friedman et al., 2010, S. 3).

⁴Diese Variable wurde aus drei Variablen Elternausbildung, HISEI und Anzahl der vorhandenen Bücher zu Hause gebildet, siehe Kapitel Hintergrundvariable (Pham et al., 2015).

Die Lösung dieses Problems sowie die Bestimmung der Regressionskoeffizienten \mathbf{b} hängen von der Auswahl des λ ab. Anhand eines Beispiels (unter Verwendung von Daten des Stratum 1 des ersten imputierten Schuldatensatzes mit Englisch-Gesamtwert als abhängige Variable) möchten wir die Bestimmung der Regressionskoeffizienten und λ illustrieren. In Abbildung 1 werden die Schätzungen der Regressionskoeffizienten (y-Achse) in Abhängigkeit von λ (x-Achse) dargestellt.

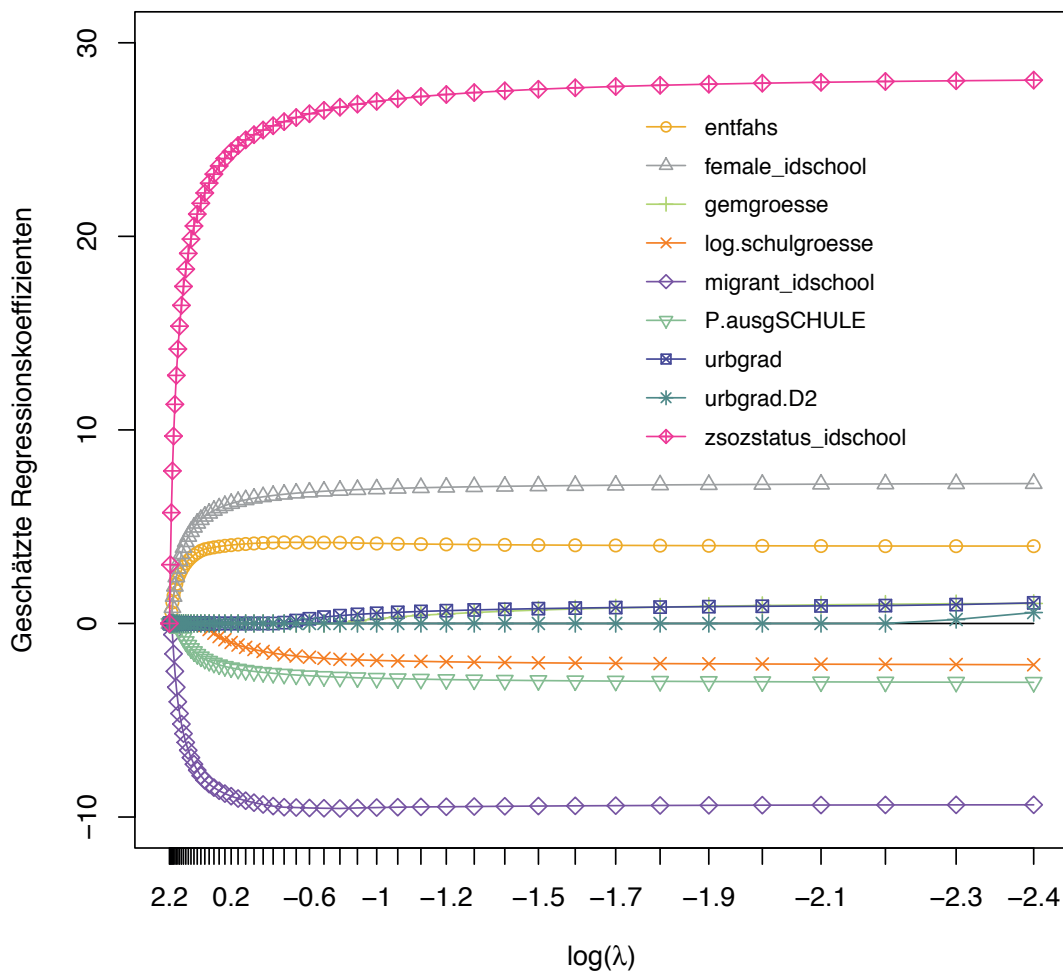


Abbildung 1: Geschätzte Regressionskoeffizienten gemäß den ersten 50 generierten λ

Zur Auswahl des optimalen λ haben wir die k -fache Kreuzvalidierungs-Methode verwendet. Der λ -Wert, mit dem das kleinste *mittlere Fehlerquadrat der Kreuzvalidierung* (cvm) erreicht wurde, wurde gewählt. Abbildung 2 veranschaulicht diesen Auswahlprozess. Bei dieser Teilanalyse (erster imputierter Datensatz, Stratum 1, Analyse auf Schulebene mit Gesamtwert als AV) war das kleinste cvm in der Höhe von 789.41, entspricht $\lambda = .14$ (bzw. $\log(\lambda) = -1.98$). Der Wert λ wurde somit definiert, darauf basierend wurden die Regressionskoeffizienten geschätzt.

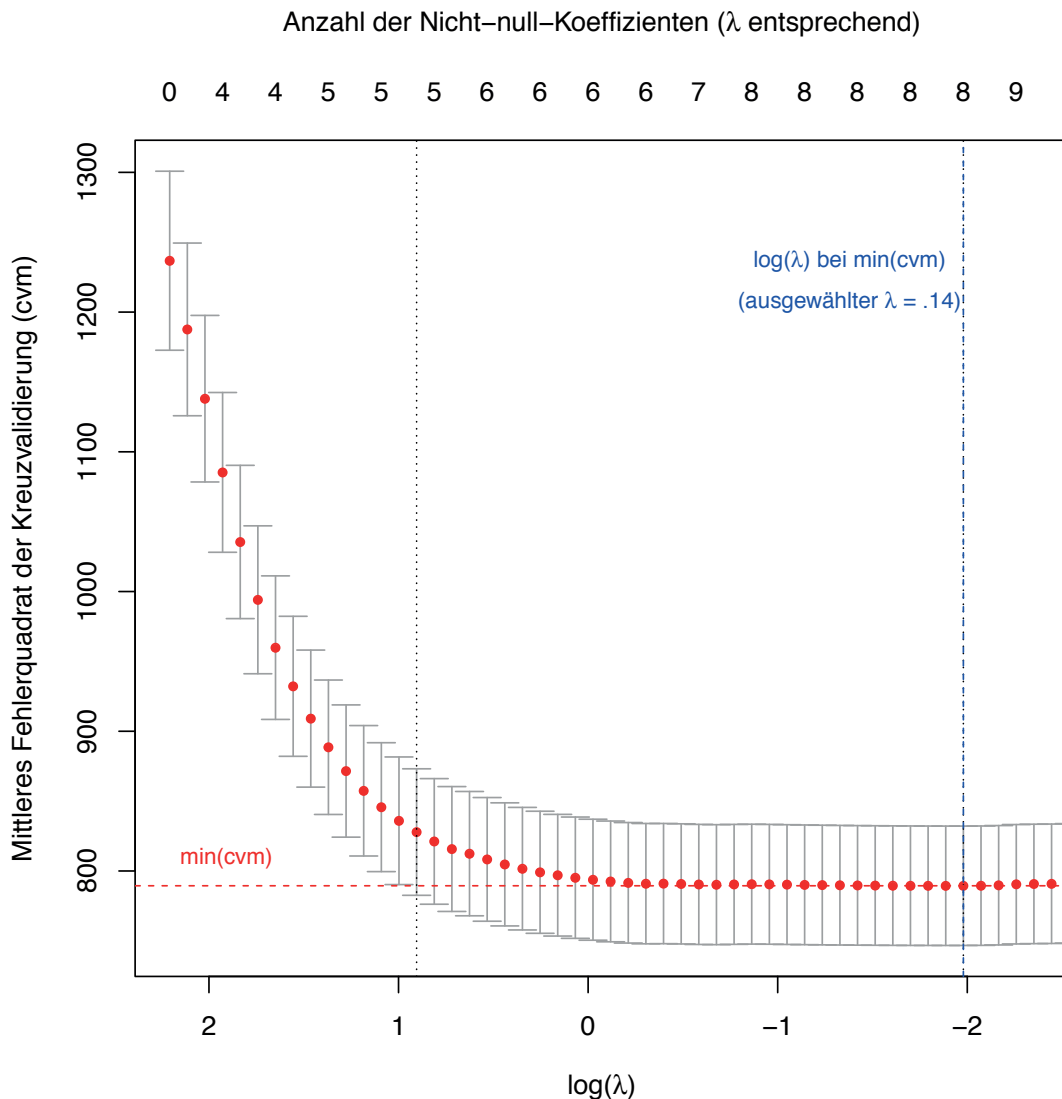


Abbildung 2: Auswahl des optimalen λ -Parameters für die Bestimmung der Regressionskoeffizienten

5.2 Modellspezifikation

Vor der Regressionsanalyse wurden die Hauptprädiktoren z-standardisiert, damit sie dieselbe Metrik besitzen. Die Dummyvariablen (siehe Tabelle 1) wurden im Modell mit einer Gewichtung $w_v = 1/K_v$ aufgenommen, wobei v der Indikator der v -ten kategorischen Variable und K_v die Anzahl der dazugehörigen Dummies bezeichnete. Mit dieser Spezifikation wollten wir sowohl den Haupteffekt als auch die dummiespezifischen Effekte gleichzeitig analysieren. Da die Dummies mit der Hauptvariable stark korrelierten, wurde im Nachhinein nur der bedeutsamste Effekt im Modell beibehalten. Um eine stabile Lösung zu gewinnen und die Replizierbarkeit der Ergebnisse zu sichern, haben wir die k -fache Kreuzvalidierungs-Methode mit dem größtmöglichen k eingesetzt mit einer eindeutigen Zuordnung der Stichprobe zu den Teilmengen. Jeder imputierte Datensatz wurde in k Teilmengen geteilt, mit $k = N_s/2$, falls N_s gerade war, bzw. $k = (N_s + 1)/2$, falls N_s ungerade war, und N_s war die Stichprobengröße des jeweiligen Stratums. Die Zuordnung

der Stichprobe zu den k -Teilmengen erfolgte durch eine wiederholte Durchnummerierung von 1 bis k .

5.3 Ergebnisse

Die resultierten Regressionskoeffizienten aus der oben beschriebenen Lasso-Schätzmethode (Arbeitsmodell) auf Schulebene mit Englisch-Gesamtleistung als abhängige Variable werden in Tabelle 4 angegeben. Zum Vergleich werden auch die Ergebnisse einer parallelen OLS-Regressionsanalyse (OLS) gegenübergestellt.

Generell wurden alle nach dem OLS-Modell signifikanten Prädiktoren (diejenige Variablen mit $p < .10$) mit vergleichbaren Regressionskoeffizienten (b) im Arbeitsmodell beibehalten. Der Anteil von Migrantinnen und Migranten und der Anteil von Schülerinnen und Schülern mit anderer Muttersprache als Deutsch wurden in keinem Fall gleichzeitig im Modell beibehalten. Erwartungsgemäß war der stärkste Prädiktor für die Schülerleistung im Fach Englisch der mittlere Sozialstatus der Schüler/innen einer Schule. Das galt für alle Schularten in allen Strata. Ebenso positiv und übergreifend prädiktiv war der Mädchenanteil in der Schule. Der Anteil ausgeschlossener Schüler/innen hing negativ mit Schülerleistung zusammen, jedoch galt das nur für APS und nicht für AHS. Die Entfernung zur nächstgelegenen AHS sagte die Schülerleistung in allen Schularten voraus, allerdings nur in ländlichen Gebieten und nicht in den Städten. Innerhalb jedes Stratum spielten die Gemeindegröße und der Urbanisierungsgrad (sowohl die Hauptvariablen als auch die Dummies) kaum eine Rolle zur Vorhersage der Schülerleistung. Insgesamt ist der aufgeklärte Varianzanteil des Arbeitsmodells pro Stratum in vergleichbarer Höhe mit dem des OLS-Modells trotz der geringeren Anzahl der Prädiktoren.

Variable	Stratum 1 (APS, Land, $N = 910$)			Stratum 2 (APS, Stadt, $N = 225$)			Stratum 3 (AHS, Land, $N = 126$)			Stratum 4 (AHS, Stadt, $N = 141$)		
	OLS		sig	Lasso		sig	OLS		sig	Lasso		sig
	b	b		b	b		b	b		b	b	
gemgroesse	1.0	2.9	-	0.0	-1.5	-	0.0	10.2	-	0.0	8.7	-
gemgroesse.D1	-	-	-	-	-	-	-	-	-	-	-	-
gemgroesse.D2	-	-	-	-	-	-	-	-	-	-	-	-
gemgroesse.D3	0.0	12.4	-	-	-	-	0.0	38.0	-	-	-	-
gemgroesse.D5	-	-	-	-	-	-	-	-	-	-	-	-
urbgrad	0.8	2.4	-	0.0	4.0	-	-1.3	-1.1	-	0.0	13.2	-
urbgrad.D2	0.0	5.2	-	-	-	-	0.0	6.6	-	-	-	-
urbgrad.D3	-	-	-	-	-	-	-	-	-	-	-	-
entfahs	4.1	4.1	***	0.0	7.3	-	3.2	4.2	*	0.0	11.2	-
privat	0.0	-0.3	-	-1.3	-2.0	-	-0.8	-0.9	-	-0.6	-1.0	-
log.schulgroesse	-2.1	-2.2	*	1.4	1.3	-	0.8	1.0	-	3.5	4.8	-
P.ausgSCHULE	-3.0	-3.1	***	-2.9	-3.6	**	0.0	-2.6	-	0.0	0.3	-
female_idschool	7.2	7.3	***	6.0	6.8	***	6.9	7.6	***	4.5	4.3	**
migrant_idschool	-9.3	-14.5	***	0.0	0.1	-	0.0	-21.7	-	0.0	13.5	-
ndtspr_idschool	0.0	5.2	-	2.7	7.0	-	0.0	27.1	-	-2.1	-11.7	-
zsozstatus_idschool	27.9	28.7	***	31.8	38.5	***	14.5	19.2	***	22.7	29.0	***
R^2	.36	.37	-	.56	.60	-	.15	.13	-	.42	.45	-

Tabelle 4: Ergebnisse der Regressionsanalyse auf Schulebene

Notiz: N ist die Stichprobengröße; b ist der unstandardisierte Regressionskoeffizient; sig ist die Signifikanz (* : $p < .10$, ** : $p < .05$, *** : $p < .01$); R^2 ist der aufgeklärte Varianzanteil des Analysemodells, fürs OLS-Modell haben wir den adjustiert R^2 gerechnet aufgrund der großen Anzahl der Prädiktoren; „-“ bedeutet nicht schätzbar.

Auf Basis der ausgewählten Prädiktoren und des Analysemodells wurde für jede Schule ein Erwartungswert geschätzt. Auch hier ähnelten die arbeitsmodellbezogenen Erwartungswerte den erwarteten Werten des OLS-Modells. Der Mittelwert der Erwartungswerte (Englisch-Gesamtwert) aller Schulen anhand des Arbeitsmodells betrug 503.22. Dasselbe Ergebnis erbrachte auch das OLS-Modell. Die Korrelation zwischen den Schätzwerten nach beiden Modellen betrug $r = .999$. Die Erwartungswerte (Schätzwerte) der einzelnen Schulen nach Lasso gegenüber denen des OLS-Modells werden in Abbildung 3 dargestellt, getrennt nach Strata.

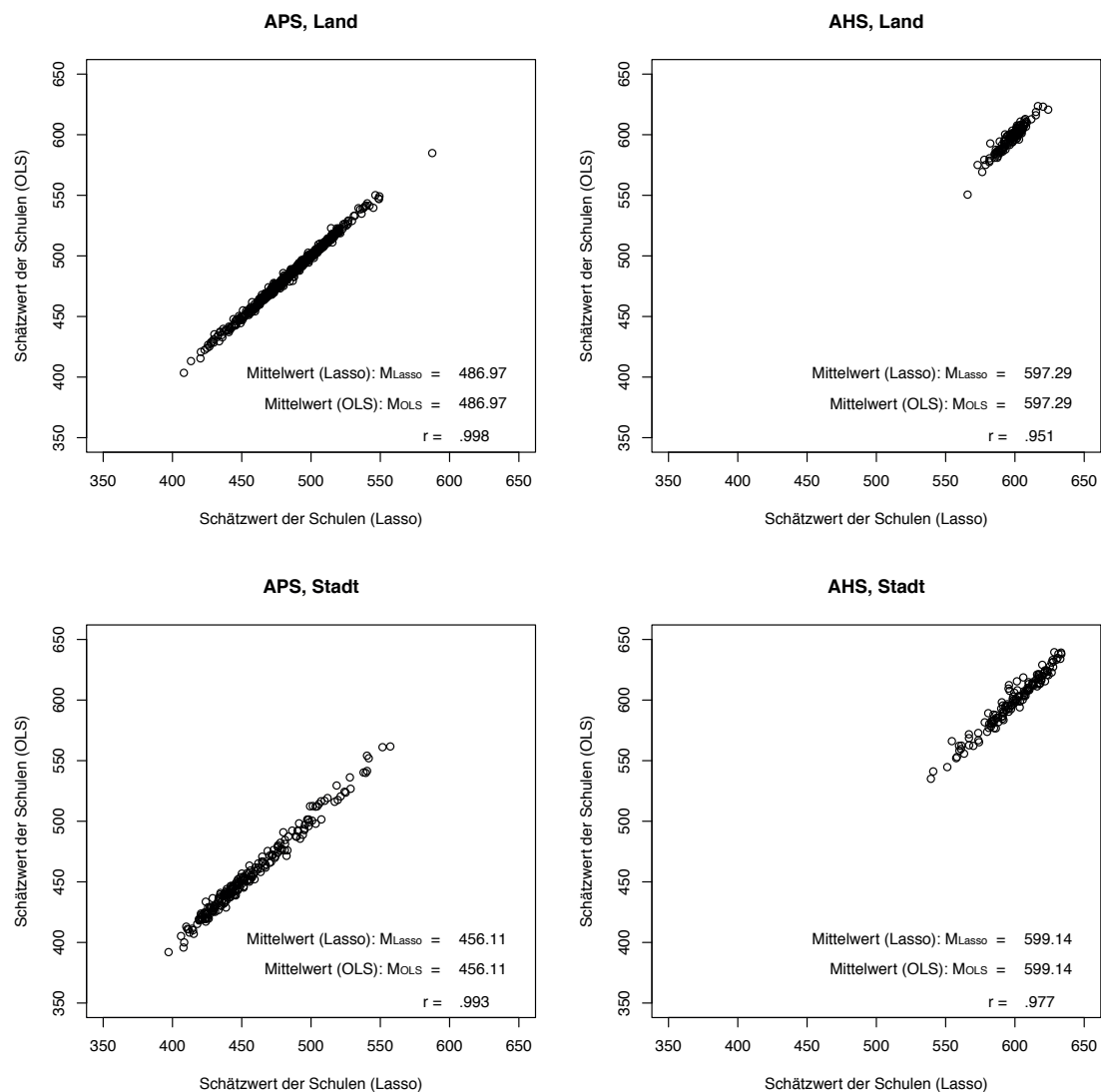


Abbildung 3: Vergleich der Schätzwerte des Arbeitsmodells und des OLS-Modells

6 Bestimmung des Erwartungsbereichs

Der letzte Schritt bei der FV-Analyse war die Bestimmung der Breite des Erwartungsbereichs (EB), der für alle Schulen und in Bezug auf alle abhängigen Variablen (Englisch-Gesamtwert und Kompetenzbereich-Werte) gültig war. Der EB jeder Schule war ein Wert,

tebereich mit dem Mittelpunkt, der dem geschätzten Erwartungswert aus dem Arbeitsmodell entsprach (s. Abschnitt 5.3). Mit der Definition des EB verfolgten wir zwei Ziele. Erstens sollte die Unsicherheit in der Bestimmung der Erwartungswerte berücksichtigt werden. Zweitens sollte die verteilungsbezogene Vorgabe erfüllt werden, dass die Anteile der Schulen mit dem erreichten Leistungswert unter-, inner- und oberhalb des eigenen Erwartungsbereichs 25 %-50 %-25 % lagen.

Eine übliche Auswahl des EB für den ersten Zweck (Berücksichtigung der Messfehler) ist in der Regel das Konfidenzintervall des Schätzwerts aus der Skalierung und aus der Regressionsanalyse. Allerdings ist die statistische Signifikanz aus unserer Sicht kein hinreichender Beleg für die praktische Signifikanz. Das Konfidenzintervall ist von der Schul- bzw. Gruppengröße abhängig, d. h. kleinere Schulen hätten einen breiteren EB gehabt, und daher eine höhere Wahrscheinlichkeit, im Erwartungsbereich zu liegen. Aus diesen Gründen wurde ein alternativer Prozess zur Bestimmung des EB gewählt. Die entsprechende Vorgehensweise war explorativ, analog zu dem Prozess in der BIST-Ü M8 (vgl. Robitzsch, in Vorbereitung). Die Breite des EB wurde zunächst so gewählt, dass die Verteilung der Schulen/Gruppen mit Augenmerk auf die Position gegenüber dem eigenen EB der Wunschverteilung 25-50-25 entsprach. Auf Schulebene betrug die optimale Breite in Bezug auf alle Kompetenzwerte 34 Punkte (siehe Abbildung 4), somit lag der EB einer Schule zwischen [Erwartungswert -17] und [Erwartungswert +17]. Auf Gruppenebene wurde eine Breite von 52 Punkten gewählt. Der EB einer Gruppe war deshalb der Bereich [Erwartungswert +/-26]. Die resultierte Verteilung auf Gruppenebene ist in Abbildung 5 dargestellt. Da die gewählte Breite des EB auf jeder Ebene hinreichend groß war, deckte der EB jeder Schule/Gruppe ebenfalls das Konfidenzintervall des Erwartungswerts ab.

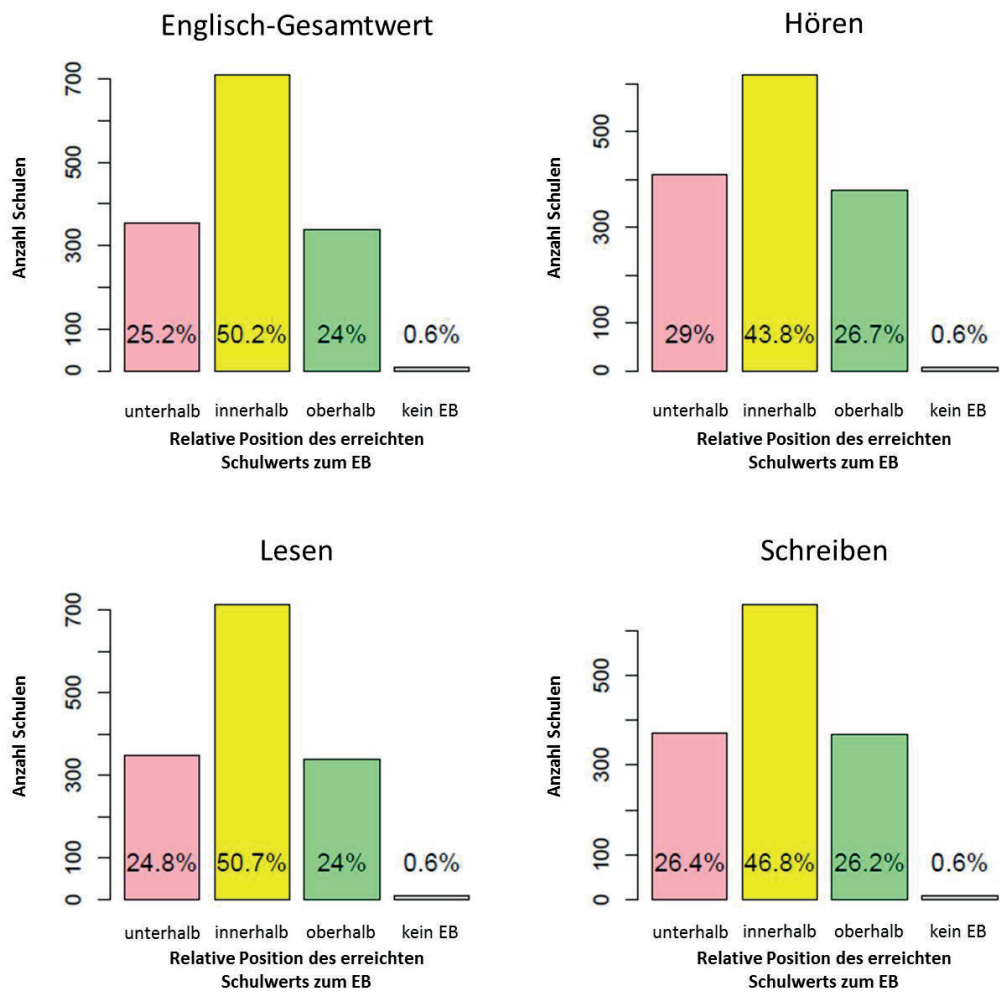


Abbildung 4: Positionsverteilung auf Schulebene: Erreichte Schulleistung gegenüber dem eigenen EB.

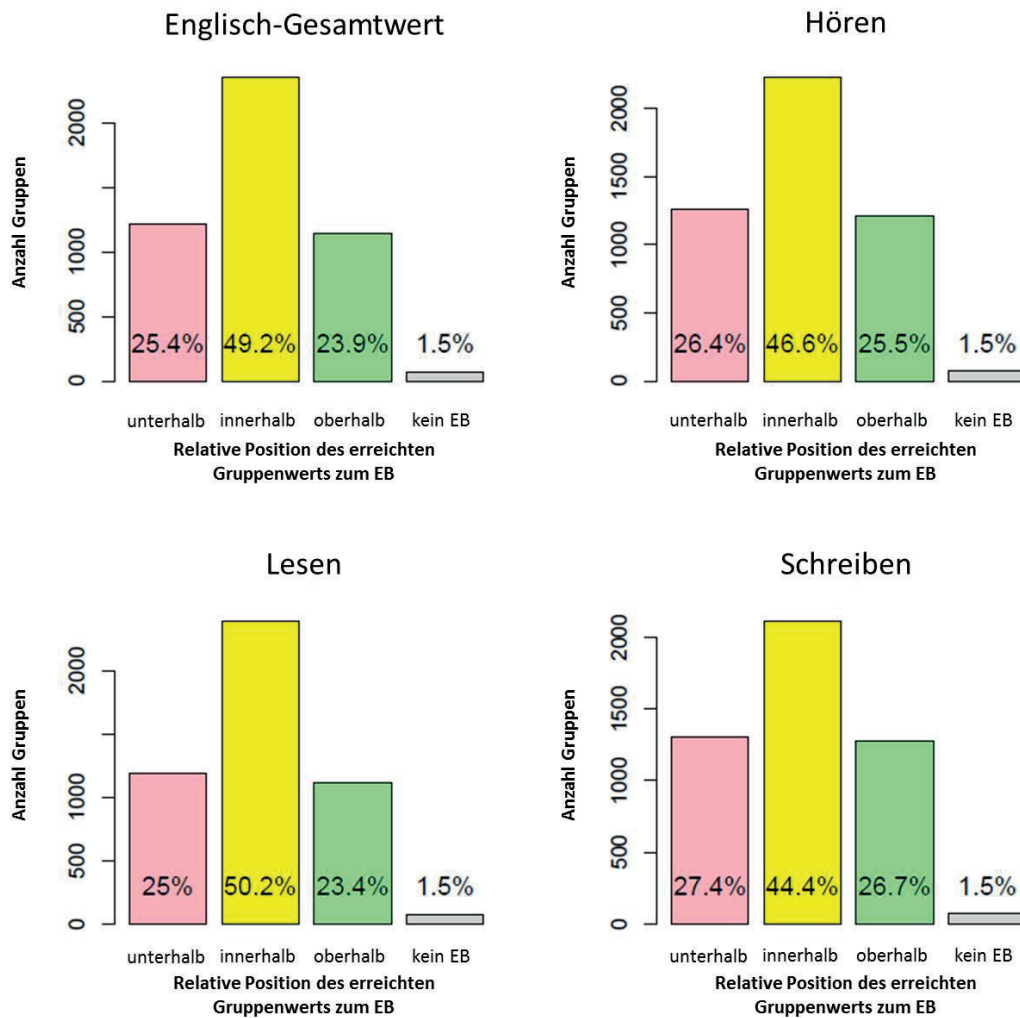


Abbildung 5: Positionsverteilung auf Gruppenebene: Erreichte Gruppenleistung gegenüber dem eigenen EB.

7 Zusammenfassung und Ausblick

Die FV-Analyse ist ein wichtiger Bestandteil der Rückmeldung im Rahmen der BIST-Ü in Österreich. Damit wird ein sozialer Vergleich zwischen Schulen mit ähnlichen Rahmenbedingungen hergestellt.

In diesem Bericht wurden die Analyse-Schritte im Rahmen der BIST-Ü E8 2013 beschrieben. Der zentrale Punkt dabei war die Verwendung des Lasso-Verfahrens bei der Regressionsanalyse, mit dem eine effiziente Kovariatenauswahl und eine bessere Interpretierbarkeit der Ergebnisse erreicht werden konnte, insbesondere aufgrund der großen Zahl an Modellprädiktoren. Die daraus resultierten Ergebnisse wurden vorgestellt und mit denen einer OLS-Regressionsanalyse verglichen. Angesichts der Identifizierbarkeit bedeutsamer Prädiktoren, die eine hohe Validität und Stabilität des Modells sichert, sowie der Reliabilität der Ergebnisse wird das Potenzial dieser Methode für die nächste FV-Analyse im Rahmen der BIST-Ü weiterverfolgt.

Literaturverzeichnis

- Fiege, C., Reuther, F. & Nachtigall, C. (2011). Faire Vergleiche? – Berücksichtigung von Kontextbedingungen des Lernens beim Vergleich von Testergebnissen aus deutschen Vergleichsarbeiten. *Zeitschrift für Bildungsforschung*, 2, 133–149.
- Friedman, J., Hastie, T. & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Grieser (Hrsg.), *Schulleitung und Schulentwicklung* (S. 119–144). Hohengehren: Schneider.
- Isaac, K. & Hosenfeld, I. (2008). Faire Ergebnisrückmeldung bei Vergleichsarbeiten. In J. Ramseger & M. Wagener (Hrsg.), *Chancenungleichheit in der Grundschule – Ursachen und Wege aus der Krise* (S. 143–146). Wiesbaden: VS-Verlag für Sozialwissenschaften.
- KMK (Hrsg.). (2006). *Gesamtstrategie der Kultusministerkonferenz zum Bildungsmonitoring*. Bonn: LinkLuchterhand.
- Little, R. J. & Rubin, D. B. (2002). *Statistical analysis with missing data*. Hoboken (Bd. 34). NJ: Wiley.
- Lüdtke, O. & Robitzsch, A. (2010). Umgang mit fehlenden Daten in der empirischen Bildungsforschung. In S. Maschke & L. Stecher (Hrsg.), *Enzyklopädie Erziehungswissenschaft Online. Fachgebiet Methoden der empirischen erziehungswissenschaftlichen Forschung, Quantitative Forschungsmethoden*. Weinheim: Juventa.
- Pham, G., Freunberger, R. & Robitzsch, A. (2015). *Hintergrundvariablen und spezielle Analysen: Technische Dokumentation - BIST-Ü Englisch, 8. Schulstufe, 2013* (Tech. Rep.). Salzburg: BIFIE.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna, Austria.
- Robitzsch, A. (in Vorbereitung). „Fairer Vergleich“. *Technische Dokumentation – BIST-Ü Mathematik, 8. Schulstufe, 2012*. (Tech. Rep.). Salzburg: BIFIE.
- Rubin, D. (1976). Inference and missing data. *Biometrika*, 63, 581–592.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58(1), 267–288.
- Tutz, G. (2012). *Regression for categorical data*. Cambridge University Press.
- Zhao, J. H. & Schafer, J. L. (2013). *pan: Multiple imputation for multivariate panel or clustered data*. (R package version 0.9)

