

Ein harmonisierter Datensatz auf der Grundlage von CILS4EU & NEPS SC4

Lena Arnold, Andreas Horr, Jörg Dollmann, Victoria Kerzner, Regine Schmidt, Hannah Soiné und Markus Weißmann

KSWD-Vortrag, 28. März 2023 Berlin







CILS4NEPS – Ein Überblick I



- Harmonisierter Datensatz, kombiniert:
 - o CILS4EU "Children of Immigrants Longitudinal Survey in Four European Countries", Welle 1-3
 - & NEPS SC4 "Nationales Bildungspanel Startkohorte 4", Welle 1-6



- Output-Harmonisierung / Ex-Post Harmonisierung
- Harmonisierung manifester und latenter Konstrukte





CILS4NEPS – Ein Überblick II

Warum?

- Erhöhte Fallzahlen für nationale Analysen Gruppen & Events
- Erhöhte Generalisierbarkeit
- Internationale Vergleichbarkeit von NEPS SC4

Wer?

- Mannheimer Zentrum für Europäische Sozialforschung (MZES) & Leibniz-Institut für Bildungsverläufe
 (LIfBi)
- KonsortSWD, RDM Projekt-Förderung







Vergleichbarkeit



CILS4EU W1-3		NEPS SC4 W1-6
Children of Immigrants Longitudinal Survey in Four European Countries	Datensatz	Nationales Bildungspanel – Startkohorte 4
2010 in Klasse 9	Start	2010 in Klasse 9
Deutschland, England, Niederlande, Schweden	Länder	Deutschland
DE: ~ 5,000 Jugendliche in Welle 1, Internat.: 19,000	Befragte	~ 15.500 Jugendliche in Welle 1
In Schul-Kontext verwaltet durch IEA-DPC bis Welle 2	Erhebung	In Schul-Kontext durchgehend verwaltet durch IEA- DPC
NEPS Stichprobenrahmen, aber andere Schulen	Stichproben- rahmen	NEPS Stichprobenrahmen
Gleicher Anteil von Personen mit und ohne Migrationshintergrund		Große Stichprobe der Zielbevölkerung in allen deutschen Bundesländern
Länderübergreifende Vergleichbarkeit: Vergleichbare Daten, Design, and Operationalisierungen in w1-3 für vier Länder	Eigenschaften	Schwerpunkt auf Kompetenztests in den Bereichen Sprache, Mathematik, Naturwissenschaften, ICT und kognitives Problemlösen





CILS4NEPS – Fallzahlen & Inhalte

DE: 20.608; Ingesamt: 35.276 Jugendliche

Welle 1: 12.372 mit Migrationshintergrund | 21.904 ohne Migrationshintergrund

• 110 harmonisierte Variablen (70 manifeste, 40 latent)

Sozio-demographie

Haushalt Situation

Migrationsgeschichte

Schulleistung und -einstellungen

Zukunftspläne

Ökonomische Situation

Momentane Situation

Romantische-/Fam.beziehungen

Sprachkenntnisse

Identität, Religion & Religiosität

Freizeitaktivitäten

Wohlbefinden/Gesundheit





Generelle Infos & Sozialer Hintergrund

Geschlecht

Geburtstag

Beschäftigungsstatus & Beruf (ISCO-08), Eltern

Höchster Bildungsabschluss, Eltern

Haushalt Situation

Haushaltsmitglieder

Haushaltsgröße

Migrationsgeschichte

Geboren in Erhebungsland

Geburtsland, Eltern

Alter bei Migration

Generationenstatus

Schulleistung & -einstellungen

Schulleistungen – Noten (Mathe, Deutsch, Englisch)

Zukunftspläne

Berufliche Ziele (ISCO-08)

Momentane Situation

Derzeitige Schulform

Romantische Beziehungen

In Partnerschaft

Migrationshintergrund Partner:in





Ex-Post Harmonisierung – Vorgehensweise

5 Schritte (nach Granda, Wolf and Hadorn, 2010; Singh, 2021):

- 1) Datensätze identifizieren
- 2) Ähnliche Fragen in Quell-Fragebögen identifizieren
- 3) Ziel-Variablen definieren
- 4) Entscheidung über- und Definition der Harmonisierungsstrategie
- 5) Routinen festhalten





Ex-Post Harmonisierung – Vorgehensweise

1) Datensätze identifizieren & vorbereiten

- CILS4EU: wide-format → long-format
- Relevante Datensätze mergen

	2010/11		2011	l/12	2012/13		
NEPS SC4	Welle 1	Welle 2	Welle 3	Welle 4	Welle 5	Welle 6	
CILS4EU	Welle 1		Welle 2		Welle 3		





Ex-Post Harmonisierung – Unsere Vorgehensweise

- 2) Ähnliche Fragen in Quell-Fragebögen identifizieren
 - Variablen-Name, Fragenformulierung, Antwortkategorien, Welle, Befragtengruppe
 - Klassifikation der Variablen

(Fast) identisch

Ähnliches Konstrukt

Nicht vorhanden

3) Ziel-Variablen definieren

- Beurteilung der Vergleichbarkeit der Variablen (Wolf et al., 2016; Granda, Wolf and Hadorn, 2010;
 Hoffmeyer-Zlotnik, 2008):
 - Ö Ähnliches Konstrukt?
 - Manifest oder latent?
 - Ö Ähnlichkeit der Antwortkategorien?



Harmonisierungsstrategie





Ex-Post Harmonisierung – Vorgehensweise

4) Entscheidung über- und Definition der Harmonisierungsstrategie

A) Matching

- Manifeste Variablen
- Zusammenführung der Antwortkategorien (Rekodierung, Zusammenfassung)
- Beispiele: *Generationenstatus*

B) Linear Equating

- Latente Konstrukte
- Durchführung: Gewichtete Häufigkeitverteilung
- Beispiele: Selbstwirksamkeit in der Schule





Beispiel: Generationenstatus

Ähnliche Logik ABER keine direkte 1:1 Anpassung möglich

- Gemeinsamkeiten:
 - o z.B. Gleiche Klassifikationslogik der 1., 2., 3. und 3,25. Generation
- Unterschiede:
 - o z.B. Sub-Klassifikationen der Generationen

Lösung?

○ → Erstellung von Varianten anstatt Anpassung





Beispiel: Akademisches Selbstkonzept Schule – gut in der Schule sein

•	_						
CILS4EU	Wie sehr stimmst du den einzelnen Aussagen zu oder nicht zu? - Ich bin mir sicher, dass ich in der Schule gut sein kann.						
	ur	rsprüngliche Ausprägung	N	harı	monisierte Ausprägung		
	1	stimme voll und ganz zu	8594	1	stimme voll und ganz zu		
	2	stimme zu	8950	2	stimme zu		
	3	weder noch	886	3	weder noch		
	4	stimme nicht zu	135	4	stimme nicht zu		
	5	stimme überhaupt nicht zu	50	5	stimme überhaupt nicht zu		
NEPS SC4	Wie schä	itzt du dich in der Schule ein? - Ic	h bin in den meist	ten Schulfäci	hern gut.		
	1	trifft völlig zu	295	0.57			
	2	trifft eher zu	3682	1.43			
	3	trifft eher nicht zu	8806	2.29			

2542

3.15





4

trifft gar nicht zu

Ex-Post Harmonisierung – Vorgehensweise

5) Routinen festhalten

Ausführliches Dokumentationsmaterial steht den Datennutzenden zur Verfügung

Dokumentation

Für die Arbeit mit CILS4NEPS stehen die nachfolgend aufgelisteten Materialien zur Verfügung. Darüber hinaus sei auf die jeweilige Datendokumentation von *CILS4EU* sowie der *NEPS-Startkohorte 4* verwiesen.

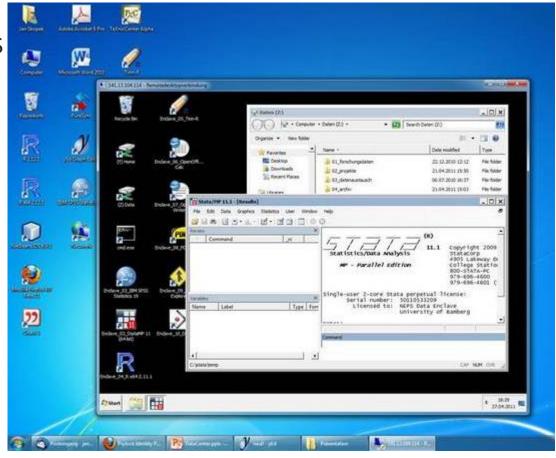
- Abschlussbericht des Harmonisierungsprojekts: PDF-de
- Codebook
- Gewichtungsreport: PDF-en
- Überblickstabelle
- Kodierungstabelle
- Technischer Bericht zum Datensatz
- Do-Files zur Erstellung der Zielitems und des harmonisierten Datensatzes
- Do-Files zu Herkunftsland und Generationenstatus
- Excel-Vorlage zu Linear Equating





CILS4NEPS - Verfügbarkeit

- CILS4NEPS Webseite unter www.neps-data.de:
 - NEPS Daten → Startkohrte Klasse 9 → CILS4NEPS
- Zugang über RemoteNEPS Gateway
 - Moderat anonymisierte Daten
- Antragstellung bei NEPS/LIfBi und CILS4EU/Gesis
 - Gemeinsame Begutachtung
 - Notwendigkeit der harmonisierten Daten







CILS4NEPS – Zusammenfassung

- Harmonisierte Daten ermöglichen erweiterte Auswertungen
 - o u.a. zu Übergängen in Bildung und Arbeitsmarkt
- Erhöhtes Analysepotential zur Untersuchung von Subgruppen
 - o v.a. Zugewanderten und ihren Nachkommen, Bildungsungleichheit
- Internationaler Vergleich mit England, Niederlanden und Schweden
- Leichter Zugang für bereits bestehende Nutzendengruppen beider Datensätze





CILS4NEPS – Ausblick

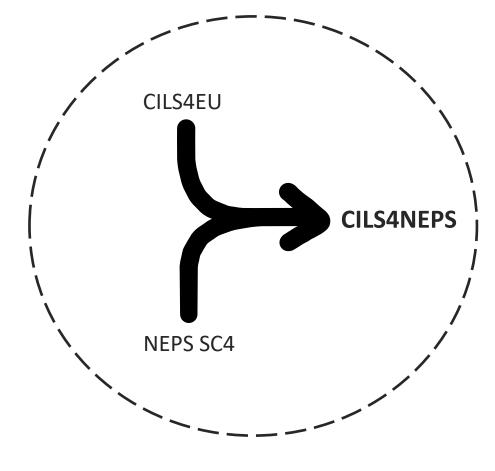
- Veröffentlichung der Daten in Kürze
- Workshop

Mögliche Erweiterungen:

- Wellen 4 bis 9 & Covid-Welle
- Weitere Befragtengruppen: Eltern, Lehrer
- Kompetenztest
- Verwendung von Standardvariablen aus anderen Harmonisierungsprojekten











- Lena Arnold, <u>lena.arnold@uni-mannheim.de</u>
- Andreas Horr, <u>andreas.horr@lifbi.de</u>



CILS4NEPS wird Rahmen der Projektförderung Forschungsdatenmanagement durch das Konsortium für die Sozial-, Verhaltens-, Bildungs-, und Wirtschaftswissenschaften (KonsortSWD) gefördert





Literatur

Granda, P., Wolf, C. and Hadorn, R. (2010). Harmonizing Survey Data. In Harkness, J. A., Braun, M., Edwards, B., Johnson, T. P., Lyberg, L. E., Mohler, P. P., Pennell, B.-E. and Smith, T. W. (Eds.). *Survey Methods in Multinational, Multiregional, and Multicultural Contexts.* Hoboken, New Jersey: Wiley, pp. 315–334.

Hoffmeyer-Zlotnik, J. H. (2008). Harmonisation of demographic and socio-economic variables in cross-national survey research. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique*, **98**, 5–24.

Singh, R. K. (2021). Harmonizing Data in the Social Sciences with Equating. In Wolbring, T., Leitgöb, H. and Faulbaum, F. (Eds.). *Sozialwissenschaftliche Datenerhebung im digitalen Zeitalter. Schriftenreihe der ASI - Arbeitsgemeinschaft Sozialwissenschaftlicher Institute.* Wiesbaden: Springer VS, pp. 123–140.

Wolf, C., Schneider, S. L., Behr, D. and Joye, D. (2016). Harmonizing survey questions between cultures and over time. In Wolf, C., Joye, D., Smith, T. and Fu, Y.-c. (Eds.). *The SAGE Handbook of Survey Methodology*. London: SAGE Publications Ltd, pp. 502–524.





CILS4NEPS - Spezialfälle

- Generationenstatus und Ethnische Herkunft
- Panel-Struktur harmonisierte Wellenstruktur
- Harmonisierte Gewichte
- Länderlisten





CILS4NEPS - Wellenstruktur

	2010/11		20:	11/12	2012/13	
NEPS SC4	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6
CILS4EU	Wave 1		Wave 2		Wave 3	
CILS4NEPS						
H_wave	Wave 1	Wave 2 (NEPS only)	Wave 3	Wave 4 (NEPS only)	Wave 5	Wave 6 (NEPS only)
H_wave2	Wave 1	-	Wave 2	-	Wave 3	-





CILS4NEPS – Missing Schema

Dataset	Value	Label	Recoding
NEPS SC4	-5/-6/-20,,-29	Item-specific missing with	
		informative value label	
CILS4EU	-33	Not available in RV	
CILS4EU	-44	Interrupted interview	
NEPS SC4	-52	Implausible value removed	
NEPS SC4	-53	Anonymised	
NEPS SC4	-54	Missing by design	
CILS4EU	-55	Other missing	
NEPS SC4	-56	Not participated	
NEPS SC4	-57	Not determinable	Previously -55 in NEPS SC4
CILS4EU	-66	Question not asked	
CILS4EU	-77	Not applicable	
CILS4EU	-88	No answer	
NEPS SC4	-90	Unknown missing	
NEPS SC4	-92	Question erroneously not asked	
NEPS SC4	-93	Does not apply	
NEPS SC4	-94	Not reached	
NEPS SC4	-95	Implausible value	
NEPS SC4	-97	Refused	
NEPS SC4	-98	Filtered	Previously -99 in NEPS SC4
Combined	-99	Don't know	Previously -98 in NEPS SC4
NEPS SC4		Filtered/system missing	





CILS4NEPS – Harmonisierte Gewichte

Würbach and Aßmann (2022):

w t CILS4NEP	Nonresponse adjusted joint panel entry weight for targets with					
<u> </u>	itomesponse adjusted joint parier entry weight for targets with					
	panel consent (unstandardized)					
w_t_CILS4NEPS_cal	Calibrated nonresponse adjusted joint panel entry weight for					
	targets with panel consent (unstandardized)					
w_t1_CILS4NEPS	Cross-sectional weight for targets participating in wave 1					
	(unstandardized)					
w_t_CILS4NEPS_std	Nonresponse adjusted joint panel entry weight for targets with					
	panel consent (standardized)					
w_t_CILS4NEPS_cal_std	Calibrated nonresponse adjusted joint panel entry weight for					
	targets with panel consent (standardized)					
w_t1_CILS4NEPS_std	Cross-sectional weight for targets participating in wave 1					
	(standardized)					





Auszug Überblickstabelle

CILS4EU

Bereich	Indicator	Variable name	Question	Categories	Students Wave	Parents Wave	Students Wave	Students Wave 3
		General information						
General information		Sex			l X	X	X	Х
		sex	Students wave 1: Are you a boy or a girl? Parents wave 1: Are you male or female?	Boy (Parents wave 1: Male) Girl (Parents wave 1: Female)	X	X	X	X
General information		Age/date of birth			Х	х	х	Х
General information	Date of birth, day	dobd (parents wave 1: dobd_nl)	When were you born?		(X)	(X)	(X)	Х
General information	Date of birth, month	dobm (parents wave 1: dobm_nl)	When were you born?		Х	(X)	(X)	Х
General information	Date of birth, year	doby	When were you born?	19	Х	Х	Х	Х
General information		Relationship to child			-	Х		
	,	reltc	What is your relationship to your child?	Biological mother or adoptive mother Biological father or adoptive father Stepmother Stepfather Another female guardian Another male guardian		х		
General information		Marital status (parents)		_		Х		
General information	Marital status	marst1	What is your marital status?	Single Married Divorced Separated Widowed		х		





Auszug Überblickstabelle

NEPS SC4

			Wa	ve 1	Wave	2	Wave 3		Wave 4
							Graduates	Pupils	Graduates
			Regular schools	Special schools	Regular schools	Special schools		Regular schools	
Variable name	Question	Categories	First-time respondent 🔻	First-time respondent *	First-time respondents	First-time responden	First-time respondents	First-time respondents	Panel respondents 🔻
General information									
Sex									
t700031	Are you	male? female?	x	x	x	x	X (t700001)	x	X (t700001)
Age/date of birth									'
gebt	First, please tell me your date						х		х
t70004m	When were you born?		Х	Х	Х	Х	X (t70000m)	х	X (t70000m)
t70004y	When were you born?		Х	х	Х	Х	X (t70000y)	Х	X (t70000y)
Relationship to child									
p731701	What is your relationship to <name child="" of="" target="">?</name>	birth mother biological father adoptive mother adoptive father foster mother foster father Partner of father Partner of mother Stepmother Stepfather Other relationship	·				·	·	·
Marital status (parents) p731110	Are you currently	married and living with your spouse, married and living separately from your spouse, divorced widowed,	·		·	·	·		





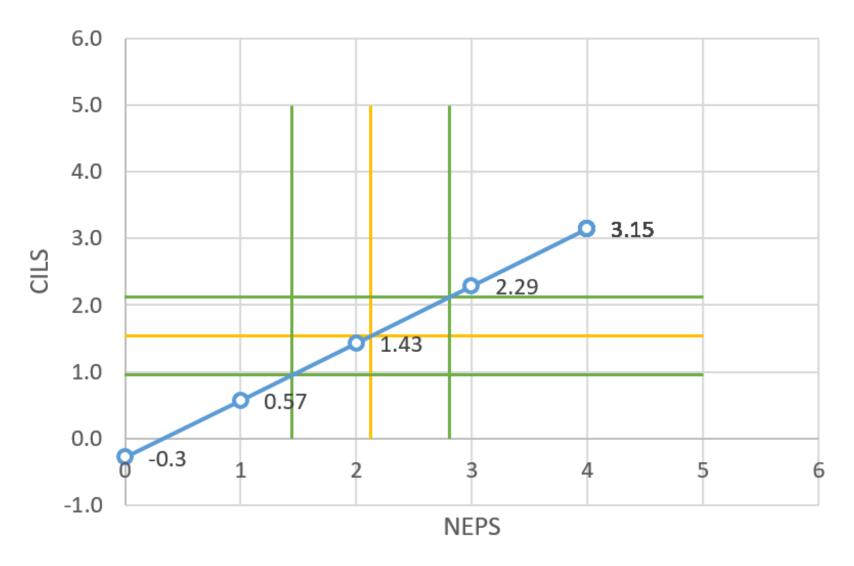
Auszug Linear Equating

H_sesch				
CILS			NEPS	
Value	Count	MATCHED	Count	Value
0	0	0.00	0	0
1	2508.8	0.57	2437.52	1
2	2338.08	1.43	8711.83	2
3	117.2	2.29	3695.52	3
4	21.71	3.15	276.23	4
5	8.86	0.00	0	5
6	0	0.00	0	6
7	0	0.00	0	7
8	0	0.00	0	8
9	0	0.00	0	9
10	0	0.00	0	10
Total	4994.65		15121.1	Total





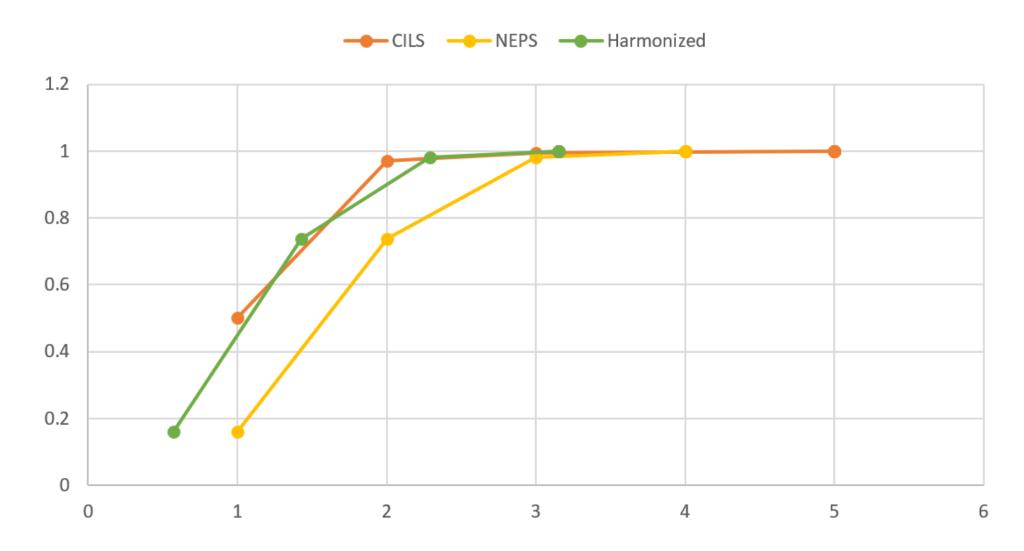
Auszug Linear Equating







Auszug Linear Equating







Content Sheet	Item	Harmonization
		Strategy
	Unique Identifier	-
Sample Information	Harmonised Wave Indicator – 6-waves	-
	Harmonised Wave Indicator – 3-waves	-
Tracking and Filter	Dataset Information	-
Variables	Filter: Participation in all waves	-
	Sex (Indicator)	Matching
General Information	Date of birth, month	Matching
	Date of birth, year	Matching
	Household members: Mother (biological, adoptive,	Matching
	foster)	
	"": Father (biological, adoptive, foster)	Matching
	"": Stepmother	Matching
Household Situation	"": Stepfather	Matching
& Composition	"": Siblings and/or stepsiblings	Matching
	"": Grandparents	Matching
	"": Other family	Matching
	"": Other persons	Matching
	"": Household size	Matching
	Mother's education (broad)	Matching
	Father's education (broad)	Matching
Social Rackovound	Mother employment status	Matching
Social Background	Father employment status	Matching
	Mother's occupation: ISCO-08	Matching
itrum Sozialjoracijang	Father's Occupation: ISCO-08	Matching

	Born in Survey Country	Matching
	Age at migration	Matching
	Mother born in survey country	Matching
Migration History	Country of birth, mother	Matching
	Father born in survey country	Matching
	Father's country of birth	Matching
	Stay tendency	Linear Equating
	Subjective school performance: Math	Linear Equating
	"": Survey country language	Linear Equating
Cahool Daufoumanaa	Grade: Math	Matching
School Performance	"": German	Matching
	"": English	Matching
	Repeated school year	Matching
	Idealistic educational aspirations	Matching
	Realistic educational aspirations	Matching
	Parents' educational aspirations	Matching
Attitudes towards	Self-efficacy: Do well at school	Linear Equating
School	"": Good grades at school	Linear Equating
School	Status attainment compared to parents	Linear Equating
	Help from teacher	Linear Equating
	Success probability at: Lower secondary school	Linear Equating
	"": Intermediate secondary school	Linear Equating





	Realistic aspiration after school year	Matching
Future Plans	Application status: job/apprenticeship/internship	Matching
r uture Puins	Idealistic occupational aspirations ISCO08	Matching
	Altruism as job selection criterion	Linear Equating
	Side job / <u>mini-job</u> / part-time job	Matching
Economic Situation	Monthly pocket money	Matching
Economic Situation	Own a computer	Matching
	Own room	Matching
	Graduation grades: Overall grade	Matching
	Current grade at school	Matching
	Current school type	Matching
Current Situation	Apprenticeship ISCO 2008	Matching
	Apprenticeship-module: Success probability	Linear Equating
	Vocational preparation number of applications	Matching
	Job/apprenticeship search	Matching
	Currently dating/partnership	Matching
Romantic	Migration background partner	Matching
Relationships	Partnership Duration	Matching
	Children in Future	Linear Equating
	Parental encouragement school achievement	Linear Equating
Family Relations	Family communication: Frequency of discussing	Linear Equating
rumity Ketations	political/social issues	
	"": Frequency of discussing movies/ TV shows	Linear Equating





Language	Language Proficiency Survey Country: Speaking	Linear Equating
	"": Comprehend	Linear Equating
	"": Read	Linear Equating
	"": Write	Linear Equating
	Second Language	Linear Equating
	Language Proficiency Second Language: Speak	Linear Equating
	"": Comprehend	Linear Equating
	"": Read	Linear Equating
	"": Write	Linear Equating
	"": Speaking	Linear Equating
Identity	Host country identification	Linear Equating
	Ethnic identification	Linear Equating
Religion	Religious Affiliation	Matching
	Religiousness	Linear Equating
Leisure Time Activities	Club membership	Matching
	Frequency of computer games on school day	Linear Equating
	Number of books	Matching
Well-Being & Health	Life satisfaction	Linear Equating
	School satisfaction	Linear Equating
	Self-esteem: good qualities	Linear Equating
	Subjective general health	Linear Equating
	Height in cm	Matching
	Height in cm (reduced version)	Matching
	Weight in kg	Matching





	Weight in kg (reduced version)	Matching
	Sports frequency	Linear Equating
Generation & Ethnic Origin - Generated	Generation status up to 3.75th generation (NEPS SC4 approach)	Linear Equating
	Generation status up to 3.5th generation (NEPS SC4 approach)	Linear Equating
	Group of origin (NEPS SC4 approach)	Linear Equating
	Missing/contradicting info: generation status (NEPS SC4 approach)	Linear Equating
	Missing/contradicting info: generation status v1	Linear Equating
	Generational status (generated) (CILS4EU approach)	Linear Equating
	Flag missing cases (generational status) (CILS4EU approach)	Linear Equating
	Flag non-trivial cases (generational status) (CILS4EU approach)	Linear Equating
	Country of origin (CILS4EU approach)	Linear Equating
	Country of origin (Germany) (CILS4EU approach)	Linear Equating
	Flag missing cases (country of origin) (CILS4EU approach)	Linear Equating
	Flag non-trivial cases (country of origin)	Linear Equating



