

Universität Leipzig

Fakultät für Sozialwissenschaften und Philosophie

Institut für Soziologie

Bachelorarbeit

Replikation von Rutkowski & Rutkowski (2013):

Measuring Socioeconomic Background in PISA: One size might not fit all

vorgelegt von

Patrick Hawlitschek

[REDACTED] April 1997 in Quedlinburg

Gutachter: Professor Dr. Roger Berger

Kontakt: **[REDACTED]@leipzig.de**

Studiengang: Bachelor of Arts, Soziologie

Leipzig, 4. August 2021

Inhaltsverzeichnis

Tabellenverzeichnis.....	3
Abbildungsverzeichnis	3
Abkürzungsverzeichnis	4
1. Einleitung	5
2. Theorie und Forschungsstand.....	7
2.1 Der Zusammenhang zwischen Schulleistung und sozialer Herkunft	7
2.2 Wie PISA die soziale Herkunft weltweit definiert	10
2.3 Konstruktäquivalenz und Messinvarianz.....	12
2.3 Vorstellung der Originalstudie (Rutkowski & Rutkowski, 2013).....	15
2.4 Ziel der Replikationsstudie und Fragenstellungen	16
3. Methode.....	17
3.1 Daten, Stichprobe und Prozedur	17
3.2 Operationalisierung	18
3.3 Statistische Analysen	20
4. Ergebnisse	23
4.1 Reliabilität	23
4.2 Konstruktvalidität	24
4.3 Kulturelle Vergleichbarkeit	24
5. Diskussion	27
Literaturverzeichnis.....	31
Appendix	34

Tabellenverzeichnis

Nummer	Titel	Seite
1	Manifeste Indikatoren zur Erfassung von Besitz an Kultur- und Wohlstands- gütern	19
2	Deskriptive Statistiken aus den konfirmatorischen Faktoranalysen (Skalenebene, n = 80)	25
3	Ergebnisse der Messinvarianzprüfung (Multigruppen-Faktoranalysen)	26
A1	Ergebnisse der Konsistenzprüfung (Cronbachs Alpha) für jedes Bildungssystem	34
A2	Ergebnisse der konfirmatorischen Faktoranalysen für CULTPOSS	36
A3	Ergebnisse der konfirmatorischen Faktoranalysen für WEALTH	38
A4	Ergebnisse der konfirmatorischen Faktoranalysen für HEDRES	40
A5	Ergebnisse der konfirmatorischen Faktoranalysen für ICTRES	42
A6	Ergebnisse der konfirmatorischen Faktoranalysen für HOMEPOS	44

Abbildungsverzeichnis

Nummer	Titel	Seite
1	Modell eines formativen Index (links) und eines reflektiven Index (rechts).	11
2	Allgemeine Repräsentation eines latenten Variablenmodells (eindimensional und kongenerisch)	14
3	Verteilung der Alphawerte auf Skalenebene getrennt nach OECD- und Partnerstaaten (n = 80)	23

Abkürzungsverzeichnis

CFI – Comparative Fit Index

CULTPOSS – Cultural Possessions (Subskala)

ESCS – Gesamtindex des sozioökonomischen und soziokulturellen Status (Index of Economic, Social, and Cultural Status)

FIML – Full Information Maximum Likelihood

HEDRES – Home Educational Resources (Subskala)

HISEI – Höchster elterlicher Berufsstatus (Highest International Socio-Economic Index of Occupational Status)

HOMEPOS – Home Possessions (Gesamtskala)

ICTRES – Information and Communication Technology Resources

OECD – Organisation für wirtschaftlichen Zusammenarbeit und Entwicklung

PARED – Elterliche Bildung (Parental Education)

PISA – Programme for International Student Assessment

RMSEA – Root Mean Square Error of Approximation

SRMR – Standardized Root Mean Square Residual

TLI – Tucker-Lewis-Index

1. Einleitung

Die soziale Herkunft ist eines der zentralen Konzepte in der soziologischen Bildungsforschung. Traditionell erhalten herkunftsspezifische Effekte eine hohe Aufmerksamkeit im Bildungskontext, um Fragen der Bildungsgerechtigkeit und Chancengleichheit zu diskutieren (Ditton & Maaz, 2011). Empirisch erweist sich die soziale Herkunft bzw. der sozioökonomische Status als relevanter Prädiktor für das Lern- und Leistungsverhalten von Schülerinnen und Schülern (Liu et al., 2020; Sirin, 2005; White, 1982). Jedoch variieren die beobachteten Effekte stark mit den eingesetzten Messinstrumenten (Sirin, 2005). Folgerichtig versuchen international angelegte Schulleistungsstudien, zum Beispiel die *Programme for International Student Assessment* (PISA), die administrierten Instrumente einheitlich zu gestalten.

Allerdings birgt dieser Ansatz neue methodologische Herausforderungen. Damit statistische Kennwerte internationale Vergleichbarkeit besitzen, also Bildungssysteme in latenten Merkmalen ihrer Schülerinnen und Schüler valide vergleichbar sind, müssen die theoretischen Konstrukte in allen Kontexten inhaltlich, operational und messmethodisch äquivalent sein. Das Konzept der Messinvarianz bezeichnet, dass ein spezifiziertes Messmodell aus manifesten und latenten Variablen in verschiedenen Teilpopulationen dasselbe Konstrukt erfasst. Höhere Werte in latenten Variablen können erst dann auf Länderunterschiede (Unterschiede im „wahren Wert“) zurückgeführt werden, wenn das eingesetzte Messmodell invariant zur Gruppenzugehörigkeit ist. Neuere Studien zeigen für ausgewählte Konstrukte, dass dies insbesondere in internationalen Studien nicht oder nur mit Einschränkung der Fall ist (Lee & Davier, 2020; Zhou et al., 2020; Lee, 2019; Asil & Brown, 2016; Rutkowski & Rutkowski, 2013).

Ungeachtet dieser Forschung werden die Ergebnisse aus PISA medial breit rezipiert und liefern Anlass zur Diskussion. Seit 2000 sammelt die Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) in nunmehr 80 Bildungssystemen Leistungsdaten 15-jähriger Schülerinnen und Schüler, die weitreichende Impulse für Bildungsreformen waren und potenziell sein können (Gorur, 2014). Obwohl das primäre Anliegen die Erfassung kognitiver Fähigkeiten darstellt, werden gleichsam sozioökonomische Indikatoren erhoben und zu einem PISA eigenen Gesamtindex des sozioökonomischen und soziokulturellen Status (Index of Economic, Social, and Cultural Status, ESCS) verrechnet (Avvisati, 2020; OECD, 2019b; Pokropek et al., 2017). Der ESCS ist das Ergebnis einer Hauptkomponentenanalyse aus den Variablen des elterlichen Berufsstatus, der elterlichen Bildung und dem Besitz an Kultur- und Wohlstandsgütern. Neben den beiden Einzelindikatoren, der Berufsstatus (HISEI) und die elterlichen Bildung (PARED), beschreibt der Besitz an Kultur- und Wohlstandsgütern ein latentes Merkmal, das mit 25 Item

operationalisiert wird. In einer Studie basierend auf Daten aus PISA 2009 konnten Rutkowski & Rutkowski (2013) zeigen, dass die Reliabilität der eingesetzten Skala in vielen Ländern nicht zufriedenstellend ist. Außerdem berichten sie, dass das theoretische Messmodell eine ungenügende Passung zu den empirischen Daten in einem Großteil der Länder aufweist. In einem letzten Schritt können sie offenlegen, dass die Subskalen nicht messinvariant gegenüber der Länderzugehörigkeit sind.

Das PISA Konsortium hat die Ergebnisse der Studie rezipiert (OECD, 2019b). Es fanden in nationalen und internationalen Gremien fortlaufend Überarbeitungen statt und auch neue Items kamen hinzu. Dieser Umstand bietet Anlass, mit neuen Daten aus PISA 2018 die Ergebnisse von Rutkowski & Rutkowski (2013) zu replizieren. Hinzu kommt, dass der Index nach eigenen Aussagen keine benennbare theoretische Fundierung aufweist, sondern eher operational im technischen Bericht konstruiert wird (Avvisati, 2020). Die statistische Absicherung des Instruments für ihren Einsatz im Sinne einer theoretisch motivierten Sozialforschung scheint daher noch dringender.

Die vorliegende Bachelorarbeit sichtet zunächst empirische Befunde zum Zusammenhang von sozialer Herkunft und Schulleistung. Anschließend wird die PISA eigene Auswertungsstrategie vorgestellt und die messtheoretischen Konzepte der Konstruktäquivalenz und Messinvarianz eingeführt. Im methodischen Teil dieser Arbeit werden auf Basis von gut 600.000 Schülerinnen und Schüler aus 21.903 Schulen zentrale Ergebnisse zur Reliabilität, Konstruktvalidität und kulturellen Vergleichbarkeit der Skala Besitz an Kultur- und Wohlstandsgütern (HOMEPOS) berichtet. Alle Analyseschritte werden für die Gesamtskala und ihre vier Subskalen, *Cultural Possessions* (CULTPOSS), *Wealth* (WEALTH), *Home Educational Resources* (HEDRES) und *Information and Communication Technology Resources* (ICTRES) durchgeführt. Die Analysen zeigen, dass nur die Gesamtskala eine zufriedenstellende Reliabilität (Cronbachs Alpha) aufweist. Die Konstruktvalidität ist größtenteils nicht sichergestellt und auch die Stufen der Messinvarianz können nicht abgesichert werden. Die Arbeit repliziert damit die Studie von Rutkowski & Rutkowski (2013) mit den Daten aus 2018 inhaltlich nahezu vollständig. Eine abschließende Diskussion benennt mögliche Verbesserungen für die Weiterentwicklung des Instruments.

2. Theorie und Forschungsstand

Die soziale Herkunft ist eine wichtige Ressource für das nationale und internationale Bildungsmonitoring. Seit Beginn der PISA-Studien legt die Berichterstattung großen Wert auf herkunftsspezifische Effekte (Avvisati, 2020; Reiss et al., 2019). Zwar ist das primäre Interesse die Erfassung domänenspezifischer Fähigkeiten, doch gilt es laut OECD weltweit als Herausforderung, allen Mitgliedern einer Gesellschaft Kompetenzerwerb zu ermöglichen und herkunftsbedingte Effekte möglichst gering zu halten (OECD, 2019a, OECD, 2019b; Ehmke & Jude, 2010).

Dafür werden Effekte der sozialen Herkunft auf die Schulleistung in einem international vergleichenden Ansatz aufbereitet und medial breit rezipiert (Kapitel 2.1). Das Auswertungsverfahren erfordert den Einsatz von Instrumenten, die reliabel und valide Aspekte der sozialen Herkunft erfassen (Kapitel 2.2), um bildungspolitische Entscheidungen zu fundieren und die Bildungspraxis zu bereichern. Jedoch bietet das aktuelle Vorgehen Anlass zur Kritik hinsichtlich der postulierten Konstruktäquivalenz und Messinvarianz (Kapitel 2.3), die in empirischen Studien zunehmend geäußert wird (Kapitel 2.4).

2.1 Der Zusammenhang zwischen Schulleistung und sozialer Herkunft

Der Zusammenhang zwischen der Schulleistung von Schülerinnen und Schülern und ihrer sozialen Herkunft (SES) ist ein über Jahrzehnte hinweg replizierbarer und international beobachtbarer Befund in der empirischen Bildungsforschung (Liu et al., 2020; Sirin, 2005; White, 1982). Metaanalysen berichten eine breite Streuung der Effektgrößen mit durchweg positiven Kennwerten ($.005 \leq r \leq .77$; Sirin, 2005). Auf Schülerebene liegt der Median mit .22 (White, 1982), .24 (Sirin, 2005) und .24 (Liu et al., 2020) im moderaten Bereich. Angesichts dessen ließe sich fragen, warum die Sozialwissenschaften ein derartig großes Interesse an dem Zusammenhang von Schulleistung und SES pflegen.

Das Interesse besteht mutmaßlich darin, dass Bildungssysteme bzw. die zugehörigen Bildungseinrichtungen (normativ) kompensatorische Arbeit leisten sollen – so genannte leistungsfremde soziale Filter werden als Gefahr für die Mobilität innerhalb der Gesellschaft betrachtet. Diverse soziologische und sozialpsychologische Theorien versuchen, diesen Umstand zu erklären (z. B. die Theorie der sozialen Reproduktion; Bourdieu, 1982). Daran anschließende Forschung untersucht beispielsweise die Urteilstendenzen der Lehrkräfte (Gentrup et al., 2020; Schneider, 2011), den Einfluss der sozio-ökonomischen Klassenkomposition (Rjosk et al., 2014; Reid &

Ready, 2013) oder die herkunftsbezogene Bildungsaspiration (Astleithner et al., 2021) auf den Bildungserfolg.

Das schlichte Vorhandensein eines herkunftsbedingten Effekts wird in der Medienberichterstattung mitunter per se als ein Signal der Ungerechtigkeit interpretiert. Jedoch ist aus methodologischer Sicht relevant, dass der Effekt mit den Operationalisierungen (z. B. Bildung, Berufsstatus oder Einkommen) oder hinzugenommenen Drittvariablen (z. B. Vorerfahrung) variiert (Sirin, 2005). Folglich ist die theoriegeleitete und valide Erfassung des SES eine unabdingbare Voraussetzung für darauf aufbauende Interpretationen oder politische Entscheidungen. In diesem Kontext verfolgt PISA eine eigene Strategie, indem die Variablen der Schulleistung und der sozialen Herkunft selbst definiert und operationalisiert werden. Mit diesem Vorteil, das gleiche Set an Variablen in vielen Bildungssystemen bereitzustellen, möchte man die Bildungspolitik bereichern.

Schulleistung ist im Rahmen von PISA ein irreführender Begriff in der deutschen Medienlandschaft, impliziert er doch die Messung von schulunterrichtlichen Inhalten. Allerdings wird in der international angelegten Studie ein selbst erstelltes Curriculum abgeprüft, dass – nach eigenen Aussagen – grundlegende Kompetenzen für das Bestehen in der modernen Welt abdeckt (OECD, 2019b). Es lässt sich also besser von *Testleistung* sprechen. Die Testaufgaben erlauben in der Form keine kriteriale Kompetenzbeschreibung für die deutschen Bildungsstandards. Das ist ein Unterschied zum deutschen Bildungsmonitoring des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB), das auf den nationalen Bildungsstandards der Kultusministerkonferenz basiert. Mit einer kriterialen Kompetenzmessung können herkunftsbedingte Aussagen hinsichtlich der erreichbaren Bildungsabschlüsse getroffen werden lassen (Henschel et al., 2019).

Die Auswertungsstrategie von PISA 2018 betrachtet Zusammenhänge innerhalb eines Landes (nationale Berichterstattung) und zwischen Ländern (internationale Berichterstattung). In knapper Form sollen nachfolgend grundlegende Einsichten vermittelt werden, wie PISA den Zusammenhang von Testleistung und sozialer Herkunft auswertet.

PISA verwendet einen eigenen *Index des sozioökonomischen und soziokulturellen Status* (ESCS), der aus der elterlichen Bildungsdauer, dem Berufsstatus und den familiären Besitztümern gebildet werden (vgl. Kapitel 2.2). Messwerte werden über alle OECD-Staaten hinweg standardisiert. Die Metrik weist einen Mittelwert von 0 mit einer Standardabweichung von 1 für den OECD-Durchschnitt auf (OECD, 2019b). In der univariaten Auswertung werden die Mittelwerte des 95%- und 5%-Perzentils eines Bildungssystem berichtet. Die Spannweite gebe Auskunft darüber, wie groß die Heterogenität innerhalb des Landes ist. Deutschland liegt leicht

unterhalb des OECD-Durchschnitts, das heißt der Abstand zwischen den oberen und unteren 5% der Daten fällt größer als in anderen Ländern aus. In dieser Auswertungsstrategie wird davon gesprochen, dass die soziale Ungleichheit stärker ausgeprägt ist. Die engste Konzentration der Daten zeigt sich in Russland, Japan und Belarus. Die breiteste Streuung weisen Kolumbien, Panama und Marokko auf (OECD, 2019b).

In der weiteren Auswertung wird eine lineare Regression zwischen der Testleistung im Lesen (Lesekompetenz) und dem ESCS geschätzt (Reiss et al., 2019).¹ Die drei Kennwerte *Achsenabschnitt*, *Steigung* der Regressionsgeraden und die *Varianzaufklärung* (R^2) dienen zur Evaluation. Der Achsenabschnitt beschreibe die durchschnittlich erreichte Lesekompetenz in Punkten, die Schülerinnen und Schüler eines Landes mit durchschnittlicher sozialer Herkunft erreichen. Die Skalierung der Kompetenz sieht einen Mittelwert von 500 und eine Standardabweichung von 100 vor. Der OECD-Durchschnitt im Achsenabschnitt liegt bei 489 Punkten. Spitzenreiter sind Estland (522), Deutschland (510) und Finnland (510). Das Schlusslicht bilden Israel (458), Mexiko (451) und Kolumbien (444).

Die Steigung interpretiert PISA als sozialen Gradienten, der aussagt, um welche Punktzahl sich die Lesekompetenz vergrößert, sobald sich der elterliche Status um eine Standardabweichung erhöht. Wenn der elterliche Status (ESCS) um eine Standardabweichung erhöht wird, dann fällt die Lesekompetenz im Durchschnitt aller Länder um 37 Punkte höher aus. Für Deutschland liegt dieser Wert bei 41.9 Punkten – einer der höchsten unter den OECD-Staaten. Noch höhere Steigungswerte weisen die Slowakei (45.6), Ungarn (45.8) und Frankreich auf (46.9). Estland besitzt den niedrigsten sozialen Gradienten mit 28.7 Punkten. Als ein weiteres Evaluationskriterium wird die Varianzaufklärung berichtet, um die Stärke des Zusammenhangs zwischen Testleistung und ESCS zu quantifizieren. Je besser der elterliche Status Varianz in der Lesekompetenz eines Landes erklärt, desto stärker sei der herkunftsbedingte Effekt zu bemessen. Erneut liegt der OECD-Durchschnitt (12 %) unter dem Niveau von Deutschland (17.2 %). Die Varianzaufklärung ist in Ungarn (19.1 %) am höchsten und in Estland (6.2 %) am niedrigsten.

Auf Basis der statistischen Kennwerte kann resümiert werden, dass der elterliche Status in den Ländern einen unterschiedlich großen Einfluss spielt. Einige der Länder sind insgesamt weniger sozial ungleich (bemessen an der Spannweite der oberen und unteren 5% der Daten) und

¹ Im aktuellen Erhebungsjahr stand die Domäne Lesen im Fokus, sodass die Zusammenhänge mit der Mathematikleistung ausgespart werden.

variieren in der Stärke des herkunftsbezogenen Einflusses (Reiss et al., 2019). Ähnlich zu allen Metaanalysen (Liu et al., 2020; Sirin, 2005; White, 1982) zeigt sich ein durchweg positiver Effekt auf die Leistung. Da PISA überall die gleichen Tests administriert, sollte der Zusammenhang frei von Unterschieden in der Operationalisierung geschätzt werden können. Doch genau an diesem Ansatz – der validen Erfassung der sozialen Herkunft mit einem einheitlichen Standardinstrument – wird immer stärker Kritik geübt (Pokropek et al., 2017; Rutkowski & Rutkowski, 2013).

2.2 Wie PISA die soziale Herkunft weltweit definiert

Das internationale Konsortium entwickelte ein Standardinstrument, mit dem die soziale Herkunft möglichst breit erfasst wird (OECD, 2019b). Der Messwert entspricht der ersten unrotierten Hauptkomponente des höchsten beruflichen Status der Eltern (HISEI), der höchsten elterlichen Bildung in Jahren (PARED) und dem Besitz an Kultur- und Wohlstandsgütern (HOME-POS). Der so ermittelte *Gesamtindex des sozioökonomischen und soziokulturellen Status* (Index of Economic, Social, and Cultural Status, ESCS) wird definiert als

a measure of students' access to family resources (financial capital, social capital, cultural capital and human capital) which determine the social position of the student's family/household. (Avvisati, 2020, S. 3; OECD, 2019a)

Obwohl dem Index keine ausformulierte Theorie zugrunde liegt und lediglich operational im technischen Bericht definiert wird (OECD, 2019a), können zentrale Programmatiken aus der soziologischen Theoriegeschichte identifiziert werden. Die soziale Herkunft wird als ein mehrdimensionales Konstrukt betrachtet, das sich an soziologischen Klassen- und Schichtkonzepten orientiert. Einflussreiche Vertreter sind an dieser Stelle Marx und Weber sowie die daran knüpfenden Forschungslinien (Ditton & Maaz, 2011). Über die materiellen und nicht-materiellen Besitzverhältnisse wird die relative Position und Teilhabe der Schülerinnen und Schüler in der Gesellschaft geschätzt, die an ökonomische und kulturelle Ressourcen sowie Interessen gebunden sei. Mit der Überführung in einen einzelnen Kennwert, dem ESCS, wird eine größere (statistische) Vorhersagekraft vermutet als die darin eingehenden drei Facetten allein (Ehmke & Jude, 2010).

Die Zusammenfassung der drei Komponenten, *Bildung*, *Berufsstatus* und *Besitz* lässt sich als eine theoretisch begründbare Auswertungsstrategie dreier, teilweise inhaltlich unabhängiger Einzelindikatoren auf Individualebene betrachten. Mehrdimensionale Modelle werden bei der

Operationalisierung der sozialen Herkunft bevorzugt (Brese & Mirazchiyski, 2013; McLoyd, 1998). Ein illustrierendes Beispiel dafür ist der hochgebildete Akademiker, der befristet über die prestigereiche Beschäftigung in der universitären Forschung kaum Eigentum akkumulieren kann. Man spricht deshalb von einer formativen Indexbildung, weil die Indikatoren die latente Variable – in diesem Fall die soziale Herkunft – bedingen (Latcheva & Davidov, 2019). Abbildung 1 zeigt das Modell einer formativen Indexbildung (links).

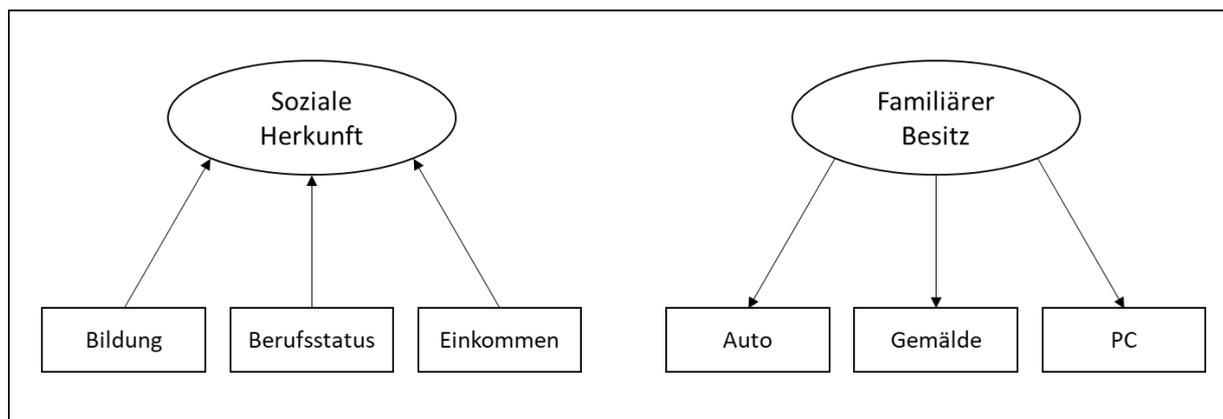


Abbildung 1: Modell eines formativen Index (links) und eines reflektiven Index (rechts).

In dem angeführten Modell gibt die Variable *Bildung* Auskunft über die verbrachten Jahre in einem Bildungssystem als ein Indikator für erworbenes Humankapital oder die an Zertifikate geknüpften Zugangsberechtigungen (Hoffmeyer-Zlotnik & Warner, 2019). Der Berufsstatus ist ein Maß für das berufliche Prestige einer ausgeübten Tätigkeit innerhalb der Gesellschaft. Dieses wird über eine rangbasierte Klassifikation aller beruflichen Tätigkeiten erreicht (Hoffmeyer-Zlotnik & Warner, 2019) und auf eine internationale Metrik gebracht (ISEI, Standard International Socio-Economic Index of Occupational Status; Ganzeboom & Treimann, 1996). Statt *Besitz* wird in der Ungleichheitsforschung als dritte Komponente vor allem das *Einkommen* betrachtet, das Aufschluss über die regelmäßig verfügbaren monetären Ressourcen gibt.

Messmethodisch birgt die direkte Einkommensabfrage einige Hindernisse für internationale Schulleistungsstudien. Lee (2019) berichtet, dass bei PISA 2015 in einigen Ländern die Frage nach dem Einkommen erhoben wurde und die *item response rate* lediglich bei 59 Prozent lag, obwohl bereits eine Kategorisierung in sechs Antwortalternativen vorlag. Generell besteht der Konsens, dass das Einkommen eine sensible Frage darstellt, die nur wenige Befragte, teils unabhängig ihres Einkommens, gewillt sind zu beantworten (Warner, 2009). Hinzu kommt der Umstand, dass Schülerinnen und Schüler Auskunft über das elterliche Einkommen geben müssen, wodurch sich die Wahrscheinlichkeit für Messungenauigkeiten erhöht (Lee, 2019). Der *familiäre Besitz* fungiert dergestalt als ein Proxy, der sich für den Einsatz in Schule eignet. Die

Abfrage von zählbaren Einheiten wie Autos oder Bädern ist, so wird in der Literatur vermutet, mit einer gesteigerten Genauigkeit verbunden (Lee, 2019).

Doch inhaltlich sind die Konzepte Einkommen und Besitz nicht deckungsgleich. Das Einkommen erfasst eine fluktuierende Größe, die monatliche bzw. jährliche Finanzmittel aus der Erwerbstätigkeit und anderen Einkommensquellen (etwa Aktien, Unterhalt oder Vermietung) einschließt. Besitz ist in den meisten Fällen eine Folge des akkumulierten Einkommens, kann aber auch über Erbe unabhängig davon sein. Aus theoretischer Perspektive werden Besitzentscheidungen (z. B. Erwerb eines Zweitautos) auf Basis des bereits vorhandenen Besitzes, der Haushaltstruktur (z. B. Anzahl der im Haushalt lebenden Personen) und der Möglichkeit, sich weiteren Besitz zu leisten (z. B. Einkommen) getätigt.

Wenn PISA den familiären Wohlstand und Besitz erfassen möchte, bedienen sie sich also nicht der direkten Frage nach dem monetären Vermögen oder Einkommen. Stattdessen konstruiert PISA einen reflektiven Index, der eine latente Variable mit manifesten Indikatoren in ein spezifiziertes Messmodell bringt. Messtheoretisch spiegeln die manifesten Indikatoren (eckig in Abbildung 1) die theoretisch angenommene Variable (rund) wider, indem sie von ihr verursacht oder beeinflusst wurden. Dadurch hängen die Indikatoren untereinander eng zusammen und weisen im formalen Sinne hohe Korrelationen miteinander auf. In der Skalenkonstruktion geht es darum, Items zu entwickeln, die das Konstrukt inhaltssvalid abdecken und dennoch empirisch abgrenzbar von anderen Konstrukten sind. Folglich setzen solche reflektiven Messmodelle formale Bedingungen voraus, die mit statistischen Methoden bei Vorliegen der empirischen Daten geprüft werden können (Schermele-Engel & Werner, 2012; Kelava & Moosbrugger, 2012).

2.3 Konstruktäquivalenz und Messinvarianz

Sobald PISA Mittelwerte auf Länderebene interpretiert, wie ungleich ein Land im Durchschnitt zu einem anderen Land ist, werden Mittelwertunterschiede auf Varianz im wahren Wert zurückgeführt. Die Anwendung eines einheitlichen Instruments über alle Bildungssysteme hinweg setzt voraus, dass „die Messung des relevanten theoretischen Konstrukts äquivalent ist, d.h. über Kontexte und Zeitpunkte invariant ist und damit das Gleiche misst“ (Latcheva & Davidov, 2019, S. 903).

Für Gruppenvergleiche dieser Art wird stillschweigend angenommen, dass das Modell in allen Ländern gleich ‚gut‘ misst. Mitunter kommt es jedoch vor, dass Konzepte nicht gleich

verstanden werden (*construct bias*), Antwort- und Stichprobenverzerrungen unterschiedlich ausgeprägt sind (*method bias*) oder dass Items in anderen Kontexten anders verstanden werden (*item bias*). Ein Beispiel für differenzielles Itemfunktionieren ist die Abfrage, ob man im Besitz eines internetfähigen Handys ist (vgl. Item 17, Tabelle 1 auf Seite 19). Während noch vor wenigen Jahren der Besitz einen hohen Status anzeigen konnte, kann sich mittlerweile beinahe jeder in Deutschland ein Smartphone leisten. Ähnlich sagt das Vorhandensein einer Klimaanlage mehr über den familiären Wohlstand in Deutschland etwas aus als in anderen, sehr viel heißeren Regionen der Welt, in denen eine Klimaanlage die Regel ist. Immer mehr Studien weisen darauf hin, dass international eingesetzt Instrumente eine eingeschränkte Vergleichbarkeit besitzen könnten (Lee & Davier, 2020; Zhou et al., 2020; Lee, 2019; Asil & Brown, 2016; Rutkowski & Rutkowski, 2013). Bei mangelnder Konstruktäquivalenz lassen sich Messunterschiede nicht eindeutig auf Unterschiede im wahren Wert, sondern ebenso auf methodische Artefakte zurückführen (Latcheva & Davidov, 2019).

Wie sieht das für den Besitz an Kultur- und Wohlstandsgütern aus? In der Messtheorie betrachtet man bei Vorliegen der empirischen Fragebogendaten die operationale bzw. *metrische Äquivalenz* von reflektiven Indizes. Es wird geprüft, ob die manifesten Indikatoren über Subgruppen hinweg gleichwertig sind (ihnen also im theoretischen Sinne die gleiche Bedeutung von den befragten Schülerinnen und Schülern aller Bildungssysteme beigemessen wird). Metrische Äquivalenz (auch: *Messinvarianz*) zeigt sich darin, dass die Daten verschiedener Subgruppen dieselbe latente Variable messen, also dass die multivariaten Verteilungsparameter der abgebildeten Faktoren invariant zur Zugehörigkeit zur Gruppe sind. Formal betrachtet ergibt sich die Gleichung (2):

$$(2) \quad f(Y | \eta, s) = f(Y | \eta)$$

Der beobachtete Skalenwert Y , gegeben dem wahren Wert η und der Gruppenzugehörigkeit s , soll gleich sein dem Skalenwert Y , gegeben dem wahren Wert η . Um dieses Phänomen genauer zu beschreiben, muss man das Messmodell der strukturgleichenden Messtheorie (*structural equation modeling*) betrachten, das zuvor als reflektives Modell eingeführt wurde (Abbildung 1, rechts). Ein solches Messmodell besteht in der Zuweisung von manifesten Indikatoren X_n zu latenten Faktorvariablen Y_n . Das theoretische spezifizierte Modell wird angemessen in ein statistisches Modell übertragen (siehe Abbildung 2).

Die manifesten, theoriebasiert ausgewählten Variablen werden direkt beobachtet, während die latente Variable nur im statistischen Sinne aus den Daten geschätzt wird (Y). Die Antwort der befragten auf ein Fragebogenitem (manifest) wird derart interpretiert, dass sie durch die theoretisch angenommene Variable verursacht wurde. Die latente Modellierung von theoretischen Konstrukten erlaubt die messfehlerbereinigte Schätzung von Zusammenhängen, etwa in Strukturgleichungsmodellen oder Pfadanalysen. Innerhalb des Variablenmodells werden ebenfalls Kennwerte geschätzt. Dabei handelt es sich um den Einfluss des Faktors auf die manifeste Variable λ_n (genannt Faktorladung) und der latenten Fehlervariablen ε_n , woraus die unerklärte Varianz in den Items ($\text{Var}(\varepsilon_n)$) gebildet wird.

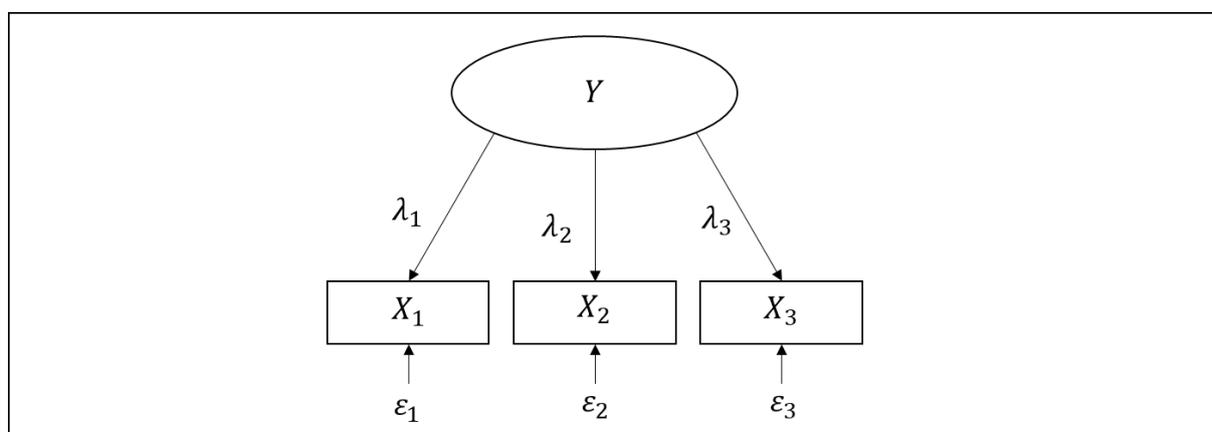


Abbildung 2: Allgemeine Repräsentation eines latenten Variablenmodells (eindimensional und kongenerisch).

Drei Stufen der Messinvarianz stellen die statistischen Voraussetzungen für valide Mittelwertvergleiche über Subgruppen in einem solchen Messmodell dar (Schulte et al., 2013). Diese müssen erfüllt sein, damit ein Instrument zur Erfassung über Subgruppen vergleichend interpretierbar ist (Fischer & Karl, 2019; Gäde et al., 2020).

In der ersten Stufe, der *konfiguralen Invarianz*, geht es auf Basis theoretischer Annahmen (intendierte Konstruktäquivalenz) um die Spezifizierung dergleichen Variablenmodelle in allen Subgruppen. Dabei sollen die Anzahl der Faktoren und das Ladungsmuster zwischen latenten Faktoren und manifesten Indikatoren invariant sein. Salopp gesagt spricht aus der Theorie nichts dagegen, dass die Variablenmodelle überall gleich sind. Die höhere Stufe der *metrischen Invarianz* prüft, ob sich die Ladungen der Faktoren auf die manifesten Variablen zwischen den Subgruppen voneinander unterscheiden. Bei Verletzung dieser Annahme hätte in Subgruppe A die latente Variable einen stärkeren oder schwächeren Einfluss auf den Indikator als in Subgruppe B. Als letzte Stufe führt die *skalare Invarianz* die Restriktion ein, dass auch die Achsenabschnitte (*intercepts*) der Items gleich sind. Inhaltlich wird damit ausgedrückt, dass sich das „das betrachtete Merkmal zwischen den Gruppen auf einer gemeinsamen Skala abbilden

lässt, indem beispielsweise ein gemeinsamer Nullpunkt vorliegt“ (Schulte et al., 2013, S. 106). Denkbar sind hier länderspezifische Unterschiede in der Zustimmungstendenz (Akquieszenz). Sobald die drei Stufen der Messinvarianz in den Daten erfüllt sind, lassen sich Niveauunterschiede in der Zielvariable statistisch abgesichert auf Varianz im wahren Wert zurückführen. Dass genau dies für die Daten aus PISA 2009 nicht zutrifft, zeigt die hier im Fokus stehende Originalstudie von Rutkowski & Rutkowski (2013).

2.3 Vorstellung der Originalstudie (Rutkowski & Rutkowski, 2013)

Die Studie „Measuring Socioeconomic Background in PISA: one size might not fit all“ erschien 2013 im Journal *Research in Comparative and International Education* der Autoren David Rutkowski und Leslie Rutkowski. Mittels Sekundäranalysen untersuchen sie die Passung der Messmodelle der sozialen Herkunft zur empirischen Datenstruktur, für jedes Teilnehmerland und über alle Länder hinweg. Mit ihrer Arbeit stellen sie infrage, dass ein einziges Standardinstrument in der Lage wäre, in so vielen Kontexten reliabel und konstruktvalide den Besitz an Kultur- und Wohlstandsgütern zu erfassen. Dazu wenden sie statistische Methoden der klassischen Testtheorie und des Strukturgleichungsparadigmas an.

Die Konsistenzanalyse zeigt eine hohe Variabilität unter den Bildungssystemen. Die Werte für Cronbachs Alpha liegen für die Subskalen WEALTH ($.57 \leq \alpha \leq .87$), CULTPOSS ($.39 \leq \alpha \leq .69$) und HEDRES ($.41 \leq \alpha \leq .71$) überwiegend nicht im zufriedenstellenden Bereich (gewünscht: $\alpha \geq .70$). Ergebnisse aus konfirmatorischen Faktoranalysen (CFA) deuten auf Fehlspezifikationen in den Messmodellen hin. Die aus der Literatur gewonnenen Cut-offs für Fitindizes (CFI/TLI $\geq .90$; RMSEA/SRMR $\leq .08$; Hu & Bentler, 1999) werden von den wenigsten Ländern erreicht. In der Subskalen *Wealth* erreichen nur Chile, Estland, Litauen und Thailand akzeptable Werte. Bei der Subskala *Home Educational Resources* ist es lediglich Deutschland. Das offenbart, dass die Faktorstruktur nicht nur für einen Großteil der Bildungssysteme fehlspezifiziert ist, sondern die Modelle auch aus theoretischer Sicht unzureichend ausgearbeitet sind.

Neben der länderspezifischen Modellevaluation schätzen Rutkowski & Rutkowski (2013) konfirmatorische Faktoranalysen im Mehrgruppendesign (*multiple-group confirmatory factor analysis*; MG-CFA). Mit der sukzessiven Hinzunahme von Restriktionen werden konfigurale (gleiche Faktorstruktur), metrische (gleiche Faktorladung) und skalare Messinvarianz (gleiche Intercepts) angenommen und in einem Goodness-of-fit-Ansatz die deskriptiven Fitindizes

3. Methode

In diesem Kapitel wird das methodische Vorgehen dargelegt, um die Forschungsfragen zu beantworten. Dies umfasst die Beschreibung der Datengrundlage, der Analytestichprobe und der statistischen Verfahren.

3.1 Daten, Stichprobe und Prozedur

Die Studie nutzt das frei zugängliche Scientific Use File aus dem Repository der OECD.² Alle Analysen werden mit *R* (R, Core Team, 2019) und den Paketen *intsvy* (Caro & Biecek, 2017) und *lavaan* (Roessel, 2012) und *tidyverse* (Wickham, 2019) durchgeführt. Ein voll reproduzierbares R-Skript der Analysen und Abbildungen liegt der Arbeit bei. Aussagen zur Stichprobenzusammensetzung werden dem technischen Bericht (OECD, 2019) entnommen.

Am aktuellen Jahrgang nahmen 79 Staaten (davon 42 Partnerstaaten) in einem zweistufigen Ziehungsdesign teil. Als Zielpopulation wurden alle im Jahr 2002 geborenen 15-jährigen Schülerinnen und Schüler definiert. Zunächst wurde auf Staatenebene nach dem *Probabilities Proportional to Size-Verfahren* (Skinner, 2014) Schulen gezogen. Tatsächlich teilgenommen haben 21.903 Schulen weltweit. Innerhalb der Schulen wurden zufällig Schülerinnen und Schüler (teilweise ungeachtet ihrer Klassenstufenzugehörigkeit) gezogen. Die Erhebung umfasst die Kompetenztests und eine Hintergrundbefragung. Es gibt eine Reihe länderspezifischer Regulierungen und Ausschlusskriterien, wie die Stichprobe ermittelt wurde, auf die hier nicht näher eingegangen werden kann (siehe dazu: *Chapter 4: Sample Design*; OECD, 2019a).

Die finale Stichprobe umfasst 612.004 Schülerinnen und Schüler (49.83 % weiblich) in 59 administrierten Erhebungssprachen aus 21.903 Schulen und 80 Bildungssystemen. 90.18 % der Befragten absolvierten die Studie computerbasiert, während der Rest eine Papierversion ausfüllte. Aufgrund heterogener Beschulungsmechanismen in den Bildungssystemen setzt sich die Stichprobe aus Schülerinnen und Schüler der Klasse 7 (0.75 %), Klasse 8 (3.79 %), Klasse 9 (31.11 %), Klasse 10 (54.75 %), Klasse 11 (7.48 %) und Klasse 12 (0.51 %) zusammen. Hinzu kommen 0.50 %, die keiner Klasse zugeordnet werden, sowie 1.11 % mit ungültigen Antworten. Etwas mehr als die Hälfte der Kinder (51.87 %) kommt nicht aus OECD-Staaten, den so genannten Partnerstaaten. Spanien ist anteilig am stärksten vertreten (35.943 bzw. 5.87 % Schülerinnen und Schüler). Das Schlusslicht bilden Island (3.296 Schülerinnen und Schüler bzw.

² <https://www.oecd.org/pisa/data/2018database/>

0.54 %) sowie das Moskauer Land (2.016 Schülerinnen und Schüler bzw. 0.33 %). Für weitere Informationen zur Stichprobenszusammensetzung auf Länderebene verweise ich auf den technischen Bericht und die zugehörigen Begleitdateien (OECD, 2019a).

3.2 Operationalisierung

PISA setzt den Index des sozioökonomischen und soziokulturellen Status (Index of Economic, Social, and Cultural Status, ESCS) zur Operationalisierung der sozialen Herkunft ein (vgl. Kapitel 2). Dieser kombiniert den beruflichen Status der Eltern (Highest International Socio-Economic Index; HISEI), die elterliche Bildung (Parental Education; PARED) und den Besitz an Kultur- und Wohlstandsgütern (Home possessions; HOMEPOS). Eine psychometrische Skalenanalyse kann nur für HOMEPOS stattfinden, da es sich bei HISEI und PARED um Einzeldikatoren handelt.

Die Erfassung der kulturellen und wohlstandsbezogenen Güter ist Teil der Hintergrundbefragung, die von den Schülerinnen und Schüler selbst ausgefüllt wird. Alle 22 international eingesetzten Items orientieren sich an direkt beobachtbaren Entitäten (z. B. Bäder, Bücher, Autos). Die Instruktion variiert, ob ein bestimmtes Gut im elterlichen Haushalt vorhanden ist („*Welche der folgenden Dinge gibt es bei dir zuhause?*“) oder in welchem Ausmaß ein Gut vorhanden ist („*Wie viele der folgenden Dinge gibt es bei dir zuhause?*“). Operational wird eine Auskunft über den Haushalt gewonnen, in dem die Befragten überwiegend leben und damit unterstellt, dass die Haushaltsmitglieder über die räumliche Nähe prinzipiell Zugang zu diesen Gütern haben. Die Antwort erfolgt über eine dichotome Auswahl (Ja – Nein) oder behält eine vierstufige Rangskala (Keine – Ein – Zwei – Drei oder mehr) vor. Einzig die Frage, wie viele Bücher es im Haushalt gibt, nutzt eine sechsstufige Rangskala. Die Antwortkategorien für dieses Item lauten: 0-10 Bücher, 11-25 Bücher, 26-100 Bücher, 101-200 Bücher, 201-500 Bücher und Mehr als 500 Bücher. Insgesamt gibt es 13 dichotome Items, acht polytome Items mit vierstufiger Auswahl und ein polytomes Items mit sechsstufiger Auswahl. Der Anteil fehlender Angaben liegt im Median bei 1.78 und ist somit als gering einzustufen. Die Items werden den vier Subskalen Kulturgüter (Cultural possessions; CULTPOSS), lernbezogene Güter (Home educational resources; HEDRES), Reichtum (WEALTH) und technikbezogene Güter (Information and Communication Technology Resources; ICTRES) zugeordnet. Eine Übersicht findet sich in Tabelle 1. Die Bachelorarbeit folgt in der weiteren Ausführung den englischsprachigen Abkürzungen.

Tabelle 1*Manifeste Indikatoren zur Erfassung von Besitz an Kultur- und Wohlstandsgütern*

Item	Auswahl	Zugehörigkeit				
		HOMEPOS	CULTPOSS	HEDRES	WEALTH	ICTRES
Einen Schreibtisch zum Lernen	2	X		X		
Ein Zimmer für dich allein	2	X			X	
Einen ruhigen Platz zum Lernen	2	X		X		
Einen Computer, den du für Schularbeiten verwenden kannst	2	X		X		
Lernsoftware	2	X		X		X
Einen Internet-Anschluss	2	X			X	X
Klassische Literatur (z. B. von Goethe)	2	X	X			
Bücher mit Gedichten	2	X	X			
Kunstwerke (z. B. Bilder)	2	X	X			
Bücher, die dir bei Schularbeiten helfen	2	X		X		
Technische Nachschlagewerke	2	X		X		
Ein Wörterbuch	2	X		X		
Bücher über Kunst, Musik oder Design	2	X	X			
Fernseher	4	X			X	
Auto	4	X			X	
Zimmer mit Badewanne oder Dusche	4	X			X	
Handy mit Internetzugang (z. B. Smartphones)	4	X			X	X
Computer (PC, Laptop oder Notebook)	4	X			X	X
Tablet-Computer (z. B. iPad, BlackBerry, PlayBook)	4	X			X	X
E-Book-Reader (z. B. Kindle, Kobo, Sony)	4	X			X	X
Musikinstrumente (z. B. Gitarre, Klavier)	4	X	X			
Wie viele Bücher gibt es bei dir zuhause?	6	X				

Anmerkung. Länderspezifische Items sind nicht aufgeführt.

3.3 Statistische Analysen

3.3.1 Reliabilität

Unter der Reliabilität wird die Genauigkeit oder Verlässlichkeit einer Messung verstanden. Als wesentliche Voraussetzung für das Gütekriterium der Validität betrachtet die Reliabilitätsprüfung die Messfehlerfreiheit. Die Reliabilität selbst ist an die Objektivität hinsichtlich der Testsituation, -durchführung und -auswertung gebunden. Das Gütekriterium der Objektivität wird angesichts des hochstrukturierten und standardisierten Vorgehens in PISA als angenommen.

Designbedingt kann keine Retest-Reliabilität ermittelt werden, daher eignet sich zur Schätzung der Reliabilität die Angabe der internen Konsistenz. Alle Items einer zugehörigen Skala erfassen das gleiche Merkmal und sollten hoch miteinander korrelieren (Schermelleh-Engel & Werner, 2012). Der Koeffizient Cronbachs Alpha (α) gibt das Ausmaß an, in dem eine nicht beobachtbare Variable die von den Items geteilte Varianz reflektiert.

$$(3) \quad Rel(x) = \alpha = \frac{m}{m-1} \times \left(1 - \frac{\sum_{i=1}^m Var(x_i)}{Var(x)} \right)$$

In der Gleichung (3) bezeichnet m die Anzahl der Items, $Var(x_i)$ ist die Varianz des i -ten Items und $Var(x)$ steht für die Gesamtvarianz des Tests. Cronbachs Alpha stellt als Schätzer für die Reliabilität eine Eigenschaft des Datensatzes und nicht des Messinstruments per se dar. Jedoch wird jede beobachtete Kovarianz zwischen den m -teiligen Testteilen als wahre Varianz bzw. als Varianz im wahren Wert interpretiert (Schermelleh-Engel & Werner, 2012). Im Allgemeinen werden Skalen mit Werten $\geq .70$ als für die Sozialforschung hinreichend reliabel beurteilt. Zum Zwecke der Konsistenzprüfung werden die im technischen Bericht dokumentierten Werte herangezogen. Für jedes Land und jede Skala wird die Reliabilität berichtet.

3.3.2 Konstruktvalidität

Die Skala HOMEPOS mit den vier Subskalen erfassen direkt beobachtbare Haushaltsmerkmale und bilden ein theoretisches Messmodell. Ob das spezifizierete Messmodell mit der empirischen Datenstruktur übereinstimmt, kann eine konfirmatorische Faktoranalyse (CFA) schätzen (Hartig, Frey & Jude, 2012; Moosbrugger & Schermelleh-Engel, 2012).

Die konfirmatorische Faktoranalyse ist eine spezielle Form eines Strukturgleichungsmodells, das im Allgemeinen prüft, ob eine hinreichende Übereinstimmung (Modellanpassung) zwischen den empirischen Daten und dem Messmodell vorliegt oder ob das Modell verworfen

werden muss (Gäde et al., 2020; Moosbrugger & Schermelleh-Engel, 2012). Sie dient den Belangen der Bachelorarbeit dazu, die Konstruktvalidität im Sinne der faktoriellen Validität zu untersuchen. Es wird erwartet, dass die Messmodelle, die explizit als Hypothese im Rahmen einer CFA behandelt werden (Gäde et al., 2020), in jedem Land eine hinreichend gute Modellpassung aufweisen. Dies wird mit ausgewählten Gütekriterien evaluiert, für die verschiedene Schwellenwerte eingeführt werden. Anhand der Kriterien kann eine akzeptable bzw. unzureichende Passung des Modells attestiert werden kann. In Anlehnung an die Originalstudie werden die folgenden Verfahren zur Modellüberprüfung ausgewählt: χ^2 -verteilte Teststatistik, Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR). Als ein weiteres Maß wird der Tucker-Lewis Index (TLI) berichtet. Dabei handelt es sich um deskriptive Indizes (bis auf den χ^2 -Test), für die es heuristische Cut-off-Werte in der Forschung gibt. Generell wird für eine gute (akzeptable) Passung erwartet für $CFI/TLI \geq .95$ (.90) und $RMSEA/SRMR \leq .08$ (Hu & Bentler, 1999).

Hu & Bentler (1999) liefern zur Bestimmung der Cut-offs eine einflussreiche Arbeit, indem sie Raten der Fehler erster Ordnung und Fehler zweiter Ordnung für verschiedene Modellspezifikationen dokumentieren. In neuerer Forschung werden die Limitationen der berichteten Werte diskutiert (Perry et al., 2015) und die Vorteile modellspezifischer Diagnostik hervorgehoben (McNeish & Wolf, 2020). Die Kritik bezieht sich darauf, dass die Empfehlungen von Hu & Bentler (1999) für komplexe Modelle (mehr als drei Faktoren, Doppelladungen, hohe Faktorinterkorrelationen) unrealistisch sind (Perry et al., 2015) und die Variabilität der Faktoranzahl, Faktorreliabilität, Itemanzahl und Stichprobengröße unberücksichtigt bleibt (McNeish & Wolf, 2020). Im Rahmen der Bachelorarbeit werden die in der Forschung diskutierten Nachteile lediglich in der Diskussion berücksichtigt.

Alle Modellparameter werden mittels der Maximum Likelihood-Methode mit robusten Standardfehlern geschätzt (estimator = "MLR"). Das Verfahren ist robust gegenüber nicht-normalverteilten Daten und korrigiert eine Überschätzung der Standardfehler (Gäde et al., 2020), z. B. bei der Verletzung der Annahme der statistischen Unabhängigkeit von Beobachtungen (Sandwich-Schätzer). Schuldaten sind typischerweise hierarchisch geschachtelt (Schülerinnen und Schüler in Schulen/Klassen). Aus diesem Grund beziehe ich den Identifier der Schulinstitution als Clustervariable ein (cluster = "CNTSCHID"). Eine Klassenvariable liegt aufgrund des Studiendesign nicht vor (vgl. Kapitel 3.1). Fehlende Werte werden über FIML (missing = "FIML") berücksichtigt. Die Originalstudie (Rutkowski & Rutkowski, 2013) ist in dieser Hinsicht nicht transparent, welche Parameterschätzung angewandt wurde. Nach dem beschriebenen Vorgehen wird für jede Subskala und jedes Land eine eigene konfirmatorische Faktoranalyse geschätzt.

Das resultiert in fünf Ergebnistabellen zu den vier Subskalen und der Gesamtskala mit den Fitindizes der Bildungssysteme. Um die Ergebnisinterpretation zu erleichtern, werden diejenigen Länder farblich markiert, deren Modellanpassung in allen vier Kriterien (grau) zufriedenstellende Kennwerte liefern. Für den Ergebnisbericht werden die Mediane, Minima und Maxima auf Skalenebene berichtet.

3.3.3 Kulturelle Vergleichbarkeit

Mit den bisherigen Analysen kann getrennt für jedes Bildungssystem bestimmt werden, ob das Messmodell zur empirischen Datenstruktur passt. Dies liefert einen Hinweis auf Konstruktvalidität. Die dritte Fragestellung beschäftigt sich damit, ob das spezifiziertere Messmodell überhaupt ländervergleichend eingesetzt werden kann. Erst wenn das Instrument invariant zur Gruppenzugehörigkeit ist (vgl. Kapitel 2.3), dann lassen sich latente Mittelwerte auf Länderebene sinnvoll miteinander vergleichen.

Die Prüfung auf Messinvarianz kann erst mit dem Vorliegen der Daten realisiert werden. Dafür kommen konfirmatorische Faktoranalysen im Mehrgruppendesign zum Einsatz. Zuerst wird ein Baseline-Modell geschätzt, das die Gruppenzugehörigkeit ignoriert. Um konfigurale Invarianz zu testen, wird anschließend die Gruppenzugehörigkeit berücksichtigt (gleiche Faktorstruktur in den Gruppen). In einem stärker restringierten Modell, der metrischen Invarianz, werden die Faktorladungen in allen Gruppen gleichgesetzt. Im letzten Modell, der skalaren Messinvarianz, werden zusätzlich die Intercepts fixiert. Wenn die deskriptiven Fit-Statistiken keine Fehlspezifikation anzeigen, dann sind die Annahmen der konfiguralen, metrischen und skalaren Invarianz gestützt. Die entsprechenden Aspekte des statistischen Modells sind damit invariant zur Ländervariable. Mit der schrittweisen Einführung von strengeren Restriktionen werden Veränderungen des Modellfits (siehe oben) in allen drei Subskalen betrachtet. Kleinere Differenzen ($\leq .03$) sind akzeptabel. Das Vorgehen zur Parameterschätzung und dem Umgang mit fehlenden Werten entspricht der Beschreibung in Kapitel 3.3.2 (siehe oben). Es wird erwartet, dass sich die Indizes zur Evaluation der Modellanpassung auch mit strengeren Restriktionen nicht bedeutsam verändern.

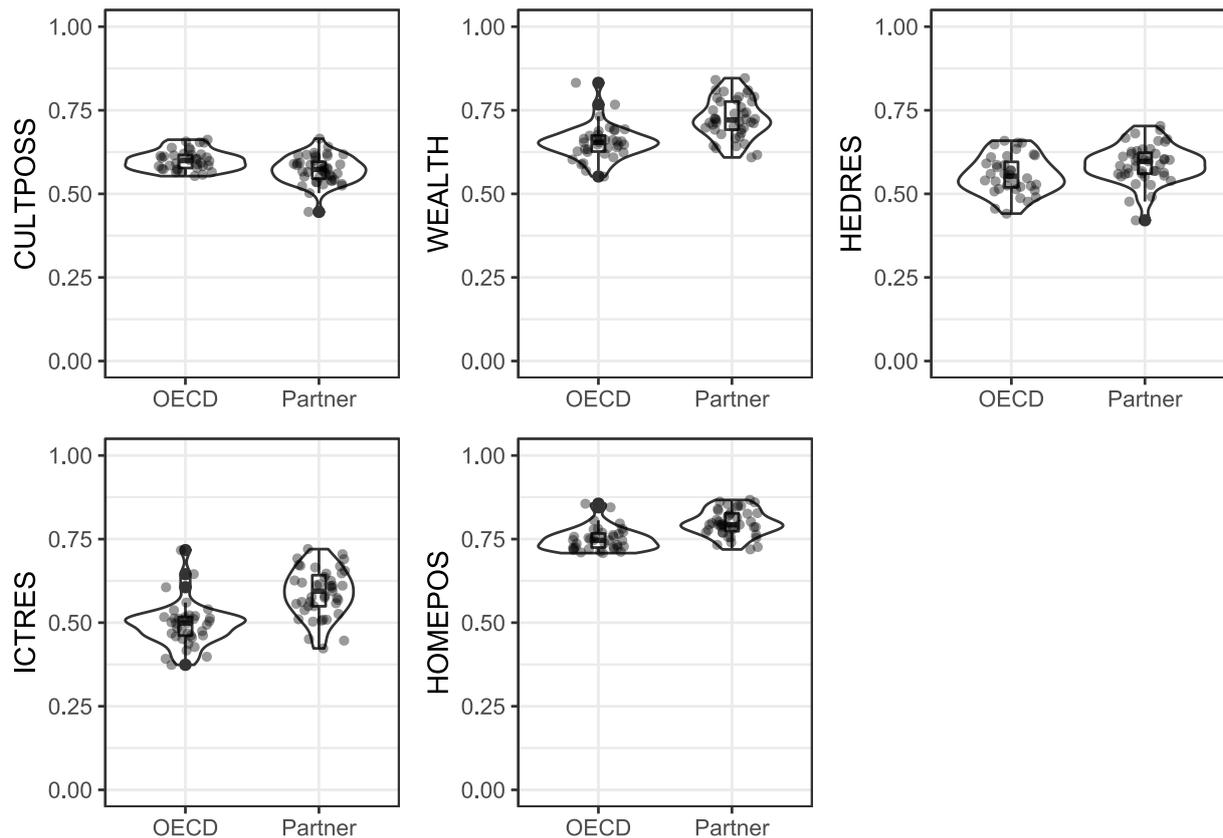


Abbildung 3: Verteilung der Alphawerte auf Skalenebene getrennt nach OECD- und Partnerstaaten (n = 80).

4. Ergebnisse

4.1 Reliabilität

Die Konsistenzprüfung zeigt für die Gesamtskala HOMEPOS zufriedenstellende Werte über alle Bildungssysteme hinweg. In der Abbildung 3 sind die Verteilungen der Alphawerte zu sehen. Der Median liegt bei .77. Selbst der kleinste Wert (Finnland) ist mit .71 noch akzeptabel. Das Maximum beträgt .87 (Peru). Die Variabilität über die Länder hinweg ist eher gering. Anders sieht es auf Subskalenniveau aus, für die die Alphawerte breit streuen. Die internen Konsistenzen liegen für CULTPOSS (Mdn = .59, Min = .45, Max = .67), HEDRES (Mdn = .57, Min = .42, Max = .70), WEALTH (Mdn = .69, Min = .55, Max = .85) und ICTRES (Mdn = .54, Min = .37, Max = .72) bei dem Großteil der Länder nicht über dem gewünschten Schwellenwert von .70 (vgl. Abbildung 3). Im direkten Vergleich mit der replizierten Studie schneidet die WEALTH Skala erneut am besten ab. In dieser Subskala lassen sich für die meisten Teilnahmestaaten hinreichende Reliabilitäten berichten. Die Tabelle A1 im Appendix berichtet die Alphawerte für jedes Bildungssystem im Detail.

4.2 Konstruktvalidität

Der überwiegende Anteil der teilnehmenden Bildungssysteme weist keine gute Modellanpassung auf. Auf Skalenebene schneidet CULTPOSS am besten ab. In diesem latenten Merkmal zeigen 26 von 80 Bildungssysteme eine zufriedenstellende Passung in allen Kriterien (grau hinterlegt, siehe Tabelle A2 im Appendix). Weiterhin erreichen nur acht von 79 in der Subskala WEALTH³ (Tabelle A3), vier von 80 in der Subskala HEDRES (Tabelle A4) und drei von 80 Bildungssystemen in der Subskala ICTRES (Tabelle A5) die gewünschten Kennwerte in allen betrachteten Indizes. Die Tabelle 2 berichtet zur besseren Übersicht die Mediane, Minima und Maxima als deskriptive Statistiken. Es ist ersichtlich, dass bereits die Mediane teilweise unter den Cut-Offs liegen und damit anzeigen, dass 50 % der Bildungssysteme Hinweise auf mangelnde Konstruktvalidität liefern. Bei der Gesamtskala HOMEPOS (*Home Possessions*) erreichen selbst die maximal beobachteten Werte nicht das forschungspraktisch übliche Niveau ($CFI \leq .900$, $TLI \leq .900$).

4.3 Kulturelle Vergleichbarkeit

In Tabelle 3 werden die Ergebnisse der Mehrgruppen-Faktoranalysen berichtet, die die Konstrukte auf Messinvarianz zwischen den beteiligten Bildungssystemen prüfen. Keine der Skalen hält den statistischen Restriktionen stand, die die Voraussetzung für valide Mittelwertvergleiche bilden. Skalare Messinvarianz ist nicht gegeben und auch die metrische Messinvarianz muss für WEALTH, HEDRES und ICTRES verworfen werden ($CFI/TLI \leq .90$ und $RMSEA/SRMR \geq .080$). Einzig bei CULTPOSS verbessert sich mit Hinzunahme der Gruppenzugehörigkeit die Modellanpassung (konfigurale Messinvarianz) und auch die metrische Messinvarianz ist ohne bedeutsame Differenzwerte für ΔCFI und $\Delta RMSEA$ gegeben. Die Messinvarianzprüfung für die Gesamtskala HOMEPOS ist nicht konvergiert.

³ In Israel (ISR) wurde diese Skala nicht administriert.

Tabelle 2*Deskriptive Statistiken aus den konfirmatorischen Faktoranalysen (Skalenebene, n = 80)*

Skala	Kennwert	Kriterium	Median	Minimum	Maximum
CULTPOSS	CFI	$\geq .900$	0.939	0.814	0.981
	TLI	$\geq .900$	0.879	0.627	0.962
	RMSEA	$\leq .080$	0.090	0.038	0.165
	SRMR	$\leq .080$	0.033	0.017	0.067
WEALTH	CFI	$\geq .900$	0.877	0.731	0.961
	TLI	$\geq .900$	0.837	0.641	0.948
	RMSEA	$\leq .080$	0.066	0.037	0.024
	SRMR	$\leq .080$	0.039	0.024	0.056
HEDRES	CFI	$\geq .900$	0.853	0.716	0.944
	TLI	$\geq .900$	0.780	0.574	0.915
	RMSEA	$\leq .080$	0.073	0.043	0.108
	SRMR	$\leq .080$	0.040	0.024	0.063
ICTRES	CFI	$\geq .900$	0.895	0.698	0.971
	TLI	$\geq .900$	0.825	0.496	0.952
	RMSEA	$\leq .080$	0.076	0.027	0.114
	SRMR	$\leq .080$	0.038	0.016	0.062
HOMEPOS	CFI	$\geq .900$	0.649	0.473	0.787
	TLI	$\geq .900$	0.612	0.417	0.764
	RMSEA	$\leq .080$	0.074	0.058	0.100
	SRMR	$\leq .080$	0.065	0.050	0.101

Anmerkung. CFI = Comparative Fit Index, TLI = Tucker-Lewis-Index, RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual.

Tabelle 3*Ergebnisse der Messinvarianzprüfung (Mehrgruppen-Faktoranalysen)*

Skala	Modell	CFI	RMSEA	Δ CFI	Δ RMSEA
CULTPOSS	Baseline	.893	.119	.	.
	Konfigural	.930	.097	.	.
	Metrisch	.900	.087	.031	.010
	Skalar	.409	.177	.491	.089
WEALTH*	Baseline	.926	.066	.	.
	Konfigural	.883	.068	.	.
	Metrisch	.782	.082	.101	.014
	Skalar	.186	.142	.596	.061
HEDRES	Baseline	.860	.079	.	.
	Konfigural	.857	.076	.	.
	Metrisch	.707	.091	.149	.015
	Skalar	.023	.145	.684	.055
ICTRES	Baseline	.921	.085	.	.
	Konfigural	.898	.077	.	.
	Metrisch	.754	.097	.144	.019
	Skalar	.178	.152	.576	.055
HOMEPOS*	Baseline	.641	.086	.	.
	Konfigural ⁺
	Metrisch ⁺
	Skalar ⁺

Anmerkung. *Israel wurde aus dieser Analyse ausgeschlossen, weil die WEALTH Skala nicht administriert wurde. Ergebnis und Interpretation beziehen sich auf 79 Bildungssysteme.

⁺Modell ist nicht konvergiert.

5. Diskussion

Die vorliegende Bachelorarbeit macht auf zwei sehr aktuelle Probleme der empirischen Sozialforschung aufmerksam. (1) Auf der einen Seite ist der Zusammenhang von Testleistung und sozialer Herkunft ein populärer Forschungsgegenstand. Jedoch erlauben heterogene Ansätze der Operationalisierung beider Variablen kaum ein vollumfassendes theoretisches wie empirisches Verständnis. Die gefundenen Effekte streuen sehr stark und variieren mit Kompetenzdomäne und theoretischer Fundierung (Sirin, 2005). Das erschwert die erkenntnistheoretische Rückführung auf Unterschiede im wahren Wert. Intuitiv plausibel erscheint daher der Ansatz von PISA, international vergleichend Instrumente zu administrieren. (2) Auf der anderen Seite führt dies zu neuen Problemen hinsichtlich der kulturellen und statistischen Vergleichbarkeit (Messinvarianz). Mit neuen statistischen Methoden wird immer mehr Forschung publiziert, die Kritik an international vergleichenden Ansätzen übt (Fischer & Karl, 2019; Gorur, 2014; Rutkowski & Rutkowski, 2013). Kritikpunkte können sein, ob das theoretische Konstrukt selbst über verschiedene Kontexte bzw. Gruppen vergleichbar verstanden wird (Konstruktäquivalenz) oder ob die statistischen Voraussetzungen für einen validen Mittelwertvergleich gegeben sind (Messinvarianz). Gleichzeitig findet kaum Kommunikation seitens über mangelnde Vergleichbarkeit statt: „any cross-country [...] comparisons using SES data from international large-scale assessments should be preceded by a careful examination of the psychometric properties of the scale used to measure SES, a topic which is rarely addressed by researchers.“ (Lee & Davier, 2020, S. 56). Forschende müssen sich bewusst sein, dass Vergleiche von Teilpopulationen in Merkmalen ohne invariante Faktorstruktur, Faktorladung und Nullpunkt inhaltslos sind (Fischer & Karl, 2019).

Innerhalb dieses Spannungsfeldes hat die Arbeit empirische Befunde zur Reliabilität, Konstruktvalidität (faktorielle Validität) und kulturellen Vergleichbarkeit des Instruments zur Erfassung von Besitz an Kultur- und Wohlstandsgütern (HOMEPOS) und ihren Subskalen geliefert. Die untersuchte Skala ist ein wichtiger Bestandteil des PISA eigenen Gesamtindex des sozioökonomischen und soziokulturellen Status (ESCS). In zentralen Berichten wird der Index in Zusammenhang mit der domänenspezifischen Testleistung gesetzt (OECD, 2019b), findet von dort mediale Beachtung und beeinflusst potenziell politische Entscheidungen. In diesem Sinne erweitert die Bachelorarbeit den stattfindenden Diskurs über die angemessene Interpretation von internationalen Schulleistungstudien.

Im Ergebnis offenbart eine Auswertung der Reliabilitätskoeffizienten (Cronbachs Alpha) für alle beteiligten Bildungssysteme eine starke Heterogenität auf Skalen- und Systemebene.

Während die Gesamtskala größtenteils reliabel zu sein scheint, sind die Subskalen nur eingeschränkt nutzbar. Auch eine Verlagerung des Schwellenwerts auf .600 schließt immer noch Bildungssysteme aus. Beispielsweise liegen für ICTRES noch 57 von 80 Bildungssystemen unter einem Alphawert von .600, weshalb die Befunde deutliche Hinweise liefern, dass die gebrauchten Items nicht unbedingt dasselbe Merkmal messen. Hinzu kommt, dass der Alphawert formal von der Itemzahl abhängt, sodass die höheren Werte für die Gesamtskala (22 internationale Items) wenig verwunderlich sind (Schermelleh-Engel & Werner, 2012; Streiner, 2003). Obwohl die Konsistenzanalyse von PISA selbst vorgenommen wird, verzichten sie stillschweigend auf eine Einordnung dieser Befunde (OECD, 2019a). Im technischen Bericht finden sich durchaus die Schwellenwerte „coefficient ranges between 0 and 1, with higher values indicating higher internal consistency meaning the set of items closely measures a common dimension. Commonly accepted cut-off values are 0.9 to signify excellent, 0.8 for good, and 0.7 for acceptable internal consistency“ und Aussagen wie „[s]imilar and high values across countries are a good indication of having measured reliably across countries.“ (S. 8).

Nach dieser üblichen Einführung verweist der technische Bericht auf eine externe Excel-Datei, in der alle Koeffizienten gelistet sind (also ohne im Bericht selbst die Kennwerte zu zeigen oder zu interpretieren). Forschenden ist es so nur schwer möglich, in transparenter Weise Ergebnisbericht und -interpretation zu vereinen. Im Hinblick auf die ermittelten Kennwerte bleibt eine kritische Reflexion des Gütekriteriums der Reliabilität aus. Um diesen Umstand zu verbessern, sollte geprüft werden, welche der Indikatoren eher schwach mit den übrigen assoziiert sind. Im Sinne der klassischen Testtheorie könnte auch eine Analyse der Trennschärfen hilfreich sein. Innerhalb des Strukturgleichparadigmas können geringe Faktorladungen dienlich sein, auf welche Indikatoren die latente Variable nur einen schwachen Einfluss besitzt.

Im Hinblick auf die Konstruktvalidität und kulturellen Vergleichbarkeit bestätigen sich die Befunde, die bereits Rutkowski & Rutkowski (2013) auf Basis von PISA 2009 publizieren. Für wenige, ausgewählte Bildungssysteme scheinen die Messmodelle brauchbar zu sein. Doch in den meisten Teilnahmestaaten können mit Hilfe von konfirmatorischen Faktoranalysen für Länder einzeln und im Mehrgruppen-Design Missspezifikationen in den Modellen vermutet werden. Das bestärkt die Annahme, dass latente Merkmale der Sozialposition eher ungeeignet sind, in einer Vielzahl an Kontexten standardisiert erfasst zu werden. Es dürfte ratsam sein, dass nationale Konsortien sich die Frage stellen, welche Indikatoren tatsächlich materiellen und kulturellen Wohlstand anzeigen. Mit den vorgenommenen Analysen kann erstmal nur auf Skalenebene argumentiert werden – über die Items selbst liefert diese Arbeit keine Aussage. Auffällig ist, dass die Skala CULTPOSS, also der Besitz an hochkulturellen Gütern, wie schon in der

Originalstudie, am besten funktioniert. Metrische Invarianz ist ebenfalls wie im Erhebungsjahr 2009 gegeben. Es deutet sich an, dass Klassische Literatur, Kunstwerke und Musikinstrumente so universell sind, dass ihr Besitz in vielen Ländern zwischen Befragten mit hoher und niedriger Merkmalsausprägung unterscheiden und sie einen gemeinsamen Faktor ausbilden. Anders sieht das bei Skalen/Items zu lern- und technikbezogenen Gütern aus. Man könnte vermuten, dass das Item, ein internetfähiges Handy zu besitzen, womöglich in Deutschland nicht ausreichend zwischen Schülerinnen und Schüler differenziert. Die Ausstattung mit Kommunikationstechnik hat sich bei jungen Erwachsenen in den vergangenen Jahren auf hohem Niveau stabilisiert. Nach Ergebnissen der JIM-Studie, die die Situation in Deutschland beschreibt, besitzt nahezu jede(r) ein eigenes Smartphone (99 %) und hat über WLAN (97 %) Zugang zum Internet (Feierabend et al., 2019). Folglich dürfte eine Abfrage technikbezogener Güter zu einer starken Varianzeinschränkung führen und damit empirisch unbrauchbar sein.

Ein Aspekt, der im Rahmen dieser Arbeit zu kurz kam, ist die dezidierte Analyse auf Bildungssystemebene. Die Ergebnispräsentation und -interpretation fand überwiegend auf Skalenebene statt. Jedoch ist denkbar, dass sich Bildungssysteme bzw. Länder in einigen Eigenschaften ähneln und vergleichbare Reliabilitäts- und Fitwerte zeigen. Beispielsweise ermöglichen großzügigere Finanzmittel, wie sie in den OECD-Staaten üblich sind, umfangreiche Pilotierungsstudien. Skalen können so im Feld genügend erprobt und ggf. Übersetzungsmängel angepasst werden. Die Originalstudie hat bereits formuliert, dass sich OECD- und Partnerstaaten unterscheiden könnten, prüft dies jedoch nicht systematisch. Lee (2019) lässt auf Länderebene einen Kennwert zum Modelmisfit der HOMEPOS-Skala und einem Wohlstandsindikator eines Landes korrelieren, findet jedoch nur eine schwache Korrelation mit dem Bruttoinlandsprodukt ($r = .11$). Eine weiterführende Arbeit könnte die in dieser Bachelorarbeit beobachteten Kennwerte mit den Ergebnissen der Originalstudie in einer Datentabelle zusammenführen. Aus methodischer Sicht ist interessant, ob sich eine gewisse Rangreihenstabilität der Bildungssysteme zeigt. Dies kann ein erster Ansatzpunkt sein, systematische Länderunterschiede, z. B. eine gemeinsame Erhebungssprache, explorativ zu untersuchen.

Zusammenfassend lässt sich festhalten, dass das Instrument zur Erfassung von Besitz an Kultur- und Wohlstandsgütern mit insgesamt 20 Entwicklungsjahren noch verbessert werden kann. Mit Bezug zu den drei Forschungsfragen erweisen sich die Skalen als mehrheitlich nicht reliabel, nicht faktoriell valide und können nicht über Systeme hinweg verglichen werden. Auch mit neueren statistischen Methoden, Fragebogendaten nachträglich über viele Subgruppen ‚vergleichbar‘ zu machen, gewinnen in diesem Zusammenhang an Bedeutung. Das sogenannte *partial invariance modeling* sorgt dafür, dass in einigen Items Gleichheit nicht erzwungen wird,

weil länderspezifische Nuancen im Anschluss an die Datenerhebung entdeckt werden (Lee & Davier, 2020). Auf der einen Seite erhöht dies die Gebrauchstauglichkeit von Sekundärdaten, auf der anderen Seite läuft es dem Ziel einer theoretisch motivierten Sozialforschung entgegen. Sobald ein theoretisches Messmodell für alle Länder gleichsam spezifiziert wurde, man anschließend ‚schlechte‘ (also nicht vergleichbare) Daten generiert und mithilfe von Statistik die Kennwerte wie gewünscht aufarbeitet, dann konstruiert man sich selbst einen Bias. Und zwar in der Hinsicht, dass man Entwicklungsarbeit überwiegend in die statistische Modellierung als in die theoretische Fundierung investiert. Dadurch bleiben Indikatoren, die womöglich Wohlstand international vergleichend anzeigen könnten, im Dunkeln.

Literaturverzeichnis

- Asil, M., & Brown, G. T. (2016). Comparing OECD PISA reading in English to other languages: Identifying potential sources of non-invariance. *International Journal of Testing*, 16(1), 71-93.
- Astleithner, F., Vogl, S., & Parzer, M. (2021). Zwischen Wunsch und Wirklichkeit: Zum Zusammenhang von sozialer Herkunft, Migration und Bildungsaspirationen. *Österreichische Zeitschrift für Soziologie*, 1-24.
- Avvisati, F. (2020). The measure of socio-economic status in PISA: a review and some suggested improvements. *Large-Scale Assessments in Education*, 8(8), 1-37.
- Bourdieu, P. (1986). "The forms of capital." In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241–258). New York, NY: Greenwood.
- Brese, F., & Mirazchiyski, P. (2013). Measuring students' family background in largescale international education studies (IERI monograph series - Special issue 2). Princeton, NJ: IEA-ETS Research Institute (IERI).
- Caro, D. H., & Biecek, P. (2017). intsvy: An R package for analyzing international large-scale assessment data. *Journal of Statistical Software*, 81(1), 1-44.
- Ditton, H., & Maaz, K. (2011). Sozioökonomischer Status und soziale Ungleichheit. In *Empirische Bildungsforschung* (S. 193-208). VS Verlag für Sozialwissenschaften.
- Ehmke, T. & Jude, N. (2010). Soziale Herkunft und Kompetenzerwerb In E. Klieme, C. Artelt, Cordula, J. Hartig, N. Jude, O. Köller, Olaf, M. Prenzel, W. Schneider, & P. Stanat, Petra (Hrsg.), *PISA 2009. Bilanz nach einem Jahrzehnt* (S. 231-254). Münster : Waxmann.
- Feierabend, S., Rathgeb, T., & Reuter, T. (2019). *JIM-Studie 2019. Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12- bis 19-Jähriger*. Medienpädagogischer Forschungsverbund Südwest.
- Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*, 10, 1507.
- Gäde, J. C., Schermelleh-Engel, K., & Brandt, H. (2020). Konfirmatorische Faktorenanalyse (CFA). In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 615-659). Springer, Berlin, Heidelberg.
- Ganzeboom, Harry B.G./Treiman, Donald J. (1996): Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. In: *Social Science Research* 25: 201-239.
- Gentrup, S., Lorenz, G., Kristen, C., & Kogan, I. (2020). Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, 66, 101296.
- Gorur, R. (2014). Towards a sociology of measurement in education policy. *European Educational Research Journal*, 13(1), 58-72.
- Hartig, J., Frey, A. & Jude, N. (2012). Validität. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 144-171). Berlin, Heidelberg: Springer.
- Henschel, S., Heppt, B., Weirich, S., Edele, A., Schipolowski, S. & Stanat, P. (2019). Zuwanderungsbezogene Disparitäten. In P. Stanat, S. Schipolowski, N. Mahler, S. Weirich & S.

- Henschel (Hrsg.), *IQB-Bildungstrend 2018. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich* (S. 295-336). Münster: Waxmann.
- Hoffmeyer-Zlotnik, J. H., & Warner, U. (2019). Nationale soziodemographische Standards und international harmonisierte soziodemographische Hintergrundvariablen. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 875-892). Springer VS, Wiesbaden.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.
- Kelava, A. & Moosbrugger, H. (2012). Deskriptivstatistische Evaluation von Items (Itemanalyse) und Testwertverteilungen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 75-102). Berlin, Heidelberg: Springer.
- Latcheva, R., & Davidov, E. (2019). Skalen und Indizes. In N. Baur & J. Blasius (Hrsg.), *Handbuch Methoden der empirischen Sozialforschung* (S. 893-905). Springer VS, Wiesbaden.
- Lee, S. S. & von Davier, M. (2020). Improving measurement properties of the PISA home possessions scale through partial invariance modeling. *Psychological Test and Assessment Modeling*, 62(1), 55-83.
- Lee, S. S. (2019). *Longitudinal And Cross-Country Measurement Invariance Of The Pisa Home Possessions Scale* [Dissertation]. University of Pennsylvania.
- Liu, J., Peng, P., & Luo, L. (2020). The relation between family socioeconomic status and academic achievement in China: a meta-analysis. *Educational Psychology Review*, 32(1), 49-76.
- McLoyd, V. C. (1998). Socioeconomic disadvantage and child development. *American Psychologist*, 53, 185–204.
- McNeish, D., & Wolf, M. G. (2020). Dynamic fit index cutoffs for confirmatory factor analysis models [PsyArxiv] <https://doi.org/10.31234/osf.io/v8yru>.
- OECD (2019a). *PISA 2018 Technical report*. Paris: OECD. Online verfügbar unter <https://www.oecd.org/pisa/data/pisa2018technicalreport/> (10.07.2021).
- OECD (2019b). *PISA 2018 Results (Volume II): Where All Students Can Succeed*, Paris: OECD. Online verfügbar unter: <https://doi.org/10.1787/b5fd1b8f-en> (10.07.2021).
- Perry, J. L., Nicholls, A. R., Clough, P. J., & Crust, L. (2015). Assessing model fit: Caveats and recommendations for confirmatory factor analysis and exploratory structural equation modeling. *Measurement in Physical Education and Exercise Science*, 19(1), 12-21.
- Pokropek, A., Borgonovi, F., & McCormick, C. (2017). On the cross-country comparability of indicators of socioeconomic resources in PISA. *Applied Measurement in Education*, 30(4), 243-258.
- R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Reid, J. L., & Ready, D. D. (2013). High-quality preschool: The socioeconomic composition of preschool classrooms and children's learning. *Early Education & Development*, 24(8), 1082-1111.

- Reiss, K., Weis, M., Klieme, E., & Köller, O. (Eds.). (2019). *PISA 2018: Grundbildung im internationalen Vergleich*. Waxmann.
- Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E., & Stanat, P. (2014). Socioeconomic and language minority classroom composition and individual reading achievement: The mediating role of instructional quality. *Learning and Instruction, 32*, 63-72.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software, 48*(2), 1-36.
- Rutkowski, D., & Rutkowski, L. (2013). Measuring socioeconomic background in PISA: One size might not fit all. *Research in Comparative and International Education, 8*(3), 259-278.
- Schermelleh-Engel, K. & Werner, C.S. (2012). Methoden der Reliabilitätsbestimmung. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (S. 119-141). Berlin, Heidelberg: Springer.
- Schneider, T. (2011). Die Bedeutung der sozialen Herkunft und des Migrationshintergrundes für Lehrerurteile am Beispiel der Grundschulempfehlung. *Zeitschrift für Erziehungswissenschaft, 14*(3), 371-396.
- Schulte, K., Nonte, S., & Schwippert, K. (2013). Die Überprüfung von Messinvarianz in international vergleichenden Schulleistungsstudien am Beispiel der Studie PIRLS. *Zeitschrift für Bildungsforschung, 3*(2), 99-118.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453.
- Skinner, C. J. (2014). *Probability proportional to size (PPS) sampling*. Wiley StatsRef: Statistics Reference Online, 1-5.
- Streiner, D. L. (2003). Being inconsistent about consistency: When coefficient alpha does and doesn't matter. *Journal of Personality Assessment, 80*(3), 217-222.
- Warner, U. (2009). *Die Messung von Einkommen in der international vergleichenden Umfrageforschung*. Dissertation. Online verfügbar unter <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-59187> (10.07.2021).
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*(3), 461-481.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D. A., François, R., ... & Yutani, H. (2019). Welcome to the Tidyverse. *Journal of open source software, 4*(43), 1686.
- Zhou, J., He, J., & Lafontaine, D. (2020). Cross-cultural comparability and validity of meta-cognitive knowledge in reading in PISA 2009: a comparison of two scoring methods. *Assessment in Education: Principles, Policy & Practice, 27*(6), 635-654.

Appendix

Tabelle A1

Ergebnisse der Konsistenzprüfung (Cronbachs Alpha) für jedes Bildungssystem

Land	HOMEPOS	CULT- POSS	HED- RES	WEALTH	ICTRES	Status
Albania	.822	.563	.607	.744	.692	Partner
Argentina	.795	.579	.633	.692	.549	Partner
Australia	.726	.565	.648	.639	.455	OECD
Austria	.726	.591	.487	.661	.468	OECD
Azerbaijan	.802	.588	.568	.759	.637	Partner
Belarus	.726	.538	.421	.609	.510	Partner
Belgium	.728	.616	.540	.645	.512	OECD
Bosnia Herzegovina	.756	.61	.554	.670	.509	Partner
Brazil	.835	.535	.602	.781	.620	Partner
Brunei Darussalam	.770	.446	.539	.701	.615	Partner
Bulgaria	.802	.591	.667	.724	.577	Partner
Canada	.747	.565	.634	.664	.510	OECD
Chile	.806	.573	.571	.732	.606	OECD
China	.818	.561	.552	.750	.604	Partner
Chinese Taipei	.793	.665	.620	.666	.526	Partner
Colombia	.860	.574	.597	.811	.705	Partner
Costa Rica	.847	.590	.584	.791	.647	Partner
Croatia	.733	.623	.477	.644	.451	Partner
Cyprus	.779	.614	.604	.711	.568	Partner
Czech Republic	.732	.638	.580	.633	.468	OECD
Denmark	.713	.598	.530	.569	.374	OECD
Dominican Rep.	.851	.535	.620	.802	.647	Partner
Estonia	.747	.578	.507	.657	.437	OECD
Finland	.710	.653	.547	.552	.392	OECD
Macedonia	.755	.565	.568	.642	.509	Partner
France	.736	.654	.525	.654	.500	OECD
Georgia	.791	.587	.561	.716	.562	Partner
Germany	.769	.608	.557	.665	.511	OECD
Greece	.754	.583	.514	.692	.537	OECD
Hong Kong	.814	.62	.634	.713	.558	Partner
Hungary	.781	.662	.549	.693	.480	OECD
Iceland	.725	.557	.618	.609	.398	OECD
Indonesia	.828	.549	.569	.790	.673	Partner
Ireland	.722	.585	.569	.591	.462	OECD
Israel	.720	.632	.615	.625	.517	OECD
Italy	.746	.583	.524	.647	.512	OECD
Japan	.712	.553	.537	.603	.517	OECD
Jordan	.826	.618	.703	.741	.665	Partner
Kazakhstan	.769	.586	.604	.637	.423	Partner
Korea	.779	.637	.586	.623	.504	OECD

Tabelle A1*Ergebnisse der Konsistenzprüfung (Cronbachs Alpha) für jedes Bildungssystem (Forts.)*

Kosovo	.763	.502	.527	.682	.563	Partner
Latvia	.717	.572	.441	.634	.459	OECD
Lebanon	.807	.551	.560	.721	.594	Partner
Lithuania	.746	.657	.510	.637	.475	OECD
Luxembourg	.772	.601	.560	.699	.540	OECD
Macao (China)	.793	.599	.575	.725	.506	Partner
Malaysia	.793	.530	.563	.725	.626	Partner
Malta	.719	.555	.602	.617	.446	Partner
Mexico	.856	.585	.560	.832	.717	OECD
Montenegro	.769	.573	.612	.698	.575	Partner
Morocco	.862	.567	.681	.810	.690	Partner
Netherlands	.708	.564	.528	.588	.417	OECD
New Zealand	.757	.575	.659	.647	.520	OECD
Norway	.758	.618	.613	.690	.484	OECD
Panama	.850	.523	.588	.841	.670	Partner
Peru	.867	.525	.613	.846	.720	Partner
Philippines	.840	.528	.599	.776	.655	Partner
Poland	.738	.593	.456	.661	.462	OECD
Portugal	.766	.609	.49	.663	.513	OECD
Qatar	.786	.569	.691	.776	.580	Partner
Rep. of Moldova	.786	.540	.551	.715	.609	Partner
Romania	.786	.578	.626	.692	.574	Partner
Russian Federation	.777	.556	.529	.713	.556	Partner
Saudi Arabia	.793	.552	.664	.707	.611	Partner
Serbia	.754	.617	.559	.678	.503	Partner
Singapore	.787	.616	.583	.698	.549	Partner
Slovak Republic	.781	.615	.653	.674	.501	OECD
Slovenia	.724	.620	.477	.627	.427	OECD
Spain	.757	.605	.515	.657	.522	OECD
Sweden	.749	.612	.591	.654	.451	OECD
Switzerland	.720	.573	.521	.620	.498	OECD
Thailand	.855	.615	.658	.806	.669	Partner
Turkey	.845	.613	.587	.767	.645	OECD
Ukraine	.740	.542	.491	.649	.537	Partner
U. A. E.	.798	.593	.656	.787	.612	Partner
United Kingdom	.762	.604	.610	.688	.494	OECD
United States	.797	.589	.653	.677	.560	OECD
Uruguay	.829	.641	.628	.740	.611	Partner
Viet Nam	.791	.587	.562	.724	.625	Partner

Tabelle A2*Ergebnisse der konfirmatorischen Faktoranalysen für CULTPOSS*

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Overall Model	42.498.002	5	.893	.786	.119	.046
ALB	61.67	5	.977	.954	.043	.018
ARE	865.713	5	.924	.848	.096	.036
ARG	341.547	5	.950	.900	.075	.028
AUS	948.997	5	.897	.794	.121	.047
AUT	104.09	5	.98	.960	.054	.019
BEL	262.116	5	.96	.920	.078	.027
BGR	234.22	5	.939	.879	.095	.036
BIH	156.251	5	.966	.933	.069	.025
BLR	305.011	5	.890	.780	.102	.041
BRA	413.155	5	.907	.815	.088	.037
BRN	192.262	5	.910	.819	.074	.031
CAN	1255.38	5	.922	.843	.108	.041
CHE	139.232	5	.965	.93	.068	.024
CHL	300.651	5	.939	.879	.089	.032
COL	151.244	5	.958	.915	.064	.025
CRI	216.333	5	.948	.896	.077	.029
CZE	223.406	5	.965	.929	.079	.027
DEU	111.629	5	.969	.938	.067	.024
DNK	185.945	5	.970	.940	.070	.023
DOM	45.001	5	.978	.955	.038	.017
ESP	789.97	5	.969	.939	.067	.024
EST	225.065	5	.943	.887	.092	.034
FIN	166.15	5	.973	.945	.076	.025
FRA	275.65	5	.952	.904	.093	.032
GBR	1.818.274	5	.845	.689	.165	.067
GEO	317.499	5	.910	.821	.108	.040
GRC	184.357	5	.958	.917	.075	.030
HKG	399.776	5	.909	.819	.116	.041
HRV	270.941	5	.952	.904	.090	.033
HUN	397.98	5	.93	.860	.124	.047
IDN	489.991	5	.909	.819	.090	.036
IRL	217.683	5	.939	.879	.088	.032
ISL	59.867	5	.978	.956	.058	.022
ISR	361.631	5	.936	.871	.105	.041
ITA	252.216	5	.968	.935	.066	.025
JOR	480.842	5	.905	.81	.103	.040
JPN	194.521	5	.943	.887	.079	.030
KAZ	614.016	5	.944	.887	.079	.030
KOR	335.778	5	.946	.893	.100	.032
KSV	89.612	5	.947	.894	.058	.025
LBN	131.439	5	.938	.877	.068	.028
LTU	220.223	5	.966	.933	.080	.027

Tabelle A2*Ergebnisse der konfirmatorischen Faktoranalysen für CULTPOSS (Fortsetzung)*

LUX	119.738	5	.970	.940	.066	.024
LVA	377.251	5	.908	.816	.119	.050
MAC	424.151	5	.840	.680	.149	.056
MAR	287.187	5	.898	.796	.092	.039
MDA	246.779	5	.901	.802	.096	.039
MEX	222.203	5	.938	.876	.082	.031
MKD	126.455	5	.957	.914	.067	.026
MLT	304.581	5	.873	.746	.135	.059
MNE	81.798	5	.981	.962	.048	.018
MYS	369.145	5	.857	.714	.110	.044
NLD	213.195	5	.930	.861	.094	.035
NOR	105.002	5	.980	.959	.06	.022
NZL	335.504	5	.921	.842	.104	.039
PAN	107.259	5	.945	.890	.059	.026
PER	335.979	5	.872	.743	.105	.042
PHL	117.564	5	.958	.915	.056	.024
POL	103.692	5	.975	.949	.059	.022
PRT	192.563	5	.955	.909	.081	.032
QAT	477.258	5	.933	.865	.084	.032
QAZ	709.236	5	.836	.672	.146	.054
QCI	1.174.145	5	.858	.716	.140	.055
QMR	186.634	5	.841	.682	.135	.059
QRT	468.268	5	.868	.736	.128	.054
ROU	509.645	5	.848	.697	.142	.058
RUS	605.768	5	.872	.745	.127	.054
SAU	486.974	5	.814	.627	.127	.058
SGP	487.659	5	.909	.817	.120	.044
SRB	393.066	5	.931	.862	.109	.041
SVK	264.597	5	.949	.899	.094	.036
SVN	250.526	5	.961	.922	.088	.032
SWE	111.718	5	.974	.948	.063	.022
TAP	700.272	5	.906	.813	.139	.047
THA	1.013.112	5	.858	.716	.153	.056
TUR	291.276	5	.933	.865	.091	.034
UKR	237.724	5	.921	.843	.089	.038
URY	129.154	5	.967	.935	.070	.026
USA	202.583	5	.946	.892	.091	.035
VNM	221.06	5	.919	.838	.090	.038

Anmerkung. Grau hinterlegte Bildungssysteme erreichen in allen Indizes gute Werte.

Tabelle A3*Ergebnisse der konfirmatorischen Faktoranalysen für WEALTH*

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Overall Model	70.918.267	27	.927	.903	.066	.035
ALB	763.015	27	.905	.873	.066	.039
ARE	4673.34	27	.862	.816	.096	.049
ARG	1.071.497	27	.902	.870	.057	.035
AUS	1.269.783	27	.850	.800	.060	.036
AUT	967.208	27	.821	.761	.072	.043
BEL	601.919	27	.891	.855	.050	.031
BGR	990.471	27	.827	.769	.083	.055
BIH	690.775	27	.867	.823	.062	.039
BLR	242.667	27	.934	.912	.037	.024
BRA	1.416.724	27	.906	.874	.070	.040
BRN	525.907	27	.924	.899	.052	.030
CAN	2.567.133	27	.861	.815	.066	.039
CHE	753.055	27	.812	.749	.068	.042
CHL	1174.86	27	.887	.850	.075	.043
COL	1.150.973	27	.917	.890	.076	.039
CRI	1069.78	27	.904	.872	.073	.042
CZE	869.971	27	.841	.788	.067	.042
DEU	491.53	27	.889	.853	.060	.035
DNK	1.002.224	27	.825	.766	.069	.042
DOM	764.442	27	.913	.883	.070	.039
ESP	4.039.735	27	.845	.793	.065	.040
EST	327.25	27	.904	.872	.046	.029
FIN	326.975	27	.876	.835	.045	.030
FRA	788.865	27	.878	.838	.067	.039
GBR	1.489.923	27	.849	.799	.064	.038
GEO	618.558	27	.892	.857	.063	.040
GRC	882.356	27	.825	.767	.071	.045
HKG	872.318	27	.859	.812	.073	.043
HRV	648.259	27	.849	.799	.059	.038
HUN	519.422	27	.870	.827	.060	.038
IDN	1.164.099	27	.937	.916	.059	.033
IRL	591.708	27	.809	.745	.061	.039
ISL	719.602	27	.731	.641	.089	.056
ISR*	NA	NA	NA	NA	NA	NA
ITA	1.506.999	27	.793	.724	.069	.043
JOR	1.109.465	27	.917	.889	.067	.038
JPN	955.22	27	.738	.651	.075	.048
KAZ	1.201.846	27	.918	.891	.047	.029
KOR	680.003	27	.826	.768	.060	.038
KSV	606.407	27	.873	.830	.065	.041
LBN	762.778	27	.885	.846	.070	.044

Tabelle A3*Ergebnisse der konfirmatorischen Faktoranalysen für WEALTH (Fortsetzung)*

LTU	696.271	27	.876	.835	.060	.038
LUX	582.165	27	.895	.860	.063	.037
LVA	359.941	27	.903	.871	.049	.031
MAC	343.288	27	.911	.881	.056	.034
MAR	672.496	27	.941	.922	.06	.034
MDA	534.888	27	.920	.893	.059	.035
MEX	1.025.041	27	.922	.896	.076	.039
MKD	633.097	27	.870	.827	.064	.041
MLT	527.503	27	.770	.694	.075	.049
MNE	938.634	27	.874	.832	.071	.043
MYS	454.711	27	.940	.919	.051	.029
NLD	427.77	27	.845	.793	.056	.035
NOR	960.755	27	.849	.798	.078	.048
NZL	700.156	27	.849	.799	.064	.040
PAN	694.494	27	.946	.928	.064	.034
PER	618.255	27	.957	.943	.06	.030
PHL	422.956	27	.961	.948	.045	.024
POL	784.933	27	.802	.736	.071	.045
PRT	721.52	27	.856	.808	.067	.042
QAT	3.225.793	27	.857	.810	.094	.051
QAZ	804.91	27	.913	.884	.066	.039
QCI	2842.39	27	.800	.733	.093	.056
QMR	343.063	27	.823	.764	.077	.05
QRT	551.531	27	.895	.859	.058	.038
ROU	626.414	27	.897	.863	.066	.04
RUS	690.549	27	.900	.866	.057	.035
SAU	931.655	27	.877	.836	.074	.044
SGP	832.22	27	.864	.818	.067	.04
SRB	860.967	27	.834	.779	.069	.044
SVK	632.819	27	.882	.842	.062	.038
SVN	745.283	27	.837	.783	.065	.041
SWE	530.732	27	.887	.849	.059	.037
TAP	695.925	27	.858	.811	.059	.036
THA	1.055.112	27	.938	.917	.067	.034
TUR	851.073	27	.902	.869	.067	.04
UKR	383.687	27	.919	.892	.047	.03
URY	642.587	27	.893	.858	.067	.041
USA	628.341	27	.842	.79	.068	.042
VNM	622.102	27	.915	.887	.064	.037

Anmerkung. *Skala wurde nicht administriert. *Anmerkung.* Grau hinterlegte Bildungssysteme erreichen in allen Indizes gute Werte.

Tabelle A4*Ergebnisse der konfirmatorischen Faktoranalysen für HEDRES*

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Overall Model	52.161.248	14	.860	.789	.079	.041
ALB	536.652	14	.856	.784	.077	.043
ARE	1.171.601	14	.915	.872	.066	.034
ARG	984.442	14	.861	.792	.077	.040
AUS	1439.61	14	.854	.782	.089	.046
AUT	269.249	14	.892	.838	.052	.030
BEL	666.671	14	.811	.717	.075	.042
BGR	439.748	14	.894	.841	.077	.038
BIH	561.907	14	.820	.730	.078	.042
BLR	209.021	14	.838	.758	.049	.030
BRA	566.095	14	.893	.839	.061	.034
BRN	432.251	14	.842	.763	.067	.038
CAN	2.078.037	14	.870	.805	.083	.043
CHE	377.483	14	.844	.766	.067	.038
CHL	546.554	14	.839	.758	.071	.039
COL	615.33	14	.853	.780	.077	.040
CRI	1.088.982	14	.737	.605	.103	.056
CZE	380.709	14	.900	.851	.061	.035
DEU	215.855	14	.912	.869	.055	.030
DNK	824.965	14	.766	.650	.088	.049
DOM	241.396	14	.924	.886	.054	.031
ESP	2.469.722	14	.818	.727	.070	.040
EST	575.578	14	.773	.659	.088	.056
FIN	249.399	14	.893	.839	.055	.032
FRA	395.34	14	.876	.814	.066	.037
GBR	772.601	14	.906	.859	.064	.033
GEO	228.939	14	.904	.855	.053	.031
GRC	378.351	14	.848	.772	.064	.037
HKG	711.11	14	.829	.743	.092	.048
HRV	223.918	14	.878	.817	.048	.028
HUN	415.418	14	.845	.767	.075	.043
IDN	1.406.886	14	.828	.742	.091	.055
IRL	324.138	14	.882	.824	.063	.034
ISL	230.605	14	.900	.850	.069	.041
ISR	858.404	14	.797	.695	.096	.056
ITA	403.143	14	.904	.856	.049	.027
JOR	458.118	14	.944	.915	.060	.030
JPN	492.699	14	.811	.717	.075	.042
KAZ	2.333.212	14	.818	.726	.092	.052
KOR	444.473	14	.875	.812	.068	.037
KSV	204.845	14	.892	.838	.052	.031
LBN	285.867	14	.868	.802	.059	.034
LTU	370.974	14	.851	.776	.061	.036
LUX	270.146	14	.906	.859	.059	.033

Tabelle A4*Ergebnisse der konfirmatorischen Faktoranalysen für HEDRES (Fortsetzung)*

LVA	428.633	14	.722	.582	.075	.045
MAC	473.575	14	.778	.666	.093	.051
MAR	336.77	14	.936	.904	.059	.029
MDA	147.999	14	.939	.909	.043	.024
MEX	1.067.325	14	.716	.574	.108	.062
MKD	312.81	14	.879	.819	.063	.036
MLT	261.914	14	.875	.813	.073	.040
MNE	780.677	14	.813	.720	.091	.048
MYS	885.132	14	.771	.656	.101	.063
NLD	289.045	14	.872	.808	.065	.037
NOR	469.468	14	.880	.820	.076	.043
NZL	545.484	14	.880	.820	.079	.041
PAN	532.16	14	.829	.744	.078	.047
PER	864.35	14	.798	.697	.100	.053
PHL	668.563	14	.837	.756	.081	.048
POL	287.466	14	.855	.783	.059	.036
PRT	372.399	14	.821	.731	.067	.038
QAT	770.324	14	.932	.898	.063	.032
QAZ	887.6	14	.745	.617	.097	.055
QCI	1.117.345	14	.823	.735	.081	.044
QMR	170.532	14	.814	.721	.075	.046
QRT	487.289	14	.808	.711	.077	.045
ROU	262.288	14	.919	.878	.060	.032
RUS	562.453	14	.844	.766	.073	.042
SAU	225.416	14	.943	.915	.051	.027
SGP	596.607	14	.846	.769	.079	.043
SRB	554.631	14	.825	.738	.077	.042
SVK	397.216	14	.914	.871	.068	.035
SVN	369.077	14	.845	.768	.063	.036
SWE	518.603	14	.834	.751	.082	.046
TAP	965.798	14	.807	.711	.097	.051
THA	835.708	14	.886	.829	.083	.042
TUR	453.601	14	.878	.817	.068	.037
UKR	345.708	14	.837	.756	.063	.036
URY	531.223	14	.844	.766	.086	.045
USA	508.617	14	.858	.788	.086	.045
VNM	529.306	14	.794	.692	.083	.046

Anmerkung. Grau hinterlegte Bildungssysteme erreichen in allen Indizes gute Werte.

Tabelle A5*Ergebnisse der konfirmatorischen Faktoranalysen für ICTRES*

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Overall Model	39.229.988	9	.921	.868	.085	.040
ALB	447.335	9	.921	.869	.088	.040
ARE	1.267.865	9	.910	.850	.086	.041
ARG	572.805	9	.894	.824	.073	.036
AUS	411.142	9	.911	.852	.059	.030
AUT	123.564	9	.938	.897	.043	.023
BEL	182.937	9	.937	.896	.048	.025
BGR	412.302	9	.838	.730	.093	.050
BIH	392.44	9	.842	.737	.082	.043
BLR	149.169	9	.922	.871	.052	.027
BRA	1.063.676	9	.844	.740	.106	.051
BRN	304.707	9	.931	.885	.070	.034
CAN	752.859	9	.912	.853	.062	.032
CHE	177.615	9	.909	.848	.057	.030
CHL	516.599	9	.895	.825	.087	.041
COL	789.735	9	.910	.850	.110	.045
CRI	549.82	9	.898	.830	.091	.043
CZE	244.974	9	.879	.798	.061	.033
DEU	123.654	9	.929	.882	.052	.027
DNK	226.567	9	.904	.839	.057	.030
DOM	467.49	9	.887	.812	.096	.045
ESP	1.043.622	9	.926	.877	.057	.030
EST	156.283	9	.870	.783	.056	.030
FIN	180.268	9	.839	.731	.058	.033
FRA	260.002	9	.919	.865	.067	.033
GBR	454.056	9	.899	.831	.061	.032
GEO	406.495	9	.843	.739	.090	.048
GRC	400.129	9	.842	.737	.083	.043
HKG	263.246	9	.905	.842	.069	.036
HRV	266.16	9	.852	.754	.066	.035
HUN	213.249	9	.862	.770	.067	.035
IDN	729.524	9	.941	.902	.082	.035
IRL	46.561	9	.971	.952	.027	.016
ISL	184.207	9	.839	.731	.078	.041
ISR	674.755	9	.744	.573	.114	.062
ITA	325.809	9	.908	.846	.055	.029
JOR	627.678	9	.918	.863	.088	.042
JPN	170.267	9	.923	.872	.054	.029
KAZ	848.508	9	.886	.810	.069	.036
KOR	340.166	9	.850	.750	.074	.038
KSV	385.193	9	.857	.762	.092	.046
LBN	382.087	9	.883	.805	.087	.047
LTU	460.017	9	.792	.653	.086	.046
LUX	172.384	9	.926	.876	.059	.031

Tabelle A5*Ergebnisse der konfirmatorischen Faktoranalysen für ICTRES (Fortsetzung)*

LVA	259.063	9	.812	.687	.073	.039
MAC	133.526	9	.907	.845	.061	.030
MAR	264.312	9	.958	.930	.065	.031
MDA	238.874	9	.930	.883	.069	.036
MEX	706.34	9	.903	.838	.109	.046
MKD	275.874	9	.870	.783	.074	.039
MLT	338.007	9	.698	.496	.105	.059
MNE	646.497	9	.823	.706	.104	.052
MYS	340.931	9	.925	.875	.078	.036
NLD	157.219	9	.871	.785	.059	.031
NOR	315.246	9	.884	.807	.078	.039
NZL	260.791	9	.912	.853	.068	.034
PAN	383.34	9	.928	.880	.083	.039
PER	370.743	9	.953	.922	.082	.034
PHL	319.374	9	.947	.911	.069	.031
POL	301.787	9	.805	.675	.076	.039
PRT	384.887	9	.841	.735	.086	.044
QAT	1.044.702	9	.879	.799	.092	.047
QAZ	567.929	9	.887	.811	.097	.047
QCI	1.092.731	9	.872	.787	.100	.049
QMR	125.759	9	.871	.786	.081	.046
QRT	282.638	9	.871	.784	.073	.038
ROU	370.391	9	.862	.771	.089	.045
RUS	346.752	9	.894	.824	.071	.036
SAU	466.433	9	.886	.810	.092	.048
SGP	335.536	9	.899	.831	.074	.037
SRB	395.891	9	.820	.700	.081	.044
SVK	269.872	9	.891	.819	.070	.037
SVN	340.675	9	.786	.643	.076	.041
SWE	309.211	9	.809	.682	.078	.041
TAP	352.642	9	.878	.797	.073	.036
THA	788.072	9	.909	.848	.100	.043
TUR	390.824	9	.927	.879	.079	.037
UKR	180.297	9	.923	.872	.056	.029
URY	333.183	9	.897	.829	.084	.043
USA	193.167	9	.920	.866	.065	.034
VNM	317.228	9	.914	.856	.080	.038

Anmerkung. Grau hinterlegte Bildungssysteme erreichen in allen Indizes gute Werte.

Tabelle A6*Ergebnisse der konfirmatorischen Faktoranalysen für HOMEPOS*

Model	χ^2	df	CFI	TLI	RMSEA	SRMR
Overall Model	920.346.181	209	.641	.604	.086	.077
ALB	6.962.733	209	.708	.677	.072	.060
ARE	35.664.412	209	.540	.492	.095	.090
ARG	12.867.063	209	.690	.657	.071	.060
AUS	16.639.588	209	.590	.547	.078	.070
AUT	8.291.215	209	.603	.561	.076	.070
BEL	8.818.155	209	.625	.586	.070	.064
BGR	7.072.236	209	.664	.629	.079	.071
BIH	7.439.962	209	.609	.568	.074	.066
BLR	4.307.848	209	.667	.632	.058	.050
BRA	12.095.985	209	.703	.671	.073	.062
BRN	5.128.748	209	.715	.685	.059	.051
CAN	29.766.904	209	.588	.544	.081	.073
CHE	6.944.848	209	.572	.527	.075	.069
CHL	9.649.938	209	.677	.643	.078	.063
COL	9480.83	209	.737	.709	.078	.063
CRI	9.237.856	209	.701	.670	.078	.062
CZE	9407.78	209	.611	.570	.079	.073
DEU	5.897.975	209	.640	.603	.076	.068
DNK	8.897.054	209	.609	.568	.074	.070
DOM	5.717.142	209	.738	.710	.069	.056
ESP	35.608.189	209	.660	.624	.069	.060
EST	5.659.138	209	.649	.612	.071	.066
FIN	4.996.182	209	.709	.678	.064	.060
FRA	9.515.548	209	.609	.568	.085	.078
GBR	18.503.319	209	.590	.546	.081	.070
GEO	5.836.684	209	.688	.655	.070	.061
GRC	7.255.919	209	.599	.556	.073	.064
HKG	7.969.925	209	.649	.612	.079	.067
HRV	6.168.058	209	.648	.611	.066	.061
HUN	6.361.145	209	.681	.647	.076	.065
IDN	12605.64	209	.743	.716	.070	.059
IRL	5.541.081	209	.636	.597	.068	.060
ISL	4.065.041	209	.651	.614	.076	.071
ISR*	NA	NA	NA	NA	NA	NA
ITA	11.177.646	209	.622	.583	.068	.059
JOR	9.630.033	209	.754	.728	.071	.058
JPN	6.079.906	209	.588	.544	.068	.060
KAZ	21.239.996	209	.657	.621	.072	.059
KOR	6.681.877	209	.711	.681	.068	.057
KSV	4.311.634	209	.660	.624	.063	.056
LBN	4.296.617	209	.745	.718	.059	.053
LTU	8.096.407	209	.635	.596	.074	.069

Tabelle A6*Ergebnisse der konfirmatorischen Faktoranalysen für HOMEPOS (Fortsetzung)*

LUX	7.055.018	209	.634	.595	.079	.071
LVA	6.029.815	209	.582	.538	.073	.066
MAC	5.309.016	209	.593	.55	.080	.070
MAR	6.176.413	209	.787	.764	.065	.054
MDA	5.608.419	209	.688	.655	.069	.059
MEX	10.042.526	209	.689	.656	.085	.071
MKD	5.519.771	209	.657	.621	.068	.063
MLT	4.392.514	209	.563	.517	.078	.069
MNE	9.458.484	209	.601	.559	.082	.074
MYS	5.954.434	209	.712	.682	.067	.056
NLD	4.932.442	209	.605	.564	.069	.064
NOR	8.974.158	209	.587	.543	.086	.082
NZL	7.961.298	209	.618	.578	.078	.069
PAN	6.764.912	209	.745	.718	.072	.063
PER	8.486.546	209	.743	.716	.081	.067
PHL	5.777.353	209	.786	.763	.061	.054
POL	6.344.897	209	.608	.567	.072	.064
PRT	6.153.629	209	.660	.624	.071	.061
QAT	28.753.221	209	.473	.417	.100	.101
QAZ	8.512.445	209	.646	.609	.077	.067
QCI	15.596.057	209	.660	.625	.078	.065
QMR	2.557.404	209	.593	.550	.075	.069
QRT	7.198.131	209	.573	.528	.076	.069
ROU	6.510.508	209	.668	.633	.077	.066
RUS	9.065.776	209	.602	.56	.075	.066
SAU	6.898.311	209	.673	.639	.072	.063
SGP	7.975.582	209	.644	.606	.075	.062
SRB	8.506.747	209	.601	.559	.078	.072
SVK	7.471.249	209	.666	.630	.077	.067
SVN	7.617.198	209	.614	.573	.075	.070
SWE	7.038.367	209	.610	.569	.078	.076
TAP	9.076.263	209	.666	.630	.077	.065
THA	12.837.419	209	.726	.698	.084	.067
TUR	7.787.158	209	.738	.710	.073	.056
UKR	6218.39	209	.597	.555	.069	.062
URY	5.826.596	209	.725	.696	.073	.061
USA	6.570.414	209	.635	.596	.080	.068
VNM	5.216.668	209	.704	.673	.070	.060

Anmerkung. *Skala wurde nicht administriert. *Anmerkung.* Grau hinterlegte Bildungssysteme erreichen in allen Indizes gute Werte.