

Schulz, Maike

Working Paper

Messartefakte bei der Erfassung der Gesundheit von Migranten in Deutschland: Zur interkulturellen Äquivalenz des SF-12-Fragebogen im Sozio-oekonomischen Panel (SOEP)

SOEPPapers on Multidisciplinary Panel Data Research, No. 447

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Schulz, Maike (2012) : Messartefakte bei der Erfassung der Gesundheit von Migranten in Deutschland: Zur interkulturellen Äquivalenz des SF-12-Fragebogen im Sozio-oekonomischen Panel (SOEP), SOEPPapers on Multidisciplinary Panel Data Research, No. 447, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/59013>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

SOEPpapers

on Multidisciplinary Panel Data Research

SOEP – The German Socio-Economic Panel Study at DIW Berlin

447-2012

Messartefakte bei der Erfassung der Gesundheit von Migranten in Deutschland: Zur interkulturellen Äquivalenz des SF-12- Fragebogen im Sozio-oekonomischen Panel (SOEP)

Maike Schulz

SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
<http://www.diw.de/soeppapers>

Editors:

Jürgen **Schupp** (Sociology, Vice Dean DIW Graduate Center)
Gert G. **Wagner** (Social Sciences)

Conchita **D'Ambrosio** (Public Economics)
Denis **Gerstorff** (Psychology, DIW Research Professor)
Elke **Holst** (Gender Studies)
Frauke **Kreuter** (Survey Methodology, DIW Research Professor)
Martin **Kroh** (Political Science and Survey Methodology)
Frieder R. **Lang** (Psychology, DIW Research Professor)
Henning **Lohmann** (Sociology, DIW Research Professor)
Jörg-Peter **Schräpler** (Survey Methodology, DIW Research Professor)
Thomas **Siedler** (Empirical Economics)
C. Katharina **Spieß** (Empirical Economics and Educational Science)

ISSN: 1864-6689 (online)

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | soeppapers@diw.de

Messartefakte bei der Erfassung der Gesundheit von Migranten in Deutschland

Zur interkulturellen Äquivalenz des SF-12-Fragebogen
im Sozio-oekonomischen Panel (SOEP)

Maike Schulz, M.A.*

Zusammenfassung

Nur wenige Studien geben Aufschluss über den Gesundheitszustand der in Deutschland lebenden Migrantenbevölkerung und beschränken sich meist auf rein medizinische Kriterien. Die vorliegende Arbeit vergleicht auf Grundlage von Daten des Sozio-oekonomischen Panels (SOEP) den Gesundheitszustand spezifischer Migrantengruppen mit dem der einheimischen Bevölkerung anhand des SF-12, einem mehrdimensionalen Messinstrument der gesundheitsbezogenen Lebensqualität. Der Schwerpunkt der Analysen liegt allerdings darauf, über die methodische Integration von gruppenvergleichender Faktoranalyse (MGCF), der Schätzung latenter Mittelwerte (MACS) und dem Einbezug von Kovariaten (MIMIC) zu testen, ob der SF-12 interkulturelle Äquivalenz aufweist und dementsprechend gültige Vergleiche zwischen der einheimischen Bevölkerung und verschiedenen Migrantengruppen zulässt.

Die Analysen zeigen, dass sich der SF-12 zwar als interkulturell valide erweist, allerdings finden sich unabhängig vom kulturellen Hintergrund Verzerrungen aufgrund der Frageformulierungen und der Antwortskala, die eine Modifikation des Messmodells nahelegen. Weiterhin lässt sich ein hinsichtlich einzelner soziostruktureller Merkmale unterschiedliches Antwortverhalten (Differential Item Functioning (DIF)) identifizieren. Unter Berücksichtigung dieser Drittvariableneffekte und DIF lassen sich schließlich keine signifikanten Unterschiede hinsichtlich des Gesundheitszustandes zwischen Migranten und Nichtmigranten feststellen. Es bedarf jedoch weiterer Forschung zum Einfluss alternativer Kovariaten auf die Messäquivalenz und zu den konkreten Ursachen der gefundenen DIF-Effekte.

Schlüsselwörter: measurement equivalence, Migration, Gesundheit, SF-12, health, differential item functioning

* Bremen International Graduate School of Social Sciences (BIGSSS)
Universität Bremen
Wiener Straße/Celsiusstraße (FVG)
28334 Bremen
mschulz@bigsss-bremen.de

Inhaltsverzeichnis

1	Hintergrund	4
1.1	Measurement Equivalence	6
1.2	Ursachen geringer Messäquivalenz.....	6
1.3	Methodik.....	8
1.4	SF-36 und SF-12.....	9
1.5	Forschungsfragen und Hypothesen.....	10
2	Datenbasis und Messinstrument	16
2.1	Überprüfung der Grundannahmen.....	17
3	Analyse der Äquivalenz des SF-12 zwischen der Bevölkerung mit und ohne Migrationshintergrund	18
3.1	Soziodemografische und -ökonomische Indikatoren.....	18
3.2	Der SF-12 Fragebogen.....	18
3.3	Modellierung des SF-12-Gesundheitskonstrukts.....	19
3.4	Psychometrische Eigenschaften des Grundmodells	21
3.5	Test des Grundmodells auf konzeptuelle Äquivalenz	22
3.6	Einführung von Kovariaten	25
3.7	Test auf Differential Item Functioning (DIF) der Kovariaten.....	27
3.8	Äquivalenzrestriktionen und Modellfit.....	32
4	Zusammenfassung	34
5	Ausblick	37
6	Literatur	39
7	Anhang	49

Tabellen

Tabelle 1: Verteilung der Befragten nach Herkunftsregion	18
Tabelle 2: Modellfit des Grundmodells nach Herkunftsregion	23
Tabelle 3: Modellfit des modifizierten Modells nach Herkunftsregion	24
Tabelle 4: Endgültig spezifizierte DIF-Effekte der Kovariaten.....	29
Tabelle 5: Schrittweise Äquivalenzrestriktion der Parameter zwischen den Gruppen.....	32
Tabelle 6: Schätzung der latenten Mittelwert der physischen und mentalen Dimension.....	34
Tabelle 7: Kriterien zur Beurteilung der Modellgüte.....	49
Tabelle 8: Lokale Kriterien zur Beurteilung des Basismodells	49
Tabelle 9: Verteilungen der SF-12-Variablen nach Herkunftsregion	50
Tabelle 10: Mittelwerte der SF-12-Variablen nach Herkunftsregion.....	53
Tabelle 11: Einfluss der Kovariaten auf die physische und mentale Gesundheit nach Herkunftsregion.....	54
Tabelle 12: Einfluss der Kovariaten auf die physische und mentale Gesundheit nach Herkunftsregion.....	55
Tabelle 13: Schätzung der latenten Mittelwerte der physischen und mentalen Gesundheit	57
Tabelle 14: Schätzungen der gleichheitsrestringierten Faktorladungen, Varianzen und Achsenabschnitte und Mittelwerte	58

Abbildungen

Abbildung 1: Grundmodell mit zwei latenten Konstrukten und 12 Indikatoren.....	21
Abbildung 2: Modifiziertes Grundmodell mit zwei Methodenfaktoren,	25
Abbildung 3: Modifiziertes Grundmodell mit signifikanten Kovariaten und spezifizierten DIF- Effekten.....	31

1 Hintergrund

Die Migrantenpopulation in Deutschland umfasste im Jahr 2008 mit 15,6 Millionen insgesamt knapp 19% der gesamten Bevölkerung Deutschlands (Statistisches Bundesamt 2010: 48). Personen mit Migrationshintergrund unterscheiden sich in ihrer Migrationserfahrung (direkt oder indirekt), den Migrationsmotiven (Oltmer 2009; Statistisches Bundesamt 2010: 51), dem rechtlichen Status (Eingebürgerte, Ausländer, Spätaussiedler, Asylsuchende)(Statistisches Bundesamt 2010: 51), dem Herkunftsland (von Gostomski 2010), soziodemografischen und sozioökonomischen Merkmalen (Bundesministerium für Familie 2008; Gesundheitsberichterstattung des Bundes 2008: 15; von Gostomski 2010) und dem Grad der der kulturellen Integration (Haug und Sauer 2007; von Gostomski 2008). Zudem verändert sich die Migrantenbevölkerung aufgrund der bestehenden starken Fluktuation ständig (Statistisches Bundesamt 2010: 67).

Der Migrationsprozess einerseits wie auch der Migrantensstatus andererseits prägen den gesamten Lebenslauf: Direkte Migrationserfahrungen sind oft mit dem Verlust und Neuaufbau sozialer Kontakte verbunden, und auch die nachfolgenden Generationen ohne direkte Migrationserfahrung sind mit dem der Trennung von Familienangehörigen konfrontiert.

Aufgrund geringer Bildung der ersten Migrantengeneration und geringer Deutschkenntnisse der zweiten Generation sowie auch teilweise durch Fremdenfeindlichkeit und Diskriminierung weist die Migrantenbevölkerung einen niedrigeren sozialen Status als die deutsche Bevölkerung ohne Migrationshintergrund auf (Gesundheitsberichterstattung des Bundes 2008: 17ff.; Lampert und Ziese 2005: 194ff.).

All diese Faktoren – der Migrationsprozess wie auch der Migrationsstatus selbst, die soziale Ungleichheit sowie kulturelle Barrieren und Fremdenfeindlichkeit – beeinflussen in komplexer Wechselwirkung den Gesundheitszustand. Es existieren verschiedene Modelle zur Erklärung der Gesundheit von Migranten wie die Theorie der positiven Selektion bzw. der Healthy-migrant-Effekt, das Modell des gesundheitlichen Übergangs (Razum 2006; Razum und Twardella 2002; Spallek und Razum 2008), die Migrations-Stress-Hypothese (Faltermeier 2005), die Theorie der sozialen Unterprivilegierung (Groenemeyer 2003) und die Erklärung durch Zugangsbarrieren zur Gesundheitsversorgung (Razum et al. 2004; Salman und Djomo 2009: 558). Im Hinblick auf das Auftreten der verschiedenen Einflussfaktoren über den Lebenslauf hinweg ist auch die Berücksichtigung der zeitlichen Dimension wichtig (Spallek und Razum 2008: 285).

Allerdings sind die bestehenden Daten zur gesundheitlichen Situation der Migrantenbevölkerung bislang ungenau, lückenhaft und berücksichtigen nicht die Heterogenität der Migrantenpopulation (Gesundheitsberichterstattung des Bundes 2008: 131; Kirkcaldy et al. 2006: 881; Kohls 2008: 11f.; Zeeb und Razum 2006: 850): Die Gesundheitsberichterstattung des Bundes wird in der Regel nicht separat für die Bevölkerung mit Migrationshintergrund

ausgewertet und beinhaltet meist nur Teilgruppen der Migrantenpopulation (vgl. Cornelißen 2005: 517; Gesundheitsberichterstattung des Bundes 2008: 31). Auch sind kaum epidemiologische Daten verfügbar, die eine Analyse der Gesundheitsdeterminanten der Migrantenbevölkerung ermöglichen (Schenk und Neuhauser 2005; Zeeb et al. 2004). Bestehende Untersuchungen zeigen auf, dass Migranten gegenüber Nichtmigranten eine höhere Tuberkulose-, Hepatitis B- und HIV-Prävalenz haben (Gesundheitsberichterstattung des Bundes 2008: 107ff.), häufiger an Magen- oder Darmkrebs erkranken, eine schlechtere Mundgesundheit (Schenk und Knopf 2007; van Steenkiste 2004; van Steenkiste et al. 2004) aufweisen und Fettleibigkeit im Jugendalter deutlich ausgeprägter ist. Migranten sind zudem oftmals psychischen Belastungen ausgesetzt und häufiger von psychosomatische Beschwerden betroffen (Gesundheitsberichterstattung des Bundes 2008: 107ff.; Wittig et al. 2004: 90). Sie sind durchweg weniger zufrieden mit der Gesundheit, wobei die Unterschiede zur Bevölkerung ohne Migrationshintergrund im Alter weiter zunehmen. Außerdem weisen Migranten aufgrund von Kommunikationsproblemen und einem unterschiedlichen Krankheitsverständnis eine schlechtere Gesundheitsversorgung auf (Gesundheitsberichterstattung des Bundes 2008: 107ff.). Allerdings zeigen sich bei Kindern und Jugendlichen mit Migrationshintergrund seltener Allergien (Gesundheitsberichterstattung des Bundes 2008: 107ff.), die Migrantenbevölkerung leidet seltener an Herz-Kreislauf-Erkrankungen (Kohls 2008: 25ff.) und weist eine geringere Mortalität als die Bevölkerung des Einwanderungslands auf (Razum 2006).

Informationen zum subjektiven Gesundheitszustand der Migrantenbevölkerung finden sich im Vergleich zu objektiven, diagnostischen Daten bislang eher selten. Dennoch gilt es, in solchen Fällen zu prüfen, inwiefern sich deren kultureller Hintergrund auf die Beantwortung gesundheitsbezogener Messinstrumente auswirkt und ob demnach die Daten von Migranten und Nichtmigranten vergleichbar sind: Kulturelle Unterschiede im Umgang mit Gesundheit und Krankheit werden zwar erkannt, bleiben jedoch in der Regel unberücksichtigt. So heißt es im Schwerpunktbericht der Gesundheitsberichterstattung des Bundes „Migration und Gesundheit“: „In Selbstangaben zur subjektiven Gesundheit gehen die kulturell geprägte Wahrnehmung von Symptomen, eine Einschätzung des Wohlbefindens, der Funktionsfähigkeit und der Anfälligkeit für Krankheiten (Vulnerabilität) ein.“ (Gesundheitsberichterstattung des Bundes 2008: 50).

Diese Studie widmet sich daher der Messung des subjektiven Gesundheitszustandes und überprüft deren kulturübergreifende Validität bei Befragten mit und ohne Migrationshintergrund.

1.1 Measurement Equivalence

Im Bereich der kulturvergleichenden Forschung hat sich in diesem Zusammenhang das Konzept der „cross-cultural equivalence“² (Johnson 1998) etabliert: Vergleichende Analysen zwischen Bevölkerungsgruppen unterschiedlicher Nation, Kultur und Religion sind nur dann sinnvoll, wenn das entsprechende Messinstrument kulturübergreifende Validität aufweist. Allgemeiner kann das Äquivalenzkonzept unter den Begriff „measurement equivalence“ (Singh 1995: 603) gefasst werden.

In der Praxis wird selten überprüft, ob die Indikatoren eines Fragebogens unabhängig vom kulturellen Hintergrund und den soziodemografischen und -ökonomischen Charakteristika eines Befragten gleich valide sind; meist werden bestehende Messinstrumente für den Einsatz in einer anderen Kultur lediglich übersetzt und die Äquivalenz der inhaltlichen Bedeutung für alle Befragtengruppen unhinterfragt angenommen (Vandenberg und Lance 2000: 4). In den letzten Jahren finden sich zunehmend Studien – allerdings überwiegend aus dem anglo-amerikanischen Raum – die eine empirische Prüfung dieser Äquivalenzannahme durchführen (Schmitt und Kuljanin 2008: 212).

1.2 Ursachen geringer Messäquivalenz

Mangelnde Messäquivalenz lässt sich einerseits auf kulturelle Unterschiede in der Validität des operationalisierten Konstrukts, andererseits Antwortstile, individuelle Befragtenmerkmale, Interviewereffekte³ und Eigenschaften des Erhebungsinstruments zurückführen.

In Bezug auf die unterschiedliche Validität eines gesundheitsbezogenen Messinstruments belegen zahlreiche Studien das kulturell geprägte Verständnis von Gesundheit und Krankheit (Flick et al. 1998; Mattes 1998). In den Herkunftskulturen der nach Deutschland Eingewanderten werden Gesundheit und Krankheit ganzheitlicher betrachtet und sind teilweise auch religiös konnotiert (Gesundheitsberichterstattung des Bundes 2008: 110). So verstehen türkische Migranten Krankheiten eher als external verursacht und den Körper ganzheitlich betreffend. Vermutungen über die Krankheitsursachen sind oft mystischer Art, sie werden zum Beispiel als überirdische Bestrafung angesehen. Psychische Belastungen und Erkrankungen werden häufig in körperliche Beschwerden verpackt, da nicht-somatische Beschwerden mit Geisteskrankheit in Verbindung gesetzt und deshalb verleugnet werden (Yildirim-Fahlbusch 2003). Folglich

² Der Begriff „cross-cultural equivalence“ fokussiert die Vergleichbarkeit bereits auf das Kriterium der Kultur. Allgemeiner wird das Konzept unter der Bezeichnung „measurement equivalence“ oder „measurement invariance“ gefasst (vgl. Vandenberg und Lance 2000: 4).

³ Etwaige Interviewereffekte (vgl. Kroh 2005) werden in dieser Arbeit nicht weiter berücksichtigt, da diese im Rahmen der Analysen nicht ermittelt werden können.

zeigen Migranten einen anderen Umgang mit Krankheitssymptomen, eine andere Offenlegung von Beschwerden und Problemen (Kirkcaldy et al. 2006: 880) und sie zeigen vermutlich auch eine andere Beurteilung des eigenen Gesundheitszustandes.

Cerda-Hegerl (2008) weist auf deutliche kulturelle Unterschiede hinsichtlich der Strategien zur Krankheitsbewältigung (individuell vs. kollektiv), der Kommunikation (verbal vs. nonverbal) von Beschwerden, der Expressivität von Schmerz und Unwohlsein sowie hinsichtlich des Ausmaßes, Empathie und Interesse zu zeigen.

Zu den verschiedenen Formen von Antwortstilen zählen Effekte sozialer Erwünschtheit, sogenannte Response-Sets (Aquieszenz, extremes Antwortverhalten, Tendenz zur Mitte), und Meinungslosigkeit (vgl. Diekmann 2007: 447ff.; Schnell et al. 2008: 353).

In zahlreichen Studien wurde gezeigt, dass Antwortstile im Ländervergleich systematisch variieren (Clarke 2001; Harzing 2005; Hui und Triandis 1989b; Johnson et al. 2005; Johnson und van de Vijver 2003; Locke und Baik 2009; Robert et al. 2006; Smith 2004). Ursachen für diese Unterschiede werden insbesondere in kulturspezifischen Wertvorstellungen gesehen, die sich in der Kommunikation ausdrücken: So weisen Befragte aus südeuropäischen Ländern mit kollektivistisch geprägter Kultur häufiger aquieszentes und extremes Antwortverhalten auf (Marin et al. 1992; van Herk et al. 2004). In den USA weisen Hispanics gegenüber Nicht-Hispanics eher extremes Antwortverhalten auf (Hui und Triandis 1989a), wobei ein hoher Grad an Akkulturation dieser Tendenz entgegenwirkt (Marin et al. 1992). Japaner und Koreaner neigen hingegen eher zur Wahl der Mittelkategorie (Chen et al. 1995; Lee und Green 1991). Um kulturelle Integration und Integrationswillen zu beweisen, orientieren sich Migranten in ihrem Antwortverhalten teilweise auch an der Öffentlichkeitsmeinung und geben dadurch sozial erwünschte und aquieszente Antworten (Aschauer 2009: 307).

Unterschiedliche Antwortstile können außerdem in individuellen Eigenschaften, z.B. der Soziodemografie oder der Persönlichkeit begründet sein: Unabhängig vom Frageninhalt belegen Studien für verschiedene Länder negative Effekte der Bildung und des Einkommens auf das Auftreten extremen oder aquieszenten Antwortverhaltens (Marin et al. 1992; Meisenberg und Williams 2008). Männer neigen eher zu extremen Antworten, wohingegen Frauen eine Tendenz zur Mitte aufweisen (Harzing 2006; Johnson et al. 1997). Durch die im Alter abnehmende Gedächtnisleistung geben ältere Menschen bei Erinnerungsfragen eher verzerrte Antworten, sind aber weniger anfällig für Fragereihenfolgeeffekte als jüngere Menschen (Schwarz et al. 2008: 26). Es zeigt sich auch ein deutlicher Zusammenhang zwischen Extraversion und der Tendenz, extrem positive Antworten zu geben (Harzing 2006).

Auch auf gesundheitsbezogene Fragen treffen diese Befunde zu: Durch den „scale of

reference bias“ (Groot 2003), d.h. im konkreten Fall durch die Orientierung am Gesundheitszustand von Personen der eigenen Altersgruppe geben ältere eine unverhältnismäßig bessere subjektive Gesundheit (Fleishman und Lawrence 2003; Groot 2003) und körperliche Vitalität bzw. Beeinträchtigung (Fleishman et al. 2002; Groenvold et al. 1995) an als jüngere Befragte. Das Auftreten physischer Symptome wird bei älteren Menschen allerdings weniger verzerrt wiedergegeben als bei jüngeren Menschen, da die Erinnerung bei Älteren durch das häufigere Auftreten präsenter ist (Schwarz et al. 2008: 26). Es besteht eine positivere Beurteilung des Gesundheitszustandes bei Männern (d'Uva et al. 2008; Jylhä et al. 1998; vgl. Little 1997), bei Personen mit geringem Einkommen und mit städtischem Wohnort (d'Uva et al. 2008).

Zum Einfluss von Eigenschaften des Fragebogens auf das Antwortverhalten wurde gezeigt, dass eine komplizierte Frageformulierung mit der Tendenz zur Wahl der Mittelkategorie zusammenhängt (Velez und Ashworth 2007) und Antwortvorgaben oftmals als Anker fungieren und somit zu Über- oder Unterschätzung führen (Meadows et al. 2000). Außerdem führt die sprachliche Übersetzung von Antwortkategorien nicht zwangsläufig zu einer äquivalenten Verwendung der Skala (Harzing 2006). Breite Antwortskalen können teilweise das Ausmaß verzerrender Antwortstile verringern (Hui und Triandis 1989a; Kieruj und Moors 2010). Auch wurden Interaktionseffekte zwischen Fragenbogen- und demografischen Merkmalen nachgewiesen (Stern et al. 2007). Sprachbarrieren führen zu missverstandenen Fragebogeninhalten (Robert et al. 2006: 69) und rufen verzerrende Antwortstile hervor (Harzing 2006; Jenkinson et al. 2001).

1.3 Methodik

Mean and Covariance Structure Analysis (MACS)-Modelle erweitern einfache Strukturgleichungsmodelle um die Schätzung der Mittelwerte der manifesten Indikatoren sowie der messfehlerbereinigten Mittelwerte der latenten Konstrukte (vgl. Ployhart und Oswald 2004: 29).

Ein anderer Strukturgleichungsansatz, das so genannte „Multiple Indicators Multiple Causes“ (MIMIC)-Modell, ermöglicht die Analyse von Kovariaten sowohl hinsichtlich ihres substantiellen Effekts auf das latente Konstrukt als auch in Bezug auf verzerrende Effekte der Kovariaten auf die Beantwortung der Items (sog. Differential Item Functioning (DIF)). Eine wichtige Einschränkung von MIMIC-Modellen im Gegensatz zu MACS-Modellen liegt darin, lediglich uniforme Verzerrungen aufdecken zu können (Marsh et al. 2006: 805).

Um die Vorteile von MACS- und MIMIC-Modellen zu vereinen, werden in der vorliegenden Arbeit beide Ansätze integriert. Erweisen sich aufgrund der Analysen die Items zwischen den Gruppen als äquivalent und wird sowohl für den Einfluss der Drittvariablen auf die latenten Konstrukte als auch für DIF im Modell kontrolliert, können latente Mittelwerte der

Gesundheitskonstrukte geschätzt und gültig verglichen werden. Die mathematischen Grundlagen der MACS- und MIMIC-Modelle sind in der folgenden Darstellung ersichtlich. Die entsprechenden Gütekriterien zur Beurteilung der Modellanpassung sind in Tabelle 7 (s. Anhang) ersichtlich.

$$Y \text{ Messmodell: } y^{(g)} = \tau^{(g)} + \lambda^{(g)} \eta^{(g)} + \varepsilon^{(g)}$$

$$\text{Strukturmodell: } \eta^{(g)} = \alpha^{(g)} + \beta^{(g)} \eta^{(g)} + \gamma^{(g)} \xi^{(g)} + \zeta^{(g)}$$

$$\text{Einfluss Kovariate - latentes Konstrukt: } \eta^{(g)} = \gamma^{(g)} X^{(g)} + \zeta^{(g)}$$

$$\text{Einfluss Kovariate - Item mit DIF: } y^{(g)} = \tau^{(g)} + \lambda^{(g)} \eta^{(g)} + \beta^{(g)} X^{(g)} + \varepsilon^{(g)}$$

wobei

y eine $r \times 1$ Matrix der q einbezogenen Variablen,
 τ ein $r \times 1$ Vektor der Achsenabschnitte der Variablen,
 λ eine $r \times n$ Matrix der Faktorladungen von n Faktoren,
 η ein $n \times 1$ Vektor des Summenindexes des latenten unabhängigen Konstrukts,
 ε ein $r \times 1$ Vektor der Messfehlerkomponente ist

und

α ein $m \times 1$ Vektor des latenten Achsenabschnitts,
 η ein $m \times 1$ Vektor der m latenten abhängigen Variablen,
 β eine $m \times m$ Matrix, die die Verbindung (Regressionskoeffizienten) zwischen den latenten abhängigen Konstrukten widerspiegelt,
 γ eine $m \times n$ Matrix, die die Verbindung zwischen den latenten abhängigen und den latenten unabhängigen Konstrukten darstellt,
 ζ die latente Messfehlerkomponente widerspiegelt

sowie

X eine Kovariate,
 β die Ladung der Kovariaten auf das endogene Item y ist.
 (g) indiziert jeweils die Gruppenzugehörigkeit.

Quelle: vgl. Hancock et al. (2000: 536); Ployhart und Oswald (2004: 30f.)

1.4 SF-36 und SF-12

1960 wurde im Rahmen der Medical Outcome Study des Forschungsinstituts RAND (Research and Development) in den Vereinigten Staaten ein standardisiertes, generisches Messinstrument entwickelt, das die „gesundheitsbezogene Lebensqualität“ in seiner Vielschichtigkeit abbilden soll: Der Short Form 36 Health Survey (SF-36) erfasst über 36 Items auf acht Subskalen den aktuellen, selbstbeurteilten Gesundheitszustand. Den acht Subskalen liegt eine zweidimensionale Aufteilung in eine physische und eine psychische Dimension zugrunde (vgl. Bullinger 2000: 192). Um die Bearbeitungszeit zu reduzieren und somit die Praktikabilität des Fragebogens zu erhöhen, wurde unter anderem eine Kurzform mit 12 Items (SF-12 und die überarbeitete Version SF-12v2) entwickelt (Ware et al. 1996).

Dem SF-12 wird größtenteils eine gute Validität für die Anwendung in verschiedenen Ländern

(Amir et al. 2002; Gandek et al. 1998; Kontodimopoulos et al. 2007; Lam et al. 2005; Montazeri et al. 2009), für verschiedene Behandlungsgruppen (Gandhi et al. 2001; Hoffmann et al. 2005; Lim und Fisher 1999; Luo et al. 2003) sowie Stabilität im Zeitverlauf (Salyers et al. 2000) bescheinigt sowie eine gute Responsivität gegenüber objektiven Veränderungen im Gesundheitszustand (Cheak-Zamora et al. 2009; Luo et al. 2003). Allerdings weisen einige Studien auf Korrelationen zwischen den Dimensionen physische und mentale Gesundheit hin, die ursprünglich als unkorreliert angenommen wurden (Farivar et al. 2007) sowie auf Mischladungen zwischen den latenten Gesundheitsdimensionen und den Indikatoren der gegensätzlichen Dimension (Gandhi et al. 2001; Maurischat und Krüger-Bödeker 2004). Morfeld et al. (2003) weisen außerdem auf systematisch fehlende Werte von Älteren und höher Gebildeten hin. Jenkinson et al. (2001) finden signifikante Unterschiede in der Interpretation der Fragen und der Antwortkategorien zwischen verschiedenen ethnischen Gruppen in Großbritannien mit unterschiedlichen Englischkenntnissen. Fleishman et al. (2003) identifizieren „Differential Item Functioning“ (DIF) hinsichtlich des Geschlechts und des Alters der Befragten: Männer und Befragte im Alter von 60 bis 69 Jahren schätzen ihren mentalen Gesundheitszustand unverhältnismäßig gut ein.

1.5 Forschungsfragen und Hypothesen

Die zugrundeliegende Fragestellung dieser Arbeit ist dementsprechend inhaltlicher wie auch methodologischer Art: Grundsätzlich wird untersucht, inwieweit sich die in Deutschland lebende Bevölkerung mit Migrationshintergrund hinsichtlich ihres Gesundheitszustandes, gemessen über den SF-12 Fragebogen, von der Bevölkerung ohne Migrationshintergrund unterscheidet.

Da solcherlei Gruppenvergleiche erst dann gültig sind, wenn vorab die Konstruktvalidität des SF-12 sowie dessen kulturübergreifende Vergleichbarkeit bestätigt werden kann, untersucht die zweite Forschungsfrage, ob Verzerrungen in der Beantwortung des SF-12 aufgrund des kulturellen Hintergrundes oder sozioökonomischer und soziodemografischer Charakteristika bestehen. Den ersten Schritt der Analyse stellt die Überprüfung der angenommenen dimensional Struktur des SF-12 Messinstruments dar.

H ₁ : Die theoretische Modellstruktur des SF-12 wird durch die empirischen Daten über alle Befragten hinweg repräsentiert.

Daraufhin prüft der Test auf konzeptuelle Äquivalenz, ob das Modell in allen Gruppen eine annähernd gleiche dimensionale bzw. konzeptuelle Struktur aufweist.

$$H_2: \Sigma^{g^1} = \Sigma^{g^2} = \dots = \Sigma^{g^6}$$

Es besteht konzeptuelle Äquivalenz des SF-12-Basismodells für alle nach Herkunftsregion spezifizierten Gruppen.

Es wurden in Studien bislang unter Anderem Effekte des Geschlechts, des Alters, der Bildung des Einkommen sowie des Erwerbsstatus' auf den Gesundheitszustand belegt und auch in der konkreten Analyse vermutet. So weisen Männer zwar eine niedrigere Lebenserwartung als Frauen auf, Frauen führen jedoch häufiger Krankheiten und Beschwerden an als Männer (Hammelstein et al. 2006). Da diese empirischen Befunde für zahlreiche Länder und Kulturen übereinstimmen (Aquino et al. 1992; Hraba et al. 1996; Krause et al. 1998; Lahelma et al. 1999; Rahman et al. 1994), wird im Folgenden angenommen, dass in allen Herkunftsgruppen Frauen einen schlechteren Gesundheitszustand aufweisen als Männer.

H₃: Frauen weisen eine schlechtere physische und mentale Gesundheit auf als Männer.

Zusätzlich zum direkten Effekt des Geschlechts auf den Gesundheitszustand wurden auch DIF-Effekte auf die subjektive Beurteilung des Gesundheitszustands entdeckt: Männer beurteilen bei gleichem Gesundheitszustand diesen verhältnismäßig besser als Frauen (Fleishman und Lawrence 2003), die Beschwerden stärker und häufiger wahrnehmen (Celentano et al. 1990; Nathanson 1975).

H_{3a}: Frauen beurteilen den physischen und den mentalen Gesundheitszustand negativer als Männer.

Bedingt durch den biologische Altersprozess sollte zu erwarten sein, dass sich die physische Dimension von Gesundheit bei Älteren im Vergleich zu Jüngeren als schlechter erweist (Hurrelmann 2006: 61ff.). Zudem kann angenommen werden, dass sich die Verschlechterung des physischen Zustandes belastend auf die psychische Gesundheit auswirkt.

H₄: Das Alter hat einen negativen Einfluss auf den physischen und den mentalen Gesundheitszustand.

Hinsichtlich altersbezogener DIF-Effekte bestehen widersprüchliche Befunde: In manchen Studien sind es ältere Menschen (Groot 2000, 2003), in anderen junge Menschen, die weniger kritisch ihrer Gesundheit gegenüber sind (d'Uva et al. 2008). Es wird die Annahme vertreten, dass ältere Menschen ihren Gesundheitszustand positiver einschätzen als jüngere Menschen mit gleichem Gesundheitszustand (Idler 1993; Leinonen et al. 2001).

H_{4a}: Das Alter bewirkt unabhängig von Migrationsstatus eine positivere Beurteilung des physischen und des mentalen Gesundheitszustandes.

Es lassen sich außerdem unterschiedliche Effekte für die Gruppen mit und ohne Migrationshintergrund vermuten: Da – mit Ausnahme der Spätaussiedler – jüngere Befragte mit Migrationshintergrund bessere Deutschkenntnisse aufweisen als ältere (Diehl und Schnell 2006), könnte es in den Migrantengruppen zusätzlich einen stärkeren Alters-DIF geben als in der Gruppe der westdeutschen und ostdeutschen Befragten. Im Fall der Spätaussiedler besteht womöglich ein positiver DIF-Effekt.

H_{4b}: Bei Befragten mit Migrationshintergrund wird der Einfluss des Alters auf die Beurteilung des physischen und des mentalen Gesundheitszustand über die Deutschkenntnisse mediert und dadurch verstärkt oder abgeschwächt.

Der Zusammenhang zwischen dem sozialen Status und dem Gesundheitszustand ist länderübergreifend stark positiv (Mielck 2008). Bildung und Einkommen als zentrale Indikatoren des sozialen Status einer Person stehen in engem Zusammenhang mit der Gesundheitsversorgung, dem Gesundheitsverhalten und der sozialen Teilhabe, was sich auf die körperliche und psychische Gesundheit auswirkt (Hurrelmann 2006: 32ff.).

H₅: Die Anzahl der Bildungsjahre hat einen positiven Einfluss auf den physischen und den mentalen Gesundheitszustand.

H₆: Das Einkommen hat einen positiven Einfluss auf den physischen und den mentalen Gesundheitszustand.

Es wird weiterhin vermutet, dass geringe Bildung einen extremen Antwortstil fördert (Meisenberg und Williams 2008). Der Effekt kann somit negativ oder auch positiv ausfallen, d.h. die Befragten geben einen übermäßig guten oder schlechten Gesundheitszustand an.

H_{5a}: Das Bildungsniveau hat entweder einen positiven oder einen negativen Einfluss auf die Beurteilung des physischen und des mentalen Gesundheitszustandes.

Wie in anderen Studien belegt wird angenommen, dass der Effekt des Einkommens negativ ist, sprich mit niedrigem Einkommen weniger Erwartungen und Ansprüche an die eigene Gesundheit gestellt werden und die Beurteilung eher positiv ausfällt (d'Uva et al. 2008).

H_{6a}: Das Einkommen hat einen negativen Einfluss auf die Beurteilung des physischen und des mentalen Gesundheitszustandes.

Den Effekt der Erwerbstätigkeit ohne eine Konkretisierung des Berufsfeldes abzuschätzen, erweist sich als schwierig. Es wird daher angenommen, dass für Berufe mit überwiegend sitzenden Tätigkeiten ein positiver Effekt und körperlich belastende Berufe einen negativen Effekt ausüben. Allerdings heben sich diese gegenläufigen Effekte je nach Berufsfeld vermutlich gegenseitig auf und es ist insgesamt kein Effekt zu verzeichnen.

H₇: Die Erwerbstätigkeit hat keinen Einfluss auf den physischen und den mentalen Gesundheitszustand.

Es können darüber hinaus gruppenspezifische Unterschiede auftreten: Da Personen mit Migrationshintergrund häufiger in handwerklichen und körperlich belastenden Berufen arbeiten und ein höheres Arbeitslosigkeitsrisiko aufweisen (Elkeles und Seifert 1996), ist ein negativer Effekt der Erwerbstätigkeit auf die Gesundheit zu vermuten.

H_{7a}: Bei Befragten mit Migrationshintergrund hat die Erwerbstätigkeit einen negativen Einfluss auf den physischen und auf den mentalen Gesundheitszustand.

Es bestehen keine empirischen Befunde darüber, ob und wie sich der Erwerbstätigkeitsstatus auf die Beurteilung des Gesundheitszustandes auswirkt. Im Vergleich zu Rentnern und Arbeitslosen sind die Antworten in der Gruppe der Vollzeit- und Teilzeiterwerbstätigen vermutlich eher weniger verzerrt.

H_{7b}: Das Ausüben einer Vollzeit- oder Teilzeittätigkeit hat keinen Einfluss auf die Beurteilung des physischen und des mentalen Gesundheitszustandes.

Der Ruhestand kann zweierlei Effekte auf den Gesundheitszustand haben: Einerseits können Personen in Rente einer höheren Isolation ausgesetzt sein, andererseits fallen berufsbezogener Stress und Belastungen weg. Es wird daher ebenfalls von sich aufwiegenden Effekten des Ruhestandes ausgegangen.

H₈: Ruheständler weisen im Vergleich zu Nicht-Ruheständlern keinen unterschiedlichen Zustand physischer und mentaler Gesundheit auf.

Da Rentner weniger regulären, verpflichtenden Aufgaben nachgehen und da soziale Kontakte abnehmen, wird vermutet, dass speziell die Items zur sozialen, physischen und emotionalen Rollenfunktion Verzerrungen unterliegen.

H_{8a}: Ruheständler beurteilen spezifische Aspekte des physischen und des mentalen Gesundheitszustandes negativer.

Arbeitslosigkeit führt oftmals zu einer Zunahme von psychischen Beschwerden, die insbesondere durch den Verlust der sozialen und psychischen Identität bedingt sind (Elkeles und Seifert 1993). Auch bezüglich des physischen Gesundheitszustandes kann angenommen werden, dass Erwerbslose einen schlechteren Gesundheitszustand aufweisen.

H₉: Arbeitslosigkeit hat einen negativen Einfluss auf den physischen und den mentalen Gesundheitszustand.

Darüber hinaus führt Arbeitslosigkeit auch zu einer geringeren Zufriedenheit mit der eigenen Gesundheit (Elkeles und Seifert 1993; Gordo 2006), wodurch eine negativere Beurteilung des Gesundheitszustandes folgen kann.

H_{9a}: Arbeitslosigkeit hat einen negativen Einfluss auf die Beurteilung des physischen und des mentalen Gesundheitszustandes.

Da sich gerade im Hinblick auf gesundheitliche Ungleichheit die Effekte von Indikatoren des sozioökonomischen Status für verschiedene Länder bestätigen lassen (Doorslaer und Koolman 2004; Eikemo et al. 2008; Mackenbach et al. 2008), kann vermutet werden, dass die Hypothesen weitgehend für alle Herkunftsgruppen gelten, wobei lediglich leichte Unterschiede in der Stärke der Effekte zu vermuten sind.

H₁₀: Die in den Hypothesen H₃ bis H_{9a} (mit Ausnahme der Hypothesen H_{4b} und H_{7a}) angenommenen Effekte bestehen für alle untersuchten Herkunftsgruppen in ähnlicher Form und in vergleichbarem Ausmaß.

Gleichheitsrestriktionen der Parameter des SF-12

Die Überprüfung spezifischer Modellstrukturen auf Äquivalenz zwischen den Gruppen beginnt mit den λ -Faktorladungen. Sie spiegeln wider, wie gut die Itemskala mit derjenigen latenten Skala übereinstimmt, die der Befragte bei der Beantwortung des Items vor Augen hat. Deutliche Unterschiede in den gruppenspezifischen Faktorladungen weisen auf eine geringe Differenzierung der Antworten und eine Tendenz zu den extremen Antwortkategorien hin.⁴ Aufgrund unterschiedlicher Sprachkenntnisse und einem unterschiedlichen Verständnis von Gesundheit und Krankheit kann angenommen werden, dass zwischen den Herkunftsgruppen keine metrische Äquivalenz besteht.

H₁₉: $\lambda_i^{g1} \neq \lambda_i^{g2} \neq \lambda_i^{g3} \neq \dots \neq \lambda_i^{gk}$
Es besteht keine metrische Äquivalenz zwischen den Herkunftsgruppen.

Weiterhin geben die Itemkonstanten τ an, inwiefern Befragte die unterste Kategorie der Antwortskala nutzen. Eine Konstante $\tau > 0$ kann im Sinne der Regressionskonstanten so verstanden werden, dass selbst wenn die tatsächliche Ausprägung auf der inneren latenten Skala des Befragten bei null liegt, dieser nicht die unterste Kategorie auf der Antwortskala wählt. Dieser Umstand weist auf ein aquieszentes Antwortverhalten hin (Cheung und Rensvold 1998: 190). Metrische und skalare Äquivalenz werden zusammen als „strong factorial

⁴ Streng genommen liegt extremes Antwortverhalten jedoch erst vor, wenn die Annahme sowohl metrischer Äquivalenz als auch skalarer Äquivalenz (s. H₁₉) verworfen werden muss.

invariance“ bezeichnet und als die wichtigsten Bedingungen für den gültigen Vergleich latenter Mittelwerte angesehen (Gregorich 2006: 16).

$H_{20}: \tau_i^{g^1} \neq \tau_i^{g^2} \neq \tau_i^{g^3} \dots \neq \tau_i^{g^k}$
 Es besteht keine skalare Äquivalenz zwischen den Herkunftsgruppen.

Ein weiterer Test ist die Gleichsetzung der Messfehlerkomponente der einzelnen Items zwischen den Gruppen. Dies wird als Test auf vergleichbare Reliabilität der Messungen zwischen den Gruppen angesehen und zeigt somit Unterschiede hinsichtlich der Güte der Messungen auf (Schmitt und Kuljanin 2008: 212). Unterscheidet sich die Indikatorreliabilität zwischen den Gruppen signifikant, werden Fragen in manchen Gruppen schlechter verstanden als in anderen und die Daten enthalten einer höhere Zufallsfehlerkomponente. Es wird angenommen, dass aufgrund schlechterer Deutschkenntnisse bei den Migrantengruppen geringere Indikatorreliabilitäten bestehen. Lassen sich hingegen metrische, skalare sowie Messfehler-Äquivalenz bestätigen, spricht man von „strict factorial invariance“ (Gregorich 2006: 16).

$H_{21}: \epsilon_i^{g^1} \neq \epsilon_i^{g^2} \neq \epsilon_i^{g^3} \dots \neq \epsilon_i^{g^k}$
 Es besteht keine Äquivalenz der Item-Messfehler zwischen den Herkunftsgruppen.

Erweist sich die Anpassung eines restringierten Modells als signifikant schlechter im Vergleich zum Modell mit frei geschätzten Parametern, sollten einzelne Parameter genauer betrachtet werden. Es ist zwar möglich, dass gar keiner der Indikatoren zwischen den Gruppen äquivalent ist, andererseits besteht möglicherweise nur in einer Gruppe keine Äquivalenz, zwischen den anderen Gruppen allerdings schon. Partielle Restriktionen können in jedem Schritt und für jede Matrix eingeführt werden, indem ein Parameter beispielsweise lediglich zwischen vier Gruppen gleichgesetzt wird, in zwei Gruppen hingegen frei geschätzt wird. Konkret würde dies bedeuten, dass Personen mit Migrationshintergrund verschiedene Gesundheitsvorstellungen oder gesundheitsbezogene Antwortstile untereinander oder im Vergleich zu den Gruppen ohne Migrationshintergrund aufweisen.

$H_{22}: x_i^{g^1} \neq x_i^{g^2} \neq x_i^{g^3} \dots \neq x_i^{g^k} \quad x \leftarrow \lambda, \tau, \epsilon$
 Hinsichtlich der Parameter λ , τ und ϵ besteht zwischen den Herkunftsgruppen partielle Äquivalenz, wobei einzelne Gruppen keine äquivalenten Parameter aufweisen.

Lässt sich in den vorigen Modellen Invarianz oder zumindest partielle Invarianz bestätigen, können zuletzt die latenten Mittelwerte der Gruppen gleichgesetzt werden. Der Globaltest kann allerdings nicht zeigen, ob sich eine Gruppe von allen anderen oder alle Gruppen voneinander unterscheiden; dies muss in paarweisen Gruppentests geprüft werden. Wie sich bereits in anderen Studien gezeigt hat, weisen Migranten in Deutschland andere und unterschiedlich

starke gesundheitlich Belastungen im Vergleich zur deutschen Bevölkerung ohne Migrationshintergrund auf. Es wird daher angenommen, dass unter Kontrolle von Kovariaten und möglichen Messfehlerkomponenten dennoch faktische Unterschiede zwischen Migranten festgestellt werden können.

H₂₃: $\alpha_i^{g^1} \neq \alpha_i^{g^2} \neq \alpha_i^{g^3} \dots \neq \alpha_i^{g^k}$
 Es bestehen Unterschiede der physischen und mentalen Gesundheit zwischen den Herkunftsgruppen.

H_{23a}: $\alpha_i^{g^1} = \alpha_i^{g^2} \neq \alpha_i^{g^3} \dots = \alpha_i^{g^k}$
 Es bestehen Unterschiede der physischen und mentalen Gesundheit zwischen den Gruppen ohne Migrationshintergrund und denen mit Migrationshintergrund.

2 Datenbasis und Messinstrument

Zur Durchführung der Analysen werden die Daten des Sozio-oekonomischen Panels (SOEP) aus den Jahren 2002, 2004, 2006 und 2008 gepoolt. Es handelt sich beim SOEP um die erste Bevölkerungsumfrage, die seit 2002 eine modifizierte Version des SF-12 in den Standard-Personenfragebogen integriert hat und in zweijährigem Abstand erhebt. Die Modifikation bezieht sich auf die Ersetzung des SF-12-Items „Beeinträchtigung in den Tätigkeiten aufgrund körperlicher Schmerzen“ durch das Item „starke körperliche Schmerzen“ aus dem ursprünglichen SF36 Fragebogen. Darüber hinaus gibt es geringfügige Veränderungen in der Frageformulierung und der Fragereihenfolge (Andersen et al. 2007: 172). Die psychometrischen Eigenschaften dieser SOEP-spezifischen Version des SF-12 wurden bislang lediglich von Grabka und Schupp (2005) untersucht.

Für die Analyse werden lediglich diejenigen Personen über 17 Jahren in Privathaushalten ausgewählt, die den SF-12 Fragebogen in den genannten Erhebungswellen zum ersten Mal ausfüllen, um so potentielle Paneffekte zu vermeiden (vgl. Schnell et al. 2008: 230). Aus dieser Auswahl resultiert eine Stichprobe von 28064 Befragten.⁵

Die Auswahl der Befragten des SOEP erfolgt über eine mehrstufige, regional geklumpte Zufallsauswahl. Es werden jeweils alle Haushaltsmitglieder ab 16 Jahren des gezogenen Haushalts befragt. Personen mit Migrationshintergrund sind eine Bevölkerungsgruppe, die in der Stichprobe überrepräsentiert ist, um migrationspezifische Analysen mit ausreichend hoher Fallzahl zu ermöglichen: Zum Einen wurde im ersten Erhebungsjahr 1984 eine Ausländerstichprobe gezogen, die Privathaushalte mit türkischer, griechischer, jugoslawischer, spanischer

⁵ Dabei verteilen sich die für die Analysen verwendeten Fälle wie folgt auf die vier Erhebungszeitpunkte:
 Welle 2002: n=22325, Welle 2004: n=1568, Welle 2006: n=3183, Welle 2008: n=988.

oder italienischer Staatsangehörigkeit beinhaltet. Zum Anderen wurden 1994 und 1995 zwei Migrantenstichproben aus der Grundgesamtheit aller derjenigen Privathaushalte gezogen, in denen mindestens ein Haushaltsmitglied nach 1984 nach Deutschland eingewandert ist (DIW Berlin 2005).

Die Datenerhebung erfolgt prinzipiell über ein persönliches Interview, im Ausnahmefall füllen Befragte den Fragebogen selbst aus (DIW Berlin 2008). In der Regel erhalten alle Befragten einen Fragebogen in deutscher Sprache bzw. wird das Interview auf Deutsch geführt.

Fehlende Werte auf den 12 Gesundheitsvariablen wurden Werte durch regressionsbasierte Schätzungen ersetzt.⁶

2.1 Überprüfung der Grundannahmen

Da Strukturgleichungsmodelle eine Kombination aus Faktoren- und Regressionsanalyse darstellen, stellen sie zur korrekten Parameterschätzung dieselben Anforderungen wie Faktoren- und Regressionsanalyse: Hierunter fallen der Test auf univariate sowie auf multivariate Normalverteilung, die Überprüfung auf Multikollinearität sowie die Betrachtung etwaiger Ausreißerwerte.

Die Items weisen durchweg eine positive Kurtosis auf und sind aufgrund der Antworttendenz zur Kategorie „nie“ bzw. „gar nicht“ (Beschwerden) mehrheitlich linksschief verteilt. Diese Kombination lässt auf deutliche Deckeneffekte (Tendenz zur Kategorie mit dem höchsten Wert) schließen. In der Konsequenz zeigen die Tests auf univariate und multivariate Normalverteilung, dass die SF-12 Items signifikant von der Normalverteilung abweichen. Aus diesem Grund, wird in den allen Analysen über das Ziehen von Bollen-Stine-Bootstrap-Stichproben (Bühner 2006: 227) ein entsprechend korrigierter p-Wert geschätzt.

Die Überprüfung auf Multikollinearität zeigt, dass für zwei Paare der SF-12 Items hohe Korrelationen bestehen: Zum Einen korreliert die Variable „starke körperliche Schmerzen in den letzten 4 Wochen“ stark mit der Variable „wegen gesundheitlicher Probleme körperlicher Art in Arbeit oder Beschäftigungen weniger geschafft“ ($r=0,87$). Zum Anderen besteht eine Korrelation von $r=0,84$ zwischen den Items „wegen seelischer oder emotionaler Probleme weniger geschafft“ und „wegen seelischer oder emotionaler Probleme in Tätigkeiten eingeschränkt“. Die Korrelationen zwischen den Kovariaten sind hingegen nicht bedenklich. Die Prüfung auf Ausreißerwerte über Mahalanobis d^2 deutet auf keine signifikanten Ausreißerwerte in den SF-12 Variablen hin.

⁶ Dabei wird die jeweilige Gesundheitsvariable jeweils auf alle anderen Variablen der SF-12-Itematterie sowie das Geschlecht, das Alter und die Anzahl der Arztbesuche regrediert.

3 Analyse der Äquivalenz des SF-12 zwischen der Bevölkerung mit und ohne Migrationshintergrund

Die Stichprobe wird nach der Herkunftsregion der Befragten aufgeteilt (Tab. 1): Die größten beiden Gruppen sind Deutsche ohne Migrationshintergrund, die in diejenigen mit ostdeutscher Herkunft und diejenigen mit westdeutscher Herkunft unterteilt sind.⁷ Die drittgrößte Gruppe in der Stichprobe sind (Spät-)Aussiedler osteuropäischer Herkunft, die vordergründig aus Rumänien, Russland, Polen und Ukraine nach Deutschland kamen. An vierter Stellen folgen Befragte türkischer Herkunft, an fünfter Stelle Migranten aus den Anwerbeländern der ehemaligen Europäischen Gemeinschaft (EG) (v.a. Griechenland, Italien und Spanien) und schließlich Einwanderer aus Jugoslawien und dessen Nachfolgestaaten.

Tabelle 1: Verteilung der Befragten nach Herkunftsregion

Herkunft	n	gültige % ⁸
Westdeutsche	17380	61,9
Ostdeutsche	7047	25,1
Türkei	934	3,3
(Ex-)Jugoslawien	503	1,8
EG-Anwerbeländer	731	2,6
erweit. osteuropäische Länder	1469	5,2
total	28064	100,0

3.1 Soziodemografische und -ökonomische Indikatoren

Wichtige soziodemografische und sozioökonomische Variablen, die sich auf den Gesundheitszustand wie auf die Interpretation der Items des SF-12 Fragebogen auswirken können, sind Geschlecht, Alter, Netto-Äquivalenzeinkommen, Anzahl der Bildungsjahre und die Erwerbssituation. Diesbezüglich treten die türkischen Befragten im Vergleich zu den anderen Herkunftsgruppen besonders durch ein niedriges Altersniveau (37 Jahre), eine geringere Anzahl an Bildungsjahren (10 Jahre) und ein sehr niedriges Netto-Äquivalenzeinkommen (990€) hervor. Auch sind sie häufiger von Arbeitslosigkeit (11,2%) betroffen als Befragte in den anderen Gruppen. Auch die anderen Gruppen mit nicht-deutscher Herkunft sind durchschnittlich jünger und weisen einen niedrigeren sozialen Status auf als Befragte West- und Ostdeutschland.

3.2 Der SF-12 Fragebogen

Die Items des SF-12 zeichnen sich durch unterschiedliche Frageformulierung und Antwortskala

⁷ Die ost- bzw. westdeutsche Herkunft wird über den Aufenthaltsort im Jahr 1989 identifiziert. Personen, die nach 1989 geboren sind, wurden in die Gruppe der Westdeutschen eingeordnet.

⁸ In allen nachfolgenden Tabellen werden ausschließlich gültige % angegeben.

len aus: Die Items „Beeinträchtigung beim Treppensteigen“ und „Beeinträchtigung bei anstrengenden Tätigkeiten“ verfügen lediglich über drei Antwortkategorien. Alle anderen Variablen geben fünf Antwortkategorien vor. Außerdem sind drei Items gegenüber den anderen neun Items negativ formuliert.⁹ Dies sind die Variablen „gegenwärtiger Gesundheitszustand“, „ruhig und ausgeglichen in den letzten 4 Wochen“ sowie „jede Menge Energie verspürt in den letzten vier Wochen“.

Die spaltenweise Prozentuierung gibt erste Hinweise auf potentielle kulturspezifische Antwortstile (s. Anhang, Tab. 9): Im Gruppenvergleich wählen türkische Befragte die Kategorie „nie“ (Beschwerden/Beeinträchtigungen) am häufigsten. Dies trifft interessanterweise auch bei zwei negativ formulierten Items zu. Hier geben türkische Befragte am häufigsten an, nie ruhig und ausgeglichen zu sein und auch der gegenwärtige Gesundheitszustand wird von türkischen Befragten im Vergleich zu Befragten anderer Herkunftsregionen am häufigsten als schlecht beurteilt.

Ostdeutsche Befragte hingegen wählen am seltensten die extreme Antwortkategorie „nie“. Bei dieser Gruppe zeigt sich am häufigsten die Wahl der Mittelkategorie verglichen mit den anderen Gruppen. Für die Gruppe der Westdeutschen finden sich derartige Tendenzen nicht und auch sind die Verteilungen der Antworten deutlich unterschiedlich von denen der Ostdeutschen.

Alle Befragten mit Migrationshintergrund zusammengenommen geben einen schlechteren generellen Gesundheitszustand an als West- und Ostdeutsche zusammen. Insgesamt zeigen die Befragten mit Migrationshintergrund beim Großteil der SF-12-Items häufiger extremes Antwortverhalten als Befragte ohne Migrationshintergrund.

Die Mittelwerte der einzelnen Items (s. Anhang, Tab. 10) zeigen allerdings, dass die Unterschiede zwischen den Gruppen nur gering sind. Der maximale Unterschied beträgt 1,4 Punkte zwischen den Ostdeutschen mit 42,5 Punkten auf dem Summenscore und Befragten aus den EG-Anwerbeländern mit 43,9 Punkten bei einer Spannweite von 12 bis 56. Die einzelnen Itemmittelwerte liegen lediglich einen Zehntel Skalenpunkt auseinander. Auch hinsichtlich der Varianz bestehen nur geringfügige Unterschiede. Die ostdeutschen Befragten weisen die geringste Varianz (7,9) auf, Befragte aus Jugoslawien und deren Nachfolgestaaten die höchste (8,7).

3.3 Modellierung des SF-12-Gesundheitskonstrukts

Im Grundmodell laden jeweils sechs Items auf die beiden Hauptdimensionen „physische Gesundheit“ und „mentale Gesundheit“. Die beiden Hauptdimensionen werden jedoch – entgegen dem theoretischen Modell – aufgrund der bestehenden empirischen Befunde als

⁹ Für die Analysen wurde die Skala rekodiert, sodass für alle SF-12-Items hohe Werte einen guten Gesundheitszustand implizieren.

korreliert angenommen (vgl. Grabka und Schupp 2005; Maurischat et al. 2005).

Den latenten Subdimensionen wird über jeweils ein Item („Marker“) eine Metrik zugewiesen, indem über mehrere Tests diejenigen Items mit der höchsten Faktorladung auf das jeweilige Konstrukt identifiziert werden (vgl. Kline 2004: 42ff.; Little et al. 2006). Dies sind die Items „in Tätigkeiten eingeschränkt“ und „weniger geschafft“. Die latenten Mittelwerte der beiden Subdimensionen können somit gemäß den Markern Werte von 1 bis 5 annehmen.

Aus Identifikationsgründen können die Achsenabschnitte der beiden Marker-Items nicht geschätzt werden, sondern müssen auf null gesetzt werden. Dieses erste Modell wird über alle Befragten spezifiziert. Die Gesamtabweichung des empirischen vom theoretischen Modell wird in χ^2 -Einheiten ausgedrückt und beträgt für das Modell rund 30671. In Relation zur Anzahl der genannten Freiheitsgrade führt dies zu einem Signifikanzniveau von $p=0,001$ ¹⁰. Der CFI beträgt 0,862, der RMSEA=0,143 und der SRMR=0,082. Gemäß Tabelle 7 (s. Anhang) mit den gegebenen Grenzwerten bedeutet dies, dass das Modell auf Grundlage des χ^2 -Tests bereits zu verwerfen ist. Auch liegen der CFI- und der RMSEA-Wert unter den Grenzwerten für akzeptable Modellanpassung.

Als mögliche Ursachen dieses „Misfits“ sind mehrere Faktoren in Betracht zu ziehen: Zum Einen der Verstoß gegen die Normalverteilung aufgrund starker Boden- und Deckeneffekten, zum Anderen die Tendenz des χ^2 -Test, bei hoher Fallzahl schon geringe Abweichungen von der theoretischen Modellstruktur als signifikant zu werten.¹¹ Es ist auch möglich, dass das theoretisch angenommene Modell schlicht nicht durch die empirische Struktur der Daten abgebildet wird. In diesem Fall wäre eine Modellmodifikation notwendig, um die Passung zwischen dem theoretischen und dem empirischen Modell zu erhöhen. Es wird nun zusätzlich auf einzelne psychometrische Kriterien eingegangen, um die Modellgüte detaillierter zu beurteilen.

¹⁰ Hier wurde ein Bollen-Stine-Bootstrap mit 750 Stichproben durchgeführt, um den nicht-normalverteilten Daten Rechnung zu tragen. Über die reguläre Schätzung erhält man einen Wert $p=0,000$. Auch in allen folgenden Analysen wird ein solcher Bootstrap durchgeführt.

¹¹ „Indeed, rejection of basically any overidentified model based on χ^2 requires only a sufficiently large number of cases.“ (Kline 2004: 136).

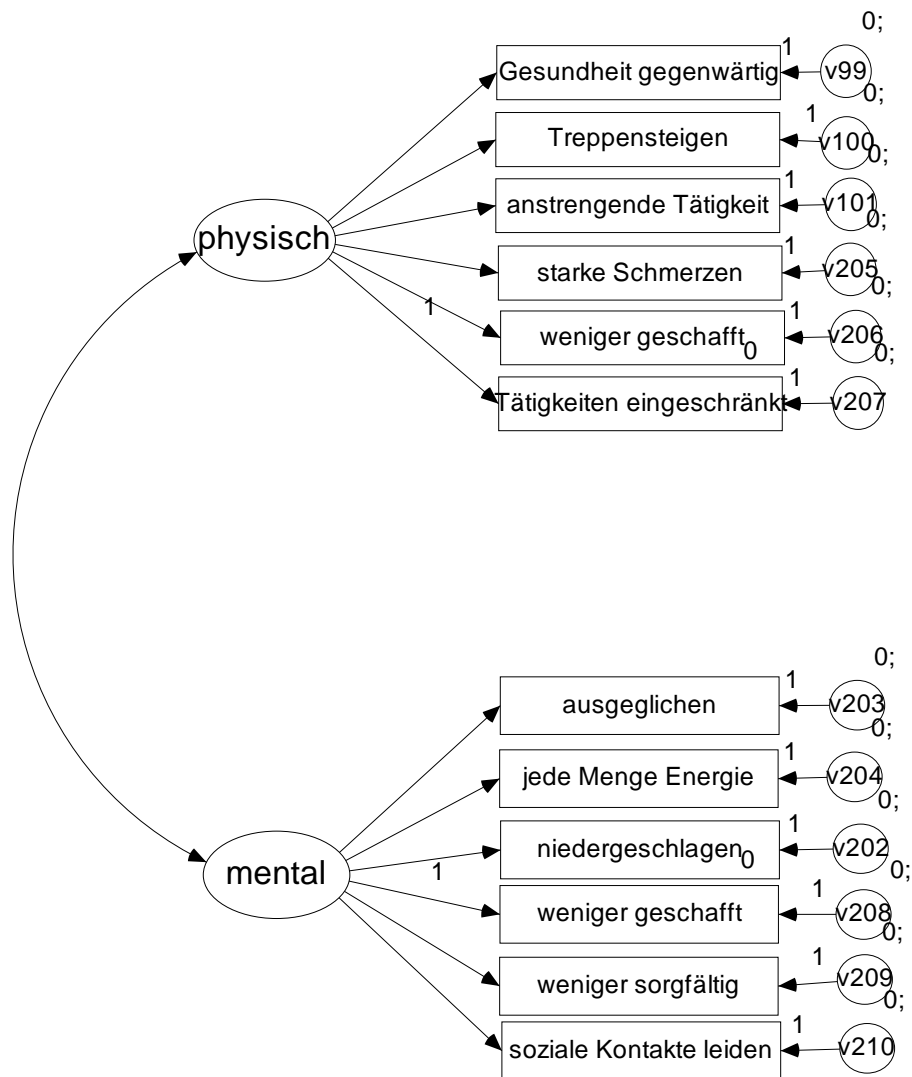


Abbildung 1: Grundmodell mit zwei latenten Konstrukten und 12 Indikatoren
Modellfit: $\chi^2=30670,9$; $df=53$, $p=0,001^{12}$ (0,000), CFI=0,862, RMSEA=0,143, SRMR=0,082

3.4 Psychometrische Eigenschaften des Grundmodells

Zur Beurteilung der einzelnen Modellstrukturen werden als lokale Gütekriterien die Indikatorreliabilität, Cronbach's Alpha und die Konstruktvalidität hinzugezogen: Alle Faktorladungen sind auf dem 0,1%-Niveau signifikant und die höchste unstandardisierte Faktorladung liegt bei $\lambda > 0,9$, was auf eine hohe Indikatorreliabilität hinweist. In zwei Fällen liegen die Ladung jedoch nur knapp über dem Grenzwert von $\lambda = 0,4$ (o. Abb.).¹³ Die Items „ruhig und ausgeglichen“ und „jede Menge Energie“ der Subdimension „mentale Gesundheit“ weisen dementsprechend eine hohe Fehlervarianz auf.

¹² In allen folgenden Analysen wird ein Bollen-Stine-Bootstrap mit 750 Stichproben gezogen und ausschließlich der korrigierte p-Wert angegeben.

¹³ Zwar weisen auch die Items „Treppensteigen“ und „anstrengende Tätigkeiten“ der physischen Subdimension nied-

Cronbach's Alpha ($\alpha=0,91$) als Maß interner Konsistenz weist einen hohen Zusammenhang der einzelnen Items mit der gesamten Skala aus, wobei die mentale Subdimension ($\alpha=0,84$) einen etwas geringeren Wert aufweist als die physische Subdimension ($\alpha=0,90$) (s. Anhang, Tab. 7). Die korrigierte Item-Skala-Korrelation als Indikator der Item-Trennschärfe zeigt, dass die Items zum Teil sehr gut den gesamten SF-12 Fragebogen widerspiegeln. Lediglich die Variable „ausgeglichen“ weist eine geringe Korrelation unter 0,5 mit den anderen Items des Fragebogens auf. Die Konstruktvalidität, gemessen über die Faktorreliabilität und die durchschnittlich erfasste Varianz (Fornell und Larcker 1981) der Subskalen sowie der gesamten SF-12-Skala kann ebenfalls bestätigt werden.

Es zeigt sich ein unklares Bild: Einerseits besteht keine offenkundige Fehlspezifizierung, da die jeweils sechs Items alle in angemessener Form die jeweilige Subdimension zu repräsentieren scheinen. Mit den Items „ausgeglichen“ und „jede Menge Energie“ konnten zwar Items identifiziert werden, die eine eher geringe Faktorladung aufweisen, allerdings geben diese Werte keinen Anlass zur Entfernung der Items aus dem Fragebogen.¹⁴ Andererseits muss aufgrund der schlechten Modellanpassung vermutet werden, dass die empirischen Daten nicht die theoretische Faktorstruktur des Modells widerspiegeln.

3.5 Test des Grundmodells auf konzeptuelle Äquivalenz

Es wird daraufhin der Test auf konzeptuelle Äquivalenz zwischen den Gruppen durchgeführt, um die angenommene Faktorstruktur für alle Gruppen zu überprüfen (s.u., Tab. 2). Zwar unterscheiden sich die Gruppen aufgrund der unterschiedlichen Fallzahlen stark im Hinblick auf das χ^2/df -Verhältnis, die anderen Fitmaße weisen allerdings auf eine ähnliche Modellanpassung in allen Gruppen hin. Gegenüber dem Grundmodell über die gesamte Stichprobe hat sich die Modellanpassung durch die gruppenspezifische Analyse nicht verbessert.

rige unstandardisierte Faktorladungen auf, diese sind jedoch aufgrund ihrer dreistufigen Skala nicht mit den anderen fünfstufigen Items vergleichbar. Hier sollte zusätzlich ein Blick auf die standardisierten Lösungen geworfen werden.

¹⁴ Hildebrandt und Temme (2006: 10) plädieren in sinnvoller Weise für ein nicht rein von statistischen Kriterien geleitetes Entfernen von Items und betonen die inhaltliche Überprüfung der Qualität des Messinstruments und seiner Interpretation durch die Befragten.

Tabelle 2: Modellfit des Grundmodells nach Herkunftsregion

Fit-Indizes	West- deutschland	Ost- deutschland	Türkei	(Ex) Jugoslawien	EG- Anwerbeländer	erw. osteurop. Länder
χ^2	19161,5	7395,0	1327,2	676,3	980,6	1727,3
p	0,001	0,001	0,001	0,001	0,001	0,001
df	53	53	53	53	53	53
χ^2/df	361,5	139,5	25,0	12,8	361,5	32,6
RMSEA	0,144	0,140	0,161	0,153	0,155	0,147
CFI	0,861	0,864	0,839	0,865	0,848	0,864
SRMR	0,084	0,078	0,093	0,091	0,087	0,083

Der Blick auf die standardisierte Residualkovarianzmatrix und die Modifikationsindizes (o. Abb.) weist auf deutliche, signifikante Abweichungen der theoretischen von der empirischen Modellstruktur in allen Gruppen hin. Es handelt sich dabei in allen sechs Gruppen um hohe Kovarianzen zwischen den Items „ausgeglichen“ und „jede Menge Energie“ sowie zwischen „ausgeglichen“ und „niedergeschlagen“. Diese Items sind Teil der mentalen Subdimension, werden aufeinanderfolgend erfasst und haben einen einheitlichen Frageeinleitungstext.

Darüber hinaus wird in allen Gruppen eine hohe standardisierte Residualkovarianz zwischen dem Item „jede Menge Energie“ und der Frage nach dem „generellen Gesundheitszustand“ aus der physischen Subdimension ausgewiesen.

Die Gemeinsamkeit der Items „ausgeglichen“, „jede Menge Energie“ und „genereller Gesundheitszustand“ liegt in der negativen Formulierung, die - eingebettet in positiv formulierte Fragen - anfällig für verzerrende Effekte ist (DiStefano und Motl 2009; Horan et al. 2003; Podsakoff et al. 2003; Roszkowski und Soven 2010; Tomas und Oliver 1999; Watson 1992).

Auch bestehen hohe unberücksichtigte Kovarianzen zwischen den Items „Treppensteigen“ und „anstrengende Tätigkeiten“. Die Items werden nacheinander erfasst und weisen beide lediglich eine dreistufige Antwortskala auf, was ebenfalls zu Verzerrungen führen kann (Chen et al. 1995; Greenleaf 1992; van Herk et al. 2004; Welkenhuysen-Gybels et al. 2003).

Da solche hohen, nicht frei geschätzten Kovarianzen im Modell zu einer verzerrten Schätzung der Parameter führen (vgl. Temme und Hildebrandt 2008: 28f.), sollte das Modell entsprechend modifiziert werden; da angenommen werden kann, dass die hohen Kovarianzen durch einen latenten, nicht spezifizierten Methodenfaktor verursacht werden (vgl. Bühner 2006: 268).

Es werden zwei weitere latente Faktoren modelliert (s. Abb. 2), die Methodeneffekte darstellen: Einer der latenten Faktoren spezifiziert einen Effekt durch die negative Formulierung der Items „genereller Gesundheitszustand“, „ausgeglichen“ und „jede Menge Energie“, die teilweise die Beantwortung der Frage beeinflussen. Der zweite latente Faktor stellt den Effekt der auf drei Kategorien beschränkten Antwortskala der Items „Treppensteigen“ und

„anstrengenden Tätigkeiten“ dar.¹⁵ Im Übrigen bleibt das Grundmodell unverändert.

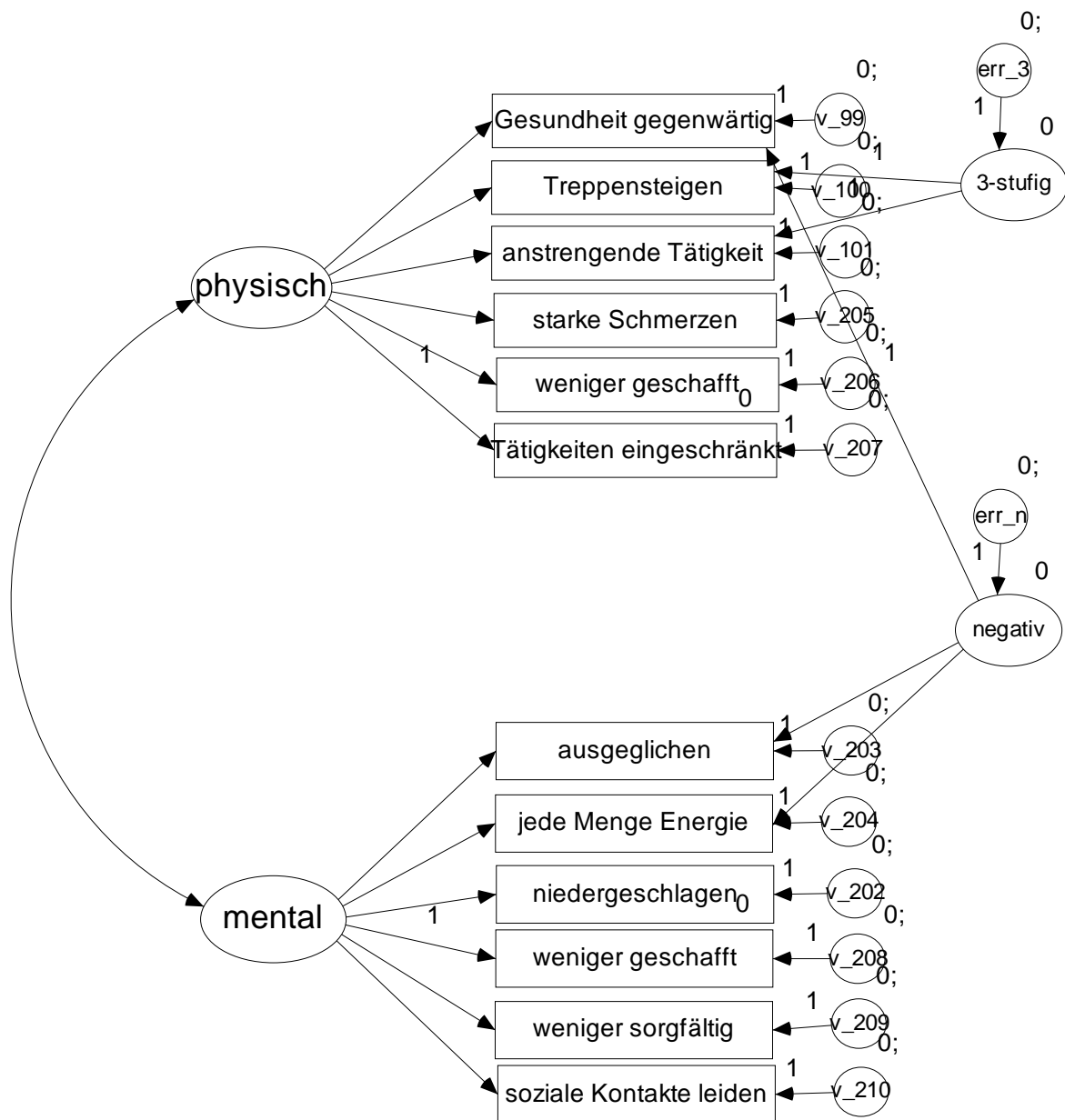
Die Modellanpassung verbessert sich durch die Einführung der latenten Faktoren deutlich: Zwar weist der p-Wert noch immer auf eine signifikante Abweichung der empirischen Daten vom theoretischen Modell hin, die anderen Fitmaße haben sich jedoch stark verbessert. CFI, RMSEA und SRMR liegen nun innerhalb der Grenzen für eine gute Modellanpassung (s. Abb. 2).

Auch die separate Berechnung der Fitmaße für die einzelnen Gruppen zeigt eine gute Modellanpassung und es zeigen sich keine starken Unterschiede der Modellanpassung zwischen den Gruppen (s. Tab. 3). Die frei geschätzten Faktorladungen des Methodenfaktors der Items mit negativer Frageformulierung sind allesamt signifikant. Im Hinblick auf die Modellsparsamkeit führt die vorgenommene Modifikation zu einem guten Modellfit und es handelt sich dabei um inhaltlich begründbare Spezifikationen für die jeweils höchsten standardisierten Residuen.

Tabelle 3: Modellfit des modifizierten Modells nach Herkunftsregion

Fit-Indizes	West-deutschland	Ost-deutschland	Türkei	(Ex) Jugoslawien	EG-Anwerbeländer	erw. osteurop. Länder
χ^2	7537,4	3057,9	361,2	280,1	336,8	689,2
p	0,001	0,001	0,001	0,001	0,001	0,001
df	47	47	47	47	47	47
χ^2/df	160,4	65,1	7,7	6,0	7,2	14,7
RMSEA	0,096	0,095	0,085	0,099	0,092	0,096
CFI	0,948	0,944	0,960	0,950	0,953	0,948
SRMR	0,066	0,061	0,066	0,074	0,068	0,063

¹⁵ Dieser latente Faktor ist allerdings nicht identifiziert, da dem latenten Faktor lediglich zwei Indikatoren zugewiesen werden. Kline (2004) verweist auf Bollen (1989) „two-indicator rule“, die zwar nur zwei Indikatoren zur Identifikation eines Konstruktes voraussetzt. Er betont gleichzeitig jedoch auch die Tatsache, dass solche Zwei-Indikatoren-Konstrukte wesentlich anfälliger für Schätzprobleme sind und empfiehlt daher, eher Faktoren mit drei Indikatoren zu modellieren (Kline 2004: 172). Eine Ausweglösung ist im konkreten Fall die Fixierung der zwei Items „Treppensteigen“ und „anstrengende Tätigkeiten“ auf eine Faktorladung von 1. Allerdings ermöglicht dies nicht mehr die Überprüfung der Signifikanz der Pfade. Es kann somit allenfalls über die Fitverbesserung geprüft werden, ob dieser latente Faktor gültig ist.



**Abbildung 2: Modifiziertes Grundmodell mit zwei Methodenfaktoren,
Berechnung über alle Befragte**

Modellfit: $\chi^2=14506$; $df=49$; $p=0,001$; $CFI=0,935$; $RMSEA=0,103$; $SRMR=0,067$

3.6 Einführung von Kovariaten

Da die Migrantenbevölkerung sich soziostrukturell von der Bevölkerung ohne Migrationshintergrund deutlich unterscheidet und auch in sich sehr starke Heterogenität aufweist, wird im Folgenden für Effekte des Geschlechts, des Alters, des Einkommen und der Bildung sowie des Erwerbsstatus auf den Gesundheitszustand und auf verzerrende Antwortstile kontrolliert.¹⁶

¹⁶ Es werden hierzu die Kovariaten Geschlecht, Alter, monatliches Netto-Äquivalenzeinkommen und Anzahl an

Die Einführung der ersten Kovariaten „männlich“ führt zu einer geringfügigen Verschlechterung des Modellfit ($\chi^2=14781$; $df=59$; $CFI=0,934$; $RMSEA=0,094$; $SRMR=0,063$, über alle Befragten).

Die Kovariaten Bildung und Arbeitslosigkeit üben nur begrenzt Einfluss auf die Gesundheit aus. Im Fall der Bildung besteht nur für die Gruppen ohne Migrationshintergrund ein positiver Effekt. Bei Befragten türkischer Herkunft besteht ein negativer Einfluss von Arbeitslosigkeit auf die physische Gesundheit, wohingegen bei ostdeutschen Befragten ein positiver Effekt besteht.

Bei westdeutschen, ostdeutschen und osteuropäischen Befragten besteht ein negativer Effekt des Ruhestands auf die physische Gesundheit und bei den Westdeutschen auch auf die mentale Gesundheit.

Betrachtet man das χ^2/df -Verhältnis, verschlechtert sich die Modellanpassung mit Einführung dieser drei Kovariaten. Die alternativen Fit-Indizes liefern hingegen ein uneinheitliches Bild. So springt der RMSEA aufgrund der Einführung von Kovariaten unsystematisch auf und ab, liegt aber durchweg im Bereich guter Modellanpassung.¹⁷ Der CFI sinkt durch die Einführung der Kovariaten stets geringfügig, wobei im Fall der Erwerbsvariablen eine starke Verringerung stattfindet. Der SRMR springt ebenso wie der RMSEA, bleibt aber im Bereich guten Fits.

Aufgrund ihres geringen Einflusses werden die Kovariate Bildung und die Dummyvariablen Arbeitslosigkeit und Rentner aus dem Modell entfernt. Durch diese Reduzierung der Kovariaten wurde daraufhin im Fall des Netto-Äquivalenzeinkommens die Kategorisierung aufgehoben. Eine Übersicht über die eingeführten Kovariaten und deren Effekte bietet Tabelle 11 (s. Anhang).

Zentrale Aspekte werden kurz diskutiert: Das Alter hat in allen Gruppen einen durchweg negativen Effekt in ähnlicher Höhe, wobei der Effekt auf die physische Gesundheit höher ist als

(Aus-)Bildungsjahren eingeführt sowie eine Dummyvariable, die die Gruppe der Arbeitslosen, der Rentner und der Vollzeit- und Teilzeiterwerbstätigen identifiziert.

Im Falle der metrischen Kovariaten Alter, Einkommen (in 1000€) und Ausbildungszeit in Jahren wurden die Variablen auf Grundlage des 25% Perzentils, des Medians und des 75% Perzentils der Befragten aus (Ex-)Jugoslawien künstlich kategorisiert, um so stabilere Ergebnisse zu erhalten, da die Fallzahl mit rund 500 Befragten in dieser Gruppen sehr klein ist.

Die Grenzwerte zur Kategorienbildung lauten wie folgt:

Alter: 1) unter 26 Jahren; 2) 26,01-39 Jahre; 3) 39,51-54 Jahre; 4) 54,01 Jahre und älter

Netto-Äquivalenzeinkommen in 1000€: 1) <0,78; 2) 0,781-1,11; 3) 1,111-1,45; 4) 1,451 und mehr

Ausbildungszeit in Jahren: 1) bis 9 Jahre; 2) 9,01-10,5 Jahre; 3) 10,51-11,5 Jahre; 4) 11,51 Jahre und mehr

Durch Einführung der Kovariaten ergibt sich eine formale Veränderung des Modells, die durch die Grundannahmen des Verfahrens festgelegt wird: Die Hauptdimensionen physische und mentale Gesundheit werden definitionsgemäß von exogenen zu endogenen Konstrukten. Dadurch lässt sich die Kovarianz zwischen den beiden Subdimensionen nicht mehr frei schätzen, weil angenommen wird, dass jegliche empirische Kovarianz der Konstrukte auf die Kovariaten zurückgeführt werden kann. Es wird diesem Aspekt Rechnung getragen, indem die Kovarianz der Fehlervarianzen („err_p“ und „err_m“) freigesetzt wird, um so weitere, unberücksichtigte Ursachen der Kovarianz in das Modell aufzunehmen.

Alternativ ließe sich ein weiterer latenter Faktor einführen, der auf die beiden Subdimensionen lädt und somit die residuale Varianz erfasst, die die Kovariaten nicht erfassen. Dies sind äquivalente Modelle, beide Modellierungen führen in diesem Fall zu denselben Schätzungen.

¹⁷ Die gute Anpassung ist unter anderem auch der Tatsache geschuldet, dass der RMSEA relativ sensibel gegenüber der Komplexität eines Modells reagiert: Die Einführung zusätzlicher Variablen führt in der Regel zu einer Verbesserung des Modellfits (vgl. Bühner 2006: 254f.).

auf die mentale Gesundheit. Frauen haben gegenüber Männern in allen Gruppen mit Ausnahme der Befragten türkischer Herkunft einen schlechteren Gesundheitszustand. Das Einkommen hat einen signifikanten positiven Einfluss auf die Gesundheit, wobei sich bei Befragten aus den EG-Anwerbeländern und aus (Ex-)Jugoslawien nur bei jeweils einer Gesundheitsdimension signifikante Effekte finden. Der Effekt der Erwerbstätigkeit ist generell positiv, in der Gruppe der türkischen Befragten und der aus (Ex-)Jugoslawien für die mentale Gesundheit jedoch nicht signifikant. Die Effekte zwischen den Herkunftsgruppen sind mit Ausnahme der Türkei ähnlich hoch. Insgesamt ist der Einfluss der Kovariaten auf die gesundheitsbezogenen Subdimensionen eher gering.

3.7 Test auf Differential Item Functioning (DIF) der Kovariaten

Der Befund, dass Befragte mit westdeutscher Herkunft die niedrigsten Effekte und solche türkischer Herkunft die stärksten Effekte der Kovariaten auf den Gesundheitszustand aufweisen, kann zweierlei Erklärungen haben: Einerseits wirken Charakteristika des sozialen Status je nach Herkunftsland womöglich tatsächlich unterschiedlich auf den Gesundheitszustand, andererseits sind womöglich verzerrende Antwortstile verantwortlich. In letzterem Fall können es zum Einen kulturelle Eigenschaften sein, zum Anderen Effekte individueller Charakteristika, die verzerrte Antworten bewirken: Männliche Befragte mit westdeutscher Herkunft bewerten womöglich ihren Gesundheitszustand positiver als die männlichen Befragten aus anderen Gruppen und verursachen somit die Unterschiede in der Effekthöhe. Ebenso könnten männliche und weibliche Befragte derselben Herkunftsgruppe ihre Gesundheit unterschiedlich kritisch beurteilen. Die Kovariaten werden also im Folgenden zusätzlich zu ihrem direkten Effekt auch auf „Differential Item Functioning“ (DIF) getestet.¹⁸ Um alle 40 möglichen DIF-Pfade (vier Kovariaten multipliziert mit zehn Items¹⁹) zu testen, werden in jedem Schritt simultan jeweils 10 Pfade pro Kovariate auf die Items spezifiziert (o. Abb).

Es zeigen sich dabei einerseits starke Unterschiede der vier Kovariaten hinsichtlich DIF, andererseits bestehen auch einige Unterschiede zwischen den Herkunftsgruppen: Die

¹⁸ Um die Items auf DIF zu überprüfen, bestehen drei Vorgehensweisen (Woods 2009: 4ff.): Einerseits können die Modifikationsindizes daraufhin überprüft werden, ob und welche zu spezifizierenden Ladungen zwischen Kovariaten und endogenen Indikatoren zur Verbesserung des Modellfit beitragen können. Hohe Modifikationsindizes weisen hierbei auf DIF hin. Alternativ lässt sich nacheinander jede endogene Variable einzeln auf DIF testen, wobei stets alle anderen Items als „Anker“ fungieren. Anker stellen den Vergleichsstandard zu den auf DIF zu testenden Items dar und werden als frei von Differential Item Functioning angenommen. Der Modellfit sowie der Test auf Signifikanz der Ladung gibt Aufschluss darüber, welche Items von DIF betroffen sind. Eine dritte Methode legt aufgrund theoretischer Überlegung oder statistischer Prüfung Anker fest, die in den folgenden beibehalten werden. In den Modellen werden schrittweise DIF-Ladungen auf die Items ergänzt bzw. entfernt und der Modellfit dient dem Test auf vorliegendes DIF. Letztgenannte Strategie wird im konkreten Fall verfolgt.

¹⁹ Hier werden die bereits bekannten Marker-Items „in Tätigkeiten eingeschränkt“ und „weniger geschafft“ zusätzlich als Anker verwendet.

signifikanten DIF-Effekte des **Alters** lassen sich in allen Herkunftsgruppen und bei nahezu allen Items finden. Interessanterweise erweist sich das Item „starke Schmerzen“ bei den Befragten west- und ostdeutscher Herkunft hinsichtlich des Alters nicht als DIF-behaftet, in allen vier Migrantengruppen hingegen schon. Der Effekt auf dieses Item ist in den vier Gruppen durchweg negativ, in der Gruppe (Ex-)Jugoslawien am höchsten und in der Gruppe Osteuropa am geringsten. Das bedeutet, dass mit steigendem Alter die Beurteilung von Schmerzsymptomen in den Migrantengruppen negativer ausfällt als die Schmerzen mit steigendem Alter objektiv zunehmen. Bei den west- und ostdeutschen Befragten taucht dieses Phänomen nicht auf, allerdings offenbaren sich in diesen beiden Gruppen zwei andere Items DIF-Effekte, die wiederum in den Migrantengruppen nicht erkennbar sind: Westdeutsche Befragte geben im Alter eher an, „weniger geschafft“ (physische Gesundheit) zu haben. Bei den ostdeutschen Befragten ist dies bei dem Item „weniger sorgfältig“ (mentale Gesundheit) der Fall, d.h. mit zunehmendem Alter sind die Befragten unkritischer was die Beeinträchtigung ihrer Tätigkeiten aufgrund der Gesundheit angeht. Die Effekte sind jedoch sehr gering. Auch für das Item „niedergeschlagen“ zeigt sich ein positiver Effekt bei west- und ostdeutschen Befragten, solchen aus den Anwerbeländern und aus Osteuropa. Diese Personen unterschätzen im Alter also ihren Zustand von Niedergeschlagenheit. Alle Befragtengruppen überschätzen ihren Zustand der Ausgeglichenheit. In solchen Fällen, in denen DIF-Effekte in allen Gruppen auftreten, sind diese in den Migrantengruppen oftmals stärker ausgeprägt als in den beiden deutschen Gruppen ohne Migrationshintergrund.

Neben dem Alter zeigen auch bei **voll- und teilzeiterwerbstätigen** Personen zahlreiche DIF-Effekte in allen Gruppen. Es zeigt sich jedoch, dass Deutsche ohne Migrationshintergrund aufgrund ihrer Erwerbstätigkeit ihren Gesundheitszustand eher verzerrt beurteilen als die Migrantenbevölkerung. Ein starker Effekt, der mit Ausnahme der Befragten aus dem ehemaligen Jugoslawien in allen anderen Gruppen auftaucht, besteht für das Item „ausgeglichen“: Die negativen Effekte weisen darauf hin, dass voll- oder teilzeiterwerbstätige Personen unverhältnismäßig seltener angeben, ausgeglichen zu sein als die Referenzgruppen bestehend aus Rentnern, Arbeitslosen, und sonstigen Erwerbstätigen. Dieser Effekt ist bei Ostdeutschen und Personen aus den EG-Anwerbeländern am deutlichsten und bei türkischen Befragten am schwächsten. Das Item „Treppensteigen“ wird von allen Gruppen verzerrt beurteilt. Es handelt sich dabei um einen positiven Effekt, sprich die erwerbstätigen Befragten geben eher an, beim Treppensteigen durch den Gesundheitszustand belastet zu sein als Befragte mit gleichem Gesundheitszustand, die nicht voll- oder teilzeiterwerbstätig sind.

Bezüglich des **Einkommens** zeigen sich kaum signifikante DIF-Effekte, wobei west- und ostdeutsche Befragte häufiger betroffen sind als Befragte mit Migrationshintergrund. Bei west- und ostdeutschen Befragten hat das Einkommen einen durchweg positiven Effekt, wobei die

Effekte marginal sind. Einen erwähnenswerten positiven Effekt zeigen Befragte aus den EG-Anwerbeländern bei dem Item „jede Menge Energie“. Das bedeutet, dass Personen mit höherem Einkommen eher angeben, viel Energie zu verspüren als Personen mit geringerem Einkommen. Bei Befragten aus dem ehemaligen Jugoslawien finden sich negative Effekte des Einkommens: Mit steigendem Einkommen wird eher seltener angegeben, nicht niedergeschlagen zu sein und durch aufgrund emotionaler Probleme in den sozialen Kontakten eingeschränkt zu sein.

Das **Geschlecht** hat insgesamt eher geringen Einfluss auf die Beurteilung des Gesundheitszustandes, wobei in den Migrantengruppen seltener verzerrende Effekte auftreten als in der Gruppe der Ost- und Westdeutschen. Entgegen der bestehenden empirischen Befunde wird der „generellen Gesundheitszustand“ von west- und ostdeutschen Männern geringfügig schlechter angegeben. Auch geben west- und ostdeutsche Männer sowie solche aus den Anwerbeländern trotz Niedergeschlagenheit seltener an, „niedergeschlagen“ zu sein.

Da nur dann sinnvolle Multigruppen-Modellvergleiche durchgeführt werden können, wenn es sich um so genannte hierarchische Modelle (Kline 2004: 145ff.) handelt, können keine gruppenspezifischen DIF-Spezifikationen vorgenommen werden. Es werden daher nur diejenigen DIF-Effekte modelliert, die in allen Gruppen signifikant sind (s. Tab. 4 & Tab. 12, Anhang).

Tabelle 4: Endgültig spezifizierte DIF-Effekte der Kovariaten

Kovariate	DIF-Effekt
Vollzeit/Teilzeit	→ Treppensteigen
Alter	→ Gesundheit gegenwärtig
Alter	→ Treppensteigen
Alter	→ anstrengende Tätigkeiten
Alter	→ ausgeglichen
Alter	→ jede Menge Energie
Männlich	→ niedergeschlagen
Einkommen	→ jede Menge Energie

Durch die Modellierung der DIF-Effekte verbessert sich der Fit des Gesamtmodells (s. Anhang, Tab. 12), sodass CFI, SRMR und RMSEA eine gute bis sehr gute Anpassung ausweisen. Die Effekte der Kovariaten auf den Gesundheitszustand verändern sich durch die Einführung der DIF-Effekte nur marginal und sind in allen Gruppen ähnlich hoch. Es gibt bei den türkischen Befragten weiterhin keinen signifikanten Effekt des Alters auf die physische Gesundheit und keinen Effekt der Voll- oder Teilzeiterwerbstätigkeit auf den mentalen Gesundheitszustand. Auch bei der Gruppe (Ex-)Jugoslawien lässt sich letzterer Effekt nicht finden sowie kein Effekt des Einkommens auf die körperliche Gesundheit. Bei Befragten aus den EG-Anwerbeländern

hingegen hat das Einkommen keinen Einfluss auf die mentale Gesundheit.

Im Falle des Geschlechts und des Einkommens werden die Effekte auf den physischen Gesundheitszustand durch die Berücksichtigung von DIF größer und die Effekte auf den mentalen Gesundheitszustand kleiner. Für die Kovariate Alter ist es umgekehrt: Der Effekt auf die physische Gesundheit wird durch den Einbezug von DIF kleiner und auf die mentale Gesundheit größer. Durch den bestehenden DIF-Effekt wären diese Kovariaten-Effekte somit jeweils über- oder unterschätzt worden.

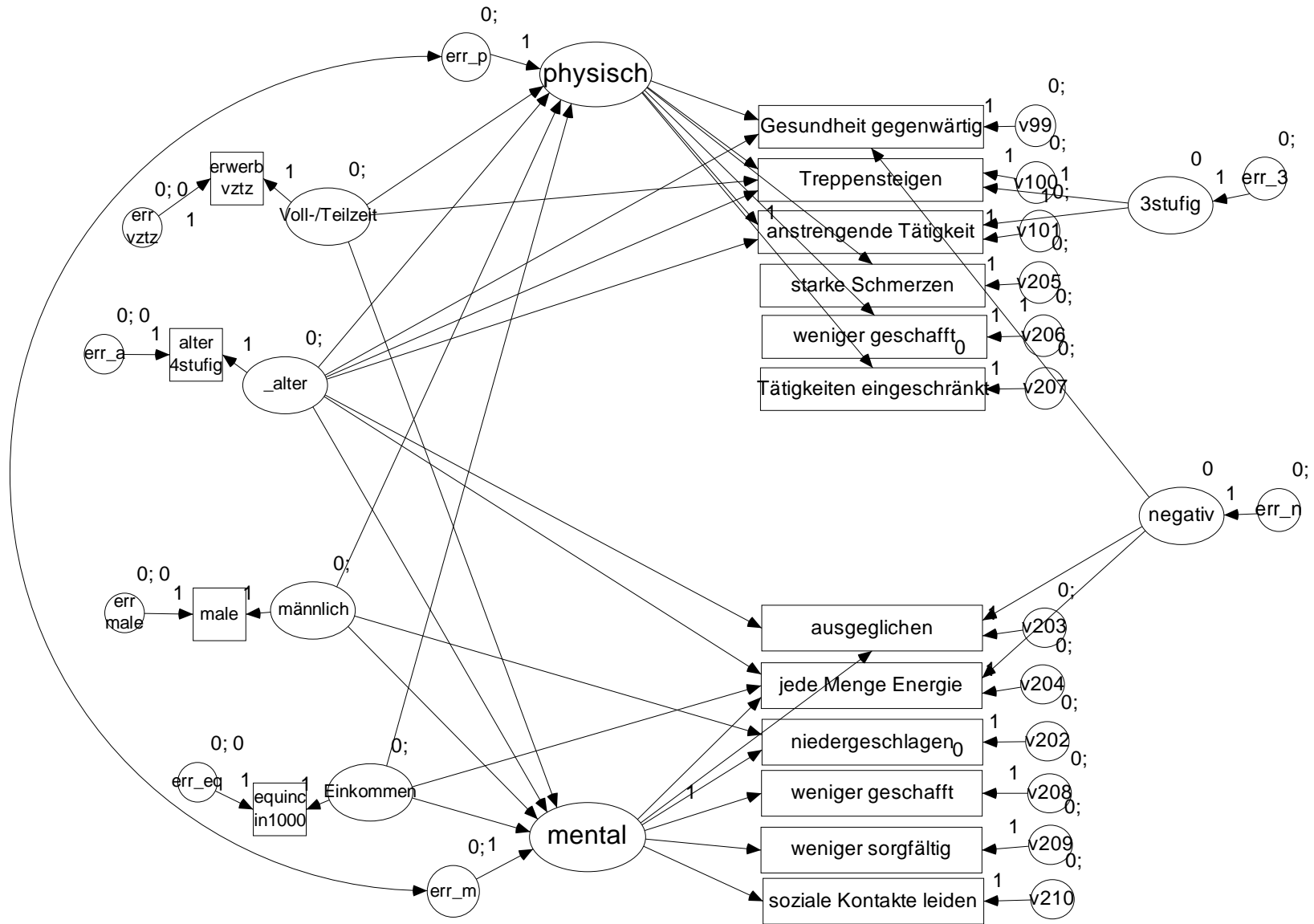


Abbildung 3: Modifiziertes Grundmodell mit signifikanten Kovariaten und spezifizierten DIF-Effekten

3.8 Äquivalenzrestriktionen und Modellfit

Das in Abbildung 3 dargestellte Modell mit acht spezifizierten DIF-Effekten wird nun über verschiedene Gleichheitsrestriktionen in den sechs Gruppen auf Äquivalenz geprüft (s. Tab. 5). Über den Test auf metrische Äquivalenz kann festgestellt werden, ob Befragte die Antwortskala in Gänze oder nur bestimmte Kategorien verwenden. Hierzu werden alle Faktorladungen des Gesundheitskonstrukts zwischen den Gruppen gleichgesetzt. Der χ^2 -Wert weist eine signifikante Modellverschlechterung aus. Allerdings zeigen die alternativen Fitindizes mit Ausnahme des RMSEA, der sich verbessert, weder eine Verbesserung noch eine Verschlechterung des Modellfit an.

Tabelle 5: Schrittweise Äquivalenzrestriktion der Parameter zwischen den Gruppen

	Modell 1 un- restringiert	Modell 2 metrische Äquivalenz	Modell 3 skalare Äquivalenz	Modell 4 Messfehler- äquivalenz	Modell 5 Äquivalenz latenter Mittelwerte
Fit-Indizes					
χ^2	17737,9	17848,3	18099,1	18628,4	18766,7
p	0,000	0,000	0,000	0,000	0,000
df	522	572	622	682	692
χ^2/df	34,0	31,2	29,1	27,3	27,1
$\Delta\chi^2$	-	110,4***	250,8 ***	529,3 ***	138,3 ***
$\Delta\chi^2/\Delta df$	-	2,2	5,0	8,8	13,8
RMSEA	0,036	0,034	0,033	0,032	0,032
$\Delta RMSEA$	-	-0,002	-0,001	-0,001	0,000
CFI	0,921	0,921	0,920	0,918	0,917
ΔCFI	-	0,000	-0,001	-0,002	-0,001
SRMR	0,066	0,066	0,066	0,066	0,066
$\Delta SRMR$	0,000	0,000	0,000	0,000	0,000

Anmerkungen: Die χ^2 -Differenz ist in allen Fällen auf dem 0,1%-Niveau signifikant.

Weiterhin werden auch die Achsenabschnitte der Items zwischen den Gruppen restringiert. Wiederum zeigt der χ^2 -Wert eine signifikante Modellverschlechterung an. Der CFI verschlechtert sich um 0,001, der RMSEA zeigt eine leichte Verbesserung an und der SRMR stagniert. Gemäß diesen Fitindizes liegt guter Modellfit vor.

Fasst man die beiden Tests auf metrische und skalare Äquivalenz nun inhaltlich zusammen, lassen sich hierdurch gruppenspezifische Unterschiede in der Verwendung der Skalenbreite und extremer Antwortkategorien aufdecken. Da der χ^2 -Differenztest ein Verwerfen der Äquivalenzannahme nahelegt, die anderen Fitindizes allerdings keine signifikante Verschlechterung anzeigen, wird die Entscheidung zugunsten der alternativen Fitindizes gefällt.

Der χ^2 wurde für das analysierte Modell vermutlich aufgrund nicht-normalverteilter Daten und einer sehr hohen Fallzahl durchweg überschätzt, sodass auch die χ^2 -Differenz wenig gültige Er-

gebnisse liefert. Diese Entscheidung wird auch durch die Betrachtung der Modifikationsindizes (M.I.)²⁰ gestützt: Es lassen sich keine konkreten Faktorladungen angeben, die vorwiegend allein für die Verschlechterung des Modellfit verantwortlich sind und demnach freigesetzt werden sollten. Es spricht somit nichts für die Existenz partieller metrischer oder skalarer Äquivalenz.

Auch für den Test auf Äquivalenz der Item-Messfehler zeigt sich auf Grundlage der χ^2 -Differenz eine signifikante Verschlechterung, wohingegen sich der RMSEA wiederum verbessert, der SRMR stagniert und der CFI eine Verschlechterung anzeigt. Diese liegt mit 0,002 jedoch weit unter dem Grenzwert von 0,01 für signifikante Verschlechterung. Die Gruppen unterscheiden sich somit nicht hinsichtlich des Ausmaßes an Zufallsfehler und weisen dieselbe Indikatorreliabilität auf. Auf Basis dieser drei Tests kann dem SF-12 „strong factorial invariance“ (Gregorich 2006: 16) zwischen den sechs Gruppen bescheinigt werden. Dies stellt die notwendige Grundlage dafür dar, gültige Mittelwertvergleiche durchführen zu können.

Die Gleichsetzung der Mittelwerte der physischen und der mentalen Gesundheit zwischen den sechs Gruppen führt zu einer Verringerung des CFI um 0,001 Einheiten und die Werte des RMSEA und SRMR bleiben gleich. Die χ^2 -Differenz ist wiederum signifikant. Da auch hier die alternativen Fitindizes zur Testentscheidung herangezogen werden, zeigen sich somit keine signifikanten Unterschiede sowohl bezüglich der physischen als auch der mentalen Gesundheit zwischen der Bevölkerung mit und ohne Migrationshintergrund sowie innerhalb der Migrantenbevölkerung je nach Herkunftsregion.

Betrachtet man die auf Grundlage dieser Restriktion geschätzten Mittelwerte, liegt der Mittelwert der physischen Gesundheit in allen Gruppen bei 3,978 und der Mittelwert der mentalen Gesundheit für alle Gruppen bei 4,190 (s. Tab. 6)²¹, was auf einen sehr guten selbstbeurteilten Gesundheitszustand der Bevölkerung unabhängig vom Bestehen eines Migrationshintergrundes hinweist.

²⁰ Der Modifikationsindex indiziert restringierte Parameter, durch deren Freisetzung sich der gesamte Modellfit verbessern lässt (vgl. Backhaus et al. 2006: 386). Dabei geben per Konvention Werte >4 Hinweis auf signifikante Abweichungen (vgl. Bühner 2006: 268), wobei sich auch ein Grenzwert von 7 etabliert hat.

²¹ Hier ist an die zugewiesene Metrik zu erinnern: Die latenten Mittelwerte bewegen sich auf der Skala von 1 (gesundheitlich eher beeinträchtigt) bis 5 (gesundheitlich weitgehend unbeeinträchtigt).

Tabelle 6: Schätzung der latenten Mittelwert der physischen und mentalen Dimension

latente Mittelwerte	West-deutschland	Ost-deutschland	Türkei	(Ex-) Jugoslawien	EG-Anwerbeländer	erw. osteurop. Länder
ohne Kovariaten						
physisch	4,018 (0,008)	3,948 (0,013)	4,119 (0,034)	4,011 (0,049)	4,135 (0,038)	4,057 (0,028)
mental	4,241 (0,007)	4,099 (0,012)	4,216 (0,033)	4,164 (0,045)	4,293 (0,034)	4,186 (0,026)
mit 4 Kovariaten						
physisch	3,984 (0,009)	3,913 (0,013)	4,126 (0,036)	3,990 (0,053)	4,126 (0,040)	4,011 (0,030)
mental	4,230 (0,008)	4,087 (0,012)	4,217 (0,036)	4,166 (0,047)	4,274 (0,037)	4,163 (0,028)
mit DIF						
physisch	3,984 (0,009)	3,913 (0,013)	4,126 (0,036)	3,990 (0,053)	4,126 (0,040)	4,011 (0,030)
mental	4,230 (0,008)	4,087 (0,012)	4,217 (0,036)	4,166 (0,047)	4,274 (0,037)	4,163 (0,028)
mit Äquivalenzrestriktionen λ , τ , ϵ						
physisch	3,990 (0,009)	3,907 (0,013)	4,102 (0,035)	3,957 (0,052)	4,091 (0,039)	4,020 (0,029)
mental	4,229 (0,008)	4,094 (0,012)	4,219 (0,034)	4,138 (0,047)	4,239 (0,036)	4,168 (0,027)
+ Äquivalenz-Restriktion α						
physisch				3,978 (0,007)		
mental				4,190 (0,006)		

Anmerkungen: Standardfehler sind in Klammern angegeben, signifikante Effekte ($p < 0,05$) sind fett gedruckt, nicht-signifikante Effekte kursiv.

4 Zusammenfassung

Die Analyse lässt sich grob in vier Schritte zusammenfassen: Zu Beginn wurde die Faktorstruktur des SF-12 überprüft, um etwaige Unterschiede in den Gesundheitsvorstellungen zu prüfen. Daraufhin wurden Kovariaten in das Modell eingeführt und deren Effekt auf den Gesundheitszustand untersucht. In einem dritten Schritt erfolgte die Modellierung der DIF-Effekte der Kovariaten zur Analyse der potentiellen Verzerrungen durch soziodemografische und sozio-ökonomische Charakteristika; im letzten Schritt wurde der Fokus auf die Äquivalenztests zwischen den sechs Herkunftsgruppen gelegt, an deren Ende die latenten Mittelwertvergleiche standen.

Im Großen und Ganzen zeigen die empirischen Daten der gesamten Stichprobe eine eher schlechte Anpassung an die theoretisch angenommene Faktorstruktur, wobei weitere psychometrische Analysen keine konkreten Ursachen dieses Misfit aufdecken konnten. Unter

der Annahme, dass dies bereits ein Indiz auf fehlende konzeptuelle Äquivalenz sein könnte, wurde die SF-12 Faktorstruktur daraufhin separat nach Herkunftsgruppen analysiert. Die Modellanpassung erwies sich allerdings in allen Gruppen als ähnlich schlecht.

Die standardisierten Residualkonvarianzen zeigten korrelierende Elemente in der Modellstruktur auf, die inhaltlich als zwei Methodeneffekte gedeutet wurden: Im Vergleich zu den positiv formulierten Items mit fünf Antwortkategorien geben Befragte bei den drei negativ formulierten Items und den Items mit dreistufiger Antwortskala verzerrte Antworten.

Durch Berücksichtigung dieser Methodeneffekte in allen Gruppen verbesserte sich die Modellanpassung wesentlich. Auf Grundlage des modifizierten Messmodells konnten die Annahme konzeptueller Äquivalenz wie auch der Test der Faktorstruktur insofern bestätigt werden, als nicht die gesundheitsbezogenen Items im Modell umstrukturiert wurden, sondern vielmehr eine Messfehlerkomponente in das Modell integriert wurde, die durch die Antwortskala des SF-12 verursacht wurde.

Es wurden daraufhin Kovariaten in das Modell eingeführt, für die ein Einfluss auf den Gesundheitszustand angenommen wurde. Diese Annahme wurde für die Bildung, den Zustand der Pensionierung und der Arbeitslosigkeit widerlegt, wobei Bildung teilweise in den Migrantengruppen den angenommenen Effekt ausübt. Arbeitslosigkeit hat lediglich bei Migranten türkischer Herkunft den angenommenen negativen Effekt. Der angenommene negative Effekt auf die physische Gesundheit bei Personen in Ruhestand zeigt sich lediglich in drei der sechs Gruppen. Damit muss die Hypothese gleicher Kovariateneffekte über alle Herkunftsgruppen verworfen werden.

Erwerbstätigkeit hat entgegen der Annahmen einen durchweg positiven Effekt in nahezu allen Gruppen. Dass in den Gruppen mit Migrationshintergrund ein zusätzlicher negativer Effekt der Erwerbstätigkeit auf die Gesundheit bestehe, konnte nicht belegt werden. Für die Kovariaten Alter, Geschlecht und Einkommen zeigten sich signifikante Effekte in der erwarteten Richtung.

Aus dem dritten Schritt, der Spezifizierung der DIF-Effekte, gehen zahlreiche DIF-Effekte für die Kovariaten Alter und die Erwerbstätigkeit, aber nur wenige Effekte für das Geschlecht und das Einkommen hervor²²: Hinsichtlich des Einkommens zeigt sich nur bei Befragten aus dem ehemaligen Jugoslawien bei zwei Items ein nennenswerter Effekt der vermuteten negativen Richtung. In den anderen Herkunftsgruppen resultieren teilweise positive Effekte.

Frauen beurteilen entgegen der Annahmen nur wenige Items negativer als Männer; bei der Frage nach dem generellen Gesundheitszustand zeigt sich sogar einen gegenläufigen Effekt.

²² Ein Test auf DIF der Kovariaten Bildung, Ruhestand und Arbeitslosigkeit wurde nicht mehr durchgeführt, da bereits die Haupteffekte nicht signifikant sind.

Für Befragte mit Migrationshintergrund spielt das Geschlecht diesbezüglich keine Rolle.

Erwerbstätigkeit erzeugt in den Gruppen mehrheitlich einen verzerrenden Effekt bei der Beurteilung des Items „ausgeglichen“: Erwerbstätige Befragte haben gruppenübergreifend stärker das Gefühl, nicht ausgeglichen zu sein und geben hier ein extremeres Urteil ab. Demgegenüber wird das Item „Treppensteigen“ in allen Herkunftsgruppen positiver beurteilt, wenn die Befragten erwerbstätig sind. Die Resultate sind somit eher inkonsistent und inhaltlich schwer zu deuten.

Am häufigsten besteht eine verzerrte Beantwortung aufgrund des Alters. Hier zeigen sich auch kulturspezifische Effekte: Die Frage nach dem Auftreten von Schmerzen wird lediglich in den Migrantengruppen verzerrt beantwortet. Dabei erfolgt mit zunehmendem Alter eine zu negative Beurteilung der Schmerzhäufigkeit. Im Gegensatz dazu treten bei den Befragten ohne Migrationshintergrund Verzerrungen bezüglich der physischen oder mentalen Rollenfunktion auf. Ältere Personen beurteilen dabei die Beeinträchtigung der Rollenfunktion weniger kritisch als der tatsächliche Gesundheitszustand vermuten lässt. Im Hinblick auf die Ausgeglichenheit besteht in allen Gruppen mit zunehmendem Alter ebenfalls ein eher unkritisches Urteil.

Generell weisen die Migrantengruppen seltener, dafür aber höhere DIF-Effekte auf als west- und ostdeutsche Befragte. Dies könnte auf die angenommenen schlechteren Deutschkenntnisse zurückzuführen sein, wobei sich keine systematischen Unterschiede in der Stärke der DIF-Effekte zwischen den Migrantengruppen erkennen lassen.

Es muss angemerkt werden, dass DIF-Analysen lediglich uniforme Verzerrungen aufdecken können. Wenn die Analysen somit seltener DIF-Effekte in den Migrantengruppen feststellen, dann können die Migrantengruppen dennoch nicht-uniform verzerrte Antworten abgeben. Personen ohne Migrationshintergrund weisen demnach häufiger uniformen DIF auf, d.h. dass sich beispielsweise extremes Antwortverhalten durch den gesamten SF-12-Fragebogen zieht.

Der Effekt von DIF auf die Schätzung der direkten Kovariateneffekte ist in allen Gruppen vergleichsweise gering und die Mittelwertschätzungen zeigen durch die DIF-Modellierung überhaupt keine Veränderung an. Auch sind kompensatorische Effekte vorstellbar, da sich die DIF-Effekte einzelner Kovariaten als teilweise gegenläufig erweisen.

Schlussendlich wurde Äquivalenz hinsichtlich der Verwendung der Antwortskala bestätigt, sowohl hinsichtlich der Skalenbreite, der Verwendung der Extremkategorien als auch hinsichtlich des Ausmaßes an Indikatorreliabilität zwischen den sechs Herkunftsgruppen. Auf dieser Grundlage konnten zuletzt die Mittelwerte verglichen werden, wobei dieser Test keine signifikanten Unterschiede ausgab. Demnach lassen sich keine Unterschiede im Gesundheitszustand zwischen den Gruppen und dementsprechend auch nicht zwischen Migranten und Nichtmigranten feststellen.

5 Ausblick

Da die quantitativ orientierte sozialepidemiologische Forschung zur Gesundheit der Migrantenbevölkerung in Deutschland aufgrund der schlechten Datenlage bislang nur sehr begrenzt ist, leistet diese Arbeit hierzu ihren Beitrag: Auf Grundlage der Daten des Sozio-oekonomischen Panels wurde mit einer verhältnismäßig großen Stichprobe von Personen mit Migrationshintergrund deren Gesundheitszustand analysiert und mit der einheimischen Bevölkerung verglichen. Hierzu wurde mit dem SF-12 ein aktuelles und umfassendes Messinstrument der gesundheitsbezogenen Lebensqualität verwendet, wohingegen bisherige Studien eher mortalitäts- und morbiditätsbezogene Indikatoren referieren und somit psychische und soziale Elemente von Gesundheit ausblenden.

Es kann als positive Entwicklung verzeichnet werden, dass umfassende gesundheitsbezogene Messinstrumente zunehmend auch in Bevölkerungsumfragen erhoben werden, da dies ein großes Potential für die Untersuchung gesundheitlicher Fragestellungen bietet. Gleichzeitig besteht dadurch die Gefahr, dass umfassende Itembatterien an Validität einbüßen, wenn sie nicht von allen potentiellen Befragten in gleicher Weise verstanden und beantwortet werden. Dies steht insbesondere im Falle der stetig wachsenden Bevölkerung mit Migrationshintergrund zu vermuten, die einen anderen kulturellen Hintergrund aufweist. Insofern trägt diese Arbeit auch zur methodologischen Forschung bei, indem die Vergleichbarkeit des Messinstruments SF-12 von Personengruppen mit und ohne Migrationshintergrund geprüft wird. In Deutschland besteht diesbezüglich noch wenig Bewusstsein für migrantensensible Forschung und die Vergleichbarkeit der Daten von Befragten verschiedener Kulturkreise wird allzu oft vernachlässigt und unhinterfragt angenommen.

Auf Basis von gepoolten SOEP-Daten aus den Jahren 2002, 2004, 2006 und 2008 wurden vier Migrantengruppen gebildet, die bezüglich ihres kulturellen und migrationsbiografischen Hintergrunds homogen sind. Diese wurden mit den beiden Personengruppen verglichen, die westdeutscher und ostdeutscher Herkunft sind. Über die Kombination aus „Multiple Group Confirmatory Factor Analysis“ (MGCFA), „Mean and Covariance Structure Analysis“ (MACS) und „Multiple Indicator Multiple Cause Models“ (MIMIC) ließen sich drei Fragestellungen untersuchen: Zum Einen die Überprüfung der interkulturellen Äquivalenz der SF-12-Items zwischen den sechs Gruppen, zum Anderen der Einfluss von Kovariaten sowohl auf den Gesundheitszustand als auch auf die Beantwortung der SF-12-Items. Schließlich wurde drittens der Test auf Mittelwertunterschiede der SF-12 Punktscores zwischen den Gruppen durchgeführt.

In der Analyse konnte gezeigt werden, dass der international angewandte SF-12 Fragebogen für die gesamte deutsche Bevölkerung sowohl mit als auch ohne Migrationshintergrund gültige und vergleichbare Daten liefert. Die Mittelwertvergleiche zeigten, dass auch unter Kontrolle

verschiedener Gesundheitsdeterminanten zwischen den sechs herkunftsspezifischen Gruppen keine substantiellen Unterschiede hinsichtlich ihrer physischen und mentalen Gesundheit bestehen. In allen Gruppen wird ein relativ guter selbstbeurteilter Gesundheitszustand angegeben.

Bei den psychometrischen Analysen zeigten sich durchweg starke Boden- und Deckeneffekte, sodass die SF-12-Items eine geringe Trennschärfe aufweisen. Außerdem rufen die negative Formulierung dreier Items sowie die beiden Items mit lediglich drei Antwortkategorien zusätzliche Antwortverzerrungen hervor.

Die Einführung der Kovariaten Geschlecht, Alter, Einkommen und Erwerbstätigkeit ergab nur schwache Drittvariableneffekte. Allerdings zeigte sich über die Analyse von „Differential Item Functioning“ (DIF), dass Befragte mit bestimmten soziodemografischen und sozioökonomischen Merkmalen verschiedene Items verzerrt, d.h. zu positiv oder zu negativ beantworten. Diese Effekte ließen sich jedoch nur teilweise sinnvoll erklären.

Zusammenfassend belegen die Analysen die äquivalente Bedeutung des SF-12 Fragebogen und seine Verwendbarkeit in vergleichenden Analysen zwischen Personen mit und ohne Migrationshintergrund. Allerdings zeigen sich generelle Schwächen des Messinstruments, was die Trennschärfe und die Effekte der Negativformulierung und teilweise kürzeren Antwortskala einzelner Items anbelangt. Darüber hinaus ließ sich eine unterschiedliche Interpretation von Frageformulierungen und Verwendung von Antwortskalen einzelner Items angesichts soziostruktureller Merkmale feststellen. Das Ausmaß an Verzerrung in der Analyse stellte sich jedoch als eher marginal heraus.

Es müssen jedoch einige Grenzen der Analyse aufgezeigt werden: Zum Einen handelt es sich bei dem im Sozio-oekonomischen Panel verwendeten SF-12-Fragebogen um eine abgewandelte Form, des SF-12, d.h. die Befunde können nur begrenzt auf die Standardform des SF-12 übertragen werden.

Außerdem wurden lediglich sieben Kovariaten getestet. Es sind zahlreiche weitere Kovariaten vorstellbar, beispielsweise integrations- und assimilationsbezogene Indikatoren, sowie detaillierte Informationen zur beruflichen Tätigkeit und zum Familienstand, die den Gesundheitszustand beeinflussen und DIF-Effekte hervorrufen können: Diesbezüglich waren jedoch nicht ausreichend Daten verfügbar. Es wäre außerdem sinnvoll, zwischen Migranten mit direktem und indirektem Migrationshintergrund zu differenzieren. Dabei muss jedoch stets bedacht werden, dass solche Analysen nur mit einer großen Fallzahl möglich sind.

Die bestehenden DIF-Effekte lassen sich kaum inhaltlich sinnvoll interpretieren. So ergaben sich oftmals nicht ausschließlich Effekte in der angenommenen Richtung, sondern bei einzelnen Items häufig gegenläufige Effekte. Hinsichtlich Erklärung der gefundenen Effekte stößt das eingesetzte Analyseverfahren an seine Grenzen; es kann vielmehr nur deskriptiv und explorativ

Unterschiede testen und unter Rückgriff auf migrationsbezogene Kontextinformationen lassen sich Vermutungen über deren Ursachen äußern.

Für eine ausführlichere Untersuchung der Ursachen fehlender Vergleichbarkeit von Messinstrumenten ist somit die Anwendung qualitativer Verfahren wie behavior coding und think aloud-Interviews (Johnson 1998) oder „experimenteller“ Verfahren mit verschiedenen Fragebogenversionen und Multi-Trait-Multi-Method-Analysen (Saris 2003) unersetzlich.

6 Literatur

Amir, Marianne, Noah Lewin-Epstein, Becker Gideon und Dan Buskila, 2002: Psychometric Properties of the SF-12 (Hebrew Version) in a Primary Care Population in Israel. Medical Care 40(10): S.918-928.

Andersen, H. H., A. Mühlbacher, M. Nübling, J. Schupp und G. G. Wagner, 2007: Computation of Standard Values for Physical and Mental Health Scale Scores Using the SOEP Version of SF-12v2. Journal of Applied Social Science Studies 127(1): S.171-182.

Aquino, EM, GM Menezes und MB Amoedo, 1992: Gender and Health in Brazil: Considerations based on the National Household Sampling Survey. Revista de Saúde Pública 26(3): S.195-202.

Aschauer, Wolfgang, 2009: Besonderheiten und Problemlagen der quantitativen Befragung bei MigrantInnen. S. 293-314, in: Weichbold, Martin, Johann Bacher und Christof Wolf (Hg.), Umfrageforschung: Herausforderungen und Grenzen. Österreichische Zeitschrift für Soziologie Sonderheft 9/2009. Wiesbaden: VS.

Backhaus, Klaus, Bernd Erichson, Wulff Plinke und Rolf Weiber, 2006: Multivariate Analysemethoden: Eine anwendungsorientierte Einführung. Berlin: Springer.

Bollen, Kenneth A., 1989: Structural Equations with Latent Variables. New York: Wiley.

Bühner, Markus, 2006: Einführung in die Test- und Fragebogenkonstruktion. München: Pearson Studium.

Bullinger, Monika, 2000: Erfassung der gesundheitsbezogenen Lebensqualität mit dem SF-36-Health Survey. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 43(3): S.190-197.

Bundesministerium für Familie, Senioren, Frauen und Jugend,, 2008: Der Mikrozensus im Schnittpunkt von Geschlecht und Migration. Möglichkeiten und Grenzen einer sekundär-analytischen Auswertung des Mikrozensus 2005. Baden-Baden: Nomos.

Celentano, David D., Martha S. Linet und Walter F. Stewart, 1990: Gender Differences in the Experience of Headache. Social Science & Medicine 30(12): S.1289-1295.

Cerda-Hegerl, Patricia, 2008: Interkulturelle Aspekte in der medizinischen Versorgung nichtdokumentierter Migranten. *Psychotherapie - Psychosomatik - Medizinische Psychologie* 58(03/04): S.136-145.

Cheak-Zamora, N. C., K. W. Wyrwich und T. D. McBride, 2009: Reliability and Validity of the SF-12v2 in the Medical Expenditure Panel Survey. *Quality of Life Research* 18: S.727-735.

Chen, Chuansheng, Shin-ying Lee und Harold W. Stevenson, 1995: Response Style And Cross-Cultural Comparisons Of Rating Scales Among East Asian And North American Students. *Psychological Science* 6(3): S.170-175.

Cheung, Gordon W. und Roger B. Rensvold, 1998: Cross-Cultural Comparisons using Non-Invariant Measurement Items. *Applied Behavioral Science Review* 6(1): S.93-110.

Clarke, Irvine, 2001: Extreme Response Style in Cross-Cultural Research. *International Marketing Review* 18(3): S.301-324.

Cornelißen, Waltraud (Hg.), 2005: Gender Datenreport - 1. Datenreport zur Gleichstellung von Frauen und Männern in der Bundesrepublik Deutschland. München.

d'Uva, Teresa Bago, Eddy van Doorslaer, Maarten Lindeboom und Owen O'Donnell, 2008: Does Reporting Heterogeneity Bias the Measurement of Health Disparities? *Health Economics* 17(3): S.351-375.

Diehl, Claudia und Rainer Schnell, 2006: "Reactive Ethnicity" or "Assimilation"? Statements, Arguments, and First Empirical Evidence for Labor Migrants in Germany. *International Migration Review* 40(4): S.786-816.

Diekmann, Andreas, 2007: Empirische Sozialforschung: Grundlagen, Methoden, Anwendungen. Reinbek: Rowohlt Taschenbuch.

DiStefano, Christine und Robert W. Motl, 2009: Self-Esteem and Method Effects Associated With Negatively Worded Items: Investigating Factorial Invariance by Sex. *Structural Equation Modeling: A Multidisciplinary Journal* 16(1): S.134-146.

DIW Berlin, 2005: Desktop Companion to the German Socio-Economic Panel (SOEP). <http://www.diw.de/sixcms/detail.php/38951> (1.7.2011).

DIW Berlin, 2008: SOEP Personenfragebogen 2008. http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.85359.de/personen_2008.pdf (1.7.2011).

Doorslaer, Eddy van und Xander Koolman, 2004: Explaining the Differences in Income-Related Health Inequalities across European Countries. *Health Economics* 13(7): S.609-628.

Eikemo, T. A., C. Bambra, K. Joyce und Espen Dahl, 2008: Welfare state regimes and income-related health inequalities: a comparison of 23 European countries. *The European Journal of Public Health* 18(6): S.593-599.

Elkeles, Thomas und Wolfgang Seifert, 1993: Arbeitslose und ihre Gesundheit: Langzeit-

analysen für die Bundesrepublik Deutschland. Sozial- und Präventivmedizin(38): S.148-155.

Elkeles, Thomas und Wolfgang Seifert, 1996: Immigrants and Health: Unemployment and Health-Risks of Labour Migrants in the Federal Republic of Germany, 1984-1992. Social Science and Medicine 43(7): S.1035-1047.

Faltermeier, Toni, 2005: Migration und Gesundheit: Fragen und Konzepte aus einer salutogenetischen und gesundheitspsychologischen Perspektive. S. 93-112, in: *P., Marschalck und Wiedl K. H.* (Hg.), Migration und Krankheit. Göttingen: V&R unipress.

Farivar, S., W. E. Cunningham und R. D. Hays, 2007: Correlated Physical and Mental Summary Scores for the SF-36 and SF-12 Health Survey, V.1. Health and Quality of Life Outcomes 5: S.54-62.

Fleishman, John A. und William F. Lawrence, 2003: Demographic Variation in SF-12 Scores: True Differences or Differential Item Functioning? Medical Care 41(7): S.III75-III86.

Fleishman, John A., William D. Spector und Barbara M. Altman, 2002: Impact of Differential Item Functioning on Age and Gender Differences in Functional Disability. Journal of Gerontology 57B(5): S.275-284.

Flick, Uwe, Beate Hoose und Petra Sitta, 1998: Gesundheit und Krankheit gleich Saúde und Doença? Gesundheitsvorstellungen bei Frauen in Deutschland. S. 32-47, in: *Flick, Uwe* (Hg.), Wann fühlen wir uns gesund? Subjektive Vorstellungen von Gesundheit und Krankheit. Weinheim: Juventa.

Fornell, Claes und David F. Larcker, 1981: Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 18(1): S.39-50.

Gandek, B., J. E. Ware, N. K. Aaronson, G. Apolone, J. B. Bjorner, J. E. Brazier, S. Kaasa, A. Lepleg, L. Prieto und M. Sullivan, 1998: Cross-Validation of Item Selection and Scoring for the SF-12 Health Survey in Nine Countries: Results from the IQOLA Project. Journal of Clinical Epidemiology 51: S.1171-1178.

Gandhi, Sanjay K., J. Warren Salmon, Sean Z. Zhao, Bruce L. Lambert, Prasanna R. Gore und Kendon Conrad, 2001: Psychometric Evaluation of the 12-Item Short-Form Health Survey (SF-12) in Osteoarthritis and Rheumatoid Arthritis Clinical Trials. Clinical Therapeutics 23(7): S.1080-1098.

Gesundheitsberichterstattung des Bundes, 2008: Schwerpunktbericht der Gesundheitsberichterstattung des Bundes – Migration und Gesundheit. Robert-Koch-Institut und Statistisches Bundesamt. Berlin.

Gordo, Laura Romeu, 2006: Effects of Short- and Long-Term Unemployment on Health Satisfaction: Evidence from German Data. Applied Economics 38(20): S.2335-2350.

Grabka, Markus und Jürgen Schupp, 2005: Erste Erfahrungen mit den neuen SOEP Gesundheitsindikatoren (2002-2004). S. 67-82, in: *Schupp, Jürgen* (Hg.), Befragungsgestützte Messung von Gesundheit - Bestandsaufnahme und Ausblick. Berlin: DIW Berlin.

Greenleaf, Eric A., 1992: Improving Rating Scale Measures by Detecting and Correcting Bias Components in Some Response Styles. *Journal of Marketing Research* 29(2): S.176-188.

Gregorich, Steven E., 2006: Do Self-Report Instruments Allow Meaningful Comparisons Across Diverse Population Groups? Testing Measurement Invariance Using the Confirmatory Factor Analysis Framework. *Med Care* 44(11 Supplement 3): S.78–94.

Groenemeyer, A., 2003: Kulturelle Differenz, ethnische Identität und die Ethnisierung von Alltagskonflikten. Ein Überblick sozialwissenschaftlicher Thematisierungen. S. 11-46, in: *Groenemeyer, A. und J. Mansel* (Hg.), *Die Ethnisierung von Alltagskonflikten*. Opladen: Leske & Budrich.

Groenvold, Mogens, Jakob B. Bjorner, Marianne Carol Klee und Svend Kreiner, 1995: Test for Item Bias in a Quality of Life Questionnaire. *Journal of Clinical Epidemiology* 48(6): S.805-816.

Groot, Wim, 2000: Adaptation and Scale of Reference Bias in Self-Assessments of Quality of Life. *Journal of Health Economics* 19(3): S.403-420.

Groot, Wim, 2003: Scale of Reference Bias and the Evolution of Health. *European Journal of Health Economics* 4: S.176-183.

Hammelstein, Philipp, Johannes Pohl, Swantje Reimann und Marcus Roth, 2006: Persönlichkeitsmerkmale. S. 61-106, in: *Renneberg, Babette und Philipp Hammelstein* (Hg.), *Gesundheitspsychologie*. Springer-Lehrbuch. Berlin/Heidelberg: Springer.

Hancock, Gregory R., Frank R. Lawrence und Jonathan Nevitt, 2000: Type I Error and Power of Latent Mean Methods and MANOVA in Factorially Invariant and Noninvariant Latent Variable Systems. *Structural Equation Modeling: A Multidisciplinary Journal* 7(4): S.534-556.

Harzing, Anne-Wil, 2005: Does the Use of English-language Questionnaires in Cross-national Research Obscure National Differences? *International Journal of Cross Cultural Management* 5(2): S.213-224.

Harzing, Anne-Wil, 2006: Response Styles in Cross-national Survey Research: A 26-country Study. *International Journal of Cross Cultural Management* 6(2): S.243-266.

Haug, Sonja und Lenore Sauer, 2007: Zuwanderung und Integration von (Spät-)Aussiedlern - Ermittlung und Bewertung der Auswirkungen des Wohnortzuweisungsgesetzes. Forschungsstudie im Auftrag des Bundesministeriums des Innern. Nürnberg. Bundesamt für Migration und Flüchtlinge.

Hildebrandt, Lutz und Dirk Temme, 2006: Probleme der Validierung mit Strukturgleichungsmodellen. SFB 649 Discussion Paper 82.

<http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2006-082.pdf> (1.7.2011).

Hoffmann, Ch, B. H. McFarland, D. Kinzie, L. Bresler, D. Rakhlin, S. Wolf und A. E. Kovas, 2005: Psychometric Properties of a Russian Version of the SF-12 Health Survey in a Refugee Population. *Comprehensive Psychology* 46: S.390-397.

Horan, Patrick M., Christine DiStefano und Robert W. Motl, 2003: Wording Effects in Self-

Esteem Scales: Methodological Artifact or Response Style? *Structural Equation Modeling: A Multidisciplinary Journal* 10(3): S.435-455.

Hraba, Joseph, Frederick Lorenz, Gang Lee und Zdenka Pechacová, 1996: Gender Differences in Health: Evidence from the Czech Republic. *Social Science & Medicine* 43(10): S.1443-1451.

Hui, C. Harry und Harry C. Triandis, 1989a: Effects of Culture and Response Format on Extreme Response Style. *Journal of Cross-Cultural Psychology* 20: S.296-309.

Hui, C. Harry und Harry C. Triandis, 1989b: Effects of Culture and Response Format on Extreme Response Style. *Journal of Cross-Cultural Psychology* 20(296-309).

Hurrelmann, Klaus, 2006: *Gesundheitssoziologie: Eine Einführung in sozialwissenschaftliche Theorien von Krankheitsprävention und Gesundheitsförderung*. Weinheim und München: Juventa.

Idler, Ellen L., 1993: Age Differences in Self-Assessments of Health: Age Changes, Cohort Differences, or Survivorship? *Journal of Gerontology* 48(6): S.S289-S300.

Jenkinson, Crispin, Tarani Chandola, Angela Coulter und Stephen Bruster, 2001: An Assessment of the Construct Validity of the SF-12 Summary Scores across Ethnic Groups. *Journal of Public Health* 23(3): S.187-194.

Johnson, Timothy, Patrick Kulesa, Isr Llc, Young Ik Cho und Sharon Shavitt, 2005: The Relation Between Culture and Response Styles: Evidence From 19 Countries. *Journal of Cross-Cultural Psychology* 36: S.264-277.

Johnson, Timothy, D. O'Rourke, N. Chavez, S. Sudman, R. Warnecke, L. Lacey und J. Horm, 1997: Social Cognition and Responses to Survey Questions among Culturally Diverse Populations. S. 87-113, in: *Lyberg, L., P. Biemer, M. Collins, E. de Leeuw, C. Dippo, N. Schwarz und D. Trewin* (Hg.), *Survey Measurement and Process Quality*. New York: Wiley.

Johnson, Timothy P., 1998: Approaches to Equivalence in Cross-Cultural and Cross-National Survey-Research. *ZUMA Nachrichten Spezial Band 3: Cross-cultural Survey Equivalence* S.1-40.

Johnson, Timothy R. und Fons J. R. van de Vijver, 2003: Social Desirability in Cross-Cultural Research. S. 195-204, in: *Harkness, Janet A., Fons J. R. van de Vijver und Peter Ph. Mohler* (Hg.), *Cross-Cultural Survey Methods*. Hoboken: Wiley.

Jylhä, M., J. M. Guralnik, L. Ferrucci, J. Jokela und E. Heikkinen, 1998: Is Self-Rated Health Comparable across Cultures and Genders? *Journal of Gerontology: Social Sciences* 53(3): S.144-152.

Kieruj, Natalia D. und Guy Moors, 2010: Variations in Response Style Behavior by Response Scale Format in Attitude Research. *International Journal of Public Opinion Research* 22(3): S.320-342.

Kirkcaldy, B., U. Wittig, A. Furnham, M. Merbach und R. Siefen, 2006: Migration und Gesundheit. *Psychosoziale Determinanten*. *Bundesgesundheitsblatt - Gesundheitsforschung -*

Gesundheitsschutz 49(9): S.873-883.

Kline, Rex B., 2004: Principles and practice of structural equation modeling. New York: Guilford Press.

Kohls, Martin, 2008: Healthy-Migrant-Effect, Erfassungsfehler und andere Schwierigkeiten bei der Analyse der Mortalität von Migranten. Eine Bestandsaufnahme. Working Paper 15. Bundesamt für Migration und Flüchtlinge. Nürnberg.

Kontodimopoulos, Nick, Evelina Pappa, Dimitris Niakas und Yannis Tountas, 2007: Validity of SF-12 summary scores in a Greek general population. Health and Quality of Life Outcomes 5: S.55-65.

Krause, Neal, Jersey Liang, Arvind Jain und Hidehiro Sugisawa, 1998: Gender Differences in Health among the Japanese Elderly. Archives of Gerontology and Geriatrics 26(2): S.141-159.

Kroh, M., 2005: Intervieweffekte bei der Erhebung des Körpergewichts in Bevölkerungsumfragen. Gesundheitswesen 67(8/9): S.646-655.

Lahelma, Eero, Pekka Martikainen, Ossi Rahkonen und Karri Silventoinen, 1999: Gender Differences in Illhealth in Finland: Patterns, Magnitude and Change. Social Science & Medicine 48(1): S.7-19.

Lam, C. L., E. Y. Tse und B. Gandek, 2005: Is the Standard SF-12 Health Survey Valid and Equivalent for a Chinese Population? Quality of Life Research 14: S.539-547.

Lampert, Thomas und Thomas Ziese, 2005: Armut, soziale Ungleichheit und Gesundheit. Expertise des Robert Koch-Institutes zum 2. Armuts- und Reichtumsbericht der Bundesregierung. Robert-Koch-Institut. Berlin.

Lee, C. und R.T. Green, 1991: Cross-cultural Examination of the Fishbein Behavioral Intentions Model. Journal of International Business Studies 25(2): S.289-305.

Leinonen, Raija, Eino Heikkinen und Marja Jylhä, 2001: Predictors of Decline in Self-Assessments of Health among Older People -- A 5-Year Longitudinal Study. Social Science & Medicine 52(9): S.1329-1341.

Lim, L. L. Y. und J. D. Fisher, 1999: Use of the 12-item Short-Form (SF-12) Health Survey in an Australian Heart and Stroke Population. Quality of Life Research 8(1): S.1-8.

Little, Todd D., 1997: Mean and covariance structures (MACS) analyses of Cross-Cultural Data: Practical and Theoretical Issues. Multivariate Behavioral Research 32(1): S.53-76.

Little, Todd D., David W. Slegers und Noel A. Card, 2006: A Non-arbitrary Method of Identifying and Scaling Latent Variables in SEM and MACS Models. Structural Equation Modeling: A Multidisciplinary Journal 13(1): S.59-72.

Locke, Kenneth D. und Kyoung-Deok Baik, 2009: Does an Acquiescent Response Style Explain Why Koreans are Less Consistent than Americans? Journal of Cross-Cultural Psychology 40(2): S.319-323.

Luo, X., M. Lynn George, I. Kakouras, C. L. Edwards, R. Pietrobon, W. Richardson und L. Hey, 2003: Reliability, Validity and Responsiveness of the Short Form 12-Item Survey (SF-12) in Patients with Back Pain. *Spine* 28: S.1739-1745.

Mackenbach, Johan P., Irina Stirbu, Albert-Jan R. Roskam, Maartje M. Schaap, Gwenn Menvielle, Mall Leinsalu und Anton E. Kunst, 2008: Socioeconomic Inequalities in Health in 22 European Countries. *The New England Journal of Medicine* 358(23): S.2468-2481.

Marin, Gerardo, Raymond J. Gamba und Barbara V. Marin, 1992: Extreme Response Style and Acquiescence among Hispanics: The Role of Acculturation and Education. *Journal of Cross-Cultural Psychology* 23(4): S.498-509.

Marsh, Herbert W., Danielle K. Tracey und Rhonda G. Craven, 2006: Multidimensional Self-Concept Structure for Preadolescents With Mild Intellectual Disabilities: A Hybrid Multigroup-MIMIC Approach to Factorial Invariance and Latent Mean Differences. *Educational and Psychological Measurement* 66(5): S.795-818.

Mattes, Petra, 1998: Gesundheit und Krankheit im internationalen Vergleich. Einstellungen in Großbritannien und der Bundesrepublik. S. 160-170, in: *Flick, Uwe* (Hg.), Wann fühlen wir uns gesund? Subjektive Vorstellungen von Gesundheit und Krankheit. Weinheim: Juventa.

Maurischat, C. und Anja Krüger-Bödeker, 2004: Analysen zum Strukturmodell des SF-36/SF-12 - eine Übersicht. S. 29-48, in: *Maurischat, Carsten, Matthias Morfeld und Thomas Kohlmann* (Hg.), Lebensqualität: Nützlichkeit und Psychometrie des Health Survey SF-36/SF-12 in der medizinischen Rehabilitation. Lengerich: Pabst Science Publishers.

Maurischat, C., I. Ehlebracht-König, A. Kühn und M. Bullinger, 2005: Strukturelle Validität des Short Form 36 (SF-36) bei Patienten mit entzündlich-rheumatischen Erkrankungen. *Zeitschrift für Rheumatologie* 64(4): S.255-264.

Meadows, Keith A., Tim Greene, Lorrain Foster und Steven Beer, 2000: The Impact of Different Response Alternatives on Responders' Reporting of Health-Related Behaviour in a Postal Survey. *Quality of Life Research* 9(4): S.385-391.

Meisenberg, Gerhard und Amandy Williams, 2008: Are Acquiescent and Extreme Response Styles Related to Low Intelligence and Education? *Personality and Individual Differences* 44(7): S.1539-1550.

Mielck, A., 2008: Soziale Ungleichheit und Gesundheit in Deutschland. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 51(3): S.345-352.

Montazeri, Ali, Mariam Vahdaninia, Sayed Mousavi und Speideh Omidvari, 2009: The Iranian Version of 12-Item Short Form Health Survey (SF-12): Factor Structure, Internal Consistency and Construct Validity. *BMC Public Health* 9(1): S.341-350.

Morfeld, Matthias, Stefan Dietsche, Wolfgang Bürger und Uwe Koch, 2003: Der SF-12 – Das Problem der Missing Data. *Diagnostica* 49(3): S.129-135.

Nathanson, Constance A., 1975: Illness and the Feminine Role: A Theoretical Review. *Social Science & Medicine* 9(2): S.57-62.

Oltmer, Jochen, 2009: Zuwanderung und Integration in Deutschland seit dem Zweiten Weltkrieg. S. 151-170, in: *Gesemann, Frank und Roland Roth* (Hg.), Lokale Integrationspolitik in der Einwanderungsgesellschaft. Migration und Integration als Herausforderung von Kommunen. Wiesbaden: VS.

Ployhart, Robert E. und Frederick L. Oswald, 2004: Applications of Mean and Covariance Structure Analysis: Integrating Correlational and Experimental Approaches. *Organizational Research Methods* 7(1): S.27-65.

Podsakoff, Philip M., Scott B. MacKenzie, Jeong-Yeon Lee und Nathan P. Podsakoff, 2003: Common Method Biases in Behavioral Research: A Critical Review of the Literature and Recommended Remedies. *Journal of Applied Psychology* 88(5): S.879-903.

Rahman, Omar, John Strauss, Paul Gertler, Deanna Ashley und Kristin Fox, 1994: Gender Differences in Adult Health: An International Comparison. *The Gerontologist* 34(4): S.463-469.

Razum, O. und D. Twardella, 2002: Time Travel with Oliver Twist - Towards an Explanation for a Paradoxically Low Mortality among Recent Immigrants. *Tropical Medicine & International Health* 7: S.4-10.

Razum, Oliver, 2006: Migration, Mortalität und Healthy-Migrant-Effekt. S. 255-270, in: *Richter, Matthias und Klaus Hurrelmann* (Hg.), Gesundheitliche Ungleichheit. Grundlagen, Probleme, Perspektiven. Wiesbaden: VS.

Razum, Oliver, Ingrid Geiger, Hajo Zeeb und Ulrich Ronellenfitsch, 2004: Gesundheitsversorgung von Migranten. *Deutsches Ärzteblatt* 101(43): S.A 2882-A 2887.

Robert, Christoph, Wayne C. Lee und Kim-Yin Chan, 2006: An Empirical Analysis of Measurement Equivalence with the Indcol Measure of Individualism and Collectivism: Implications for Valid Cross-Cultural Inference. *Personnel Psychology* 59: S.65-99.

Roszkowski, Michael J. und Margot Soven, 2010: Shifting Gears: Consequences of Including Two Negatively Worded Items in the Middle of a Positively Worded Questionnaire. *Assessment & Evaluation in Higher Education* 35(1): S.113-130.

Salman, Ramazan und Katja Nagasso Djomo, 2009: Migration, Integration und Gesundheit in Deutschland. S. 555-572, in: *Gesemann, Frank und Roland Roth* (Hg.), Lokale Integrationspolitik in der Einwanderungsgesellschaft. Migration und Integration als Herausforderung von Kommunen. Wiesbaden: VS.

Salyers, Michelle P., Hayden B. Bosworth, Jeffrey W. Swanson, Jerilynn Lamb-Pagone und Fred C. Osher, 2000: Reliability and Validity of the SF-12 Health Survey among People with Severe Mental Illness. *Medical Care* 38(11): S.1141-1150.

Saris, Willem E., 2003: Multitrait-Multimethod Studies. S. in: *Harkness, Janet A., Fons J. R. van de Vijver und Peter Ph. Mohler* (Hg.), Cross-Cultural Survey Methods. Hoboken: Wiley.

Schenk, L. und H. Neuhauser, 2005: Methodische Standards für eine migrantensensible Forschung in der Epidemiologie. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 48(3): S.279-286.

Schenk, L. und H. Knopf, 2007: Mundgesundheitsverhalten von Kindern und Jugendlichen in Deutschland. Erste Ergebnisse aus dem Kinder- und Jugendgesundheitsurvey (KiGGS). Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 50(5/6): S.653-658.

Schmitt, Neal und Goran Kuljanin, 2008: Measurement Invariance: Review of Practice and Implications. Human Resource Management Review 18(4): S.210-222.

Schnell, Rainer, Paul Bernhard Hill und Elke Esser, 2008: Methoden der empirischen Sozialforschung. München: Oldenbourg.

Schwarz, Norbert, Bärbel Knäuper, Daphna Oysermann und Christine Stich, 2008: The Psychology of Asking Questions. S. 18-34, in: *De Leeuw, Edith Desirée, Joop Hox und Don A. Dillman* (Hg.), International Handbook of Survey Methodology. New York: Erlbaum.

Singh, Jagdip, 1995: Measurement in Cross-National Research. Journal of International Business Studies 26(3): S.597-619.

Smith, Peter B., 2004: Acquiescent Response Bias as an Aspect of Cultural Communication Style. Journal of Cross-Cultural Psychology 35(1): S.50-61.

Spallek, Jacob und Oliver Razum, 2008: Erklärungsmodelle für die gesundheitliche Situation von Migrantinnen und Migranten. S. 271-290, in: *Bauer, Ullrich, Uwe H. Bittlingmayer und Matthias Richter* (Hg.), Health Inequalities: Determinanten und Mechanismen gesundheitlicher Ungleichheit. Wiesbaden: VS.

Statistisches Bundesamt, 2010: Statistisches Jahrbuch 2010.

<http://www.destatis.de/jetspeed/portal/cms/Sites/destatis/SharedContent/Oeffentlich/B3/Publikation/Jahrbuch/StatistischesJahrbuch,property=file.pdf>

Stern, Michael J., Don A. Dillman und Jolene D. Smyth, 2007: Visual Design, Order Effects, and Respondent Characteristics in a Self-Administered Survey. Survey Research Methods 1(3): S.121-138.

Temme, Dirk und Lutz Hildebrandt, 2008: Gruppenvergleiche bei hypothetischen Konstrukten – Die Prüfung der Übereinstimmung von Messmodellen mit der Strukturgleichungsmethodik. Eine Untersuchung am Beispiel der Markenforschung. SFB 649 Discussion Paper 42. <http://sfb649.wiwi.hu-berlin.de/papers/pdf/SFB649DP2008-042.pdf> (1.7.2011).

Tomas, Jose M. und Amparo Oliver, 1999: Rosenberg's Self-Esteem Scale: Two Factors or Method Effects. Structural Equation Modeling: A Multidisciplinary Journal 6(1): S.84-98.

van Herk, Hester, Ype H. Poortinga und Theo M. M. Verhallen, 2004: Response Styles in Rating Scales. Evidence of Method Bias in Data from six EU countries. Journal of Cross-Cultural Psychology 35(3): S.346-360.

van Steenkiste, M., 2004: Zugang zu zahnärztlichen Leistungen und Einstellung zum Zahnarzt bei deutschen und türkischen Eltern. Gesundheitswesen 66(2): S.93-101.

van Steenkiste, M., A. Becher, R. Banschbach, S. Gaa, S. Kreckel und C. Pocanschi, 2004: Prävalenz von Karies, Fissurenversiegelungen und Füllungsmaterial bei deutschen Kindern und

Kindern von Migranten. *Gesundheitswesen* 66(11): S.754-758.

Vandenberg, Robert J. und Charles E. Lance, 2000: A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods* 3(1): S.4-70.

Velez, Pauline und Steven D. Ashworth, 2007: The Impact of Item Readability on the Endorsement of the Midpoint Response in Surveys. *Survey Research Methods* 1(2): S.69-74.

von Gostomski, Christian Babka, 2008: Türkische, griechische, italienische und polnische Personen sowie Personen aus den Nachfolgestaaten des ehemaligen Jugoslawien in Deutschland. Erste Ergebnisse der Repräsentativbefragung „Ausgewählte Migrantengruppen in Deutschland 2006/2007“ (RAM). Working Paper 11. Nürnberg. Bundesamt für Migration und Flüchtlinge.

von Gostomski, Christian Babka, 2010: Fortschritte der Integration. Zur Situation der fünf größten in Deutschland lebenden Ausländergruppen. Forschungsbericht 8. Berlin. Bundesministeriums des Innern.

Ware, John E., Mark Kosinski und Susan D. Keller, 1996: A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. *Medical Care* 34(3): S.220-233.

Watson, Dorothy, 1992: Correcting for Acquiescent Response Bias in the Absence of a Balanced Scale: An Application to Class Consciousness. *Sociological Methods Research* 21(1): S.52-88.

Welkenhuysen-Gybels, Jerry, Jaak Billiet und Bart Cambre, 2003: Adjustment for Acquiescence in the Assessment of the Construct Equivalence of Likert-Type Score Items. *Journal of Cross-Cultural Psychology* 34(6): S.702-722.

Wittig, U., M. Merbach, R. G. Siefen und E. Brähler, 2004: Beschwerden und Inanspruchnahme des Gesundheitswesens von Spätaussiedlern bei Einreise nach Deutschland. *Gesundheitswesen* 66(2): S.85-92.

Woods, Carol M., 2009: Evaluation of MIMIC-Model Methods for DIF Testing With Comparison to Two-Group Analysis. *Multivariate Behavioral Research* 44(1): S.1-27.

Yildirim-Fahlbusch, Y., 2003: Kulturelle Missverständnisse. *Deutsches Ärzteblatt* 100: S.1179-81.

Zeeb, H. und O. Razum, 2006: Epidemiologische Studien in der Migrationsforschung. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 49(9): S.845-852.

Zeeb, H., B. T. Baune, W. Vollmer, D. Cremer und A. Krämer, 2004: Gesundheitliche Lage und Gesundheitsversorgung von erwachsenen Migranten - ein Survey bei der Schuleingangsuntersuchung. *Gesundheitswesen* 66(2): S.76-84.

7 Anhang

Tabelle 7: Kriterien zur Beurteilung der Modellgüte

Kriterium	Anforderung	Quelle
χ^2/df	$\leq 2,5$	Bühner (2006: 292f.)
p	$> 0,05$	Bühner (2006: 292f.)
Signifikanz der standardisierten Residualkovarianz (5%-Niveau)	$-1,96 < \varepsilon_{ik} < 1,96$	Bühner (2006: 292f.)
CFI	$\geq 0,90$ akzeptabler Fit $\geq 0,95$ guter Fit	Bühner (2006: 292f.)
RMSEA	$\leq 0,08$ akzeptabler Fit $\leq 0,05$ guter Fit	Bühner (2006: 292f.)
SRMR	$\leq 0,11$	Bühner (2006: 292f.)
$\Delta\chi^2$ -Differenztest	$\Delta\chi^2(\Delta df) \leq \chi^2_{krit}$	(Kline 2004: 146ff.)
ΔCFI	$\Delta CFI \leq 0,01$ mögliche Verschlechterung $0,01 < \Delta CFI \leq 0,02$ eindeutige Verschlechterung	Vandenberg und Lance (2000: 46)
Indikatorreliabilität	$\geq 0,4$	Fornell und Larcker (1981: 45ff.)
Faktorreliabilität	$\geq 0,6$	Fornell und Larcker (1981: 45ff.)
durchschnittlich erfasste Varianz	$\geq 0,5$	Fornell und Larcker (1981: 45ff.)

Tabelle 8: Lokale Kriterien zur Beurteilung des Basismodells

Skala	Cronbach´s Alpha	Faktorreliabilität	durchschnittlich erfasste Varianz
SF-12 gesamt	0,91	0,94	0,60
physische Gesundheit	0,84	0,91	0,65
mentale Gesundheit	0,90	0,87	0,55

Tabelle 9: Verteilungen der SF-12-Variablen nach Herkunftsregion

SF-12 Variablen		West-deutschland		Ost-deutschland		Türkei		(Ex-) Jugoslawien		EG-Anwerbeländer		erw. osteuropäische Länder	
		n	%	n	%	n	%	n	%	n	%	n	%
gegenwärtiger Gesundheitszustand (Item p99)	sehr gut	570	3,3	215	3,1	36	3,9	13	2,6	22	3,0	53	3,6
	gut	2002	11,5	836	11,9	126	13,5	83	16,5	93	12,8	206	14,1
	zufriedenstellend	5431	31,3	2328	33,1	235	25,2	136	27,1	205	28,1	437	29,8
	weniger gut	7258	41,8	2952	42,0	399	42,7	209	41,6	313	42,9	565	38,5
	schlecht	2092	12,1	698	9,9	138	14,8	61	12,2	96	13,2	205	14,0
	total	17353	100,0	7029	100,0	934	100,0	502	100,0	729	100,0	1466	100,0
Beeinträchtigung beim Treppen- steigen durch Gesundheitszustand (Item p100)	stark	1570	9,1	641	9,1	71	7,6	45	9,0	56	7,7	141	9,6
	ein wenig	4471	25,8	2144	30,5	215	23,1	151	30,1	180	24,7	364	24,8
	gar nicht	11300	65,2	4245	60,4	643	69,2	306	61,0	493	67,6	961	65,6
	total	17341	100,0	7030	100,0	929	100,0	502	100,0	729	100,0	1466	100,0
Beeinträchtigung bei anstrengende Tätigkeiten (Item p101)	stark	2189	12,6	847	12,1	86	9,3	55	11,0	78	10,8	177	12,1
	ein wenig	5353	30,9	2625	37,4	247	26,6	167	33,4	201	27,7	462	31,6
	gar nicht	9765	56,4	3547	50,5	594	64,1	278	55,6	446	61,5	825	56,4
	total	17307	100,0	7019	100,0	927	100,0	500	100,0	725	100,0	1464	100,0
niedergeschlagen und trübsinnig in den letzten 4 Wochen (Item p10202)	immer	242	1,4	103	1,5	21	2,3	13	2,6	9	1,3	18	1,2
	oft	2120	12,2	991	14,1	123	13,2	88	17,6	108	15,0	218	14,9
	manchmal	5714	33,0	2608	37,1	290	31,2	161	32,3	236	32,8	478	32,7
	fast nie	5430	31,4	2126	30,3	281	30,2	144	28,9	193	26,8	403	27,5
	nie	3803	22,0	1199	17,1	215	23,1	93	18,6	173	24,1	346	23,7
	total	17309	100,0	7027	100,0	930	100,0	499	100,0	719	100,0	1463	100,0

FortsetzungTabelle 9: Verteilung der SF-12-Variablen nach Herkunftsregion

SF-12 Variablen		West-deutschland		Ost-deutschland		Türkei		(Ex-) Jugoslawien		EG- Anwerbeländer		erw. osteuropäische Länder	
		n	%	n	%	n	%	n	%	n	%	n	%
ruhig und ausgeglichen in den letzten 4 Wochen (Item p10203)	immer	411	2,4	174	2,5	43	4,6	22	4,4	24	3,3	35	2,4
	oft	2100	12,2	1004	14,3	120	13,0	74	14,8	98	13,6	179	12,3
	manchmal	5792	33,5	2526	36,0	333	36,0	178	35,6	255	35,4	511	35,0
	fast nie	7652	44,3	2876	41,0	344	37,1	186	37,2	264	36,7	609	41,8
	nie	1328	7,7	432	6,2	86	9,3	40	8,0	79	11,0	124	8,5
	total	17283	100,0	7012	100,0	926	100,0	500	100,0	720	100,0	1458	100,0
jede Menge Energie verspürt in den letzten 4 Wochen (Item p10204)	immer	847	4,9	316	4,5	61	6,6	26	5,3	23	3,2	73	5,0
	oft	2915	16,9	1324	18,9	174	18,8	92	18,6	124	17,3	247	17,0
	manchmal	7282	42,2	3168	45,2	370	39,9	186	37,6	296	41,2	640	44,0
	fast nie	5397	31,3	1926	27,5	252	27,2	153	30,9	208	29,0	407	28,0
	nie	817	4,7	275	3,9	70	7,6	38	7,7	67	9,3	88	6,0
	total	17258	100,0	7009	100,0	927	100,0	495	100,0	718	100,0	1455	100,0
starke körperliche Schmerzen in den letzten 4 Wochen (Item p10205)	immer	378	2,20	132	1,90	19	2,0	16	3,20	16	2,2	20	1,4
	oft	1542	8,90	713	10,2	98	10,6	58	11,6	81	11,3	151	10,4
	manchmal	3293	19,1	1525	21,8	199	21,4	117	23,4	135	18,8	308	21,1
	fast nie	4804	27,8	2024	28,9	233	25,1	136	27,2	206	28,7	358	24,6
	nie	7255	42,0	2617	37,3	379	40,8	173	34,6	279	38,9	620	42,6
	total	17272	100,0	7011	100,0	928	100,0	500	100,0	717	100,0	1457	100,0
wegen gesundheitlicher Probleme körperlicher Art in Arbeit oder Beschäftigungen weniger geschafft (Item p10206)	immer	400	2,30	135	1,90	15	1,6	12	2,4	14	2,0	26	1,8
	oft	1595	9,30	655	9,40	65	7,0	50	10,1	52	7,3	119	8,2
	manchmal	3548	20,6	1742	24,9	195	21,1	98	19,8	133	18,5	341	23,5
	fast nie	4349	25,3	1913	27,4	216	23,4	126	25,5	181	25,2	325	22,4
	nie	7320	42,5	2545	36,4	431	46,7	208	42,1	337	47,0	643	44,2
	total	17212	100,0	6990	100,0	922	100,0	494	100,0	717	100,0	1454	100,0

FortsetzungTabelle 9: Verteilung der SF-12-Variablen nach Herkunftsregion

SF-12 Variablen		West- deutschland		Ost- deutschland		Türkei		(Ex-) Jugoslawien		EG- Anwerbeländer		erw. osteuropäische Länder	
		n	%	n	%	n	%	n	%	n	%	n	%
wegen gesundheitlicher Probleme	immer	496	2,9	190	2,7	16	1,7	10	2,0	12	1,7	37	2,6
körperlicher Art in Arbeit oder	oft	1336	7,8	546	7,8	56	6,1	51	10,3	48	6,7	98	6,8
Tätigkeiten eingeschränkt	manchmal	3258	19,0	1510	21,6	170	18,5	84	17,0	119	16,7	303	20,9
(Item p10207)	fast nie	4235	24,7	1897	27,2	226	24,6	124	25,2	179	25,1	309	21,3
	nie	7844	45,7	2835	40,6	451	49,1	224	45,4	355	49,8	701	48,4
	total	17169	100,0	6978	100,0	919	100,0	493	100,0	713	100,0	1448	100,0
wegen seelischer/emotionaler	immer	206	1,2	70	1,0	13	1,4	7	1,4	6	0,8	20	1,4
Probleme in Arbeit oder	oft	881	5,1	446	6,4	51	5,5	31	6,3	34	4,8	89	6,1
Beschäftigungen weniger geschafft	manchmal	2709	15,7	1368	19,5	159	17,2	81	16,4	95	13,3	256	17,6
(Item p10208)	fast nie	4116	23,9	1957	27,9	190	20,6	126	25,6	187	26,2	323	22,2
	nie	9310	54,1	3165	45,2	509	55,2	248	50,3	393	55,0	767	52,7
	total	17222	100,0	7006	100,0	922	100,0	493	100,0	715	100,0	1455	100,0
wegen seelischer/emotionaler	immer	154	0,9	55	0,8	8	0,9	5	1,0	6	0,8	15	1,0
Probleme weniger Sorgfalt	oft	566	3,3	263	3,8	34	3,7	22	4,5	23	3,2	58	4,0
bei Arbeiten oder Tätigkeiten	manchmal	2291	13,3	1099	15,7	131	14,3	73	14,8	93	13,0	231	15,9
(Item p10209)	fast nie	4366	25,4	2125	30,4	201	21,9	129	26,2	186	26,0	338	23,3
	nie	9801	57,1	3448	49,3	544	59,3	263	53,5	408	57,0	809	55,8
	total	17178	100,0	6990	100,0	918	100,0	492	100,0	716	100,0	1451	100,0
wegen gesundheitlicher oder	immer	188	1,1	69	1,0	12	1,3	6	1,2	4	0,6	14	1,0
seelischer Probleme in	oft	759	4,4	372	5,3	26	2,8	30	6,0	32	4,5	74	5,1
sozialen Kontakten eingeschränkt	manchmal	2182	12,6	1026	14,6	121	13,1	68	13,7	87	12,2	230	15,8
(Item p10210)	fast nie	3602	20,9	1767	25,2	200	21,6	108	21,8	160	22,4	290	19,9
	nie	10528	61,0	3783	53,9	565	61,1	284	57,3	432	60,4	848	58,2
	total	17259	100,0	7017	100,0	924	100,0	496	100,0	715	100,0	1456	100,0

Tabelle 10: Mittelwerte der SF-12-Variablen nach Herkunftsregion

	min	max	West- deutschland	Ost- deutschland	Türkei	(Ex) Jugoslawien	EG- Anwerbeländer	erw. osteurop. Länder
gegenwärtiger Gesundheitszustand*	1	5	3,5	3,4	3,5	3,4	3,5	3,5
Beeinträchtigung beim Treppensteigen durch Gesundheitszustand	1	3	2,6	2,5	2,6	2,5	2,6	2,6
Beeinträchtigung bei anstrengenden Tätigkeiten	1	3	2,4	2,4	2,5	2,4	2,5	2,4
in den letzten 4 Wochen niedergeschlagen und trübsinnig	1	5	3,6	3,5	3,6	3,4	3,6	3,6
in den letzten 4 Wochen ruhig und ausgeglichen*	1	5	3,4	3,3	3,3	3,3	3,4	3,4
in den letzten 4 Wochen jede Menge Energie verspürt*	1	5	3,1	3,1	3,1	3,2	3,2	3,1
in den letzten 4 Wochen starke körperliche Schmerzen	1	5	4,0	3,9	3,9	3,8	3,9	4,0
in den letzten 4 Wochen wg. gesundheitlicher Probleme körperlicher Art in Arbeit o. Tätigkeiten weniger geschafft	1	5	4,0	3,9	4,1	3,9	4,1	4,0
wegen gesundheitlicher Probleme körperlicher Art in Arbeit oder Tätigkeiten eingeschränkt	1	5	4,0	3,9	4,1	4,0	4,1	4,1
wegen seelischer/emotionaler Probleme in Arbeit oder Beschäftigungen weniger geschafft	1	5	4,2	4,1	4,2	4,2	4,3	4,2
wegen seelischer/emotionaler Probleme weniger Sorgfalt bei Arbeiten oder Tätigkeiten	1	5	4,3	4,2	4,3	4,3	4,4	4,3
wegen gesundheitlicher oder seelischer Probleme in sozialen Kontakten eingeschränkt	1	5	4,4	4,3	4,4	4,3	4,4	4,3
Summenskala (alle Items)	12	56	43,5	42,5	43,7	42,7	43,9	43,4
Standardabweichung der Summenskala			8,0	7,9	8,2	8,7	8,0	8,2

Anmerkung: Die Variablen mit * waren ursprünglich umgekehrt gepolt und wurden entsprechend rekodiert, sodass generell eine hoher Wert auf den Items eine positive Bewertung der Gesundheit bedeutet.

Tabelle 11: Einfluss der Kovariaten auf die physische und mentale Gesundheit nach Herkunftsregion

Gruppe	Dimension	männlich		Alter		Netto-Äquivalenzeinkommen ²³		Vollzeit/Teilzeit	
		unstand.	stand.	unstand.	stand.	unstand.	stand.	unstand.	stand.
Westdeutschland	physisch	0,147 (0,016)	0,072	-0,309 (0,008)	-0,306	0,057 (0,005)	0,088	0,342 (0,016)	0,166
	mental	0,154 (0,015)	0,086	-0,083 (0,007)	-0,094	0,047 (0,005)	0,083	0,176 (0,015)	0,098
Ostdeutschland	physisch	0,166 (0,023)	0,083	-0,369 (0,011)	-0,397	0,145 (0,020)	0,086	0,354 (0,023)	0,177
	mental	0,210 (0,023)	0,118	-0,139 (0,010)	-0,168	0,164 (0,019)	0,109	0,215 (0,023)	0,121
Türkei	physisch	<i>0,035</i> <i>(0,063)</i>	<i>0,019</i>	-0,373 (0,032)	-0,388	0,217 (0,073)	0,099	0,156 (0,064)	0,082
	mental	0,160 (0,066)	0,085	-0,237 (0,034)	-0,249	0,223 (0,077)	0,102	<i>0,042</i> <i>(0,067)</i>	<i>0,022</i>
(Ex-)Jugoslawien	physisch	0,381 (0,094)	0,183	-0,312 (0,043)	-0,326	<i>0,087</i> <i>(0,074)</i>	<i>0,053</i>	0,224 (0,094)	0,108
	mental	0,340 (0,089)	0,179	-0,147 (0,041)	-0,168	0,153 (0,071)	0,102	<i>0,124</i> <i>(0,089)</i>	<i>0,065</i>
EG-Anwerbeländer	physisch	0,235 (0,070)	0,123	-0,358 (0,034)	-0,394	0,188 (0,057)	0,123	0,213 (0,071)	0,111
	mental	0,191 (0,068)	0,113	-0,122 (0,033)	-0,150	<i>0,093</i> <i>(0,055)</i>	<i>0,069</i>	0,229 (0,068)	0,135
erw. osteurop. Länder	physisch	0,114 (0,052)	0,055	-0,435 (0,024)	-0,460	0,151 (0,036)	0,106	0,295 (0,052)	0,143
	mental	0,178 (0,052)	0,095	-0,199 (0,024)	-0,233	0,096 (0,036)	0,075	0,218 (0,052)	0,117

Modellfit: $\chi^2=23503,9$; $df=570$; $p=0,001$; $CFI=0,895$; $RMSEA=0,040$; $SRMR=0,074$

Anmerkungen: Standardfehler sind in Klammern angegeben, signifikante Effekte ($p<0,05$) sind fett gedruckt, nicht-signifikante Effekte kursiv.

²³ gemessen in 1000 €

Tabelle 12: Einfluss der Kovariaten auf die physische und mentale Gesundheit nach Herkunftsregion

Dimension	Westdeutschland				Ostdeutschland				Türkei			
	ohne DIF		mit DIF		ohne DIF		mit DIF		ohne DIF		mit DIF	
	physisch	mental	physisch	mental	physisch	mental	physisch	mental	physisch	mental	physisch	mental
unstandardisiert												
männlich	0,147 (0,016)	0,154 (0,015)	0,149 (0,016)	0,147 (0,015)	0,166 (0,023)	0,210 (0,023)	0,170 (0,024)	0,195 (0,023)	<i>0,035</i> (0,063)	0,160 (0,066)	<i>0,035</i> (0,065)	0,154 (0,066)
Alter	-0,309 (0,008)	-0,083 (0,007)	-0,280 (0,008)	-0,087 (0,007)	-0,369 (0,011)	-0,139 (0,010)	-0,337 (0,011)	-0,144 (0,011)	-0,373 (0,032)	-0,237 (0,034)	-0,341 (0,033)	-0,238 (0,034)
Netto-Äquivalenzeinkommen	0,057 (0,005)	0,047 (0,005)	0,058 (0,005)	0,046 (0,005)	0,145 (0,020)	0,164 (0,019)	0,148 (0,020)	0,162 (0,019)	0,217 (0,073)	0,223 (0,077)	0,222 (0,075)	0,219 (0,077)
Vollzeit/Teilzeit	0,342 (0,016)	0,176 (0,015)	0,341 (0,016)	0,175 (0,015)	0,354 (0,023)	0,215 (0,023)	0,354 (0,024)	0,213 (0,023)	0,156 (0,064)	<i>0,042</i> (0,067)	0,158 (0,065)	<i>0,041</i> (0,067)
standardisiert												
männlich	0,072	0,086	0,072	0,082	0,083	0,118	0,084	0,109	<i>0,019</i>	0,085	<i>0,018</i>	0,082
Alter	-0,306	-0,094	-0,276	-0,099	-0,397	-0,168	-0,361	-0,174	-0,388	-0,249	-0,352	-0,250
Netto-Äquivalenzeinkommen	0,088	0,083	0,088	0,081	0,086	0,109	0,087	0,107	0,099	0,102	0,100	0,100
Vollzeit/Teilzeit	0,166	0,098	0,165	0,097	0,177	0,121	0,176	0,120	0,082	<i>0,022</i>	0,082	<i>0,021</i>
Fit												
χ^2	ohne DIF						mit DIF					
	23503,9						17737,92					
p	0,001						0,001					
df	570						522					
χ^2/df	41,2						34,0					
RMSEA	0,040						0,036					
CFI	0,895						0,921					
SRMR	0,074						0,0659					

Anmerkungen: Standardfehler sind in Klammern angegeben, signifikante Effekte ($p < 0,05$) sind fett gedruckt, nicht-signifikante Effekte kursiv.

Fortsetzung Tabelle 12: Einfluss der Kovariaten auf die physische und mentale Gesundheit nach Herkunftsregion

Dimension	(Ex-)Jugoslawien				EG-Anwerbeländer				erw. osteurop. Länder			
	ohne DIF		mit DIF		ohne DIF		mit DIF		ohne DIF		mit DIF	
	physisch	mental	physisch	mental	physisch	mental	physisch	mental	physisch	mental	physisch	mental
unstandardisiert												
männlich	0,381 (0,094)	0,340 (0,089)	0,387 (0,095)	0,336 (0,089)	0,235 (0,070)	0,191 (0,068)	0,238 (0,072)	0,182 (0,068)	0,114 (0,052)	0,178 (0,052)	0,115 (0,053)	0,169 (0,052)
Alter	-0,312 (0,043)	-0,147 (0,041)	-0,280 (0,044)	-0,148 (0,041)	-0,358 (0,034)	-0,122 (0,033)	-0,326 (0,035)	-0,122 (0,033)	-0,435 (0,024)	-0,199 (0,024)	-0,406 (0,024)	-0,200 (0,024)
Netto-Äquivalenzeinkommen	<i>0,087</i> (0,074)	0,153 (0,071)	<i>0,091</i> (0,075)	0,154 (0,071)	0,188 (0,057)	<i>0,093</i> (0,055)	0,194 (0,058)	<i>0,089</i> (0,055)	0,151 (0,036)	0,096 (0,036)	0,155 (0,037)	0,093 (0,036)
Vollzeit/Teilzeit	0,224 (0,094)	<i>0,124</i> (0,089)	0,222 (0,095)	<i>0,123</i> (0,089)	0,213 (0,071)	0,229 (0,068)	0,213 (0,072)	0,229 (0,069)	0,295 (0,052)	0,218 (0,052)	0,293 (0,053)	0,217 (0,052)
standardisiert												
männlich	0,183	0,179	0,185	0,177	0,123	0,113	0,124	0,107	0,055	0,095	0,055	0,090
Alter	-0,326	-0,168	-0,291	-0,169	-0,394	-0,150	-0,357	-0,151	-0,460	-0,233	-0,428	-0,233
Netto-Äquivalenzeinkommen	<i>0,053</i>	0,102	<i>0,055</i>	0,102	0,123	<i>0,069</i>	0,066	<i>0,126</i>	0,106	0,075	0,108	0,072
Vollzeit/Teilzeit	0,108	<i>0,065</i>	0,106	<i>0,065</i>	0,111	0,135	0,134	0,111	0,143	0,117	0,141	0,116
Fit												
χ^2	ohne DIF						mit DIF					
	23503,9						17737,92					
p	0,001						0,001					
df	570						522					
χ^2/df	41,2						34,0					
RMSEA	0,040						0,036					
CFI	0,895						0,921					
SRMR	0,074						0,0659					

Anmerkungen: Standardfehler sind in Klammern angegeben, signifikante Effekte ($p < 0,05$) sind fett gedruckt, nicht-signifikante Effekte kursiv.

Tabelle 13: Schätzung der latenten Mittelwerte der physischen und mentalen Gesundheit

latente Mittelwerte	West-deutschland	Ost-deutschland	Türkei	(Ex-) Jugoslawien	EG-Anwerbeländer	erw. osteurop. Länder
ohne Kovariaten						
physisch	4,018 (0,008)	3,948 (0,013)	4,119 (0,034)	4,011 (0,049)	4,135 (0,038)	4,057 (0,028)
mental	4,241 (0,007)	4,099 (0,012)	4,216 (0,033)	4,164 (0,045)	4,293 (0,034)	4,186 (0,026)
mit 4 Kovariaten						
physisch	3,984 (0,009)	3,913 (0,013)	4,126 (0,036)	3,990 (0,053)	4,126 (0,040)	4,011 (0,030)
mental	4,230 (0,008)	4,087 (0,012)	4,217 (0,036)	4,166 (0,047)	4,274 (0,037)	4,163 (0,028)
mit DIF						
physisch	3,984 (0,009)	3,913 (0,013)	4,126 (0,036)	3,990 (0,053)	4,126 (0,040)	4,011 (0,030)
mental	4,230 (0,008)	4,087 (0,012)	4,217 (0,036)	4,166 (0,047)	4,274 (0,037)	4,163 (0,028)
mit Äquivalenzrestriktionen λ , τ , ϵ						
physisch	3,990 (0,009)	3,907 (0,013)	4,102 (0,035)	3,957 (0,052)	4,091 (0,039)	4,020 (0,029)
mental	4,229 (0,008)	4,094 (0,012)	4,219 (0,034)	4,138 (0,047)	4,239 (0,036)	4,168 (0,027)
+ Äquivalenz-Restriktion α						
physisch				3,978 (0,007)		
mental				4,190 (0,006)		

Anmerkungen: Standardfehler sind in Klammern angegeben, signifikante Effekte ($p < 0,05$) sind fett gedruckt, nicht-signifikante Effekte kursiv.

Tabelle 14: Schätzungen der gleichheitsrestringierten Faktorladungen, Varianzen und Achsenabschnitte und Mittelwerte

	Regressions- gewichte		Varianzen		Achsen- abschnitte		Mittelwerte
gegenwärtiger Gesundheitszustand (p99)	0,556	(0,005)	0,414	(0,004)	1,215	(0,019)	2,410
Treppensteigen (p100)	0,346	(0,003)	0,110	(0,002)	1,156	(0,014)	2,536
anstrengende Tätigkeiten (p101)	0,409	(0,004)	0,121	(0,002)	0,779	(0,015)	3,434
niedergeschlagen (p10202)	0,661	(0,006)	0,641	(0,006)	0,787	(0,026)	3,582
ausgeglichen (p10203)	0,411	(0,006)	0,455	(0,008)	1,671	(0,025)	3,410
jede Menge Energie (p10204)	0,384	(0,006)	0,302	(0,013)	1,494	(0,026)	3,117
starke Schmerzen (p10205)	0,776	(0,005)	0,537	(0,005)	0,839	(0,021)	3,936
weniger geschafft körperlich (p10206)	0,971	(0,004)	0,189	(0,003)	0,052	(0,016)	3,925
in Tätigkeiten eingeschränkt (p10207)	1,000		0,141	(0,003)			3,990
weniger geschafft emotional (p10208)	1,000		0,150	(0,003)			4,229
weniger sorgfältig (p10209)	0,890	(0,004)	0,160	(0,002)	0,578	(0,018)	4,340
soziale Kontakte leiden (p10210)	0,793	(0,005)	0,383	(0,004)	0,995	(0,022)	4,351

Anmerkung: dargestellt sind die unstandardisierten Regressionsgewichte, Standardfehler in Klammern, signifikante Effekte ($p < 0,05$) sind fett gedruckt, nicht-signifikante Effekte sind kursiv gedruckt