



Sprachkompetenzen von Jugendlichen mit Migrationshintergrund im deutschen Bildungssystem

Die Bedeutung des außerfamiliären
Sprachgebrauchs und methodische Hürden
der Messung von Sprachkompetenzen

Dissertation
zur Erlangung des Doktorgrades
Doctor rerum politicarum (Dr. rer. pol.)

vorgelegt von
Julian Seuring

eingereicht an der
Otto-Friedrich-Universität Bamberg
Fakultät Sozial- und Wirtschaftswissenschaften

Bamberg, 2025

Diese Arbeit hat der Fakultät Sozial- und Wirtschaftswissenschaften der Otto-Friedrich-Universität Bamberg als Dissertation vorgelegen.

Gutachter:innen: Prof. Dr. Cornelia Kristen
Prof. Dr. Claudia Diehl
Prof. Dr. Corinna Kleinert

Tag der mündlichen Prüfung: 04.06.2025

Dieses Werk ist als freie Onlineversion über das Forschungsinformationssystem (FIS; <https://fis.uni-bamberg.de>) der Universität Bamberg erreichbar.

Dieses Werk ist durch das deutsche Urheberrecht geschützt. Es steht Ihnen frei, dieses Werk auf jede Art und Weise zu nutzen, die durch die geltende Gesetzgebung zum deutschen Urheberrecht erlaubt ist. Für andere Verwendungszwecke müssen Sie die Erlaubnis der Rechteinhaberinnen und Rechteinhaber einholen.

URN: [urn:nbn:de:bvb:473-irb-108701x](https://nbn-resolving.org/urn:nbn:de:bvb:473-irb-108701x)

DOI: <https://doi.org/10.20378/irb-108701>

Danksagung

Eine Dissertation ist immer ein Produkt ihrer Zeit und der Menschen, die einen begleiten und prägen. In meinem Fall war es ein sehr langer Zeitraum und dementsprechend gab es auch viele Personen, die eine zentrale Rolle gespielt haben und spielen, manche über gewisse Etappen, andere über die komplette Strecke. Ohne deren direkten wie indirekten Beitrag würde die Arbeit in dieser Form heute nicht existieren. Jetzt ist die Zeit gekommen Danke zu sagen.

Zuallererst geht mein Dank an die drei Gutachterinnen, die drei Cs (zwei davon EC): An Cornelia Kristen für das Vertrauen, den Zuspruch und die Begeisterung, die mir über all die Jahre entgegengebracht wurde, sowie für die Freiheit, die ich genießen durfte, um meinen eigenen Weg zu finden. An Claudia Diehl für die direkte Zusage zur Zweitbegutachtung und für das Grundvertrauen unser kleines Forschungsprojekt sicher durch all die erwartbaren und unerwarteten Katastrophen zu manövrieren. An Corinna Kleinert als Drittgutachterin des Komitees für die Bereitschaft, diese Rolle „gerne“ einzunehmen, sowie für die terminliche Flexibilität. Allen dreien danke ich für die intensive und kritische Lektüre meiner Dissertation, die berechtigten und hilfreichen Kommentare, die ich sowohl in den Gutachten als auch in der Disputation erhalten habe, sowie die wertschätzende Kommunikation während des gesamten Verfahrens.

Weiterhin möchte ich mich bei den Koautorinnen der vier Artikel, die Teil der kumulativen Dissertation sind, für die großartige Zusammenarbeit bedanken. Der konstruktive Austausch mit Aileen Edele, Bernadette Strobel, Camilla Rjosk und Petra Stanat hat mich nachhaltig geprägt. Ich hoffe, dass wir in zukünftigen Projekten mal wieder daran anknüpfen können.

Mein besonderer Dank gilt auch Gisela Will und Melanie Olczyk, die von Anfang an dabei waren und vermutlich maßgeblich dazu beigetragen haben, dass ich überhaupt eine wissenschaftliche Laufbahn einschlagen durfte. Aus dem Kreis der einstigen – wenn auch schon länger nicht mehr – „anonymen“ Dissertioniker ist inzwischen ein verlässliches Netzwerk für Ideenaustausch, Unterstützung und gemeinsame Projekte entstanden. Ich bin sehr froh, dass sie meine wissenschaftliche Laufbahn bis heute in unterschiedlichen Konstellationen begleiten, und hoffe, dass es noch lange so bleibt.

Im Laufe der Jahre durfte ich im Rahmen meiner Tätigkeiten, Forschungsprojekte und Publikationen – der kleinen und großen Nebenbaustellen also – viele wunderbare Menschen kennenlernen. Sowohl der fachliche Austausch als auch die nicht weniger bedeutsamen „einfachen“ Gespräche zwischen Tür und Angel oder beim Mittagessen haben einen bleibenden Eindruck hinterlassen. Mein Dank geht insbesondere an Anna, Andreas, Christoph, Daniel, Diana, Florian, Georg, Hartmut (auch vHzH), Jasper, Jörg, Julia, Loreen, Mathias, Miriam, Petra, Regina, Regine und Susann.

Auch bei den vielen Hilfskräften, die mich in unterschiedlichen Phasen mit ihrem unermüdlichen Einsatz unterstützt haben, möchte ich mich ausdrücklich bedanken. Mit Blick auf die Artikel und das Rahmenpapier zur Dissertation ganz besonders bei Lara, Lucas und Johanna für die mehrfache Prüfung und Überarbeitung der Referenzen im Text und der Literaturverzeichnisse.

Ohne Rückendeckung, Ausgleich und Ablenkung abseits des wissenschaftlichen Zir-

kus geht es nicht. Deshalb danke ich Dani, Jo, Rick, Sigg, Steffi, sowie Mathieu und Minnie – wenn auch leider zu selten – für die wirklich wichtigen Dinge. Ihr wart immer da und werdet es auch bleiben.

Ein ganz großes Dankeschön gilt auch meiner Partnerin Miriam, die mir in jeder Lebenslage bedingungslos zur Seite steht. Ohne ihren motivierenden Zuspruch, ihre unerschütterliche Überzeugung, dass es irgendwann schon fertig wird, und die zugestandenen Freiräume, mich auch mal am Wochenende oder im Urlaub daranzusetzen – gleichzeitig aber auch die vermutlich noch viel zielführendere und gut gemeinte Überredungsleistung, gerade das nicht zu tun und auch mal Abstand zu nehmen, wäre es wohl nie etwas geworden. Ich hoffe, ich kann das in gleichem Umfang zurückgeben.

Zu guter Letzt möchte ich meiner Familie danken, insbesondere meinen Eltern, Frank und Monika. Sie haben alles gegeben, um mir den Weg ins Studium zu ermöglichen und mich dabei unterstützt, wo sie nur konnten. Sie sind vermutlich stolzer als sonst jemand auf den erfolgreichen Abschluss meiner Promotion – und das sei ihnen gegönnt. Aus tiefer Wertschätzung für ihre vielen Bemühungen widme ich auch ihnen diese Arbeit. Danke.

Inhaltsverzeichnis

1	Einleitung	13
2	Theoretischer Hintergrund, Forschungsüberblick und Beitrag der vorliegenden Arbeit	19
2.1	Außerfamiliärer Sprachgebrauch und Zweitspracherwerb	19
2.1.1	Ein allgemeines Modell des Spracherwerbs	19
2.1.2	Sprachgebrauch und Zweitspracherwerb	23
2.1.3	Beitrag der vorliegenden Arbeit	26
2.2	Validität subjektiver Einschätzungen von Sprachkompetenzen	29
2.2.1	Ein allgemeines Modell des Antwortverhaltens in Umfragen	30
2.2.2	Antwortverzerrungen bei Selbsteinschätzungen von Sprachkompetenzen	39
2.2.3	Beitrag der vorliegenden Arbeit	56
3	Datengrundlage und methodisches Vorgehen	59
4	<i>Artikel 1: Spracherhalt oder Sprachverlagerung? Erstsprachgebrauch und Zweitsprachkompetenzen bei Jugendlichen mit Migrationshintergrund</i>	65
4.1	Einleitung	67
4.2	Die Bedeutung des Sprachgebrauchs für die Zweitsprachkompetenz	69
4.3	Kontexte des Sprachgebrauchs	70
4.4	Daten und Methode	72
4.5	Ergebnisse	77
4.5.1	Sprachgebrauch verschiedener Herkunftsgruppen	77
4.5.2	Sprachgebrauch im Generationenverlauf	79
4.5.3	Individuelle Muster des Sprachgebrauchs: Sprachgebrauchsprofile	82
4.5.4	Zusammenhänge des Sprachgebrauchs mit den Sprachkompetenzen	83
4.6	Fazit und Diskussion	86
4.7	Literatur	90
4.8	Anhang	95
5	<i>Artikel 2: Ethnic classroom composition and minority language use among classmates: Do peers matter for students' language achievement?</i>	103
5.1	Introduction	105

5.2	The German context	106
5.3	Ethnic classroom composition effects	107
5.3.1	Ethno-lingual classroom composition and educational achievement	108
5.3.2	Disentangling the mechanisms associated with ethno-lingual classroom composition	110
5.4	Data, variables, and methods	112
5.4.1	Data	112
5.4.2	Variables	113
5.4.3	Methods	115
5.5	Results	117
5.5.1	Ethnic minority and ethno-lingual classroom composition	117
5.5.2	Ethno-lingual classroom composition and German reading comprehension	117
5.5.3	Ethno-lingual classroom composition and minority language proficiency	120
5.6	Limitations and robustness of findings	122
5.7	Summary and conclusion	123
5.8	References	124
5.9	Appendix	130
6	Artikel 3: Why bother with testing? The validity of immigrants' self-assessed language proficiency	135
6.1	Introduction	137
6.2	Evidence for the validity of ability self-assessments	138
6.2.1	Correlations between self-assessed and tested language proficiency	139
6.2.2	Multivariate analyses using self-assessed and tested language proficiency	141
6.2.3	Limitations of previous research	141
6.3	Analytical approach	142
6.3.1	Determinants of L1 and L2 skills	142
6.3.2	L1 skill as a determinant of L2 proficiency	143
6.3.3	L2 skill as a determinant of educational achievement	143
6.3.4	Two types of self-assessment	144
6.4	Method	144
6.4.1	Data	144
6.4.2	Instruments	145
6.4.3	Data analyses	151
6.5	Results	151
6.5.1	Correlations of self-assessed language skills with language test scores	151
6.5.2	Multivariate analyses of immigrant students' L1 and L2 skills	152
6.6	Discussion and conclusion	157
6.7	References	163

6.8	Appendix	173
7	Artikel 4: When can we (not) rely on self-reported language proficiency? An investigation of the processes underlying response bias and their implications for research on ethnic inequalities	181
7.1	Introduction	183
7.2	Background	184
7.2.1	Tendencies toward overestimation bias in self-reported language proficiency	184
7.2.2	Processes underlying response biases	185
7.2.3	Cognitive abilities	185
7.2.4	Self-concept	186
7.2.5	Frame of reference	188
7.3	Aims and expectations	189
7.4	Data and methods	189
7.4.1	Sample	189
7.4.2	Determining bias in self-reports	190
7.4.3	Explanatory and background variables	193
7.4.4	Analytical approach	194
7.5	Results	195
7.5.1	Accounting for overestimation bias in self-reported language proficiency	195
7.5.2	Sensitivity analyses	199
7.6	Discussion and conclusion	202
7.7	References	205
7.8	Appendix	212
8	Gesamtdiskussion und Fazit	219
	Literaturverzeichnis	231
	Anhang	247

Abbildungsverzeichnis

2.1	Schematische Darstellung des Antwortprozesses in Abhängigkeit des Modus der Informationsverarbeitung	33
4.1	Individuelle Profile des Sprachgebrauchs	74
4.2	Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland	78
4.3	Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland im Generationenverlauf	80
4.4	Kombinierte Profile des Sprachgebrauchs mit Eltern, Geschwistern und Peers oder Medien nach Herkunftsland	83
A4.1	Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland (nur multilinguale Schüler:innen)	98
A4.2	Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland im Generationenverlauf (nur multilinguale Schüler:innen)	99
A4.3	Kombinierte Profile des Sprachgebrauchs mit Eltern, Geschwistern und Peers oder Medien nach Herkunftsland (nur multilinguale Schüler:innen)	100
A5.1	Distribution of language use with classmates	134
A5.2	Correlation between proportion of ethnic minority students and proportion of minority language students on classroom level	134
A6.1	Ability self-assessment L1	179
A6.2	Ability self-assessment L2 (German)	179
A6.3	Performance self-assessment L1	180
A6.4	Performance self-assessment L2 (German)	180
7.1	Correlation (A) and distribution (B) of standardized self-reports and test scores	192
7.2	Conditional correlations between overestimation bias in self-reports and students' linguistic self-concept (A), self-esteem (B), and frame of reference (C) according to their general cognitive abilities	198
7.3	Summary of sensitivity analyses of the results regarding background characteristics (A), explanatory variables (B), and conditional correlations (C) using alternative operationalizations of overestimation bias	201

Tabellenverzeichnis

4.1	Randverteilungen der in den multivariaten Analysen verwendeten Variablen	76
4.2	Lineare Regression der Lese- und Wortschatzkompetenz	84
A4.1	Lineare Regression der Lese- und Wortschatzkompetenz (nur multilinguale Schüler:innen)	101
A4.2	Verteilung Generationenstatus nach Herkunftsgruppe	102
5.1	Distribution of model variables and sample characteristics	116
5.2	Predicting German reading comprehension by ethnic minority and ethnolinguistic classroom composition	118
5.3	Predicting German reading comprehension by ethno-lingual classroom composition and minority language use	119
5.4	Predicting minority language proficiency by the proportion of Russian-/Turkish-speaking classmates	121
A5.1	Distribution and description of minority language population	131
A5.2	Robustness analysis of German reading comprehension using 4-point scales of minority language use	132
A5.3	Robustness analysis of German reading comprehension with additional controls	133
6.1	Sample characteristics and distribution of model variables	146
6.2	Correlations between the self-assessment indicators and test scores for L1 and L2	152
6.3	Predicting L1 skills	154
6.4	Predicting L2 skills	156
6.5	Predicting L2 reading test scores based on L1 skills	158
6.6	Predicting mathematics achievement based on L2 skills	159
A6.1	Studies comparing self-assessments and test scores/teacher estimates of linguistic skills	174
7.1	Distribution of variables used in the analysis	194
7.2	Predicting overestimation bias in self-reported German proficiency	196
A7.1	Sensitivity analysis (S1) using the single self-report item on reading ability	213
A7.2	Sensitivity analysis (S2) using the self-reported test performance	214
A7.3	Sensitivity analysis (S3) using a stricter criterion to define the analysis sample	215
A7.4	Sensitivity analysis (S4) using the residualized difference score	216

Tabellenverzeichnis

A7.5	Average tested and self-reported German language proficiency by students' background characteristics	217
A.1	Lineare Regression getesteter und selbsteingeschätzter L2-Kompetenzen auf das globale Selbstwertgefühl	248
A.2	Lineare Regression getesteter und selbsteingeschätzter L2-Kompetenzen auf die Big Five Persönlichkeitsmerkmale	249

1 Einleitung

Kinder und Jugendliche mit Migrationshintergrund¹ schneiden im deutschen Bildungssystem gegenüber ihrer einheimischen Vergleichsgruppe insgesamt weniger erfolgreich ab. So erzielen Schüler:innen mit Migrationshintergrund im Schnitt niedrigere schulische Leistungen (z.B. Henschel et al. 2019, 2022, 2023) und besuchen seltener die anspruchsvolleren Bildungszweige der Sekundarstufe I bzw. II (z.B. Olczyk et al. 2016) als Gleichaltrige ohne Migrationshintergrund. Tendenzen ethnischer Bildungsungleichheiten lassen sich – wenn auch in unterschiedlichem Maße – für alle Etappen der Bildungslaufbahn, vom Kindergarten bis zur beruflichen Ausbildung und dem Hochschulbereich, beobachten (vgl. Diehl et al. 2016a; Dollmann 2017). Vor dem Hintergrund dieser bestehenden Ungleichheitsmuster rücken die Erklärungsfaktoren ethnischer Leistungsdisparitäten zunehmend ins Blickfeld des wissenschaftlichen und gesellschaftspolitischen Interesses. Dies gewinnt auch durch den stark gestiegenen Anteil geflüchteter Kinder und Jugendlicher im deutschen Bildungssystem in den letzten Jahren zusätzlich an Relevanz.

Mit Blick auf die Erklärung ethnischer Bildungsungleichheiten verweist die einschlägige Literatur auf die vergleichsweise geringe Ressourcenausstattung von Zuwandererfamilien, welche häufig mit einer nachteiligeren sozialstrukturellen Positionierung im Aufnahmeland in Verbindung gebracht wird (vgl. Heath und Brinbaum 2007). In Anlehnung an Boudon (1974) werden demnach Einflüsse bedeutsam, die mit den *sozialen* Bedingungen des Leistungserwerbs verknüpft sind. Der Zusammenhang zwischen ethnischen Leistungsdisparitäten und Unterschieden in der sozialen Herkunft konnte in zahlreichen Studien belegt werden (z.B. Kristen 2008b; Müller und Stanat 2006; Segeritz et al. 2010). Allerdings lassen sich ethnische Bildungsungleichheiten, zumindest mit Blick auf spezifische Herkunftsgruppen und Zuwanderungsgenerationen, nicht vollständig durch die soziale Herkunft aufklären.

Im Zusammenhang mit den verbleibenden ethnischen Bildungsunterschieden wird verstärkt die Bedeutung weiterer, *migrationspezifischer* Einflussfaktoren diskutiert (vgl. Diehl et al. 2016b; Heath und Brinbaum 2007; Heath et al. 2008; Kristen und Dollmann 2010; Van de Werfhorst und Van Tubergen 2007). Die Beherrschung der Sprache des Aufnahmelandes (L2)² gilt dabei als zentrale Schlüsselgröße für den Bil-

¹Der Begriff Migrationshintergrund, wie er in dieser Arbeit verwendet wird, orientiert sich am Geburtsland und umfasst sowohl Personen mit eigener Wanderungserfahrung (1. Generation) als auch Kinder und Jugendliche nachfolgender Zuwanderungsgenerationen, die selbst im Aufnahmeland geboren sind und deren Eltern (2. Generation) oder Großeltern (3. Generation) aus dem Ausland zugezogen sind.

²Im Rahmen der Arbeit wird nachfolgend die Herkunftssprache als L1 oder Erstsprache, und die Verkehrssprache im Aufnahmeland als L2 oder Zweitsprache bezeichnet, unabhängig davon, in

derungserfolg von Kindern und Jugendlichen mit Migrationshintergrund (Esser 2006). Die besondere Bedeutung der L2 für den schulischen Erfolg gründet sich darin, dass diese in der Regel auch die Unterrichtssprache in der Schule ist, und somit eine Grundvoraussetzung für die Vermittlung und Aufnahme von Lehrinhalten darstellt (Paetsch et al. 2015). Gleichzeitig ist Deutsch als Unterrichtsfach Teil des Lehrplans, weswegen sich L2-Kompetenzen, beispielsweise in Form des Leseverstehens, direkt in den schulischen Leistungen widerspiegeln. Der positive Einfluss von Kompetenzen in der Verkehrssprache auf den schulischen Leistungserwerb ist unumstritten und konnte empirisch wiederholt belegt werden (z.B. Baumert und Schümer 2001; Dollmann und Kristen 2010; Paetsch et al. 2015; Strobel 2016).

Kinder mit Migrationshintergrund wachsen häufig mit einer ausländischen Erstsprache (L1) auf, was mit einem verzögerten Erwerb der deutschen Sprache einhergehen kann. Empirische Befunde deuten darauf hin, dass bereits im Vorschulalter signifikante Unterschiede in den L2-Kompetenzen zwischen Kindern mit und ohne Migrationshintergrund bestehen (z.B. Dubowy et al. 2008; Becker et al. 2013; Biedinger 2007, 2011). Vergleichbare Muster zeigen sich ebenfalls für die nachgelagerten Bildungsabschnitte der Primar- und Sekundarstufe (für den Grundschulbereich, vgl. Dollmann 2010; Henschel et al. 2022; Kristen 2008b; Schwippert et al. 2007, 2012; für die Sekundarstufe, vgl. Henschel et al. 2023; Müller und Stanat 2006; Stanat et al. 2010a; Walter 2009b). Die Ausgangsunterschiede in den L2-Kompetenzen scheinen demnach über die gesamte Schullaufbahn, teilweise selbst bis in den tertiären Bildungsbereich, nicht vollständig abgebaut werden zu können (vgl. Olczyk et al. 2016).

Bei der Interpretation dieser verallgemeinerten Tendenzen gilt es allerdings zu berücksichtigen, dass ethnische Disparitäten in den L2-Kompetenzen, sowie damit verbundene Unterschiede in den schulischen Leistungen insgesamt, zwischen verschiedenen Herkunftsgruppen und Zuwanderungsgenerationen sehr heterogen ausfallen. Besonders ausgeprägte Nachteile in den L2-Kompetenzen können beispielsweise für Zugewanderte aus der Türkei, dem ehemaligen Jugoslawien und arabischen Ländern festgestellt werden, wohingegen Jugendliche aus Polen und der ehemaligen Sowjetunion vergleichsweise besser abschneiden (vgl. Haag et al. 2016, S. 455f.; Segeritz et al. 2010). Im Generationenverlauf zeichnet sich weitestgehend ein Trend zur Abnahme der sprachlich bedingten Leistungsdisparitäten ab, wobei sich auch hier Unterschiede zwischen Herkunftsgruppen bemerkbar machen (vgl. Olczyk et al. 2016).

Vor dem Hintergrund der sprachlichen Kompetenzunterschiede und den damit einhergehenden Bildungsungleichheiten zwischen Jugendlichen mit und ohne Migrationshintergrund stellt sich weiterhin die vorgelagerte Frage nach den Bedingungen, welche den Erwerb der deutschen Sprache fördern bzw. behindern. Die vorliegende Dissertation setzt an dieser übergeordneten Fragestellung an und beleuchtet sowohl inhaltliche als auch methodische Aspekte, die zur Schließung bestehender Forschungslücken beitragen sollen. Dabei wird ein theoriegeleitetes Forschungsvorhaben verfolgt, in welchem soziologische Ansätze systematisch durch Argumente aus angrenzenden Disziplinen, wie der (Sozial-)Psychologie und den Erziehungswissenschaften, ergänzt und empirisch geprüft

welcher Reihenfolge die beiden Sprachen erlernt wurden (vgl. Kristen et al. 2011).

werden. Die Dissertation setzt sich kumulativ aus *vier* Aufsätzen zusammen, welche in Kooperation mit Kolleg:innen des Lehrstuhls für Soziologie, insb. Sozialstrukturanalyse (Otto-Friedrich-Universität Bamberg) und dem Institut für Qualitätsentwicklung im Bildungswesen Berlin (Humboldt-Universität zu Berlin) entstanden sind.

Der inhaltlich motivierte Beitrag der Dissertation geht der Frage nach, inwiefern die Verwendung der Herkunftssprache (L1) außerhalb des familiären Kontexts den Kompetenzerwerb in der Sprache des Aufnahmelandes von Jugendlichen mit Migrationshintergrund bedingt. Es wird angenommen, dass durch den häufigen Gebrauch der Herkunftssprache weniger Gelegenheiten bestehen, die Verkehrssprache zu sprechen und dabei entsprechende Kenntnisse zu erwerben (Chiswick und Miller 2001; Esser 2006). Zahlreiche empirische Befunde stützen diese These mit Blick auf die Familiensprache und deuten darauf hin, dass die Verwendung der Herkunftssprache innerhalb der Familie mit durchschnittlich geringeren L2-Kompetenzen in Zusammenhang steht (z.B. Baumert und Schümer 2001; Edele und Stanat 2016; Kristen 2008b; Stanat et al. 2010a; Strobel 2016; Walter 2009b).

Im Gegensatz dazu liegen kaum methodisch belastbare Befunde zum Sprachgebrauch in anderen Kontexten, zum Beispiel mit Freund:innen oder Mitschüler:innen, vor. Die Betrachtung weiterer Kontexte erscheint aber insofern bedeutsam, als dass sich, gerade im Zusammenhang mit ethnischen Segregationstendenzen im Wohn- und Schulumfeld (vgl. Baur und Häussermann 2009; Friedrichs 2008; Kristen 2008a), auch außerhalb der Familie Gelegenheitsstrukturen und Anreize zum L1-Gebrauch vermuten lassen. So wird beispielsweise in der Literatur zu Einflüssen der ethnischen Klassenzusammensetzung auf den schulischen Leistungserwerb wiederholt die herkunftssprachliche Kommunikation zwischen Klassenkamerad:innen als vermittelnder Mechanismus angeführt (z.B. Driessen 2002; Eksner und Stanat 2011; Esser 2006; Peetsma et al. 2006; Van Ewijk und Slegers 2010), obgleich eine empirische Prüfung dieser Annahme noch aussteht.

Der erste Artikel der vorliegenden Arbeit knüpft an diese Überlegungen an und untersucht, wie verbreitet die Verwendung der Herkunftssprache innerhalb und außerhalb des familiären Kontexts ist (siehe Kapitel 4). Der Beitrag bietet zunächst einen differenzierten Überblick über die Muster des Sprachgebrauchs von Jugendlichen mit Migrationshintergrund am Ende der Sekundarstufe I. Dabei werden neben der in der Familie gesprochenen Sprache auch weitere Kontexte, wie beispielsweise der Freundeskreis, miteinbezogen sowie Unterschiede zwischen Herkunftsgruppen und Zuwanderungsgenerationen beschrieben. Darauf aufbauend werden die Zusammenhänge des außerfamiliären Sprachgebrauchs mit den Kompetenzen in der Verkehrssprache analysiert.

Da Opportunitäten zum Gebrauch der Herkunftssprache in Abhängigkeit der ethnischen Klassenzusammensetzung variieren können, beleuchtet der zweite Artikel die Bedeutung der ethnischen Klassenkomposition für den schulischen Leistungserwerb (siehe Kapitel 5). Im Vordergrund steht dabei eine empirische Prüfung der Annahme, dass negative Einflüsse der ethnischen Komposition auf den Kompetenzerwerb in der Verkehrssprache durch den L1-Gebrauch unter Klassenkamerad:innen vermittelt werden. Die Ergebnisse beider Artikel werden vor dem Hintergrund bestehender ethnischer

Bildungsungleichheiten interpretiert.

Ein Großteil der aus der bisherigen Forschung resultierenden Erkenntnisse zu den Bedingungen – wie auch den Folgen – des Spracherwerbs und den oftmals beobachteten ethnischen Unterschieden beruht auf subjektiv eingeschätzten Sprachkompetenzen der Befragten. Da die Verlässlichkeit dieser Instrumente jedoch fraglich bleibt, widmet sich der methodische Beitrag der Dissertation der Validität subjektiver Einschätzungen als Indikator für die tatsächlichen Sprachkompetenzen von Jugendlichen mit Migrationshintergrund. Obwohl sich, durch die fortschreitende Etablierung großangelegter Schulleistungsstudien und Schüler:innenbefragungen, wie PISA, PIRLS, NEPS und CILS-4EU, und der damit verbundenen Verfügbarkeit von sprachbezogenen Testleistungen, eine Kehrtwende bezüglich der Verwendung subjektiver Sprachmessungen abzeichnet, ist deren Einsatz gerade in allgemeiner ausgerichteten Umfragen sowie in kleineren Studien nach wie vor sehr verbreitet. Subjektive Einschätzungen bieten die Möglichkeit, leistungsbezogene Informationen zu erheben, ohne umfangreiche Testverfahren einsetzen zu müssen, was für viele Studien aufgrund zeitlicher und finanzieller Restriktionen gar nicht in Frage käme. Dementsprechend scheint es angebracht die Verlässlichkeit von subjektiv eingeschätzten Sprachkompetenzen und die darauf aufbauenden Ergebnisse bisheriger Forschungsbeiträge kritisch zu hinterfragen.

Die meisten Autor:innen räumen Einschränkungen ihrer Forschungsergebnisse aufgrund eventuell fehlerbehafteter Selbsteinschätzungen ein, entkräften diese Kritik aber häufig mit dem Verweis auf ausgewählte Befunde, die eine hohe Korrelation zwischen selbsteingeschätzten und getesteten Sprachkompetenzen nahelegen. Bei genauerer Betrachtung des vergleichsweise umfassenden Forschungsstands zum Zusammenhang zwischen subjektiven und objektiven Sprachmessungen erweist sich die Befundlage aber als deutlich heterogener, als dies entsprechende Verweise vermuten lassen (vgl. Li und Zhang 2021; Ross 1998). Vielmehr deuten Befunde aus Untersuchungen zu multivariaten Zusammenhängen darauf hin, dass die Verwendung von subjektiven Einschätzungen als Indikator für die tatsächliche Sprachkompetenz von Befragten zu systematischen Verzerrungen der Ergebnisse führen kann (Charette und Meng 1994; Finnie und Meng 2005). Welche Mechanismen einem verzerrten Antwortverhalten bei der subjektiven Einschätzung von Sprachkompetenzen zugrunde liegen, wurde bislang hingegen kaum untersucht. Dies scheint allerdings bedeutsam, da sich nur anhand der Kenntnis über relevante Einflussfaktoren potentielle Fehlerquellen identifizieren und gegebenenfalls beseitigen lassen, oder sich zumindest das Ausmaß und die Richtung der Verzerrungen abschätzen und ausgehend davon die Ergebnisse besser einordnen lassen.

Im Rahmen des dritten Artikels erfolgt zunächst eine systematische Auswertung bisheriger Forschungsergebnisse zum Vergleich von subjektiven und objektiven Sprachmessungen (siehe Kapitel 6). Darüber hinaus werden subjektiv eingeschätzte Sprachkompetenzen von Jugendlichen mit Migrationshintergrund einer umfassenden Validitätsprüfung unterzogen. Hierzu werden subjektive Einschätzungen der L1- und L2-Kompetenzen mit testbasierten Leistungen abgeglichen und anhand exemplarischer Analysemodelle auf differentielle Zusammenhänge hin untersucht.

Der vierte und letzte Artikel widmet sich der Frage nach den Faktoren, welche systematisch verzerrten Selbsteinschätzungen zugrunde liegen (siehe Kapitel 7). Unter

Rückbezug auf die allgemeinen Prozesse des Befragtenverhaltens in Umfragen wird dabei ein theoretischer Rahmen herangezogen, welcher es erlaubt, bisher eher für sich stehende Argumente aus verschiedenen Forschungszweigen zu integrieren und klare Hypothesen zur Wirkung der relevanten Faktoren sowie deren Zusammenspiel abzuleiten (siehe Kapitel 2.2). Anschließend werden diese theoretischen Annahmen am Beispiel der selbsteingeschätzten L2-Kompetenzen von Jugendlichen mit Migrationshintergrund empirisch geprüft. Ausgehend von den Resultaten beider Artikel werden Implikationen der Verwendung subjektiver Sprachmessungen mit Blick auf bisherige sowie zukünftige Forschungsarbeiten diskutiert.

Im nachfolgenden Kapitel 2 wird zunächst der theoretische Hintergrund separat für die zwei zentralen Themenkomplexe der Arbeit – *Außerfamiliärer Sprachgebrauch und Zweitspracherwerb* (Kapitel 2.1) und *Validität subjektiver Einschätzungen von Sprachkompetenzen* (Kapitel 2.2) – geschildert sowie die bisherige Befundlage skizziert. Um Redundanzen mit den später folgenden Aufsätzen möglichst zu vermeiden, zielt die Diskussion des Forschungsstands darauf ab, anhand ausgewählter Befunde entscheidende Forschungslücken aufzuzeigen und, daran anknüpfend, jeweils den Beitrag der vorliegenden Arbeit hervorzuheben. Nach einer Beschreibung der Datengrundlage sowie des methodischen Vorgehens (Kapitel 3), widmen sich die folgenden Kapitel den vier Forschungsartikeln, aus welchen sich die kumulative Dissertation zusammensetzt (Kapitel 4–7). Die Arbeit schließt mit einer Diskussion und Einordnung der zentralen Befunde (Kapitel 8).

2 Theoretischer Hintergrund, Forschungsüberblick und Beitrag der vorliegenden Arbeit

2.1 Außerfamiliärer Sprachgebrauch und Zweitspracherwerb

In diesem Kapitel wird die Bedeutung des Gebrauchs der Herkunftssprache für den Erwerb von Kompetenzen in der Verkehrssprache diskutiert. Die häufige Verwendung der Herkunftssprache wird als eine zentrale Ursache für die vergleichsweise geringen L2-Kompetenzen von Jugendlichen mit Migrationshintergrund gegenüber der einheimischen Referenzgruppe angesehen (vgl. Esser 2006; Kristen et al. 2011; Scheele et al. 2010; Stanat 2008). Um die konkrete Wirkungsweise des Sprachgebrauchs auf den Zweitspracherwerb aufzudecken, ist es zielführend, zunächst die grundlegenden Bedingungen des Spracherwerbs zu bestimmen. Ausgehend von einem allgemeinen handlungstheoretischen Rahmen lässt sich anschließend der Sprachgebrauch systematisch mit den übergreifenden Bestimmungsfaktoren des Zweitspracherwerbs verknüpfen. Auf diese Weise ist es möglich, die zugrundeliegenden Wirkzusammenhänge eindeutig zu spezifizieren und deren Relevanz im Lichte eines übergeordneten Kontexts des Zweitspracherwerbs zu interpretieren.

Im nachfolgenden Abschnitt (2.1.1) wird ein allgemeiner handlungstheoretischer Rahmen zur Modellierung des Spracherwerbs präsentiert. Darauf aufbauend folgt in Abschnitt 2.1.2 eine Beschreibung der Wirkungsweise des Sprachgebrauchs auf den Zweitspracherwerb von Jugendlichen mit Migrationshintergrund. In diesem Zusammenhang wird weiterhin die Bedeutung des Gebrauchs der Herkunftssprache innerhalb und außerhalb der Familie, sowie insbesondere im schulischen Umfeld, mit Blick auf die bisherige Befundlage diskutiert und bestehende Forschungslücken aufgezeigt. Im Anschluss daran betont Abschnitt 2.1.3 den Beitrag der vorliegenden Arbeit für den aktuellen Forschungsstand.

2.1.1 Ein allgemeines Modell des Spracherwerbs

Der Erwerb von Sprachkompetenzen kann als Investitionsprozess in Humankapital aufgefasst werden (Chiswick und Miller 2001; Esser 2006). Das Niveau der Sprachkompetenz zu einem gegebenen Zeitpunkt entspricht dabei dem kumulativen Ergebnis sämtlicher vorgelagerter Investitionsentscheidungen und den damit verbundenen Lern-

prozessen (Espenshade und Fu 1997; Esser 2006; Kristen 2019). Als Investition in den Spracherwerb gilt jegliche Aktivität, die potentiell mit einer Steigerung der Sprachkompetenzen verbunden ist. Dies beinhaltet sowohl strukturierte Lernprozesse wie den Besuch eines Sprachkurses oder Deutschunterricht in der Schule, als auch informelle Prozesse wie das Lesen eines Buches oder die Verwendung der Sprache im alltäglichen Kontext.

Weiterhin scheint es gleichbedeutend zu sein, ob mit der Investitionsentscheidung bewusst eine Steigerung der Kompetenzen bezweckt wird, oder dies als nicht intendierte Folge hieraus resultiert (Esser 2006, S. 62f.). Zum Beispiel kann angenommen werden, dass der aktive Gebrauch der Verkehrssprache im Freundeskreis mit gesteigerten L2-Kompetenzen einhergeht, auch wenn die vorgelagerte Entscheidung, ob nun L1 oder L2 mit den Freund:innen gesprochen wird, nicht explizit vor dem Hintergrund einer gezielten Investition für oder gegen den L2-Erwerb getroffen wurde. Entscheidend ist dabei jedoch, dass dieser vorgelagerten Entscheidung trotzdem eine Selektion von Handlungsalternativen zugrunde liegt. Dementsprechend folgen beide Lernprozesse einer bestimmten Handlungslogik.

Die Überlegungen zu einer dem Spracherwerb vorausgehenden Selektion von Handlungsalternativen lassen sich in ein allgemeines handlungstheoretisches Modell übersetzen. Im Folgenden wird dies in Anlehnung an Hartmut Essers (1999, 2006) Ausführungen zur Theorie des subjektiv erwarteten Nutzens (SEU-Theorie) expliziert. Der Spracherwerb, wie auch der Kompetenzerwerb im Allgemeinen, kann dabei als spezieller Anwendungsfall eines übergreifenden Investitionsmodells verstanden werden (Esser 2006, S. 53, 292; für die formale Anwendung auf den schulischen Kompetenzerwerb, vgl. auch Dollmann 2010).

Ausgangspunkt des allgemeinen Modells ist eine Entscheidungssituation vor dem Hintergrund verschiedener Handlungsalternativen. Dabei wird angenommen, dass Akteure die Folgen jeder Alternative rational abwägen und ihre Entscheidungen mit dem Ziel einer subjektiven Nutzenmaximierung treffen (vgl. Esser 1999, S. 247f.). In der Sprache der SEU-Theorie messen Akteure jeder Alternative einen bestimmten Gesamtnutzen bei, das sogenannte SEU-Gewicht, welches sich aus der Bewertung und der erwarteten Eintrittswahrscheinlichkeit der Folgen zusammensetzt (ebd.). Letztlich wird aus allen vorliegenden Handlungsalternativen jene mit dem höchsten SEU-Gewicht ausgewählt.

Angewendet auf den Spracherwerb ergeben sich zwei grundlegende Handlungsalternativen: Der Verbleib im Status quo (sq) und damit verbunden die Hinnahme eines bereits erreichten Kompetenzniveaus, versus die Investition in weitere kompetenzsteigernde Aktivitäten (in). Da sich der mit den bereits erworbenen Sprachkompetenzen verbundene Nutzen (U_{sq}) mit hoher Wahrscheinlichkeit und ohne anfallende Kosten realisieren lässt, ergibt sich für den Verbleib im Status quo folgendes SEU-Gewicht:

$$SEU_{sq} = U_{sq} \tag{2.1}$$

Im Gegensatz dazu ist die Realisierung des Nutzens (U_{in}), der einer Investition in weitere Sprachkompetenzen beigemessen wird, an die Wahrscheinlichkeit des erfolg-

reichen Kompetenzerwerbs (p_{in}) gekoppelt. Beispielsweise führt die Entscheidung zur Teilnahme an einem Sprachkurs nicht zwangsläufig auch zur Erreichung des angestrebten Sprachniveaus. Hier werden zusätzliche Faktoren relevant, welche sich in der Erfolgswahrscheinlichkeit niederschlagen, wie die aktive Teilnahme am Unterricht, die Vor- und Nachbereitung der Lerninhalte, oder auch das individuelle Lernvermögen. Sollte der Erwerb zusätzlicher Kompetenzen nicht gelingen, so verbleiben die Akteure trotz Investition in dem jeweiligen Status quo. Mit der entsprechenden Gegenwahrscheinlichkeit ($1 - p_{in}$) resultiert aus der Investition dann lediglich der mit den bereits erworbenen Kompetenzen verbundene Nutzen (U_{sq}). Darüber hinaus müssen, unabhängig vom Ausgang des Kompetenzerwerbs, die anfallenden Kosten der Investition (C_{in}), in Form von materiellem, zeitlichem oder kognitivem Aufwand, getragen werden. Formal lässt sich die Evaluation des entsprechenden SEU-Gewichts wie folgt darstellen:

$$SEU_{in} = U_{in} * p_{in} + U_{sq} * (1 - p_{in}) - C_{in} \quad (2.2)$$

Unter der Prämisse, dass Individuen – zumindest im Rahmen ihres subjektiven Ermessensspielraums – optimale Investitionsentscheidungen treffen, wird dann in den weiteren Spracherwerb investiert, wenn der erwartete Gesamtnutzen der Investition den des Verbleibs im Status quo übersteigt; also wenn:

$$SEU_{in} > SEU_{sq} \quad (2.3)$$

Durch Einsetzen der Gleichungen (2.1) und (2.2) sowie einer Auflösung und Umstellung der Ungleichung folgt aus (2.3) die formale Bedingung für den erfolgreichen Spracherwerb:

$$U_{in} - U_{sq} > \frac{C_{in}}{p_{in}} \quad (2.4)$$

Die Differenz aus dem erwarteten Nutzen der Investition und dem des Verbleibs im Status quo ($U_{in} - U_{sq}$) kann als Grenznutzen der zusätzlich erworbenen Sprachkompetenzen interpretiert werden. Dieser Term entspricht der *Motivation* für den Spracherwerb. Demgegenüber steht eine gewisse Investitions- bzw. Lernschwelle, welche überschritten werden muss, um zusätzliche Sprachkompetenzen zu erwerben (Esser 2006, S. 41, 75). Die zu überwindende Schwelle wird dabei durch die Relation zwischen den anfallenden *Kosten* (C_{in}) und den Opportunitäten für den Spracherwerb (p_{in}) bestimmt. Letztere ergeben sich wiederum aus dem Zusammenspiel zwischen dem *Zugang* (*exposure*) zu kompetenzsteigernden Aktivitäten und der *Effizienz*, mit der entsprechende Lerngelegenheiten genutzt werden können ($p_{in} = p_{exp} * p_{eff}$):

$$U_{in} - U_{sq} > \frac{C_{in}}{p_{exp} * p_{eff}} \quad (2.5)$$

Es zeichnen sich demnach vier übergeordnete Faktoren ab, welche den Erwerb von Sprachkompetenzen bedingen: Motivation, Zugang, Effizienz und Kosten. Damit umfasst das Modell auch die Grundgedanken linguistischer und ökonomischer Ansätze zur Erklärung des Spracherwerbs, die von vergleichbaren Einflussgrößen ausgehen (z.B. Chiswick und Miller 1995, 2001; Spolsky 1989). Gleichzeitig ermöglicht das SEU-Modell, gegenüber früheren Ansätzen, eine erweiterte Spezifizierung der zugrundeliegenden Wirkzusammenhänge. So geht klar hervor, dass die Motivation für einen weiteren Spracherwerb nicht allein durch den Anreizwert gesteigerter Kompetenzen bestimmt wird, sondern ebenfalls in Abhängigkeit des Status quo variiert. Die Motivation fällt entsprechend umso geringer aus, je gewinnbringender die bereits erworbenen Kompetenzen eingesetzt werden können, *ceteris paribus*. Hieraus lassen sich zwei zentrale Tendenzen für den Spracherwerb in Abhängigkeit des Status quo ableiten: Einerseits ist davon auszugehen, dass die Investitionsneigung mit steigenden Sprachniveau abnimmt, d.h. die Lernkurve weist einen abnehmenden Grenznutzen auf. Andererseits wird mit Blick auf den Anwendungsfall des *Zweitspracherwerbs* von Jugendlichen mit Migrationshintergrund, neben den bereits vorhandenen L2-Kompetenzen, zusätzlich die Verwertbarkeit von Kompetenzen in der Erstsprache als Status quo relevant (vgl. Esser 2006, S. 74f.).³ Wird der Erstsprache ein hoher Anreizwert zugeschrieben, beispielsweise wenn der Aufnahmekontext ausreichend Möglichkeiten zur Verwendung der L1 bietet, so kann dies – vermittelt über eine verminderte Motivation – den Zweitspracherwerb einschränken.

Eine weitere Besonderheit des allgemeinen Modells besteht darin, dass sich der Einfluss der Faktoren Motivation, Zugang und Effizienz multiplikativ zusammensetzt. Folglich stellen alle drei Faktoren eine notwendige Bedingung für den erfolgreichen Spracherwerb dar (vgl. Esser 2006, S. 77). Bei vollständig ausbleibendem Zugang können die Motivation und die Lerneffizienz noch so hoch sein, es wird dennoch nicht zum Erwerb zusätzlicher Sprachkompetenzen kommen. Die Situation ohne ein Mindestmaß an sprachlichem Zugang stellt empirisch sicherlich eher die Ausnahme als die Regel dar, dennoch verdeutlicht dieser Zusammenhang einen zentralen Aspekt: Einschränkungen im Zugang können nur bedingt durch eine erhöhte Motivation und Effizienz ausgeglichen werden, und umgekehrt.

Zusammenfassend lässt sich festhalten, dass Investitionen in den (Zweit-)Spracherwerb, und das daraus resultierende Kompetenzniveau, umso höher ausfallen, je ausgeprägter die Motivation, der Zugang sowie die Lerneffizienz sind und je weniger Kos-

³Die Überlegungen zur Relevanz des Status quo sind, zumindest implizit, bereits in den linguistischen und ökonomischen Ansätzen angelegt, etwa durch die Berücksichtigung der sprachlichen Vorkenntnisse (Spolsky 1989, S. 14ff.) oder dem Verweis auf den abnehmenden Grenznutzen bei steigenden Kompetenzen (Chiswick und Miller 2001). Allerdings bleibt eine explizite Bestimmung des Status quo, gerade auch hinsichtlich der Bedeutung der Erstsprache, aus (vgl. Esser 2006, S. 73f.).

ten anfallen. Ausgehend von den geschilderten Zusammenhängen können sprachliche Leistungsunterschiede zwischen Jugendlichen mit und ohne Migrationshintergrund, oder auch zwischen verschiedenen Herkunftsgruppen und Zuwanderungsgenerationen, grundsätzlich durch systematisch unterschiedliche Ausprägungen der vier übergreifenden Faktoren erklärt werden. Dabei gilt es, die individuellen Randbedingungen zu berücksichtigen und mit den entsprechenden Modellparametern zu verknüpfen. Nachfolgend wird dieses analytische Vorgehen am Beispiel des Sprachgebrauchs, welcher im Fokus der vorliegenden Arbeit steht, veranschaulicht.

2.1.2 Sprachgebrauch und Zweitspracherwerb

Ausgehend von dem allgemeinen theoretischen Rahmen, lässt sich die Verwendung der Herkunftssprache über zwei Mechanismen mit dem Zweitspracherwerb in Verbindung bringen.

Erstens, kann der aktive und häufige L1-Gebrauch die Lerngelegenheiten für den Zweitspracherwerb begrenzen. Dem zugrunde liegt die Annahme, dass die Zeit, die zur Verwendung der L1 aufgebracht wird, gleichermaßen den verfügbaren Zeitrahmen für den Gebrauch und damit auch den Erwerb der Verkehrssprache verringert (Hopf 2005; Scheele et al. 2010). Dieses „time-on-task“-Argument geht auf ein Lernmodell nach Carroll (1963) zurück. Demnach ergibt sich der Lernzuwachs in einer bestimmten Kompetenzdomäne aus dem Quotienten zwischen der Zeit, die Individuen in die Aneignung entsprechender Kompetenzen investieren, und dem zeitlichen Umfang, der benötigt wird, um sich entsprechende Inhalte anzueignen. Da die für Lernprozesse insgesamt aufgebrauchte oder verfügbare Zeit beschränkt ist, muss diese, je nach individuellen Präferenzen und Bedürfnissen, auf unterschiedliche Lerninhalte aufgeteilt werden. Demzufolge ist jede Investition in den Erwerb einer bestimmten Kompetenz gleichzeitig mit Opportunitätskosten gegenüber anderen Kompetenzbereichen verbunden. Ausgehend davon lässt sich für Jugendliche, die in alltäglichen Situationen ausschließlich die L2 verwenden, bei gleicher Lerneffizienz ein höherer Kompetenzzuwachs in der Verkehrssprache prognostizieren, als für diejenigen, die zumindest manchmal auch die L1 nutzen. Übertragen auf das allgemeine Modell des Spracherwerbs beschränkt die Verwendung der L1, anstelle der L2, den Zugang zur Verkehrssprache und damit auch den Erwerb entsprechender Kompetenzen.

Zweitens, können sich im Sprachgebrauch die bestehenden Gelegenheitsstrukturen zur Verwertbarkeit der Herkunftssprache widerspiegeln. Mit zunehmenden Möglichkeiten zur Verwendung der L1 innerhalb des Aufnahmekontexts ist ein erhöhter Anreizwert des Status quo verbunden. Derartige Tendenzen können beispielsweise durch eine eigenethnische Einbettung im Freundschaftsnetzwerk oder im Schulumfeld begünstigt werden (vgl. Kristen und Olczyk 2013). Besonders bedeutsam scheint hier jedoch die Situation im familiären Umfeld (vgl. Alba et al. 2002). Sollten die Eltern oder nahestehende Familienangehörige nur über eingeschränkte L2-Kompetenzen verfügen, dann besteht seitens der Kinder ein erhöhter Anreiz, Kompetenzen in der Herkunftssprache zu erhalten oder auszubauen, um so die Kommunikation innerhalb der Familie sicherzustellen. Allerdings ist nicht davon auszugehen, dass dadurch die Anreize, welche

der Erwerb zusätzlicher L2-Kompetenzen verspricht (z.B. höhere Bildungszertifikate), vollständig kompensiert werden. Dennoch kann sich der erhöhte Anreizwert des Status quo, vermittelt über eine geringere Motivation, nachteilig auf den Zweitspracherwerb auswirken.

Bringt man beide Argumente zusammen, so sollten zunehmende Kontaktmöglichkeiten zur Herkunftssprache und die aktive Nutzung dieser Gelegenheiten mit Nachteilen in den Kompetenzen der Verkehrssprache einhergehen.⁴

Die allgemeine Wirkungsweise des L1-Gebrauchs lässt sich auf unterschiedliche Kontexte der alltäglichen Sprachnutzung gleichermaßen anwenden. Neben der Kommunikation innerhalb der Familie bieten auch Medien weitreichende Möglichkeiten die Herkunftssprache zu nutzen. Darüber hinaus ist anzunehmen, dass sich auch mit Blick auf die Interaktion im Freundeskreis oder im Schulumfeld – zumindest für die quantitativ bedeutsamsten Zuwanderungsgruppen (z.B. Personen aus der Türkei oder der ehemaligen Sowjetunion) – zusätzliche Gelegenheiten ergeben die L1 zu verwenden. Insgesamt sollten die Kompetenzen in der Verkehrssprache also umso geringer ausgeprägt sein, je häufiger bzw. in je mehr Kontexten die Herkunftssprache anstelle der L2 verwendet wird. Gleichzeitig ist bei einem vorwiegenden Gebrauch der Verkehrssprache aber auch eine Zunahme der L2-Kompetenzen zu erwarten.

Entsprechende Zusammenhänge konnten für die innerhalb der Familie verwendete Sprache mehrfach belegt werden. So zeichnet sich in den Befunden zahlreicher Studien ab, dass der L1-Gebrauch in der Familie in negativem Zusammenhang mit den Deutschkompetenzen von Kindern und Jugendlichen mit Migrationshintergrund steht (z.B. Baumert und Schümer 2001; Edele und Stanat 2016; Kristen 2008b; Stanat et al. 2010a; Walter 2009b). Neben den direkten Einflüssen auf den L2-Erwerb, könnte der familiäre Kontext aber auch deswegen eine tragende Rolle einnehmen, weil dieser oftmals darüber entscheidet, ob Kinder die L1 überhaupt erlernen und dann wiederum auch in anderen Kontexten nutzen können (Alba et al. 2002; Biedinger et al. 2015). Die Nachteile, die dem L1-Gebrauch in der Familie zugeschrieben werden, könnten folglich – zumindest in Teilen – auch mit dem L1-Gebrauch in weiteren Kontexten in Verbindung gebracht werden. Derartige (konfundierende) Einflüsse des Sprachgebrauchs in weiteren Kontexten fanden in der Forschung bisher allerdings weniger Beachtung.

⁴Neben diesen negativen Wirkungsweisen des L1-Gebrauchs werden in der Literatur ebenfalls mögliche vorteilhafte Seiten einer Erhaltung und Nutzung der L1 diskutiert. Entsprechende Argumentationslinien bauen meist auf den Grundzügen der „Theory of Segmented Assimilation“ (Portes und Zhou 1993) auf und betonen die Bedeutung der Herkunftssprache für die Vermittlung (bildungs-)förderlicher Orientierungen und Ressourcen innerhalb der Familien und eigenethnischen Netzwerke (z.B. Bankston und Zhou 1995; Mouw und Xie 1999; Portes und Hao 2002). Derartige Vorteile des L1-Gebrauchs werden in erster Linie hinsichtlich der schulischen Leistungen und Bildungsbeteiligung von Kindern und Jugendlichen mit Migrationshintergrund diskutiert, nicht aber hinsichtlich der Kompetenzen in der Verkehrssprache, für die sich aus theoretischer wie empirischer Sicht eine nachteilige Wirkung des L1-Gebrauchs abzeichnet (vgl. Kristen und Olczyk 2013). Da in der vorliegenden Arbeit gerade der Zweitspracherwerb im Vordergrund steht, werden die im Zusammenhang mit förderlichen Wirkungsweisen stehenden Argumente nur am Rande aufgegriffen – etwa zur Kontrolle weiterer potentieller Mechanismen im Rahmen der Analysen zu Einflüssen der ethnischen bzw. ethno-lingualen Klassenkomposition (vgl. Kapitel 5).

Die Frage, ob der häufige L1-Gebrauch in anderen Kontexten zusätzlich zur Erklärung von ethnischen Leistungsdisparitäten in der Verkehrssprache beiträgt und welche Bedeutung dabei den jeweiligen Kontexten zukommt, hängt letztendlich auch davon ab, ob der außerfamiliäre Gebrauch der Herkunftssprache überhaupt ein quantitativ bedeutsames Phänomen oder lediglich eine empirische Randerscheinung darstellt. Sollte sich beispielsweise abzeichnen, dass Jugendliche mit Migrationshintergrund unter Freund:innen und Mitschüler:innen nahezu ausschließlich Deutsch sprechen, so wäre diesbezüglich auch kein signifikanter Erklärungsbeitrag für ethnische Leistungsunterschiede zu erwarten. Folglich gilt es zunächst deskriptiv zu klären, wie sich der Sprachgebrauch innerhalb einzelner Kontexte sowie auch im Vergleich zwischen den verschiedenen Kontexten verteilt.

Die überschaubare Befundlage hierzu liefert erste Hinweise darauf, dass der Medienkonsum (z.B. Diehl und Schnell 2006; Simon und Neuwöhner 2011; Worbs 2010) und die außerfamiliäre Kommunikation (z.B. Gärtig et al. 2011; Nauck 2007) vorwiegend durch die Nutzung der Verkehrssprache geprägt ist. Gleichzeitig verwendet aber auch ein nicht unbeträchtlicher Anteil von Personen mit Migrationshintergrund besonders bei der Fernseh- und Internetnutzung zumindest manchmal die Herkunftssprache (Simon und Neuwöhner 2011). Ein vergleichbarer Trend zeigt sich auch für die Kommunikation im Freundes- und Bekanntenkreis (Gärtig et al. 2011; Nauck 2007; Salentin und Wilkening 2003). Mit Blick auf unterschiedliche Tendenzen zwischen Herkunftsgruppen weist der Forschungsstand übereinstimmend darauf hin, dass türkischstämmige Personen, gegenüber den meisten anderen Zuwanderungsgruppen, sowohl innerhalb als auch außerhalb der Familie häufiger die Herkunftssprache nutzen (z.B. Diehl und Schnell 2006; Nauck 2007; Simon und Neuwöhner 2011; Salentin und Wilkening 2003; Strobel und Kristen 2015).

Bei der Interpretation der Befundmuster muss allerdings berücksichtigt werden, dass die bisherigen Forschungsbeiträge meist einzelne Kontexte des Sprachgebrauchs beleuchten. Ein Vergleich des Sprachgebrauchs zwischen Kontexten sowie eine übergreifende Einordnung ausgehend von diesen Befunden wird dadurch erschwert, dass diese größtenteils auf sehr unterschiedlichen Alterskohorten und Herkunftsgruppen beruhen. Dementsprechend lässt sich nicht eindeutig bestimmen, ob der Sprachgebrauch tatsächlich nach Kontexten oder lediglich in Abhängigkeit der jeweils zugrundeliegenden Analysepopulation variiert. Fraglich ist auch, ob sich die Befunde zu meist älteren Migrant:innen auf den Anwendungsfall von Jugendlichen mit Migrationshintergrund übertragen lassen. Weiterhin ist anzunehmen, dass der L1-Gebrauch in der Generationenfolge abnimmt (vgl. Hans 2010; Ramm et al. 2005; Segeritz et al. 2010). Eine Berücksichtigung des Generationenstatus bleibt jedoch in den meisten Studien zum außerfamiliären Sprachgebrauch aus. Dies ist insofern kritisch zu betrachten, als dass die generationale Zusammensetzung zwischen Herkunftsgruppen stark variiert und dementsprechend ein Vergleich unterschiedlicher Zuwanderungsgruppen nur bei zusätzlicher Differenzierung des Generationenstatus aussagekräftige Befunde zulässt (vgl. Gresch und Kristen 2011).

Weiterhin hat der Sprachgebrauch im schulischen Umfeld *empirisch* in der bisherigen Forschung kaum Beachtung gefunden. Das bestehende Forschungsdesiderat wird

insbesondere auch dadurch deutlich, dass nahezu alle Abhandlungen zu Einflüssen der ethnischen Klassenzusammensetzung auf die schulische Leistungsentwicklung den herkunftssprachlichen Gebrauch zwischen Mitschüler:innen als potentielle Ursache für nachteilige Effekte diskutieren (z.B. Driessen 2002; Esser 2006; Olczyk 2018; Peetsma et al. 2006; Stanat 2006; Stanat et al. 2010b; Van Ewijk und Sleegers 2010; Veerman und Dronkers 2016; Walter 2009a). Die ausbleibende empirische Prüfung dieser Zusammenhänge ist in erster Linie der Tatsache geschuldet, dass die wenigen Datensätze, welche überhaupt eine aussagekräftige Untersuchung von Effekten der Schul- und Klassenkomposition erlauben, meist keine Informationen zum Sprachgebrauch zwischen Mitschüler:innen enthalten.

Aber auch unabhängig davon, wurden nur selten analytische Konsequenzen aus den theoretischen Überlegungen gezogen. So sollten Gelegenheiten zur Verwendung der L1 zwischen Mitschüler:innen weniger vom Anteil Personen mit Migrationshintergrund in der Klasse abhängen als vielmehr davon wie viele Mitschüler:innen die gleiche Herkunftssprache sprechen (vgl. Stanat et al. 2010b). Im Gegensatz dazu wurden in der bisherigen Forschung eine Reihe sehr unterschiedlicher Charakteristiken zur Operationalisierung der ethnischen Klassenkomposition herangezogen (z.B. Hautfarbe, Staatsbürgerschaft, Geburtsland, Muttersprache, etc.), was eine Einschätzung der ohnehin äußerst heterogen ausfallenden Befundlage zunehmend erschwerte (vgl. Kristen und Olczyk 2013; Eksner und Stanat 2011). Darüber hinaus ist davon auszugehen, dass sich Effekte der ethnischen bzw. ethno-lingualen Klassenkomposition differentiell auf die schulischen Leistungen von Jugendlichen mit und ohne Migrationshintergrund auswirken, da letztere ausschließlich die Verkehrssprache sprechen (vgl. Bankston und Caldas 1996; Stanat 2006; Van Ewijk und Sleegers 2010). Weiterhin lässt sich die Wirkungsweise des durch den L1-Gebrauch bedingten Spracherwerbs nicht ausschließlich auf die Kompetenzen in der Verkehrssprache anwenden, sondern gleichermaßen auch auf herkunftssprachliche Kompetenzen. Demzufolge wäre zu erwarten, dass der Anteil eigenethnischer bzw. gleichsprachiger Mitschüler:innen in der Klasse positiv mit den L1-Kompetenzen zusammenhängt (vgl. Van Tubergen und Mentjox 2014). Versuche einer derartigen indirekten Prüfung der Wirkungsweise des L1-Gebrauchs im schulischen Umfeld blieben im Groß der bisherigen Forschungsbeiträge weitestgehend aus.

2.1.3 Beitrag der vorliegenden Arbeit

Die vorliegende Dissertation setzt an der bisherigen Befundlage an und versucht im Rahmen von zwei Beiträgen die Schließung der geschilderten Forschungslücken voranzutreiben.

Der erste Beitrag besteht darin, die quantitative Bedeutung des außerfamiliären Sprachgebrauchs von Jugendlichen mit Migrationshintergrund deskriptiv zu bestimmen und hieraus resultierende Zusammenhänge mit den Zweitsprachkompetenzen aufzudecken (siehe Kapitel 4). Anhand einer einheitlichen Datengrundlage werden zunächst die Muster des Sprachgebrauchs innerhalb der Familie (d.h. mit dem Vater, der Mutter und den Geschwistern), im Freundeskreis, im schulischen Umfeld und beim Medienkonsum beschrieben. Dadurch lassen sich die nur eingeschränkt vergleichbaren

Befundmuster bisheriger – auf unterschiedlichen Analysepopulationen beruhenden – Studien prüfen. Ergänzend zur bisherigen Forschung, erfolgt zusätzlich eine Differenzierung des kontextspezifischen Sprachgebrauchs nach unterschiedlichen Herkunftsgruppen, Zuwanderungsgenerationen sowie der Kombination aus beidem. Darüber hinaus wird geprüft, inwieweit der Sprachgebrauch in unterschiedlichen Kontexten mit den Zweitsprachkompetenzen in Verbindung steht und welche Kontexte hierbei besonders ausschlaggebend sind. Ausgehend davon kann abgeschätzt werden, ob sich die mit dem familiären L1-Gebrauch assoziierten Nachteile durch eine verbreitete L1-Nutzung in anderen Kontexten zusätzlich verstärken, oder sich hingegen bei häufiger Verwendung der Verkehrssprache außerhalb der Familie sogar abschwächen.

Vor dem Hintergrund der bislang kaum erforschten Mechanismen, welche Einflüssen der ethnischen Klassenzusammensetzung auf den schulischen und sprachlichen Kompetenzerwerb zugrunde liegen könnten, widmet sich der zweite Beitrag der Frage nach dem Erklärungspotential des L1-Gebrauchs im schulischen Kontext (siehe Kapitel 5). Zwar ist die Verwendung der Herkunftssprache unter Mitschüler:innen nur einer von zahlreichen Faktoren, die zur Erklärung negativer Effekte der ethnischen Klassenkomposition diskutiert werden (für einen Überblick vgl. Driessen 2002; Peetsma et al. 2006; Van Ewijk und Slegers 2010), dennoch wird diesem, gerade im Kontext der deutschen Forschungslandschaft, eine tragende Rolle zugeschrieben (vgl. Eksner und Stanat 2011; Esser 2006). Im Gegensatz zu früheren Forschungsarbeiten kann in der vorliegenden Analyse auf einen Indikator zum Sprachgebrauch mit Mitschüler:innen zurückgegriffen werden, wodurch sich die vermuteten Zusammenhänge auf direktere Weise empirisch testen lassen. Um die Befunde zusätzlich abzusichern, erfolgt darüber hinaus eine indirekte Prüfung über theoriegeleitete Forschungshypothesen. Analog zu den in Abschnitt 2.1.2 beschriebenen Erwartungen, werden hierzu Effekte der ethnischen Klassenzusammensetzung denen der ethno-lingualen Komposition gegenübergestellt, differenzielle Einflüsse auf den Leistungserwerb von Jugendlichen mit und ohne Migrationshintergrund geprüft, und letztlich Zusammenhänge zwischen der ethno-lingualen Klassenkomposition und den herkunftssprachlichen Kompetenzen untersucht. Die Befunde geben nicht nur Aufschluss über die zugrundeliegenden Wirkzusammenhänge, sondern ebenfalls über den Stellenwert der ethnischen Klassenzusammensetzung für den Kompetenzerwerb insgesamt.

2.2 Validität subjektiver Einschätzungen von Sprachkompetenzen

In diesem Kapitel wird die Verlässlichkeit subjektiv eingeschätzter Sprachkompetenzen von Jugendlichen mit Migrationshintergrund als Indikator für deren tatsächliches Sprachvermögen thematisiert. Die Erfassung von Selbstberichten der Befragten zu ihren Kompetenzen bietet die Möglichkeit auch in Umfragen leistungsbezogene Informationen (effizient) zu erheben, ohne umfassende psychometrische Testverfahren einsetzen zu müssen, was in vielen Fällen wohl ohnehin den zeitlichen und monetären Rahmen sprengen würde und dementsprechend gar nicht in Frage käme. Nicht verwunderlich also, dass der Einsatz von Selbsteinschätzungsinstrumenten sehr verbreitet ist und diese subjektive Messung in vielen Datensätzen als einziger Indikator für die Sprachkompetenzen der Befragten vorliegt. Gleichzeitig gelten Selbstangaben zum eigenen Leistungsniveau, nicht nur mit Blick auf den sprachlichen Bereich, sondern vielmehr Domänen übergreifend, als wenig akkurat und anfällig für systematisch verzerrte Antworttendenzen (vgl. Dunning 2005; Falchikov und Boud 1989; Kruger und Dunning 1999; Zell und Krizan 2014). Diese fehlerhaften Tendenzen können sich auch auf die Ergebnisse statistischer Analysen auswirken, in denen Selbsteinschätzungen als Indikator für objektive Sprachkompetenzen verwendet werden. Es besteht die Gefahr, dass dabei zentrale Zusammenhänge verschleiert werden oder künstliche Korrelationen erzeugt werden und letztendlich falsche Rückschlüsse aus den Analysen gezogen werden. Um derartige Fehlschlüsse zu vermeiden, gilt es, die maßgeblichen Bedingungen und Prozesse, die den systematischen Antwortverzerrungen zugrunde liegen, zu identifizieren. Darauf aufbauend können Strategien entwickelt werden, um potentielle Fehlerquellen bereits bei der Datenerhebung – z.B. durch verbesserte Instrumente zur Selbsteinschätzung – auszuschalten, oder zumindest das Ausmaß und die Richtung der auftretenden Verzerrung in Analysen abzuschätzen und auf dieser Basis die Ergebnisse besser einordnen zu können.

Antwortverzerrungen in Befragungssituationen stellen ein grundsätzliches Problem der Umfrageforschung dar, welches sich neben Kompetenzeinschätzungen auf nahezu jegliche Befragungsinhalte übertragen lässt. Die Muster potentiell auftretender Antwortverzerrungen sind dabei ebenso vielfältig wie die Ursachen, die derartigen Tendenzen zugrunde liegen, und können in Abhängigkeit unterschiedlichster Faktoren variieren, wie beispielsweise Charakteristika des Erhebungsinstruments, der Interviewer:innen und der Befragten (Sudman und Bradburn 1974). Die vorliegende Problemstellung der Validität subjektiv eingeschätzter Sprachkompetenzen von Jugendlichen mit Migrationshintergrund kann als ein spezieller Anwendungsfall des generellen Antwortverhaltens in Umfragen angesehen werden. Um die Ursachen systematischer Fehleinschätzungen von Sprachkompetenzen aufzudecken, bietet es sich an, zunächst die allgemeinen Prozesse des Antwortverhaltens näher zu beleuchten. Ausgehend davon lassen sich dann die speziell für die Einschätzung von Sprachkompetenzen relevanten Bedingungen identifizieren und gezielt mit dem Antwortprozess in Verbindung bringen.

Nachfolgend wird zunächst ein allgemeiner theoretischer Rahmen des Antwortpro-

zesses in Befragungssituationen skizziert, welcher unterschiedliche Ansätze aus der bisherigen Umfrageforschung integriert (Abschnitt 2.2.1). In einem weiteren Schritt werden die für Antwortverzerrungen bei der Selbsteinschätzung von Sprachkompetenzen ausschlaggebenden Bedingungen und Abläufe spezifiziert und deren empirische Relevanz im Lichte bisheriger Befunde diskutiert (Abschnitt 2.2.2). Abschließend wird in Abschnitt 2.2.3 der empirische Beitrag der vorliegenden Arbeit hervorgehoben.

2.2.1 Ein allgemeines Modell des Antwortverhaltens in Umfragen

In den vergangenen Jahrzehnten wurden zahlreiche theoretische Überlegungen zu den kognitiven Abläufen bei der Beantwortung von Fragen in Umfragesituationen und den damit verbundenen Fehlerquellen für Antwortverzerrungen erarbeitet. Nach einem übergreifenden Schema von Roger Tourangeau (1984) lässt sich der Antwortprozess in vier grundlegende Schritte unterteilen (vgl. auch Cannel et al. 1981; Strack und Martin 1987; Sudman et al. 1996; Tourangeau und Rasinski 1988; Tourangeau et al. 2000): Zunächst müssen Befragte die ihnen gestellte Frage sowohl sprachlich als auch inhaltlich verstehen und interpretieren (*comprehension*). Darauf aufbauend erfolgt der Abruf relevanter Informationen aus dem Gedächtnis (*retrieval*), welche anschließend hinsichtlich ihrer Passung zum Frageinhalt bewertet und zu einem Gesamturteil zusammengefasst werden (*judgment*). In einem letzten Schritt muss das Urteil dann noch in eine angemessene Antwort übersetzt werden (*response*), was im Fall von standardisierten Befragungen meist der Zuordnung des Urteils zu einer der vorgegebenen Antwortoptionen entspricht.

Antwortverzerrungen können grundsätzlich in jedem dieser vier Prozessschritte auftreten, etwa durch eine fehlerhafte Interpretation der Frage, dem Abruf inakkuratere oder unvollständiger Informationen, einer unzureichenden Urteilsbildung oder der Auswahl einer falschen Antwortoption (Tourangeau et al. 2000, S. 8). Verzerrungen in den ersten Phasen des Antwortprozesses können dabei auch maßgeblich die Bedingungen der nachgelagerten Schritte mitbestimmen, obwohl der Ablauf der vier übergeordneten Prozessschritte nicht notwendigerweise sequentiell erfolgen muss (Sudman et al. 1996, S. 77; Tourangeau et al. 2000, S. 16). Wird beispielsweise eine Frage fehlinterpretiert, ist davon auszugehen, dass anschließend Informationen, die nicht mit dem ursprünglich intendierten Frageinhalt zusammenhängen, abgerufen und bewertet werden.

Die Fehlerquellen für ein verzerrtes Antwortverhalten lassen sich den Charakteristika der Befragten, des Befragungsinstruments und, in Abhängigkeit des Administrationsmodus, der Interviewer:innen zuschreiben (Sudman und Bradburn 1974). Einflüsse des Befragungsinstruments sowie der Eigenschaften und Verhaltensweisen der Interviewer:innen können hauptsächlich für die Prozessschritte *comprehension* und *response* erwartet werden, beispielsweise in Form einer unpräzisen Formulierung der Fragen oder einer bewertenden Kommunikation der Antwortoptionen durch die Interviewer:innen. Die Informationsverarbeitung der Schritte *retrieval* und *judgment* ist hingegen vorwiegend von kognitiven und motivationalen Eigenschaften der Befragten geprägt (z.B.

Krosnick 1991; Krosnick et al. 1996). Da allerdings der Abruf relevanter Informationen inhaltlich durch die Fragestellung vordefiniert wird und das Urteil letztlich einem vorgegebenen Antwortschema zugeordnet werden muss, sind Antwortverzerrungen vielmehr das Resultat aus dem Zusammenspiel der drei Faktoren (Schaeffer 2004; Sudman et al. 1996, S. 66, 73).

Ein besonders prominentes Beispiel für die Interaktion der drei Einflussfaktoren sind systematische Antwortverzerrungen durch soziale Erwünschtheit.⁵ Der „social desirability bias“ (SD-Bias) ist ein weit verbreitetes Phänomen in Umfragen und konnte für ein breites Spektrum von Befragungsinhalten beobachtet werden, wie zum Beispiel bei Fragen zu Wahlbeteiligungen (Belli et al. 2001), Alkohol- und Drogenkonsum (Tourangeau und Yan 2007), fremdenfeindlichen Einstellungen (Stocké 2004) oder zum sexuellen Verhalten (Tourangeau und Smith 1996). Sozial erwünschtes Antwortverhalten entsteht in Situationen, in welchen die inhaltliche Ausrichtung der Fragen und/oder Antwortoptionen den Befragten eine klare Erwünschtheitswahrnehmung signalisieren, was in besonderem Maße bei sensitiven Befragungsinhalten der Fall ist (Tourangeau und Yan 2007). Darüber hinaus ist ein SD-Bias verstärkt dann zu erwarten, wenn Befragte gleichzeitig ein ausgeprägtes Bedürfnis nach sozialer Anerkennung verspüren; eine Eigenschaft, die weniger situationsabhängig als vielmehr in der Persönlichkeit der Befragten verankert ist (DeMaio 1984; Paulhus 2003). Die Befriedigung des Anerkennungsbedürfnisses ist wiederum daran gekoppelt, ob in der Befragungssituation direkte Reaktionen in Form sozialer Anerkennung oder Missbilligung zu erwarten sind, zum Beispiel durch die Interviewer:innen (Stocké 2004). Zusammenfassend sind die Anreize eine sozial erwünschte Antwortoption zu wählen, gegenüber der Angabe des „wahren“ Urteils, umso höher, je eindeutiger Erwünschtheitssignale das *Befragungsinstrument* bzw. der Befragungsinhalt aussendet, je stärker das Bedürfnis nach sozialer Anerkennung der *Befragten* ausgeprägt ist und je wahrscheinlicher wertende Reaktionen der *Interviewer:innen* erwartet werden.

Während potentiellen Fehlerquellen im Befragungssetting durch eine ordentliche Studienkonzeption weitgehend vorgebeugt werden kann, stellen Charakteristika der Befragten eine nur sehr bedingt kontrollierbare Variable in der Befragungssituation dar. Selbst wenn Fehlerquellen seitens des Befragungsinstruments und des Interviewer:innen-Verhaltens vollständig ausgeschlossen werden könnten, kann es aufgrund von Eigenschaften der Befragten dennoch zu Antwortverzerrungen kommen. Vor diesem Hintergrund ist es umso entscheidender, die dem Antwortverhalten von Befragten zugrundeliegenden Prozesse, sowie deren Zusammenspiel mit dem Befragungssetting, und den damit verbundenen Fehlerquellen für Antwortverzerrungen zu identifizieren.

In einer – aus Sicht der Forschenden – idealen Umfragesituation bearbeiten die

⁵Zusätzlich zur abstrakten Beschreibung der allgemeinen Prozesse des Antwortverhaltens werden in Abschnitt 2.2.1 die Kernargumente anhand eines exemplarischen Anwendungsfalls veranschaulicht. Dies geschieht am Beispiel sozial erwünschter Antwortverzerrungen. Die anwendungsbezogenen Beispiele sind visuell durch eingerückten Text in kleinerer Schriftgröße gekennzeichnet.

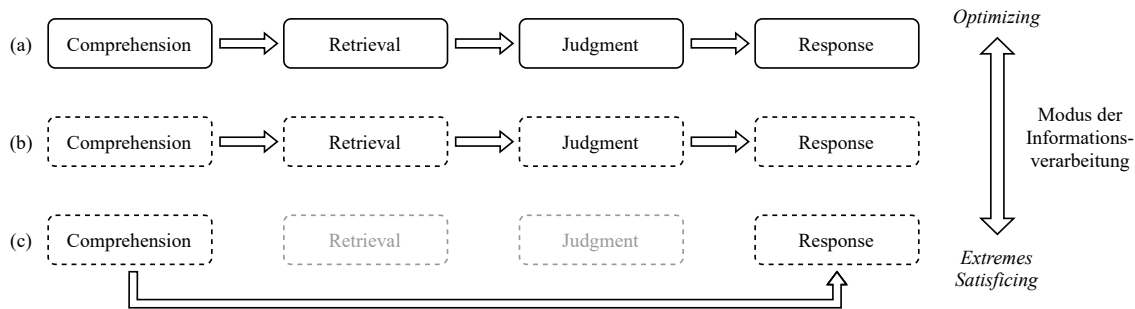
Befragten jeden der vier Schritte des Antwortprozesses vollständig und mit höchster Sorgfalt. Demnach versuchen sie die Intention der Frage korrekt zu identifizieren und alternative Interpretationen auszuschließen, rufen alle potentiell relevanten Informationen aus dem Gedächtnis ab, prüfen die Angemessenheit der abgerufenen Informationen, gleichen diese gegeneinander ab und bilden darauf basierend ihr Gesamturteil, und wählen letztlich die Antwortoption, die am besten dem Urteil entspricht (vgl. Tourangeau und Rasinski 1988; Krosnick 1991). Dieses idealtypische Antwortverhalten wird als „*optimizing*“ bezeichnet (vgl. Krosnick 1991, S. 214). Demnach nutzen Befragte ihre mentalen Ressourcen bestmöglich aus, um den kognitiven Anforderungen des Antwortprozesses gerecht zu werden. Ein solches optimierendes Antwortverhalten ist kognitiv sehr herausfordernd und setzt eine enorme Anstrengungsbereitschaft seitens der Befragten voraus (Krosnick 1999; Vannette und Krosnick 2014). Dementsprechend müssen schon besondere Anreizstrukturen gegeben sein, damit Befragte diesen Anforderungen gerecht werden, etwa in Form einer stark verankerten intrinsischen Motivation bei der Studienteilnahme oder einem besonderen Interesse an den Inhalten der Befragung oder an einzelnen Fragen (Krosnick 1999). Aber auch bei hoher Motivation können im Verlauf der Befragung kognitive Müdigkeitserscheinungen eintreten, gerade in längeren Befragungen und/oder bei anspruchsvollen Fragestellungen. Folglich ist es eher unwahrscheinlich, dass Befragte in einer realen Befragungssituation (durchgehend) den hohen kognitiven Aufwand betreiben, um optimierend zu antworten. Vielmehr ist davon auszugehen, dass Befragte ihre kognitiven Ressourcen nur bedingt ausreizen. Diese Tendenz wird in der Literatur als „(survey) *satisficing*“ bezeichnet (Krosnick 1991, S. 215).⁶

Satisficing stellt eine grundlegende Strategie des Befragtenverhaltens dar, bei der Befragte als „kognitive Geizhalse“ (Taylor 1981) versuchen den kognitiven Aufwand des Antwortprozesses möglichst gering zu halten (Sudman et al. 1996, S. 70). Demnach bearbeiten Befragte einzelne Prozessschritte – oder auch alle – weniger sorgfältig und/oder unvollständig, d.h. sie verarbeiten nicht alle relevanten und verfügbaren Informationen, sondern nur diejenigen, die leicht zugänglich sind und eine kognitiv sparsame Beantwortung der Frage versprechen (Tourangeau und Rasinski 1988; Vannette und Krosnick 2014). Die Heuristik des *satisficing* ließe sich auch als „kognitiver Energiesparmodus“ umschreiben, bei dem Befragte den Einsatz ihrer kognitiven Ressourcen im Antwortprozess auf das Nötigste herunterschrauben. Im Kern kann jede Abweichung von der optimierenden Antwortgenerierung als *satisficing* gewertet werden, welche sich allerdings in Form und Stärke unterscheiden kann.

Abbildung 2.1 bietet eine schematische Gegenüberstellung der Abläufe bei – von oben nach unten – abnehmendem Grad an Vollständigkeit und Sorgfalt der Bearbeitung des Antwortprozesses. Der Fall a repräsentiert das idealtypische Verhalten des *optimizing*, bei welchem die Befragten jeden einzelnen Schritt vollständig und mit

⁶Die Begriffe *optimizing* und *satisficing* sind angelehnt an die handlungstheoretischen Grundlagen einer eingeschränkten Rationalität bei Simon (1957), welche von Krosnick (1991) auf das Antwortverhalten in Umfragesituationen übertragen wurden. Im Rahmen der vorliegenden Arbeit werden die beiden englischen Begriffe übernommen (nachfolgend kursiv hervorgehoben), weil sich bis dato keine klare Begriffsbezeichnung im Deutschen etabliert hat.

Abbildung 2.1: Schematische Darstellung des Antwortprozesses in Abhängigkeit des Modus der Informationsverarbeitung



Anmerkung: Eigene Darstellung. Exemplarische Gegenüberstellung des Antwortprozesses im Fall von a) *optimizing*, b) schwachem *satisficing* und c) extremem *satisficing*. Gestrichelte Umrandung der Prozessschritte kennzeichnet einen reduzierten Grad an Vollständigkeit und Sorgfalt mit dem diese bearbeitet werden; hellgrau markierte Prozessschritte ohne Verbindungspfeil werden im Antwortprozess komplett ausgelassen.

höchster Sorgfalt durchlaufen. Bei einer schwach ausgeprägten Tendenz zum *satisficing* hingegen, führen Befragte zwar ebenfalls alle vier Schritte des Antwortprozesses aus, bearbeiten einzelne oder alle Schritte aber nur in reduziertem Umfang (in Abb. 2.1, Fall b, gekennzeichnet durch gestrichelte Umrandungen der Prozessschritte). Das äußert sich beispielsweise darin, dass Befragte eine Frage nach ihrer ersten Assoziation damit interpretieren, nicht alle potentiell bedeutsamen Informationen aus dem Gedächtnis, sondern nur die offensichtlichen und leicht zugänglichen Erinnerungen, abrufen und diese nur grob auf Korrektheit und Passung zur Fragestellung evaluieren (vgl. Krosnick 1991; Krosnick et al. 1996). Letztlich geben sich Befragte bei der Auswahl der Antwortoption weniger Mühe und entscheiden sich nicht unbedingt für diejenige, die ihr Urteil am besten repräsentiert, sondern für die erste vertretbare Option (ebd.). Als Konsequenz aus schwachem *satisficing* ergeben sich nicht zwangsläufig komplett „falsche“ Angaben der Befragten, allerdings nimmt die Reliabilität der Antworten ab, d.h. die generelle Tendenz von Antwortmustern wird richtig wiedergegeben, aber es entstehen größere Zufallsschwankungen um den „wahren“ Wert. Verbreitete Muster bei schwachem *satisficing* sind beispielsweise Zustimmungstendenzen oder auch Reihenfolgeeffekte bei kategorialen Antwortskalen, wie die Wahl der ersten oder letzten aufgelisteten Antwortkategorie (vgl. Vannette und Krosnick 2014).

Satisficing kann aber auch extremere Ausprägungen annehmen, bis hin zu einer Strategie bei der die Befragten den kognitiven Aufwand so weit reduzieren, dass sie keinerlei Informationen abrufen und bewerten – also die Schritte *retrieval* und *judgment* komplett überspringen (vgl. Abb. 2.1, Fall c). In diesem Fall wird die Frage nur oberflächlich interpretiert, um eine für Außenstehende sinnvoll erscheinende Antwort geben zu können, auch wenn diese losgelöst von jeglicher inhaltlichen Fundierung ist (Krosnick 1991; Krosnick et al. 2002). Beispiele für besonders starke Formen des *satisficing* ohne inhaltlichen Bezug wären rein zufällige Antworten oder die unreflektierte

Wahl der Ausweichkategorie „weiß nicht“ (Krosnick et al. 2002; Vannette und Krosnick 2014).

Das Antwortverhalten in Umfragen lässt sich als Kontinuum zwischen den beiden Ausprägungen *optimizing* und extremen *satisficing* beschreiben und spiegelt dabei den Grad der Vollständigkeit und Sorgfalt wider, mit welchen die vier Schritte des Antwortprozesses – insbesondere *retrieval* und *judgment* – bearbeitet werden (Krosnick 1991, 1999). Die Unterscheidung zwischen den beiden Endpunkten ist konzeptionell aber *nicht* gleichzusetzen mit dem Grad der Validität bzw. Verzerrung von Antworten. Zwar steigt die Wahrscheinlichkeit inakkuratere Antworten zweifelsohne mit abnehmender Sorgfalt bei der Bearbeitung des Antwortprozesses, allerdings können auch bei einem optimierendem Antwortverhalten systematische Verzerrungen auftreten, etwa in Form von sozial erwünschten Antworten, welche die Befragten wissentlich der „wahren“ Antwort vorziehen. Gleichzeitig muss das Ergebnis eines verkürzten Antwortprozesses nicht zwangsläufig inakkurat oder falsch sein, sondern kann ebenso – wenn auch eher zufällig – mit dem „wahren“ Wert übereinstimmen. Der zentrale Unterschied zwischen *optimizing* und *satisficing* besteht im jeweiligen Modus der Informationsverarbeitung, welcher dem Antwortprozess zugrunde liegt.⁷ Befragte unterscheiden sich demnach nicht nur bezüglich der Ausprägungen bestimmter Randbedingungen (z.B. Einstellungen, Wissen, Anreize), welche ausschlaggebend für die Wahl der Antwort sowie potentielle Verzerrungen sind, sondern auch darin, wie sie Informationen und verzerrende Einflüsse im Verlauf des Antwortprozesses verarbeiten. Dies ist insofern entscheidend, als dass Fehlerquellen den Antwortprozess auf unterschiedliche Art und Weise beeinflussen, woraus sich auch Implikationen für das Ausmaß der Verzerrung sowie für eine potentielle Fehlerbeseitigung ergeben.

⁷Die Unterscheidung der zwei Modi der Informationsverarbeitung orientiert sich in den Grundzügen an elaborierten Konzepten wie sie bereits in „dual-process“-Theorien aus der kognitiven Sozialpsychologie (z.B. Chaiken und Trope 1999; Fazio 1990; Strack und Deutsch 2004) sowie im Modell der Frame-Selektion aus der Soziologie (Esser 2001; Kroneberg 2005, 2011) angelegt sind. Da je nach Modus unterschiedliche kognitive Prozesse und Handlungs- bzw. Selektionslogiken ablaufen, werden teilweise auch unterschiedliche theoretische Erklärungsansätze relevant. Beispielsweise zeichnet sich in den idealtypischen Prozessabläufen bei *optimizing* implizit die Annahme einer rationalen Selektionslogik ab. Ähnlich zu den in Kapitel 2.1 geschilderten handlungstheoretischen Grundlagen, ließe sich in diesem Fall die Wert-Erwartungs-Theorie zur Erklärung des Antwortverhaltens und potentieller Antwortverzerrungen heranziehen (vgl. Esser 1990, 1991; Tourangeau et al. 2000, S. 281ff.). Eine instrumentell rationale Antwortselektion – wie sie in diesem Ansatz angenommen wird – setzt aber voraus, dass im Rahmen des Antwortprozesses *alle* zugänglichen Informationen vor dem Hintergrund der subjektiven Ziele *sorgfältig* bewertet und in die Entscheidung einbezogen werden. Ist diese Voraussetzung nicht erfüllt – wie im Fall von *satisficing*, ist hingegen von einer eingeschränkten Rationalität der Befragten auszugehen, welche entsprechend auch eine gesonderte theoretische Modellierung verlangt. Vor diesem Hintergrund erfordert die Erklärung des Antwortverhaltens einen übergeordneten theoretischen Rahmen, der in Abhängigkeit von den kognitiven Prozessen der Informationsverarbeitung auch unterschiedliche Selektionslogiken bei der Urteilsbildung und Wahl der Antwort berücksichtigt.

Im Beispiel des sozial erwünschten Antwortverhaltens können Verzerrungen in unterschiedlicher Weise auftreten. Zum einen kann der SD-Bias das Resultat von *optimizing* sein. Demnach tritt der SD-Bias erst am Ende der Prozesskette bei der Wahl der Antwort (*response*) auf (Tourangeau et al. 2000). Dem zugrunde liegt die Annahme, dass Befragte alle vorgelagerten Prozessschritte vollständig und sorgfältig durchlaufen und sich ein „korrektes“ Urteil bilden, dann aber bei der Wahl der Antwort davon abweichen (Kaminska und Foulsham 2016). Es handelt sich also um eine bewusste Täuschung seitens der Befragten in Form eines sozial erwünschten Verhaltens. Demgegenüber kann sozial erwünschtes Antwortverhalten aber auch eine kognitiv sparsame Beantwortungsstrategie darstellen (*satisficing*). Sozial erwünschte Antwortkategorien sind für die Befragten meist relativ offensichtlich und ohne großen kognitiven Aufwand identifizierbar. Dies ermöglicht es den Befragten eine plausible Antwort zu geben, ohne überhaupt weitere Informationen abrufen (*retrieval*) und beurteilen (*judgment*) zu müssen (Holtgraves 2004). Der Ausgang beider Antwortstrategien wäre hier identisch: Befragte geben sozial erwünschte Antworten. Aber die Ursachen sowie das Zustandekommen des sozial erwünschten Antwortverhaltens unterscheiden sich grundlegend in Abhängigkeit des Modus der Informationsverarbeitung. Im Falle des *optimizing* wird im Prozess der Antwortgenerierung ein „korrektes“ Urteil zur Beantwortung der Frage gebildet, bei *satisficing* hingegen nicht. Wenn man nun die bekannten Bedingungsfaktoren für sozial erwünschtes Antwortverhalten in der Befragungssituation – d.h. Erwünschtheitssignale und Reaktion auf unerwünschtes Verhalten (vgl. Beispiel weiter oben) – vollständig beseitigen könnte, so würde *optimizing* valide Antworten liefern, da die Befragten keine Anreize mehr hätten von ihrem eigentlichen Urteil abzuweichen. Im Fall von *satisficing* liegt die Ursache für den SD-Bias hingegen nicht in den genannten Bedingungsfaktoren selbst, sondern in einer vorgelagerten Entscheidung mit der Zielsetzung die Frage mit möglichst geringem kognitiven Aufwand zu beantworten. Dementsprechend ist zu erwarten, dass bei ausbleibenden SD-Anreizen die Befragten auf andere Strategien ausweichen, um den Antwortprozess abzukürzen. Auch wenn sich dadurch sozial erwünschte Antworten vermeiden lassen, würde dies vermutlich dennoch nicht zu akkurateren Antworten führen, sondern lediglich zu anders gelagerten Verzerrungen.

Welchen Modus der Informationsverarbeitung Befragte im Rahmen des Antwortprozesses wählen, hängt von drei übergeordneten Faktoren ab (Krosnick 1991; Krosnick et al. 1996): (1) der Schwierigkeiten der Aufgabe (*task difficulty*), also dem kognitiven Aufwand, der erforderlich ist um die Frage korrekt zu beantworten, (2) die Fähigkeiten der Befragten (*ability*), also die kognitiven Kapazitäten, die zur Beantwortung der Frage zur Verfügung stehen, und (3) die Motivation den zur Beantwortung der Frage erforderlichen kognitiven Aufwand aufzubringen (*motivation*). Die Wahrscheinlichkeit vom optimierenden Antwortverhalten abzuweichen, steigt mit zunehmender Schwierigkeit der Aufgabe, mit geringen kognitiven Ressourcen und mit einer geringen Bereitschaft diese auch aufzubringen.

$$p(\textit{satisficing}) = \frac{\beta_1(\textit{task difficulty})}{\beta_2(\textit{ability}) * \beta_3(\textit{motivation})} \quad (2.6)$$

Wie aus der formalen Darstellung ersichtlich wird, können die drei Bestimmungsfaktoren mit unterschiedlichem Gewicht (β_1 , β_2 und β_3) in die Modusselektion einfließen. Zudem hängen die Einflussgrößen multiplikativ zusammen, woraus sich weitere Prognosen für das Zusammenspiel der Faktoren ergeben (vgl. Krosnick 1991, S. 225). Bei geringer Aufgabenschwierigkeit sind die Fähigkeiten und Motivation der Befragten weniger ausschlaggebend. Ist eine Frage vergleichsweise einfach zu beantworten, dann ist es eher unwahrscheinlich, dass Befragte den Antwortprozess abkürzen; die kognitive Leistungsfähigkeit und/oder die Motivation müsste in diesem Fall schon enorm gering ausgeprägt sein. Umgekehrt müssen bei schwierig zu beantwortenden Fragen die Befragten besonders leistungsfähig und motiviert sein, damit diese trotz hohem Anspruchsniveau optimal beantwortet werden. Weiterhin kann eine geringe kognitive Leistungsfähigkeit nur bedingt durch eine erhöhte Motivation ausgeglichen werden, *vice versa*. Die funktionalen Zusammenhänge zwischen den drei Faktoren deuten auch auf einen allgemeinen Trend hin: Ein optimierendes Antwortverhalten ist nur dann zu erwarten, wenn alle drei Bedingungen erfüllt sind, d.h. bei angemessener Schwierigkeit der Aufgabe und entsprechender kognitiver Leistungsfähigkeit sowie Motivation seitens der Befragten. Ist nur eine dieser Voraussetzungen nicht gegeben, kann dies bereits hinreichend sein, um *satisficing* hervorzurufen (vgl. Krosnick et al. 1996, S. 33).⁸ In diesen Zusammenhängen wird erneut die zentrale Bedeutung des schon eingangs betonten Zusammenspiels von Faktoren des Befragungsinstrumentes (*task difficulty*) und den Charakteristika der Befragten (*ability* und *motivation*) deutlich.

Satisficing kann eine grundlegende Heuristik zur Beantwortung von Umfragen darstellen und sich durch die komplette Befragung ziehen. Es kann aber ebenso situativ auftreten, zum Beispiel, wenn sich nach einiger Zeit kognitive Erschöpfungserscheinungen einstellen und die Motivation für *optimizing* sinkt, oder auch nur bei der Beantwortung einzelner Fragen, etwa bei komplexen oder für die Befragten nicht interessanten

⁸Die Beschreibung der Bestimmungsfaktoren für die Modusselektion lässt vermuten, dass Befragte grundsätzlich die Antwortstrategie des *optimizing* verfolgen und nur unter bestimmten Bedingungen von diesem Standard abweichen. Es ließe sich aber ebenso gut annehmen, dass Befragte von Grund auf eher versuchen, den kognitiven Aufwand möglichst gering zu halten und nur unter bestimmten Bedingungen ein optimierendes Antwortverhalten an den Tag legen. Im Modell entspräche dies dann der exakt entgegengesetzten Prognose aus Formel 2.6:

$$p(\textit{optimizing}) = \frac{\beta_2(\textit{ability}) * \beta_3(\textit{motivation})}{\beta_1(\textit{task difficulty})} \quad (2.7)$$

Ungeachtet des Status quo, bleiben die Bestimmungsfaktoren und deren Zusammenspiel für einen Wechsel auf den jeweils alternativen Modus der Informationsverarbeitung dieselben. Welcher Modus als Status quo angenommen wird, ist letztendlich eine paradigmatische Frage. Für die Prognose des Antwortverhaltens und potentieller Antwortverzerrungen ist dies analytisch gleichbedeutend, weswegen an der Stelle auf eine weitere Auseinandersetzung mit dieser Frage verzichtet wird.

Themen (vgl. Roßmann 2017, S. 57ff.; Vannette und Krosnick 2014). Um Anreize für *satisficing* in der Umfragesituation möglichst gering zu halten, können Forscher:innen eigentlich nur an zwei Stellschrauben drehen (vgl. auch Vannette und Krosnick 2014): einerseits, die kognitiven Anforderungen (*task difficulty*) möglichst gering halten, etwa durch klare Anweisungen und Hilfestellungen bei komplexen Fragen und/oder durch eine insgesamt kürzere Befragungsdauer; andererseits, die Motivation der Befragten (*motivation*) möglichst aufrecht erhalten, beispielsweise durch eine angemessene Incentivierung der Teilnahme. Die Fähigkeiten der Befragten (*ability*) sind hingegen eine nicht kontrollierbare Größe, die gerade bei geringer Ausprägung besondere Anforderungen an das Befragungssetting stellt.

Vor dem Hintergrund der konzeptionellen Unterscheidung zwischen *optimizing* und *satisficing*, sowie der Identifikation der ausschlaggebenden Bedingungen, welche die Modusselektion bestimmen, lässt sich die Wirkungsweise von verzerrenden Einflussfaktoren auf das Antwortverhalten weiter präzisieren. Entscheidend ist hierbei, dass neben den Bestimmungsfaktoren der verzerrenden Einflüsse selbst, noch weitere und teils davon unabhängige Charakteristika der Befragungssituation sowie der Befragten relevant werden, welche den Modus der Informationsverarbeitung bestimmen und dadurch auch, ob und wie stark sich verzerrende Einflussfaktoren letztendlich auf den Antwortprozess auswirken. So können verzerrende Einflüsse, und besonders auch deren Bestimmungsgrößen, in Abhängigkeit der Aufgabenschwierigkeit, der kognitiven Leistungsfähigkeit und/oder der Motivation der Befragten sowohl verstärkt als auch abgeschwächt ins Gewicht fallen (vgl. auch Krosnick 1991, S. 225f.). Beispielsweise konnte gezeigt werden, dass klassische Antwortmuster in Befragungen, wie systematische Reaktionen auf die Antwortreihenfolge, Zustimmungstendenzen oder auch die Wahl einer Ausweichkategorie (z.B. „keine Angabe“ oder „weiß nicht“), bei Befragten mit geringeren kognitiven Fähigkeiten bzw. niedrigerem Bildungshintergrund sowie bei schwach ausgeprägter Motivation häufiger auftreten (Krosnick et al. 1996; Krosnick et al. 2002). Die Moderation verzerrender Einflüsse durch kognitive und motivationale Aspekte lässt darauf schließen, dass Befragte hierbei auf einfache Antwortmuster zurückgreifen, um den kognitiven Aufwand bei der Beantwortung der Fragen zu reduzieren; folglich also als Konsequenz des *satisficing*. Über eine Spezifikation solcher Interaktionszusammenhänge werden eine präzisere Identifikation der zugrundeliegenden Mechanismen und damit letztlich auch eine differenziertere Prognose über das Ausmaß und die Richtung von Antwortverzerrungen möglich.

Anknüpfend an das vorherige Beispiel des SD-Bias wäre zu erwarten, dass sich die drei Bestimmungsfaktoren des sozial erwünschten Antwortverhaltens – sozial erwünschter Befragungsinhalt, Bedürfnis nach sozialer Anerkennung und Reaktion auf Verhalten – in Abhängigkeit des Modus der Informationsverarbeitung unterschiedlich stark auf den Antwortprozess auswirken. So müssen im Fall von *optimizing* alle drei Voraussetzungen erfüllt sein, um sozial erwünschtes Antwortverhalten hervorzurufen (Stocké 2004). Der SD-Bias fällt dabei umso stärker aus, je eindeutiger Signale und Anreize vorliegen. Mit abnehmendem Grad der Sorgfalt und Vollständigkeit im

Antwortprozess (*satisficing*) sollten hingegen die Einflüsse des Anerkennungsbedürfnisses und der erwarteten Reaktion auf (un)erwünschtes Verhalten zunehmend geringer werden. Die grundlegende Motivation für ein sozial erwünschtes Antwortverhalten besteht hier eben nicht (zwangsläufig) in einem stark verankerten Anerkennungsbedürfnis und dem Ziel dieses zu befriedigen, wie dies beim *optimizing* der Fall ist, sondern darin, die Frage kognitiv sparsam zu beantworten. Somit kann sozial erwünschtes Antwortverhalten selbst bei fehlendem Bedürfnis nach sozialer Anerkennung ausgelöst werden. Entscheidend ist hierbei in erster Linie, ob die Fragestellung ein klares Erwünschtheitssignal aussendet oder nicht. Demzufolge sollte der Einfluss der Erwünschtheitswahrnehmung gegenüber den anderen beiden Determinanten relativ an Bedeutung gewinnen. Allerdings könnten sich hingegen auch die anderen beiden Einflussfaktoren für sozial erwünschtes Antwortverhalten bei *satisficing* verstärken, beispielsweise, wenn Befragte mit einem stark ausgeprägten Anerkennungsbedürfnis gerade dann auf Erwünschtheitsanreize reagieren, wenn sie ohnehin schon erschöpft von der Beantwortung des Fragebogens sind. Es lassen sich folglich unterschiedliche Prognosen für differentielle Wirkungsweisen aufstellen, welche zur analytischen Abgrenzung im jeweiligen Anwendungsfall theoretisch noch konkreter spezifiziert werden müssten. Die angeführten Beispiele sollen an der Stelle lediglich den zentralen Unterschied zwischen den Modi der Informationsverarbeitung verdeutlichen, welcher letztendlich darin besteht, dass die Effekte der Bestimmungsfaktoren für sozial erwünschte Antworten, und damit auch die allgemeinen Tendenzen des SD-Bias, in Abhängigkeit der begünstigenden Faktoren für *satisficing* variieren können. Dementsprechend sollte im Fall von *satisficing* der SD-Bias gerade bei schwierig zu beantwortenden Fragen, geringer kognitiver Leistungsfähigkeit und/oder geringer Motivation der Befragten besonders stark ausfallen. Im Gegensatz dazu wären bei optimierendem Antwortverhalten keinerlei Interaktionen mit diesen drei Faktoren zu erwarten. Über eine Bestimmung derartiger Interaktionen (z.B. zwischen kognitiven Fähigkeiten und Erwünschtheitswahrnehmung) ließe sich dann prüfen, ob ein vorliegender SD-Bias eher auf *optimizing* oder *satisficing* zurückzuführen ist, also ob Befragte bewusst täuschen oder lediglich versuchen den kognitiven Aufwand in der Befragung zu reduzieren. Weiterhin können die Interaktionen Hinweise darauf liefern, unter welchen Bedingungen bestimmte Einflussfaktoren stärker bzw. schwächer ins Gewicht fallen. Dies ermöglicht es die Prozesse, die dem Antwortverhalten zugrunde liegen, näher zu bestimmen.

Das geschilderte Modell zu den grundlegenden Abläufen und Bestimmungsfaktoren des Antwortverhaltens in Umfragen beansprucht für sich keineswegs den Anforderungen einer umfassenden Theorie zur Erklärung von Antwortverzerrungen gerecht zu werden. Vielmehr bietet das Modell einen allgemeinen theoretischen Rahmen, ausgehend von dem sich verschiedene Ansätze und Argumente, zu einem konkreten Anwendungsfall, systematisch integrieren und präzisieren lassen. Die einzelnen theoretischen Argumente werden dabei mit den zuvor beschriebenen Prozessschritten und Faktoren in Verbindung gebracht, wodurch gezielte Prognosen zur Wirkungsweise sowie zum

Zusammenspiel der relevanten Einflussgrößen abgeleitet werden können. Die konkreten Zusammenhänge und Prognosen müssen stets im jeweils vorliegenden Anwendungsfall entsprechend der theoretischen Argumentation noch näher spezifiziert werden. Im nachfolgenden Abschnitt wird das Modell herangezogen, um systematische Antwortverzerrungen bei Selbsteinschätzungen von Sprachkompetenzen zu erklären.

2.2.2 Antwortverzerrungen bei Selbsteinschätzungen von Sprachkompetenzen

Wie eingangs betont, wird die Verlässlichkeit von Selbsteinschätzungen zur Bestimmung der tatsächlichen Kompetenz *allgemein* stark in Frage gestellt. Dies geht aus einer inzwischen gut abgesicherten Befundlage hervor, wonach für zahlreiche Kompetenzbereiche gezeigt werden konnte, dass subjektive Einschätzungen in vielen Fällen doch vergleichsweise stark von der objektiv getesteten Leistung abweichen (vgl. Dunning 2005; Falchikov und Boud 1989; Kruger und Dunning 1999; Mabe und West 1982; Zell und Krizan 2014). Mit Blick auf die Ausprägung der Abweichungen zeichnen sich zwei klare Tendenzen ab: Zum einen schätzen Individuen ihre eigenen Fähigkeiten meist eher zu optimistisch ein, und zum anderen schätzen leistungsstarke Individuen ihre Kompetenzen akkurater ein als Individuen, die in dem entsprechenden Bereich weniger gute Leistungen aufweisen (Dunning et al. 2003; Kruger und Dunning 1999; Ehrlinger et al. 2008). Folglich ist das vorherrschende Muster der Abweichungen zwischen subjektiver und objektiver Leistungsbeurteilung durch eine systematische Überschätzung von Individuen mit tatsächlich eher geringerem Leistungsniveau geprägt.

Aufgrund dieser Tendenz ist davon auszugehen, dass Selbsteinschätzungen eine geringere Varianz in den Kompetenzen abbilden als objektiv tatsächlich vorliegt, was mit gewissen Unschärfen in empirischen Analysen verbunden sein kann. Aufgrund der geringeren Varianz in den selbsteingeschätzten Kompetenzen ist allein technisch schon eine Unterschätzung relevanter Zusammenhänge zu erwarten. Darüber hinaus können aber auch noch weitere, schwerwiegendere Verzerrungen der Ergebnisse auftreten, sofern dem Antwortverhalten selbst systematische Tendenzen zugrunde liegen. Das Ausmaß derartiger Verzerrungen lässt sich nur beurteilen, wenn die Fehlerquellen im Antwortprozess, die zur Überschätzung der eigenen Kompetenz führen, bekannt sind.

Welche Bedingungen und Faktoren zur Entstehung des speziellen Antwortmusters bei Selbsteinschätzungen beitragen, ist allerdings weniger gut erforscht. In der Literatur werden einige Faktoren, welche sich auf die Präzision bzw. die Verzerrung von Selbsteinschätzungen auswirken könnten, diskutiert und vereinzelt auch empirisch geprüft. Beispielsweise wird angenommen, dass generelle kognitive und metakognitive Fähigkeiten (z.B. Dunning 2005; Kruger und Dunning 1999; Truxillo et al. 2008; Widhiarso 2014), Aspekte des Selbstbilds sowie des Selbstkonzepts (z.B. Critcher und Dunning 2009; Ehrlinger und Dunning 2003), oder auch Referenzrahmen und soziale Vergleichsprozesse (z.B. Marsh et al. 2008; Thijs et al. 2010; Trautwein et al. 2009) bei der Bestimmung der eigenen Kompetenz bedeutend sind. Die diesen Faktoren zugrundeliegenden Argumente stehen dabei in der bisherigen Diskussion weitestgehend für

sich. Im Nachfolgenden wird daher eine systematische Zusammenführung der einzelnen theoretischen Überlegungen, die darüber hinaus auch das Zusammenspiel der Einflussgrößen berücksichtigt, vorgenommen. Dies erfolgt am Untersuchungsgegenstand der Validität subjektiv eingeschätzter L2-Kompetenzen von Jugendlichen mit Migrationshintergrund, welcher im Vordergrund der vorliegenden Arbeit steht. Es kann allerdings angenommen werden, dass die grundlegenden Zusammenhänge in vergleichbarer Weise auch auf Selbsteinschätzungen in anderen Kompetenzbereichen sowie auf andere Populationen – etwa ebenso auf Personen ohne Migrationshintergrund – zutreffen.

Die wohl verbreitetste Form der subjektiven Messung von Sprachkompetenzen, wie sie häufig auch in groß angelegten Bevölkerungsumfragen eingesetzt wird, beruht auf Selbstberichten in Bezug auf Fragen wie beispielsweise „*Wie gut beherrschen Sie die deutsche Sprache?*“⁹. In der Regel werden hierbei die Kompetenzen differenziert für die übergreifenden Dimensionen Schreiben, Lesen, Sprechen und Verstehen, oder eine Auswahl daraus abgefragt. Die Befragten können dann ihre Einschätzung anhand einer mehrstufigen Antwortskala einordnen, etwa von „*überhaupt nicht*“, über „*sehr schlecht*“, „*eher schlecht*“, „*eher gut*“, bis „*sehr gut*“. Die konkreten Frageformulierungen sowie die Anzahl und Ausprägung der Antwortoptionen weichen, je nach Studie, geringfügig von dem angeführten Beispiel ab; die grundlegende Ausrichtung der Fragen und folglich auch die damit einhergehenden Anforderungen und Schwierigkeiten beim Beantworten der Fragen sind jedoch vergleichbar.

Anders als Fragen zu faktischen Informationen, wie dem Alter oder dem höchsten erreichten Bildungsabschluss, welche vergleichsweise einfach zu beantworten sein sollten, stellt der Antwortprozess bei der subjektiven Einschätzung der eigenen Sprachkompetenz deutlich höhere Anforderungen an die Befragten. Selbsteinschätzungen von Kompetenzen werden generell als eine höchst anspruchsvolle Aufgabe angesehen (vgl. Carter und Dunning 2008), bei der die Befragten beträchtliche kognitive Anstrengungen aufbringen müssen, um angemessene Informationen abzurufen und diese zu einem verlässlichen Urteil zu verarbeiten. Eine vollständige und sorgfältige Bearbeitung des Antwortprozesses ist demnach sehr voraussetzungsreich, besonders hinsichtlich der kognitiv fordernden Prozessschritte *retrieval* und *judgment*, und deswegen auch anfällig für Störgrößen, gerade wenn die Befragten die benötigten kognitiven Kapazitäten nicht aufbringen können oder wollen. Die hohe Aufgabenschwierigkeit in Kombination mit einer gegebenenfalls geringeren Fähigkeit und/oder Motivation der Befragten begünstigen das Auftreten von *satisficing*, was sowohl die Reliabilität der Selbsteinschätzungen reduzieren dürfte als auch Tendenzen systematischer Verzerrungen im Antwortprozess auslösen bzw. verstärken könnte.

⁹Das hier angeführte Beispiel der Frageformulierung sowie der dazugehörigen Antwortkategorien ist an das Instrument, wie es in der ersten bzw. zweiten Welle der Startkohorte 4 des Nationalen Bildungspanels (NEPS; Blossfeld und Roßbach 2019) eingesetzt wurde, angelehnt. Der entsprechende Fragebogen kann unter nachfolgendem Link abgerufen werden (vgl. S. 46, Item 86): https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/1-1-0/SC4_1-1-0_Q_w1-2_de.pdf (zuletzt aufgerufen am 17.11.2024).

2.2.2.1 Bedingungen für fehlerhafte Selbsteinschätzungen im Antwortprozess

Im Folgenden werden zentrale Bedingungen diskutiert, die die Reliabilität und die Validität von selbsteingeschätzten Sprachkompetenzen beeinflussen können. Die Argumentation folgt dabei strukturell den vier Prozessschritten des Antwortverhaltens.

Bedingungen für das Verständnis der Fragen (*comprehension*): Die Aufgabenschwierigkeit wird in erster Linie durch die Fragestellung im Befragungsinstrument, sowie die damit verbundenen Anforderungen an die Interpretationsleistung der Befragten, vorgegeben. Abhängig von dem erfragten Inhalt bzw. davon, wie die Befragten die Frage interpretieren, werden hier Anhaltspunkte vordefiniert, welche Indikatoren zur Bestimmung der erfragten Kompetenz aufschlussreich sein könnten und im weiteren Verlauf des Antwortprozesses abgerufen und verarbeitet werden sollten (vgl. Tourangeau und Rasinski 1988). Im Fall von *klassischen* Fragen zur Selbsteinschätzung (z.B. „Wie gut können Sie auf Deutsch lesen?“) wird die – ohnehin schon komplexe – Aufgabe der Bestimmung der eigenen Sprachkompetenz zusätzlich dadurch erschwert, dass die relativ abstrakt gehaltene Formulierung keine eindeutigen Anhaltspunkte oder Vorgaben darüber liefert, welche Informationen zur Beantwortung herangezogen werden sollten. So lässt die Fragestellung beispielsweise offen, ob diese sich auf die Bewältigung sprachlicher Anforderungen im Alltag oder im schulischen Kontext bezieht, ob es um spezifische Kompetenzen geht oder um eine globale Bewertung, ob die Frage rein subjektiv auf die Zufriedenheit mit der eigenen Leistung abzielt oder doch eher auf eine Einordnung der Kompetenzen im Vergleich zu anderen Personen bzw. einem objektiven Kriterium. Welche Indikatoren zur Selbsteinschätzung im Rahmen des weiteren Antwortprozesses verarbeitet werden, hängt in diesem Fall maßgeblich von der Interpretation der Befragten ab.¹⁰

Die Fragen können folglich sehr unterschiedlich aufgefasst und entsprechend uneinheitlich beantwortet werden. Vor diesem Hintergrund liegt die Vermutung nahe, dass Fragen zu spezifischeren Fähigkeiten und Situationen – zum Beispiel, ob man einen persönlichen Brief verstehen kann (Brantmeier et al. 2012), oder wie gut man die zentralen Punkte eines einfachen Zeitungsartikels verstehen kann (Harder et al. 2018) – leichter und akkurater beantwortet werden können als globale Kompetenzeinschätzungen. Dies bestätigt sich auch mit Blick auf die empirische Befundlage, wonach Selbsteinschätzungen spezifischer Kompetenzen in diversen Bereichen (z.B. Ackerman

¹⁰Bei Personen, welche die Administrationssprache der Befragung nicht (ausreichend) beherrschen, kommt erschwerend hinzu, dass die Frage nicht nur inhaltlich, sondern auch *sprachlich* nicht (richtig) verstanden werden könnte. Diese zusätzliche Problemstellung ergibt sich aber eher bei der Befragung von Neuzugewanderten, die zum Zeitpunkt der Befragung eventuell noch so gut wie kein Deutsch sprechen. Für die Schüler:innen mit Migrationshintergrund, die größtenteils der zweiten Zuwanderungsgeneration angehören und um die es in der vorliegenden Arbeit hauptsächlich geht, sollte dies hingegen weniger ausschlaggebend sein. Zudem stellen Sprachprobleme eine generelle Herausforderungen in Befragungen von Migrant:innen und ihren Nachkommen dar, und keine spezielle Problematik von Fragen zur Selbsteinschätzung von Sprachkompetenzen, weswegen dieser Aspekt nachfolgend nicht weiter aufgegriffen wird.

et al. 2002; Mabe und West 1982), darunter auch Sprachkompetenzen (Brantmeier und Vanderplank 2008; Brantmeier et al. 2012; Li und Zhang 2021; Ross 1998), oftmals stärker mit objektiven Testleistungen korrelieren als globale Einschätzungsmaße.

Bedingungen für den Abruf von Informationen (*retrieval*): Es ist davon auszugehen, dass die meisten Jugendlichen keine vorgefertigte und umfassende Beurteilung ihrer Sprachkompetenzen ausgebildet haben, auf die sie im Antwortprozess zurückgreifen und direkt als Antwort weitergeben könnten. Vielmehr werden die Jugendlichen wohl erst in der Befragungssituation mit der Aufgabe, die eigenen Sprachkompetenzen zu bestimmen, konfrontiert und müssen ein entsprechendes Urteil im Rahmen des Antwortprozesses neu generieren, oder eine bereits existierende Vorstellung zumindest vor dem Hintergrund der Fragestellung mit zusätzlichen Informationen abgleichen und gegebenenfalls aktualisieren (vgl. Sudman et al. 1996, S. 69ff.; Tourangeau et al. 2000, S. 12f.). Hierzu müssen die Jugendlichen im nächsten Schritt des Antwortprozesses (*retrieval*) möglichst alle relevanten Informationen aus dem Gedächtnis abrufen, die zur (akkuraten) Bestimmung ihrer Sprachkompetenz aufschlussreich sein könnten. Beispielsweise könnten sie rekapitulieren, wie gut sie in bestimmten Situationen in der Vergangenheit abgeschnitten haben, etwa bei Prüfungen in der Schule, oder ob sie im alltäglichen Umgang auf irgendwelche Probleme gestoßen sind und bestimmte sprachliche Anforderungen nicht (zufriedenstellend) meistern konnten. Wenn sich Jugendliche zum Beispiel daran erinnern, dass sie in alltäglichen Gesprächen oftmals Schwierigkeiten hatten das richtige Wort zu finden oder bestimmte Begriffe des Gegenübers nicht kannten, deutet dies auf bestehende sprachliche Grenzen und mögliche Defizite hin, die in die Beurteilung der eigenen Kompetenz einfließen können.

Der Abruf relevanter Informationen aus dem Gedächtnis setzt gewisse kognitive Anstrengungen voraus. Jugendliche mit hohen kognitiven Fähigkeiten sollte es in diesem Zusammenhang vergleichsweise leicht fallen den Anforderungen gerecht zu werden, wohingegen diejenigen mit gering ausgeprägten kognitiven Fähigkeiten – und/oder fehlender Motivation diese Leistung aufzubringen – häufiger Schwierigkeiten haben sollten adäquate Informationen abzurufen und aufgrund dessen auch eher dazu tendieren den Antwortprozess abzukürzen (Krosnick 1991; Krosnick et al. 2002). Die Jugendlichen rufen dann vielleicht nicht alle relevanten Informationen aus dem Gedächtnis ab, sondern beschränken sich bei der Suche lediglich auf die besonders leicht zugänglichen Informationen (vgl. Sudman et al. 1996, S. 71; Tourangeau und Rasinski 1988). Zum Beispiel könnten die Jugendlichen nur an unmittelbar naheliegende Situationen denken, etwa an die Note in der letzten Klassenarbeit im Deutschunterricht oder an die Kommunikation im engsten Freundeskreis, aber nicht an sprachliche Probleme, denen sie vielleicht im Laufe des gesamten Schuljahres oder im sonstigen Alltag begegnet sind. Dabei besteht ein erhöhtes Risiko, weitere relevante und potentiell gegenläufige Hinweise zu übersehen, wodurch ein selektives Bild der eigenen Sprachkompetenz entstehen könnte.

In der Tendenz folgt daraus, dass Jugendliche mit geringeren kognitiven Fähigkeiten ihre Sprachkompetenzen weniger akkurat einschätzen. Die eher oberflächliche Bearbei-

tung des Antwortprozesses bedeutet dabei jedoch zunächst nicht, dass die Jugendlichen ihre Kompetenzen systematisch über- oder unterschätzen, sondern verstärkt lediglich das Auftreten und das Ausmaß von Zufallsschwankungen im Antwortverhalten (in beide Richtungen), wodurch die Reliabilität sinkt. Allerdings variieren diese Zufallsschwankungen nicht bloß nach den kognitiven Fähigkeiten, sondern – da kognitive Fähigkeiten eine zentrale Bestimmungsgröße der Sprachkompetenz darstellen (z.B. Esser 2006; Kristen 2019) – gleichzeitig auch in Abhängigkeit der Sprachkompetenzen selbst. Da das Kompetenzspektrum nach unten (und oben) natürlich begrenzt ist bzw. durch die Antwortskala bestimmte Grenzen vorgegeben werden, resultiert auch aus Zufallsschwankung die Tendenz einer Überschätzung im unteren Leistungsbereich, allein schon, weil Abweichungen nach unten nur sehr bedingt möglich sind. So ist es beispielsweise technisch nicht möglich, die eigene Kompetenz zu unterschätzen, wenn man eine Sprache überhaupt nicht spricht. In diesem Fall führen selbst rein *zufällige* Abweichungen vom wahren Wert zwangsläufig zu einer Überschätzung der Sprachkompetenz.

Darüber hinaus wird aber auch angenommen, dass Individuen mit geringeren (kognitiven bzw. sprachlichen) Fähigkeiten häufig ein systematisch überzogenes Bild ihrer tatsächlichen Kompetenz abrufen (z.B. Kruger und Dunning 1999, Dunning et al. 2003). Die Argumentation in diesem Zusammenhang stellt die spezifische Fähigkeit der Metakognition, als zusätzlichen kognitiven Mechanismus, in den Vordergrund. Nach Flavell (1979) beschreibt Metakognition die Fähigkeit, die eigene Leistung akkurat reflektieren und beurteilen zu können. Dabei geht es darum, Lücken und Fehler in der eigenen Leistung zu erkennen; zum Beispiel im Nachgang zu einer Prüfung die richtig und falsch gelösten Antworten korrekt zu identifizieren (Dunning et al. 2003). Individuen mit geringen metakognitiven Fähigkeiten verfügen nicht über die erforderlichen kognitiven Strategien, um die relevanten Informationen (effizient) abzurufen und zu verarbeiten, und daraus letztendlich eine akkurate Einschätzung der eigenen Kompetenz zu generieren (Kruger und Dunning 1999). Dunning et al. (2003) sprechen in diesem Zusammenhang auch von einer doppelten Benachteiligung („double curse“), weil die kognitiven Anforderungen, die eigene Leistung akkurat einschätzen zu können, sehr ähnlich zu denen sind in einer bestimmten Aufgabe gut abzuschneiden: Um beispielsweise beurteilen zu können, ob man eine Aufgabe in einem Test richtig gelöst hat, muss man die richtige Lösung kennen. Mit zunehmenden Kompetenzen und Erfahrungen entwickeln Individuen – als Teil des Lernprozesses – gleichzeitig auch (meta-)kognitive Strategien, die eigene Leistung und besonders auch die Schwächen in diesem Bereich akkurater beurteilen zu können. Jugendliche mit eher geringen Sprachkompetenzen bzw. geringen (meta-)kognitiven Fähigkeiten übersehen daher oftmals bestehende Defizite und tendieren dazu, ihre Kompetenzen überzubewerten. Dieser Prozess läuft weitestgehend *unbewusst* ab, da die Individuen schlichtweg nicht in der Lage sind, Fehler und Schwächen in der eigenen Leistung zu erkennen (Dunning et al. 2003, Ehrlinger et al. 2008). Mit Blick auf den Antwortprozess äußert sich diese Tendenz darin, dass beim Abruf relevanter Informationen aus dem Gedächtnis gerade die negativen Erfahrungen ausgelassen werden, was zur Folge hat, dass positive Erfahrungen überproportional ins Gewicht fallen und die eigenen Sprachkompetenzen

entsprechend *systematisch* überschätzt werden.

Bedingungen für die Urteilsbildung (*judgment*): Im nachfolgenden Schritt des Antwortprozesses (*judgment*) müssen die zuvor abgerufenen Informationen hinsichtlich ihrer Aussagekraft und Passung zur Beantwortung der Frage bewertet werden und ausgehend davon zu einem Urteil zusammengefasst werden. Hier geht es unter anderem darum, zu beurteilen, ob die abgerufenen Erfahrungen und Situationen repräsentativ für das „normale“ Verhalten sind, oder ob diese eventuell außergewöhnliche Abweichungen davon darstellen (Sudman et al. 1996, S. 57). Beispielsweise können auch sonst sehr gute Schüler:innen in einer Prüfung schlecht abschneiden, etwa weil sie krankheitsbedingt darauf nur unzureichend vorbereitet sind. Manche Jugendliche sind vielleicht auch in einer mündlichen Prüfung oder einer für sie besonders wichtigen Alltagssituation so angespannt, dass sie Schwierigkeiten haben sich zu artikulieren. Derartige Ausnahmefälle gilt es als solche zu identifizieren und entsprechend mit geringerer Gewichtung in die Gesamtbeurteilung der eigenen Sprachkompetenz einfließen zu lassen. Im Zuge der Beurteilung müssen Befragte ein geeignetes Kriterium als Vergleichsmaßstab definieren, mit dem sie die abgerufenen Informationen abgleichen und bewerten können (Sudman et al. 1996, S. 56). Die Wahl des Kriteriums kann die Bewertung der eigenen Leistung maßgeblich beeinflussen. Zieht man beispielsweise die Sprachkompetenzen von Personen aus dem näheren Umfeld (z.B. Freund:innen, Mitschüler:innen, Eltern) als Bezugspunkt heran, führt dies vermutlich zu einer gänzlich anderen Beurteilung als ein Vergleich mit den sprachlichen Fähigkeiten von Nachrichtensprecher:innen. Ähnlich verhält es sich im schriftlichen Bereich, etwa mit Blick auf Beiträge in sozialen Netzwerken im Gegensatz zu Artikeln in Tageszeitungen. In Abhängigkeit des angesetzten Bewertungsmaßstabs kann die Urteilsbildung folglich sehr unterschiedlich ausfallen. Das verdeutlicht, dass die (akkurate) Einschätzung der eigenen Sprachkompetenz nicht allein davon bestimmt wird, welche Informationen im Antwortprozess zunächst abgerufen werden, sondern ebenso davon, wie diese Informationen nachfolgend im Zuge der Urteilsbildung weiter verarbeitet und bewertet werden.

Eine relevante Größe, auf die Jugendliche bei der Beurteilung ihrer Sprachkompetenzen Bezug nehmen könnten, hängt mit Aspekten ihres Selbstkonzepts, also der Wahrnehmung der eigenen Person (Shavelson et al. 1976, S. 411), zusammen. Obwohl die Jugendlichen meist keine vorgefertigte Beurteilung ihrer konkreten Sprachkompetenzen besitzen, sollten sie zumindest über eine grobe, vermutlich stark verallgemeinerte und idealisierte Vorstellung über sich selbst bzw. über ihre Leistungsfähigkeit verfügen. Vor diesem Hintergrund ist davon auszugehen, dass Jugendliche sich bei der Beurteilung ihrer Sprachkompetenzen – zumindest teilweise – auch auf diese, in ihrem Selbstkonzept verankerten Informationen stützen (Ehrlinger 2008; Ehrlinger und Dunning 2003). Dabei können Wahrnehmungen der eigenen Leistungsfähigkeit in spezifischen Bereichen gleichermaßen bedeutend werden wie globale Aspekte des Selbstkonzepts, etwa das allgemeine Selbstbewusstsein (Critcher und Dunning 2009). Das Selbstkonzept kann im Antwortprozess als Richtwert zur Leistungsbeurteilung dienen, mit dem die abgerufenen Informationen verglichen und entsprechend bewertet

werden. Sofern das Selbstkonzept die tatsächlichen Kompetenzen adäquat widerspiegelt, sind dabei auch akkurate Leistungsbeurteilungen zu erwarten. Allerdings haben Individuen häufig überzogene Vorstellungen ihrer allgemeinen Leistungsfähigkeit (Carter und Dunning 2008), was eine Verzerrung der Bewertung und letztendlich auch der Selbsteinschätzung der Sprachkompetenz zur Folge haben kann. Diese Tendenz hängt mit der intrinsisch verankerten Zielsetzung zusammen, das Selbstkonzept möglichst aufrechtzuerhalten. Die Argumentation geht auf die Grundzüge der kognitiven Dissonanztheorie (Festinger 1957) sowie verwandter Ansätze (z.B. Aronson 1968; Steele 1988; Swann 1983) zurück, wonach Individuen ihre Denk- und Handlungsmuster gezielt auf Ausgänge ausrichten, die im Einklang mit ihrem Selbstkonzept stehen. Im Gegensatz dazu rufen Informationen, die dem eigenen Selbstkonzept widersprechen, kognitive Dissonanzen und Unbehagen hervor (z.B. Freund und Kasten 2012; Harmon-Jones und Harmon-Jones 2007). Um derartige psychische Kosten zu vermeiden und ein positives Selbstbild aufrecht zu erhalten, tendieren Individuen dazu, ihre Leistung so zu beurteilen, dass die Bewertung zu ihrem bestehenden Selbstkonzept passt (Wells und Sweeney 1986). Dabei werden gegenläufige Hinweise bei der Urteilsbildung geringer gewichtet oder komplett ignoriert, selbst wenn diese eine verlässlichere Beurteilung der tatsächlichen Leistung ermöglichen würden (Farh und Dobbins 1989; Kwang und Swann 2010; Swann und Read 1981). Demzufolge neigen Jugendliche mit einem stark ausgeprägten Selbstkonzept dazu, positive Ereignisse überzubewerten und letztlich ihre Sprachkompetenzen zu überschätzen. Beispielsweise stufen Jugendliche mit einem hohen allgemeinen Selbstbewusstsein ihre Sprachkompetenzen tendenziell höher ein als weniger selbstbewusste Jugendliche mit vergleichbaren Kompetenzen (Wells und Sweeney 1986). Die Tendenz der Überschätzung sollte dabei mit zunehmender Diskrepanz zwischen Selbstkonzept und tatsächlicher Leistung umso stärker ins Gewicht fallen; also besonders bei Jugendlichen mit sehr positivem Selbstkonzept, aber eher geringen Sprachkompetenzen.

Weiterhin wird angenommen, dass Individuen sich mit anderen Individuen aus ihrem sozialen Umfeld vergleichen, um zu einer Beurteilung der eigenen Leistung zu gelangen (vgl. Marsh 1986). Soziale Referenzrahmen stellen einen naheliegenden Bezugspunkt dar, den Jugendliche im Antwortprozess – aber auch ganz allgemein – als Bewertungskriterium zur Beurteilung ihrer eigenen Sprachkompetenz heranziehen. Dabei können die Jugendlichen einen Vergleich mit generalisierten Bezugsgruppen (z.B. die durchschnittliche Sprachkompetenz in der Klasse oder im Freundeskreis) sowie mit ausgewählten Individuen (z.B. dem oder der besten Freund:in) anstellen (Marsh et al. 2008). Je nachdem welcher Referenzrahmen zur Beurteilung herangezogen wird und wie dieser sprachlich zusammengesetzt ist, setzen bestimmte Bezugsgruppen inter-individuell uneinheitliche Bewertungsmaßstäbe. Demnach können Jugendliche mit vergleichbaren Sprachkompetenzen ihre Leistung in Abhängigkeit des Referenzrahmens unterschiedlich bewerten. Ein bekanntes Phänomen in diesem Zusammenhang ist der sogenannte Fischteicheffekt („big-fish-little-pond-effect“; vgl. Marsh 1987; Marsh und Parker 1984). Demzufolge tendieren Individuen dazu ihre Leistung besser zu bewerten, wenn sie sich mit vergleichsweise leistungsschwachen Bezugsgruppen vergleichen. Umgekehrt fällt die Bewertung der eigenen Kompetenz mit zunehmendem Leistungsniveau des Referenz-

rahmens aber auch umso negativer aus. Diese Zusammenhänge wurden bisher in erster Linie hinsichtlich der Einflüsse der Schulklassenzusammensetzung auf die Ausprägung des akademischen Selbstkonzepts untersucht (z.B. Shajek et al. 2006; Thijs et al. 2010; Trautwein et al. 2009). Die zugrundeliegende Argumentation lässt sich in vergleichbarer Weise aber auch auf den vorliegenden Anwendungsfall übertragen. Neben möglichen indirekten Einflüssen, die über Aspekte des Selbstkonzepts vermittelt werden, können darüber hinaus auch direkte Referenzrahmeneffekte auf die subjektive Leistungsbeurteilung erwartet werden. Demnach schätzen Jugendliche ihre Sprachkompetenzen tendenziell optimistischer ein, wenn ihr soziales Umfeld eher unvoreilhaft sprachliche Voraussetzungen bietet (z.B. bei geringem Sprachniveau der Mitschüler:innen). Dies sollte gerade auf Jugendliche mit selbst eher geringen Sprachkompetenzen zutreffen, weswegen diese stärker zu Überschätzungen neigen als Jugendliche in eher anspruchsvollen Kontexten, die ihre Leistung deutlich kritischer betrachten.

Bedingungen des Informationsabrufs (*retrieval*) und der Urteilsbildung (*judgment*) in Abhängigkeit des Modus der Informationsverarbeitung: Bewertungsmaßstäbe, die auf dem Selbstkonzept oder dem Referenzrahmen basieren, sind für die Befragten vergleichsweise einfach zugänglich und identifizierbar, weswegen es naheliegend ist, dass Jugendliche diese Aspekte bei der Beurteilung ihrer Sprachkompetenz heranziehen. Dabei können Einflüsse des Selbstkonzepts und soziale Vergleichsprozesse die Urteilsbildung systematisch verzerren. Der Grad dieser Verzerrungen kann allerdings in Abhängigkeit von weiteren Faktoren variieren. Der Bezug auf Aspekte des Selbstkonzepts sowie auf Referenzrahmen dient der Bewertung der zuvor aus dem Gedächtnis abgerufenen Informationen. Dieser Abgleich erfordert ein hohes Maß an kognitiver Anstrengung, gerade wenn sehr viele und/oder widersprüchliche Indikatoren bewertet werden müssen. Dabei gilt es nicht nur die zuvor abgerufenen Informationen im Lichte der angesetzten Kriterien zu bewerten, sondern gleichzeitig auch die Angemessenheit der jeweiligen Bewertungskriterien selbst zu reflektieren und diese gegebenenfalls anzupassen. Jugendliche, die diese kognitive Leistung im Rahmen des Antwortprozesses nicht oder nur eingeschränkt erbringen können bzw. wollen, sind den (verzerrenden) Einflüssen des Selbstkonzepts und sozialer Vergleichsprozesse in besonderer Weise ausgesetzt, weil die Bezugsgrößen in diesem Fall vergleichsweise unreflektiert und ungefiltert in die Beurteilung einfließen. Diese Tendenz verstärkt sich noch weiter, wenn Jugendliche den Antwortprozess bereits im vorhergehenden Schritt (*retrieval*) abgekürzt haben. Befragte, die nur unzureichende Indikatoren abgerufen haben, auf deren Basis keine verlässliche Beurteilung möglich ist, greifen alternativ häufig auf verallgemeinerte oder besonders naheliegende Informationen – wie das Selbstkonzept und Referenzrahmen – zurück, um sich ein Urteil zu bilden (Sudman et al. 1996, S. 57). Wurden beispielsweise gar keine (verwertbaren) Informationen abgerufen, könnte dies zur Folge haben, dass die Einschätzung der eigenen Sprachkompetenzen ausschließlich auf dem allgemeinen Selbstkonzept und dem Vergleich mit anderen beruht, wobei die damit einhergehenden Verzerrungstendenzen dann auch in maximalem Ausmaß durchschlagen. Je aufschlussreicher und eindeutiger die zuvor abgerufenen Infor-

mationen hingegen sind, desto schwächer sollten Einflüsse des Selbstkonzepts und des Referenzrahmens ausfallen.

Der Rückbezug auf Aspekte des Selbstkonzepts und des Referenzrahmens im Zuge der Urteilsbildung kann folglich Ausdruck einer kognitiv sparsamen Antwortstrategie (*satisficing*) sein; etwa, weil die Jugendlichen die erforderlichen kognitiven Anstrengungen in der Befragungssituation nicht aufbringen können oder wollen. Dieses Verhalten folgt entweder als Konsequenz einer bereits unvollständigen Bearbeitung des vorhergehenden Schrittes im Antwortprozess (*retrieval*), oder tritt – erst oder auch zusätzlich dazu – im Schritt der Urteilsbildung (*judgment*) auf, mit dem Ziel die zuvor abgerufenen Informationen mit möglichst geringem Aufwand zu verarbeiten. In beiden Fällen sollte die Tendenz für *satisficing*, und damit auch das Ausmaß potentieller Verzerrungen durch das Selbstkonzept und soziale Vergleichsprozesse, bei Jugendlichen mit eher geringen kognitiven Fähigkeiten besonders ausgeprägt sein. Mit zunehmenden kognitiven Fähigkeiten – und der Motivation diese in der Befragungssituation auch aufzubringen – fällt es den Befragten einerseits leichter zunächst aufschlussreiche Informationen abzurufen und diese andererseits dann nachfolgend auch reflektiert zu beurteilen. Aus diesen theoretischen Überlegungen lässt sich demnach eine (negative) Interaktion zwischen den kognitiven Fähigkeiten der Jugendlichen und den verzerrenden Einflüssen des Selbstkonzepts sowie des Referenzrahmens auf die Einschätzung der Sprachkompetenz prognostizieren.

Bedingungen für die Wahl der Antwort (*response*): Wenn sich die Jugendlichen schließlich ein Urteil über ihre Sprachkompetenzen gebildet haben, müssen sie dieses im letzten Schritt (*response*) noch einer der vorgegebenen Antwortkategorien zuweisen. Im Idealfall wählen sie dabei diejenige Antwortoption aus, die ihr Urteil am besten widerspiegelt. Allerdings können Befragte auch an dieser Stelle den Antwortprozess abkürzen, indem sie beispielsweise nicht alle Antwortoptionen in Betracht ziehen und gegeneinander abwägen, sondern direkt die erste akzeptabel erscheinende Option auswählen (Krosnick 1991). Befragte geben dabei zwar weniger akkurate Antworten, die grundlegende Richtung wird aber richtig wiedergegeben, d.h. komplett falsch sind die Antworten nicht. Beispielsweise könnten Befragte mit hohen Sprachkompetenzen ohne weiteres Nachdenken ihre Fähigkeiten als „gut“ oder „sehr gut“ einstufen. Sie werden hingegen nicht angeben, dass ihre Sprachkompetenzen „schlecht“ oder „sehr schlecht“ seien. Durch derartige Tendenzen sind eher keine systematischen Verzerrungen zu erwarten, wohl aber erhöhte Zufallsschwankungen, welche im Aggregat die Reliabilität der Messung verringern.

Ähnlich verhält es sich mit Charakteristika der eingesetzten Antwortskala, die an dieser Stelle von besonderer Bedeutung sind und ebenfalls gewisse Unschärfen in den Antwortprozess einführen können. So ist aus der Umfrageforschung bekannt, dass Aspekte wie die Reihenfolge der Antwortoptionen (z.B. Holbrook et al. 2007; Schwarz et al. 1992) sowie deren numerische Repräsentation (z.B. Schwarz et al. 1991) das Antwortverhalten maßgeblich beeinflussen können. Einflüsse der Antwortskala lassen sich darauf zurückführen, dass Befragte die vorgegebenen Antwortoptionen als Refe-

renzrahmen zur Verortung ihrer Einschätzung nutzen (vgl. Schwarz und Bienias 1990). Gleichzeitig können die Antwortoptionen teilweise aber auch als Orientierungshilfe für die vorhergehenden Prozessschritte dienen (vgl. Sudman et al. 1996, S. 64, 74) und Anhaltspunkte dafür liefern, wie die Frage interpretiert werden soll, welche Informationen relevant sein könnten und welche Bewertungsmaßstäbe anzusetzen sind. Entscheidend ist dabei, ob die Antwortoptionen – und vor allem die beiden Endpunkte der Skala – trennscharfe und für alle Befragte gleichbedeutende Kategorien vorgeben, die möglichst präzise dem jeweiligen Urteil entsprechen.

Mit Blick auf die bei *klassischen* Fragen zur Selbsteinschätzung von Sprachkompetenzen meist eingesetzte Antwortskala von „*überhaupt nicht*“ oder „*sehr schlecht*“ bis „*sehr gut*“ lassen sich diesbezüglich einige Problemstellen identifizieren. Die Kategorie „*überhaupt nicht*“ setzt einen eindeutigen Ankerpunkt am unteren Ende der Antwortskala. Befragte, die die Sprache überhaupt nicht sprechen, finden sich in dieser Kategorie eindeutig wieder und können die Frage präzise beantworten. Im Gegensatz dazu geben die weiteren Antwortkategorien „*sehr/eher schlecht*“ und „*eher/sehr gut*“ relativ unscharfe Kompetenzbereiche vor, die keine allgemein definierten Richtwerte repräsentieren und entsprechend von den Befragten sehr unterschiedlich interpretiert werden können, was wiederum das Auftreten von Zufallsschwankungen im Antwortverhalten begünstigen dürfte (vgl. Dustmann und Van Soest 2002). Wenn die Antwortskala, wie in diesem Fall, keine klare Referenz bietet, könnte dies die Befragten dazu veranlassen, verstärkt auch andere Indikatoren als Referenzrahmen – wie das Selbstkonzept oder soziale Vergleiche – heranzuziehen. Die Einordnung der eigenen Sprachkompetenz in das vorgegebene Antwortschema wird zudem dadurch erschwert, dass oftmals ein entsprechender Gegenpol zur Kategorie „*überhaupt nicht*“ am oberen Ende des Leistungsspektrums fehlt. Beispielsweise ließe sich durch eine zusätzliche Kategorie „*auf muttersprachlichem Niveau*“ das Leistungsspektrum zumindest an beiden Endpunkten relativ klar abstecken und somit ein einheitlicher Referenzrahmen vorgeben, den die Befragten bereits bei der Beurteilung der eigenen Kompetenzen berücksichtigen können.

Bedeutung fähigkeitsbezogener und motivationaler Bedingungen: Vor dem Hintergrund der Diskussion um die Ursachen und Mechanismen von verzerrten Selbsteinschätzungen stellt sich grundlegend die Frage, ob Individuen ihre (Sprach-)Kompetenzen schlichtweg nicht akkurat einschätzen *können* oder ob sie dies nicht *wollen* bzw. bewusst täuschen. So können systematische Verzerrungen auftreten, weil Jugendliche den Antwortprozess – und vor allem die kognitiv anspruchsvollen Schritte *retrieval* und *judgment*, entweder aufgrund nicht ausreichender kognitiver Kapazitäten oder wegen fehlender Motivation diese aufzubringen, abkürzen. Unabhängig davon können Jugendliche im letzten Schritt (*response*) aber auch bewusst von ihrem eigentlichen Urteil abweichen und falsche Angaben machen, beispielsweise weil sie gute Sprachkompetenzen als sozial erwünscht wahrnehmen und sich entsprechend darstellen möchten (vgl. Heilenman 1990).

Aufschluss darüber, ob die systematische Überschätzung der eigenen Leistung eher auf fähigkeitsbezogene oder motivationale Faktoren zurückzuführen ist, liefern Hinwei-

se aus experimentellen Studien, in denen die beiden Einflussgrößen isoliert untersucht wurden. Ehrlinger et al. (2008) haben hierzu eine Reihe von Experimenten durchgeführt, deren zentraler Versuchsaufbau jeweils darin bestand, die motivationalen Anreize für fehlerhaftes Antwortverhalten durch eine manipulierte Incentivierung der Probanden auszuschalten. Zum Beispiel wurde in einem Experiment den teilnehmenden Student:innen 100 Dollar geboten, wenn sie die Anzahl der korrekt gelösten Aufgaben in einem zuvor absolvierten Test (bestehend aus 20 Aufgaben) exakt einschätzen, und 35 Dollar, wenn ihre Einschätzung zumindest in einem gewissen Toleranzbereich liegt. Trotz der hohen Anreize, eine möglichst genaue Beurteilung abzugeben, fielen die Selbsteinschätzungen unter der Experimentalbedingung nicht wesentlich akkurater aus als in der Kontrollgruppe, der keinerlei Belohnung in Aussicht gestellt wurde. Gleichzeitig war die Genauigkeit der Einschätzung in beiden Gruppen stark an die Leistung der Probanden gekoppelt, d.h. das allgemein vorherrschende Muster systematischer Überschätzungen besonders im unteren Leistungsbereich konnte auch hier repliziert werden. Vergleichbare Befunde zeigten sich auch hinsichtlich unterschiedlicher Untersuchungspopulationen, Kompetenzbereiche und Incentives (vgl. Ehrlinger et al. 2008).

Auch wenn sich, ausgehend von diesen Hinweisen, Einflüsse einer geringen Motivation sowie der Anreize sozialer Erwünschtheit auf das Antwortverhalten zwar nicht grundsätzlich ausschließen lassen, so legen die Befunde dennoch nahe, dass die Fähigkeiten von besonderer Bedeutung sind. Das Auftreten von *satisficing* wird in diesem Zusammenhang folglich eher davon bestimmt, ob Befragte die zur Beantwortung der Frage erforderlichen kognitiven Leistung überhaupt erbringen können, und weniger davon, ob sie ausreichend motiviert sind diese Leistung aufzubringen. Dies betont einmal mehr den zentralen Stellenwert der *kognitiven Fähigkeiten* für den Ablauf des Antwortprozesses allgemein, aber auch speziell mit Blick auf den vorliegenden Anwendungsfall der Antwortverzerrungen bei Selbsteinschätzungen von Sprachkompetenzen. So ist anzunehmen, dass die mit der Einschätzung der eigenen Sprachkompetenz verbundene Aufgabenstellung sich für die Jugendlichen oftmals zu anspruchsvoll gestaltet, weswegen diese im Antwortprozess auf einfachere Heuristiken ausweichen und dabei verstärkt Indikatoren heranziehen, die ein selektives und teils systematisch überzogenes Bild ihrer Leistung entstehen lassen.

2.2.2.2 Einschränkungen der theoretischen Modellierung im Anwendungsfall

Die hier dargelegte theoretische Argumentation zu den Entstehungsbedingungen und -prozessen von verzerrten Selbsteinschätzungen der eigenen Sprachkompetenzen stützt auf der Prämisse, dass die verzerrenden Einflüsse im Rahmen des Antwortprozesses auftreten. Demgegenüber könnten Jugendliche sich aber auch bereits vorab ein Urteil über ihre Sprachkompetenzen gebildet haben, welches sie in der Befragung lediglich weitergeben. In diesem Fall würde sich der Antwortprozess vergleichsweise einfach gestalten, da nur das vorhandene Urteil abgerufen und einer Antwortoption zugewiesen werden muss; eine zusätzliche Beurteilung erübrigt sich (vgl. Sudman et al. 1996, S. 58). Die

sonst kognitiv sehr voraussetzungsreichen und fehleranfälligen Schritte *retrieval* und *judgment* entfallen hier weitestgehend. Demzufolge könnten auch die zuvor diskutierten Fehlerquellen und Mechanismen, die gerade im Zuge dieser beiden Prozessschritte bedeutsam sein sollten, das Zustandekommen von verzerrten Selbsteinschätzungen *nicht* erklären, was die Anwendung des dargelegten theoretischen Modells insgesamt in Frage stellen würde.

Zwar wird es als eher unwahrscheinlich erachtet, dass ein bereits existierendes Urteil auch exakt zur konkreten Fragestellung und den angebotenen Antwortkategorien passt und ohne jegliche kognitive Verarbeitung im Antwortprozess einfach direkt weitergegeben werden kann (vgl. Sudman et al. 1996, S. 70; Tourangeau et al. 2000, S. 13), aber vollständig ausräumen lässt sich dies nicht; vor allem auch nicht, dass die eigentliche Verzerrung eventuell schon bei der ursprünglichen Urteilsbildung im Vorfeld aufgetreten ist und sich im Rahmen des Antwortprozesses lediglich fortsetzt. Dieser Aspekt muss bei der Diskussion um die Reichweite des theoretischen Modells einschränkend berücksichtigt werden. Gleichzeitig weisen die theoretischen Modelle zu den kognitiven Abläufen des Antwortprozesses aber enge Parallelen zu allgemeinen theoretischen Ansätzen der Informationsverarbeitung und Urteilsbildung aus der Kognitions- und Sozialpsychologie auf – nicht zuletzt, weil die Modelle auch daraus hervorgingen (vgl. Schwarz 2007; Sudman et al. 1996, S. 55). In diesem Zusammenhang ist davon auszugehen, dass die grundlegenden Prozesse einer allgemeinen Logik folgen, die in vergleichbarer Weise auch auf eine vorgelagerte Urteilsbildung übertragbar ist. Der Abruf relevanter und aussagekräftiger Informationen aus dem Gedächtnis, die anschließende Integration und Bewertung dieser sowie der Grad an Sorgfalt und Vollständigkeit, mit dem die Informationen verarbeitet werden, sollten zentraler Bestandteil jeder Urteilsbildung sein. Demzufolge sind auch bei einer vorgelagerten Urteilsbildung (verzerrende) Einflüsse des Selbstkonzepts und des Referenzrahmens, deren Ausmaß in Abhängigkeit der kognitiven Fähigkeiten variiert, zu erwarten. Vor diesem Hintergrund sollte es für die Erklärung des Anwendungsfalls systematischer Verzerrungen von selbsteingeschätzten Sprachkompetenzen gleichbedeutend sein, ob die Jugendlichen ihr Urteil im Rahmen des Antwortprozesses bilden oder dieses bereits im Vorfeld generiert haben; die Prognosen zum Einfluss und Zusammenspiel der zentralen Bedingungen sind identisch. Im speziellen Fall der Befragung kommen lediglich weitere situationsabhängige Bedingungen hinzu, die zusätzliche Verzerrungen auslösen und/oder die allgemeinen Tendenzen begünstigen können, wie beispielsweise Einflüsse, die mit der Frageformulierung, der Antwortskala oder dem Verhalten der Interviewer:innen zusammenhängen.

Ein weiterer Einwand betrifft die theoretische Annahme, die Verzerrungen bei der Einschätzung der eigenen Sprachkompetenz seien – unter anderem – *ursächlich* auf allgemeinere und ebenfalls häufig überzogene Formen des Selbstkonzepts zurückzuführen (z.B. Ehrlinger und Dunning 2003). Beide Aspekte, Selbsteinschätzung und Selbstkonzept, könnten aber auch Ausdruck einer gemeinsamen zugrundeliegenden Dimension sein, die von weiteren, in der Argumentation bisher unberücksichtigten Faktoren bestimmt wird. Die Ursache der Verzerrung könnte folglich tiefer liegen und die Selbsteinschätzung sowie allgemeinere Formen des Selbstbilds in gleicher Weise beeinflussen; beispielsweise durch eine grundlegende Neigung, sich selbst möglichst positiv

zu sehen und darzustellen (z.B. Kwang und Swann 2010). Vor diesem Hintergrund stellt sich zurecht die Frage, ob der theoretische Rückbezug auf Einflüsse allgemeinerer Aspekte des Selbstkonzepts das eigentliche Problem lediglich verlagert und damit gar nicht zu dessen Klärung beiträgt. Schließlich lässt die dargelegte Argumentation in diesem Zusammenhang offen, wie diese allgemeinen und potentiell ebenfalls von der tatsächlichen Leistungsfähigkeit abweichenden Selbstbilder wiederum zustande kommen.

Das Selbstkonzept über die Leistungsfähigkeit allgemein, sowie gerade hinsichtlich der Fähigkeiten in spezifischen Bereichen, ist in den Grundzügen tatsächlich sehr ähnlich zur Selbsteinschätzung der eigenen Kompetenzen gestaltet. Beide Konstrukte basieren auf der subjektiven Wahrnehmung der eigenen Eigenschaften und Erfahrungen, wobei das Selbstkonzept ein vergleichsweise eher umfassendes und über die Zeit hinweg stabiles Bild der Leistungsfähigkeit repräsentiert (z.B. Bong und Skaalvik 2003; Freund und Kasten 2012) und, demgegenüber, die Selbsteinschätzung der eigenen Sprachkompetenz kontextspezifisch ausgerichtet ist und auch zeitlich stärkere Schwankungen aufweisen kann (z.B. Dustmann und Van Soest 2001). In beiden Fällen findet eine Beurteilung der eigenen Leistung statt – wenn auch in teils unterschiedlicher Detailtiefe, was die Vermutung nahelegt, dass hierzu vergleichbare Informationen abgerufen und verarbeitet werden, und das Urteil entsprechend auch den gleichen verzerrenden Einflüssen ausgesetzt ist. Hinsichtlich *leistungsbezogener* Aspekte des Selbstbilds, wie beispielsweise dem akademischen Selbstkonzept der Jugendlichen, ist durchaus plausibel anzunehmen, dass derartige Parallelen zur Selbsteinschätzung bestehen, wohingegen für grundlegendere und weitestgehend leistungsunabhängige Dimensionen, wie etwa dem generellen Selbstbewusstsein, eher genuine Einflüsse auf die Selbsteinschätzung zu erwarten sind.

Folgt man der Argumentationslinie einer gemeinsamen grundlegenden Dimension, spiegelt sich die Verzerrung analog in beiden Konstrukten wider. Demnach wäre zu erwarten, dass bei Personen mit positivem Selbstkonzept unweigerlich auch die Selbsteinschätzungen gleichermaßen vorteilhaft ausfallen, unabhängig von der tatsächlichen Leistung. Zum Beispiel sollten Jugendliche, die davon überzeugt sind, grundsätzlich leicht neue Sprachen zu lernen, auch ihre Deutschkompetenzen entsprechend optimistisch einschätzen. Im Gegensatz dazu prognostiziert die theoretische Argumentation der vorliegenden Arbeit eine konditionale Wirkungsweise der Aspekte des Selbstbilds auf die Selbsteinschätzung der Sprachkompetenz, in Abhängigkeit der kognitiven Fähigkeiten. Einflüsse des Selbstbilds sollten gerade bei gering ausgeprägten kognitiven Fähigkeiten bedeutend sein. Eine Berücksichtigung dieser konditionalen Beziehung in den Analysen kann Aufschluss darüber geben, ob zwischen Selbstbild und Selbsteinschätzung eine kausale Beziehung besteht. Obwohl eine tiefergehende gemeinsame Ursache der Verzerrung damit nicht ausgeschlossen werden kann, würde die Interaktion zwischen Selbstbild und kognitiven Fähigkeiten darauf hindeuten, dass verzerrende Einflüsse auf die Selbsteinschätzung zumindest teilweise durch Aspekte des Selbstbilds vermittelt werden. Somit kann das theoretische Modell zwar nicht die Ursache potentieller Verzerrungen des Selbstbilds umfassend erklären, dafür aber, unter welchen Bedingungen sich (verzerrende) Einflüsse des Selbstbilds auf die Selbsteinschätzung der Sprachkompetenzen in besonderem Maße auswirken. Das Modell ermöglicht damit

eine Prognose, wann die Einschätzung eher leistungsbasiert erfolgt, und wann verstärkt auf Basis verallgemeinerter und teils leistungsunabhängiger Einstellungen zur eigenen Person geurteilt wird.

2.2.2.3 Forschungsstand zu den Bedingungen und Auswirkungen verzerrierter Selbsteinschätzungen

Ausgehend von den theoretischen Überlegungen sowie den bisherigen Erkenntnissen zur Verlässlichkeit von Selbsteinschätzungen allgemein ist zu erwarten, dass selbsteingeschätzte Sprachkompetenzen von Jugendlichen mit Migrationshintergrund eine geringere Reliabilität als objektive Sprachmessungen aufweisen und teilweise auch systematisch davon abweichen. Zum einen können im Zuge des Antwortprozesses erhöhte Zufallsschwankungen auftreten – zum Beispiel aufgrund von Ungenauigkeiten in der Frageformulierung oder der Antwortskala, welche die Reliabilität der Messung reduzieren. Zum anderen können verzerrende Einflüsse der kognitiven Fähigkeiten, des Selbstkonzepts und des Referenzrahmens die Validität einschränken. Die damit einhergehende Tendenz zur (systematischen) Überschätzung der eigenen Leistung – gerade derjenigen mit eher geringen Fähigkeiten – führt insgesamt zu einer verringerten Varianz der selbsteingeschätzten Sprachkompetenzen gegenüber einer objektiven Messung. Vor diesem Hintergrund birgt eine unbedachte Verwendung der subjektiv gemessenen Sprachkompetenzen in empirischen Analysen das Risiko, falsche Rückschlüsse aus den Ergebnissen zu ziehen – zumindest, wenn die Selbsteinschätzung als Indikator für die tatsächliche Sprachkompetenz der Zielpersonen herangezogen wird. Dabei dürften Zusammenhänge mit weiteren interessierenden Variablen, schon allein technisch aufgrund der insgesamt verringerten Varianz der Selbsteinschätzungen, tendenziell unterschätzt werden. Hinzu kommen Auswirkungen systematisch verzerrierter Einflüsse, die sich sowohl in Form einer Unter- als auch Überschätzung von Zusammenhängen bemerkbar machen können. Aus den theoretischen Annahmen lässt sich beispielsweise ableiten, dass Zusammenhänge, die auf den kognitiven Fähigkeiten beruhen, in der Tendenz unterschätzt werden, und Beziehungen, die mit dem Selbstkonzept und dem Referenzrahmen in Verbindung stehen, eher überschätzt werden. Die Fehler, die mit der Verwendung verzerrierter Selbsteinschätzungen in statistischen Analysen potentiell auftreten können, reichen dabei von mehr oder weniger geringfügigen Abweichungen der Effektstärken nach oben und unten bis hin zu einer Verschleierung zentraler Zusammenhänge oder der Angabe von Scheinkorrelationen.

Wie schwerwiegend derartige Fehler letztendlich auch empirisch ausfallen, lässt sich aus dem bisherigen Forschungsstand nur sehr vage abschätzen. Die meisten Studien, die sich der Verlässlichkeit von subjektiv eingeschätzten Sprachkompetenzen widmen, konzentrieren sich auf eine Untersuchung der Korrelation zwischen Selbsteinschätzung und objektiven Kriterien. Aus dem doch sehr umfangreichen Forschungsstand dazu zeichnen sich allerdings keine eindeutigen Tendenzen ab. Einige Beiträge finden kaum Überschneidungen zwischen subjektiv und objektiv erhobenen Sprachkompetenzen (z.B. Brantmeier 2006; Lin et al. 2001; Peirce et al. 1993; Rauch et al. 2010), andere hingegen finden sehr hohe Korrelationen zwischen beiden Messungen an (z.B. Bachman

und Palmer 1982; LeBlanc und Painchaud 1985; Peyer und Kaiser 2010; Wilson 1999). Eine vergleichbare Einschätzung geht aus zwei Meta-Analysen hervor, die über zahlreiche Studien hinweg im Durchschnitt eine moderate Korrelation von 0,47 (Li und Zhang 2021) und 0,63 (Ross 1998) zwischen subjektiven und objektiven Sprachkompetenzen finden, wobei auch hier die große Varianz in den Befunden der herangezogenen Einzelbeiträgen auffällt (vgl. Li und Zhang 2021, S. 198). Die systematische Beurteilung der Befundlage wird zudem dadurch erschwert, dass sich die Untersuchungsanlagen der bisherigen Studien mit Blick auf die zugrundeliegende Analysepopulation sowie die herangezogenen Instrumente zur Selbsteinschätzung und objektiven Kriterien teils sehr unterschiedlich gestalten, was die Vergleichbarkeit der einzelnen Beiträge erheblich einschränkt.

Vor dem Hintergrund der unterschiedlichen Ausrichtung der Studien bleibt zudem offen, inwiefern sich aus dem bisherigen Forschungsstand überhaupt belastbare Rückschlüsse für den speziellen Anwendungsfall der Validität selbsteingeschätzter Sprachkompetenzen von Jugendlichen mit Migrationshintergrund ziehen lassen. Zwei Aspekte gilt es in diesem Zusammenhang einschränkend zu berücksichtigen: die Vergleichbarkeit der Analysepopulationen sowie die Vergleichbarkeit der Instrumente zur Selbsteinschätzung.

Die überwiegende Mehrheit der Forschungsbeiträge basiert auf relativ kleinen, selektiven Stichproben von (ausländischen) Studierenden im Rahmen des Fremdsprachenunterrichts. Studien speziell zu Jugendlichen mit Migrationshintergrund, oder auch allgemein zu Personen mit Zuwanderungserfahrung, gibt es hingegen nur wenige (z.B. Charette und Meng 1994; Finnie und Meng 2005; Janssen-van Dielen 1989; Rauch et al. 2010). Zwar sollten die allgemeinen Prozesse des Antwortverhaltens für alle Gruppen und Anwendungsbereiche gleichermaßen gelten, aber es ist anzunehmen, dass sich die Randbedingungen, sowohl des Zweitpracherwerbs selbst als auch der subjektiven Einschätzung dazu, zwischen den Untersuchungspopulationen doch vergleichsweise stark unterscheiden, was ein unterschiedliches Maß an Verlässlichkeit der Selbsteinschätzung mit sich bringen kann. Beispielsweise erhalten Teilnehmer:innen in Fremdsprachenkursen kontinuierlich Feedback im Rahmen des Unterrichts oder in Form einer Leistungsbewertungen bei Prüfungen. Zudem sind die Kurse meist auf das Erreichen eines bestimmten sprachlichen Niveaus ausgerichtet (z.B. die Kompetenzstufen A1 bis C2 nach dem gemeinsamen europäischen Referenzrahmen für Sprachen „CEFR“). Den Kursteilnehmer:innen stehen demnach einige Kriterien zur Verfügung, die einen relativ klaren Referenzrahmen bieten und anhand derer sie ihre Sprachkompetenzen vergleichsweise einfach einordnen könnten, vor allem wenn die Selbsteinschätzung ebenfalls anhand der Skala des CEFR erfolgt (z.B. Alderson 2005). Im Gegensatz dazu erlernen Jugendliche mit Migrationshintergrund die Verkehrssprache in der Regel ungesteuert als Erst- oder Zweitsprache, meist von klein auf innerhalb der Familie und des sozialen Umfelds. Vergleichbare Rückmeldungen begleitend zum Lernprozess und klare Referenzrahmen fehlen dabei weitestgehend – oder dürften zumindest deutlich heterogener ausfallen. Demzufolge ist zu erwarten, dass die Selbsteinschätzungen von Jugendlichen mit Migrationshintergrund gegenüber Sprachkursteilnehmer:innen tendenziell weniger akkurat sind und in stärkerem Maße verzerrenden Einflüssen unterliegen.

Weiterhin wurden in den bisherigen Studien vorwiegend umfangreiche und auf spezifische sprachliche Fähigkeiten und Situationen ausgerichtete Instrumente zur Selbsteinschätzung untersucht. Von zentralem Interesse für den vorliegenden Anwendungsfall sind hingegen die *klassischen* Fragen zur Erfassung der globalen Sprachkompetenzen, wie sie im Kontext der Integrationsforschung sowie in groß angelegten Bevölkerungsumfragen hauptsächlich eingesetzt werden, und auf denen einige zentrale Erkenntnisse zu ethnischen Ungleichheiten beruhen. Die Befundlage zur Selbsteinschätzung spezifischer Kompetenzen lässt sich nur bedingt auf die Verlässlichkeit der globalen Abfrage übertragen, da spezifische Kompetenzen der Annahme nach für die Befragten leichter zu beurteilen sind und die Messung entsprechend eine höhere Übereinstimmung mit objektiven Kriterien zeigen sollte (vgl. Brantmeier und Vanderplank 2008; Brantmeier et al. 2012; Li und Zhang 2021; Ross 1998). Die Ergebnisse der Meta-Analyse von Li und Zhang (2021) liefern Hinweise, die diese Annahme stützen: In ihrer Analyse über 67 Studien hinweg untersuchten sie den moderierenden Einfluss verschiedener Charakteristika des Studiendesigns auf die Korrelation zwischen subjektiver und objektiver Sprachmessung. Als besonders bedeutsam erwiesen sich hierbei diverse Aspekte der Ausgestaltung der Fragen und Skalen zur Selbsteinschätzung, darunter auch die Ausrichtung der Instrumente (spezifisch versus allgemein). Studien, die danach fragten, wie gut die Befragten konkrete sprachliche Anforderungen im Alltag meistern können, erzielten im Schnitt signifikant höhere Korrelationen als Studien, die allgemeine Sprachkompetenzen erfassten (ebd.). Diese Zusammenhänge könnten auch – zumindest indirekt – als Indiz dafür gewertet werden, dass die kognitiven Fähigkeiten in Abhängigkeit der Aufgabenschwierigkeit von zentraler Bedeutung für die Entstehung verzerrter Selbsteinschätzungen sind. Nach dem dargelegten theoretischen Modell geben Fragen zu spezifischen Kompetenzen einen vergleichsweise klaren Bezugsrahmen vor, weswegen deren Beurteilung ein geringes Maß an kognitiver Anstrengung seitens der Befragten erfordert. Demzufolge sollten die kognitiven Fähigkeiten der Befragten hier für die Beurteilung weniger ausschlaggebend sein, was gleichzeitig auch das Einflusspotential leistungsunabhängiger und verzerrender Faktoren begrenzt. Definitive Aussagen zu den konkreten Abläufen lassen sich allerdings erst anhand einer direkten empirischen Identifikation der zugrundeliegenden Mechanismen treffen, welche noch aussteht.

Insgesamt lassen sich aus den bisherigen Studien zur Korrelation zwischen Selbsteinschätzung und objektiver Messung nur bedingt Hinweise entnehmen, die zur Beantwortung der vorliegenden Forschungsfrage nach der Validität gängiger Selbsteinschätzungsinstrumente zur Messung der *globalen* Sprachkompetenzen von *Jugendlichen mit Migrationshintergrund* beitragen. Allerdings ist zu erwarten, dass die Korrelation im Anwendungsfall tendenziell noch geringer ausfällt, wobei eine genauere quantitative Einschätzung des Niveaus ausgehend von den heterogenen Befunden als Referenzpunkt schwerfällt. Darüber hinaus ist die Aussagekraft der Untersuchung von Korrelationen generell begrenzt, da dies allein noch keine uneingeschränkte Beurteilung der Validität der subjektiven Messung zulässt. Eine besonders hohe Korrelation deutet zwar an, dass die Messung sowohl reliabel als auch valide ist. Im Gegenzug lässt eine geringe Korrelation aber nicht zwangsläufig auf eine unzureichende Validität schließen, sondern könnte ebenso ausschließlich mit einer schwachen Reliabilität der Messung einhergehen.

Letzteres hätte zur Folge, dass interessierende Zusammenhänge in Analysen zwar unterschätzt, aber in der Tendenz korrekt wiedergegeben werden. Weitere systematische Verzerrungen sind nur bei zusätzlich eingeschränkter Validität zu erwarten.

Die Validität selbsteingeschätzter Sprachkompetenzen sowie das Ausmaß potentieller Verzerrungen lassen sich gezielt über einen Abgleich der Zusammenhängestrukturen in multivariaten Analysen, unter Berücksichtigung variierender Indikatoren der Sprachkompetenz (Selbsteinschätzung versus objektive Messung) bei sonst identischen Modellen, untersuchen. Hierfür bietet es sich an, etablierte Analysemodelle mit vergleichsweise gut abgesicherten Befunden als Referenz heranzuziehen. Eine derartige Herangehensweise wurde bisher nur in wenigen Ausnahmefällen verfolgt, und dies ausschließlich im Kontext von Analysen zu Integrationsprozessen erwachsener Personen mit Migrationshintergrund. Beispielsweise fand Esser (1985) keine gravierenden Abweichungen zwischen Analysemodellen, in denen selbsteingeschätzte Sprachkompetenzen und Testergebnisse verwendet wurden – weder in Modellen zu den Bedingungen des Zweitspracherwerbs (Sprachkompetenz als abhängige Variable), noch in Modellen mit L2-Kompetenzen als erklärender Variable für unterschiedliche Bereiche der Sozialintegration (vgl. auch Esser 2006). Neuere, auf umfangreicheren Stichproben basierende Studien konnten demgegenüber zum Teil erhebliche Abweichungen in den Ergebnissen feststellen, wenn selbsteingeschätzte anstatt getesteter Sprachkompetenzen verwendet wurden. Dies zeigte sich sowohl hinsichtlich einzelner Bedingungen des Zweitspracherwerbs (Charette und Meng 1994; Finnie und Meng 2005) als auch bei der Bedeutung von L2-Kompetenzen für den Erwerbsstatus (Finnie und Meng 2005).

Die Frage nach der Validität der Selbsteinschätzung als Maß für die tatsächlichen Sprachkompetenzen lässt sich anhand der wenigen, methodisch belastbaren Forschungsbeiträge hierzu keineswegs abschließend klären. Dennoch legen die bisherigen Befunde nahe, dass die Verwendung von subjektiven Sprachkompetenzen – zumindest unter gewissen Umständen – zu systematischen Verzerrungen der Ergebnisse und damit zu fehlerhaften Rückschlüssen führen kann. Es bleibt aber offen, unter welchen konkreten Bedingungen und hinsichtlich welcher Anwendungsbereiche der Einsatz subjektiver Sprachkompetenzen als (besonders) problematisch einzustufen ist. Vor diesem Hintergrund wird die Notwendigkeit einer Identifikation der zentralen Mechanismen, die den verzerrten Selbsteinschätzungen zugrunde liegen, deutlich. Erst mit Kenntnis der ausschlaggebenden Bedingungen und Prozesse lässt sich eine systematische Beurteilung zum Ausmaß potentieller Fehler in Analysen, auch in Abhängigkeit des jeweiligen Anwendungsfalls, anstellen.

Hinsichtlich der relevanten Bedingungen und Bestimmungsfaktoren für akkurate bzw. verzerrte Selbsteinschätzungen gibt es in der bisherigen Forschung jedoch nur sehr vereinzelte Anhaltspunkte. Diese beruhen meist auch nicht auf Befunden zu selbsteingeschätzten Sprachkompetenzen und/oder Jugendlichen mit Migrationshintergrund, sondern gehen aus anderen Anwendungsbereichen hervor. Dennoch lassen sich daraus erste Hinweise zur empirischen Relevanz der theoretisch angenommenen Einflussgrößen ableiten. Beispielsweise zeigen Befunde, dass Befragte mit hohen kognitiven Fähigkeiten, in Befragungen allgemein sowie auch speziell auf Fragen zur Selbsteinschätzung bestimmter Eigenschaften, verlässlicher und konsistenter antworten als Personen mit

geringer ausgeprägten kognitiven Fähigkeiten (z.B. Krosnick et al. 1996; Truxillo et al. 2008; Widhiarso 2014). Weiterhin lassen sich Hinweise darauf finden, dass Personen mit positivem Selbstkonzept (bezüglich ihrer Leistungsfähigkeit) sowie mit hohem generellen Selbstbewusstsein dazu tendieren ihre Leistungen zu überschätzen (z.B. Ehrlinger und Dunning 2003, Wells und Sweeney 1986). Auch hinsichtlich der Bedeutung des Referenzrahmens liegen Befunde vor, die Einflüsse sozialer Vergleichsprozesse und Bezugsgruppeneffekte bei der Beurteilung der eigenen Leistung nahelegen (z.B. Shajek et al. 2006; Strickhouser und Zell 2015; Thijs et al. 2010; Trautwein et al. 2009). Es ist anzunehmen, dass Einflüsse der kognitiven Fähigkeiten, des Selbstbilds und des Referenzrahmens in vergleichbarer Weise auch bei der Selbsteinschätzung der Sprachkompetenzen von Jugendlichen mit Migrationshintergrund bedeutsam werden; eine (direkte) empirische Prüfung der Zusammenhänge steht allerdings noch aus. Des Weiteren wurden die Bestimmungsfaktoren bisher vorwiegend isoliert voneinander untersucht, weswegen sich eine Beurteilung der relativen Bedeutung der einzelnen Faktoren und – noch gravierender – deren Zusammenspiel als schwierig erweist. Eine Ausnahme bildet die Untersuchung von Schild (2024), in welcher simultan der Einfluss unterschiedlicher Faktoren auf die selbsteingeschätzten Deutschkompetenzen von geflüchteten Jugendlichen empirisch geprüft wurde. Die Ergebnisse legen nahe, dass die kognitiven Fähigkeiten und Aspekte, die mit dem Selbstbild in Zusammenhang stehen, bei der Aufklärung verzerrter Selbsteinschätzungen besonders relevant sind – zumindest für die spezielle Untersuchungspopulation. Tiefgehende Untersuchungen des Zusammenspiels der relevanten Einflussgrößen und damit verbundener moderierender und bedingter Wirkungsweisen blieben in der bisherigen Forschung weitestgehend unberücksichtigt. Die Untersuchung potentieller Interaktionen zwischen den Faktoren ist jedoch von entscheidender Bedeutung, um zu prüfen, ob Einflüsse des Selbstkonzepts und des Referenzrahmens bei geringen kognitiven Fähigkeiten stärker ins Gewicht fallen, und demzufolge – wie im theoretischen Modell angenommen – Ausdruck von *satisficing* im Antwortprozess sein könnten.

2.2.3 Beitrag der vorliegenden Arbeit

Im Rahmen der vorliegenden Dissertation wird die Problematik der Validität von selbsteingeschätzten Sprachkompetenzen in zwei Aufsätzen behandelt, welche die zuvor geschilderten Forschungslücken aufgreifen und versuchen zur Schließung dieser beizutragen. Dabei wird die Validität der subjektiven Sprachmessung, zum einen, ergänzend zur bisherigen Forschung, bei Jugendlichen mit Migrationshintergrund als zusätzlichen Anwendungsfall untersucht. Zum anderen bieten die beiden Beiträge eine über die meisten bisherigen Arbeiten hinausgehende Untersuchung zu den Bedingungen und Folgen von Antwortverzerrungen bei selbsteingeschätzten Sprachkompetenzen.

Im ersten Beitrag erfolgt ein systematischer Vergleich selbsteingeschätzter Sprachkompetenzen mit objektiven Testergebnissen (siehe Kapitel 6). Der Beitrag bietet zunächst eine detaillierte Zusammenstellung bisheriger Forschungsarbeiten zum Zusammenhang von subjektiv und objektiv gemessener Sprachkompetenz, wobei zentrale Charakteristika der Studiendesigns (z.B. Analysepopulation, Instrumente zur Selbstein-

schätzung, objektive Kriterien) herausgearbeitet werden, nach denen sich die Studien und deren Befunde einordnen lassen. Die eigene empirische Untersuchung geht dann über eine Bestimmung der Korrelation zwischen Selbsteinschätzung und Testleistung hinaus und prüft die Validität anhand exemplarischer Analysemodelle mit Anwendungsbezug. Der Abgleich der beiden Messungen erfolgt an vergleichsweise gut abgesicherten Modellen zu den Bedingungen und Folgen des Spracherwerbs. Damit knüpft der Beitrag an die wenigen Forschungsarbeiten, welche die Verlässlichkeit von selbsteingeschätzten Sprachkompetenzen auch multivariat – meist im Kontext der Arbeitsmarktintegration – untersuchen, an und erweitert den bisherigen Forschungsstand um Analysen zur Bildungssituation von Jugendlichen mit Migrationshintergrund. In den Analysen wird neben der Selbsteinschätzung der Kompetenz im Deutschen zusätzlich auch die Einschätzung der Kompetenzen in den Erstsprachen Russisch und Türkisch auf ihre Verlässlichkeit hin untersucht. Die Ergebnisse des Beitrags geben Aufschluss darüber, welche Fehler mit der Verwendung selbsteingeschätzter Sprachkompetenzen – zumindest hinsichtlich der untersuchten Fallbeispiele – verbunden sind.

Eine noch gezieltere und auch allgemeinere Einschätzung zum Ausmaß der Fehler, die Selbsteinschätzungen in Analysen verursachen, bietet der zweite Beitrag der Arbeit, welcher die Bestimmungsfaktoren von Antwortverzerrungen bei der Einschätzung der eigenen Sprachkompetenzen näher beleuchtet (siehe Kapitel 7). Im Zentrum steht dabei die Frage, inwiefern sich die allgemeinen kognitiven Fähigkeiten der Jugendlichen, ihr akademisches Selbstkonzept und generelles Selbstbewusstsein, sowie ihr sprachlicher Referenzrahmen im schulischen Umfeld (verzerrend) auf die Verlässlichkeit der selbsteingeschätzten Sprachkompetenzen im Deutschen auswirken. In der empirischen Analyse wird sowohl die Relevanz der einzelnen Faktoren allgemein geprüft als auch deren relative Bedeutung bestimmt. Darüber hinaus werden Interaktionen zwischen den Faktoren geschätzt, um zu testen, ob die (verzerrenden) Einflüsse des Selbstbilds und des Referenzrahmens in Abhängigkeit der kognitiven Fähigkeiten variieren. Diese differenziellen Wirkungsweisen bilden den Kern des theoretischen Modells, weswegen die empirische Prüfung der speziellen Zusammenhänge gleichzeitig auch als weitergehender Test des Modells insgesamt gewertet werden kann. Neben der theoretischen Bedeutung des Beitrags, lassen sich aus den Ergebnissen aber auch zentrale Hinweise für die Forschungspraxis ableiten. Ausgehend von den Ergebnissen zu den Bestimmungsfaktoren von Antwortverzerrungen lassen sich genauere Prognosen anstellen, unter welchen Bedingungen selbsteingeschätzte Sprachkompetenzen (besonders) fehlerbehaftet sind und in welchen Anwendungsfällen der Einsatz als weniger problematisch einzustufen ist. Vor diesem Hintergrund lassen sich dann auch, einerseits, Versuche zur Vermeidung bzw. Verringerung von Antwortverzerrungen in Umfragen, etwa durch optimierte Selbsteinschätzungsinstrumente, und, andererseits, die Aussagekraft bisheriger Befunde, die auf selbsteingeschätzten Sprachkompetenzen beruhen, neu bewerten.

3 Datengrundlage und methodisches Vorgehen

Für die empirischen Analysen wird auf Daten des Nationalen Bildungspanels (NEPS; Blossfeld und Roßbach 2019) zurückgegriffen. Das NEPS bietet einen umfangreichen Datenbestand zu Bildungsverläufen unterschiedlicher Altersgruppen in Deutschland. Im Speziellen werden Daten der Startkohorte 4 (Jugendliche ab Klassenstufe 9) verwendet, welche sich unter verschiedenen Gesichtspunkten in besonderer Weise zur Beantwortung der Fragestellungen dieser Dissertation eignen. Die Analysen zu den vier Aufsätzen bauen damit auf einer gemeinsamen Datengrundlage auf, unterscheiden sich allerdings teilweise hinsichtlich der Definition der Analysepopulation, der Operationalisierung der Variablen sowie des methodischen Vorgehens, da diese spezifischen Details auf den jeweiligen Anwendungsfall zugeschnitten werden. Eine ausführliche Beschreibung zum konkreten Vorgehen findet sich in den einzelnen Aufsätzen selbst, weshalb sich die Beschreibung an dieser Stelle auf die zentralen Hintergrundinformationen zum übergeordneten Studiendesign sowie den grundlegenden methodischen Herangehensweisen beschränkt und, darüber hinaus, das besondere Analysepotential der Daten mit Blick auf die Untersuchung der Forschungsfragen der vorliegenden Arbeit in den Vordergrund stellt.

Stichprobe und Analysepopulation

Die Zielpopulation der Startkohorte 4 sind Schüler:innen, die im Schuljahr 2010/2011 die neunte Klasse an einer allgemeinbildenden Schule besuchten. Die Zielpersonen wurden anhand eines stratifizierten zweistufigen Stichprobendesigns ausgewählt (vgl. Steinhauer und Zinn 2016): Zunächst wurde eine Stichprobe aus allen staatlich anerkannten Schulen in Deutschland, die Unterricht für die neunte Klassenstufe anbieten, gezogen. Dies umfasste neben Regelschulen auch Förderschulen speziell für Kinder mit sonderpädagogischem Förderbedarf. Die Auswahl der Schulen erfolgte stratifiziert nach Schulform, wobei zwischen Gymnasium, Haupt- bzw. Mittelschule, Realschule, integrierte Gesamtschule, Freie Waldorfschule, Schule mit mehreren Bildungsgängen und Förderschule unterschieden wurde. In jeder der gezogenen Schulen wurden – sofern vorhanden – zufällig zwei neunte Klassen ausgewählt und alle Schüler:innen dieser Klassen zur freiwilligen Teilnahme an der Studie eingeladen. Die Befragung der Zielpersonen sowie die Durchführung der Kompetenztestung fanden innerhalb der Schulen bzw. Klassen statt. Die Testung wurde in Gruppen administriert, anschließend füllten die Schüler:innen einen Fragebogen (*Paper-Pencil*) aus (vgl. Skopek et al. 2013). Die Ausgangsstichprobe des Panels umfasste 16.425 Schüler:innen, wovon 15.672 an der

ersten Welle im Schuljahr 2010/2011 teilgenommen haben (ebd.). Neben den Zielpersonen wurden ebenfalls Befragungen mit Erziehungsberechtigten, Lehrkräften und Schulleitungen durchgeführt.

In der vorliegenden Arbeit werden Informationen aus den ersten beiden Erhebungswellen genutzt. Die erste Welle wurde im Zeitraum November 2010 bis Januar 2011 durchgeführt, und Welle 2 etwa ein halbes Jahr später von Mai bis Juli 2011 (vgl. Skopek et al. 2013). Aufgrund des vergleichsweise kurzen Abstands zwischen den Wellen und der größtenteils unterschiedlichen Befragungsinhalte sowie getesteten Kompetenzdomänen, werden die zwei Messzeitpunkte im Rahmen der Analysen zu einem erweiterten Querschnittsdatensatz zusammengeführt. Die Beschränkung auf den Querschnitt der ersten beiden Wellen hat den rein pragmatischen Hintergrund, dass zum Zeitpunkt der Fertigstellung der meisten Beiträge der Dissertation noch keine weiteren Erhebungswellen bzw. Messzeitpunkte für die relevanten Variablen veröffentlicht waren – beispielsweise wurde die für alle vier Beiträge zentrale abhängige Variable der Lesekompetenz erst in der siebten Welle im Schuljahr 2013/2014 wiederholt erhoben (und 2016 mit dem Scientific-Use-File veröffentlicht).

Für die Analysen zu den vier Aufsätzen werden Schüler:innen an Förderschulen ausgeschlossen, d.h. die Analysepopulation beschränkt sich auf Schüler:innen der neunten Klasse an allgemeinbildenden Regelschulen. Diese Einschränkung ist nicht inhaltlich begründet, sondern ergibt sich daraus, dass die Erhebungsinstrumente und Feldarbeitsprozesse an die speziellen Voraussetzungen an Förderschulen angepasst wurden und damit die Vergleichbarkeit mit Schüler:innen an Regelschulen nur bedingt gegeben ist (Skopek et al. 2013). So wurden in Förderschulen größtenteils andere Kompetenztests eingesetzt, weil die regulären Testinstrumente (z.B. zum Leseverstehen) dort keine ausreichend verlässlichen Ergebnisse liefern, die einen fairen Vergleich zwischen Förder- und Regelschulen ermöglichen würden (vgl. Gnambs 2020). Eine übergreifende Analyse, gerade auch hinsichtlich des für diese Arbeit zentralen Explanandums der sprachlichen Kompetenzen, ist demzufolge nicht umsetzbar. Nach Ausschluss von Förderschulen und wenigen weiteren Ausfällen (z.B. aufgrund temporärer Abwesenheiten oder fehlender Informationen auf zentralen Variablen) umfasst die Analytestichprobe insgesamt bis zu 14.815 Schüler:innen, darunter ca. 29 Prozent mit Migrationshintergrund (vgl. Olczyk et al. 2016; zur Identifikation der migrantischen Population im NEPS siehe Kristen et al. 2016), aus 1.006 Schulklassen und 539 Schulen. Einige Beiträge der Arbeit konzentrieren sich dabei ausschließlich auf die Population mit Migrationshintergrund oder auf Jugendliche, die mit einer anderen Sprache als Deutsch aufgewachsen sind, weswegen die Fallzahlen je nach Analyse unterschiedlich ausfallen.

Analysepotential

Die Analyse zu Schüler:innen der neunten Klassenstufe hat ganz allgemein den Vorteil, dass die Befunde damit an einem umfangreichen Forschungsstand anknüpfen, da viele groß angelegte Studien, wie beispielsweise CILS4-EU oder PISA, inkl. der nationalen Erweiterungen des IQB-Ländervergleichs bzw. Bildungstrends für Deutschland, eine vergleichbare Altersgruppe adressieren. Die Erkenntnisse der vorliegenden Arbeit lassen

sich daher gezielt im Lichte der bisherigen Befundlage einordnen. Die Vergleichbarkeit zu anderen Studien ist grundsätzlich für alle vier Teilfragestellungen von Bedeutung, besonders aber für die Prüfung der Validität selbsteingeschätzter Sprachkompetenzen, wofür anwendungsbezogene Analysemodelle aus der bisherigen Forschung herangezogen werden.

Darüber hinaus bietet die umfangreiche Stichprobe der Startkohorte 4 eine differenzierte Betrachtung der migrantischen Population. Ausgehend von den hohen Fallzahlen und den vorliegenden Indikatoren ist es möglich, verschiedene Herkunftsgruppen zu unterscheiden und den Migrationshintergrund bis in die dritte Generation zu identifizieren (vgl. Kristen et al. 2016), sowie – zumindest für die quantitativ bedeutsamsten Gruppen (z.B. ehemalige Sowjetunion, Polen, Türkei) – die Kombination aus beidem abzubilden. Eine Unterscheidung nach Herkunftsgruppen und/oder Generationenstatus erfolgt in allen Analysen der vorliegenden Arbeit, wenn auch in unterschiedlicher Detailstufe. Besonders detailliert werden diese Unterschiede im Rahmen der Beschreibung der Muster des Sprachgebrauchs in verschiedenen Kontexten herausgearbeitet (vgl. Kapitel 4), um ein möglichst umfassendes Bild des Status quo nachzuzeichnen und damit auch potentiell abweichende Trends im Generationenverlauf zwischen Herkunftsgruppen aufdecken zu können. Neben den Informationen zum Geburtsland der Jugendlichen, ihrer Eltern und Großeltern beinhalten die Daten noch weitere Indikatoren, anhand derer gezielte Analysen zu inhaltlich relevanten Subgruppen angestellt werden können. So lassen sich innerhalb der Population mit Migrationshintergrund weiterhin auch Jugendliche identifizieren, die mit einer anderen Familiensprache als Deutsch aufgewachsen sind. Diese sind für unsere Fragestellungen von besonderem Interesse, weil sich nur für mehrsprachige Schüler:innen überhaupt die Frage stellt, welche Sprache sie zum Beispiel mit Mitschüler:innen verwenden (vgl. Kapitel 5).

Als Indikator für die Sprachkompetenz im Deutschen werden die Ergebnisse aus einem eigens im NEPS entwickelten, psychometrischen Kompetenztest zum Leseverstehen herangezogen (vgl. Gehrler et al. 2012; nähere Details zum Testinstrument und dessen Administration finden sich auch in den Aufsätzen). Die Domäne Leseverstehen liefert dabei nicht nur Hinweise auf das Sprachvermögen im Deutschen, sondern bildet auch eine zentrale Dimension des schulischen Kompetenzerwerbs allgemein ab. Die Lesekompetenz wird in allen vier Aufsätzen der Dissertation als zentrale Messung der *objektiven* Sprachkompetenzen genutzt. Zum weiteren Abgleich werden vereinzelt zusätzlich noch die deutschen Wortschatzkenntnisse hinzugezogen (z.B. Kapitel 4), welche im NEPS anhand der deutschen Ausgabe des „Peabody Picture Vocabulary Test“ (Version 4; Dunn und Dunn 2007; Lenhard et al. 2015) erfasst werden. Der Einsatz psychometrischer Kompetenztests in verschiedenen bildungsbezogenen Domänen ist eine besondere Qualität des NEPS, welche zusätzlich noch durch die Testung des Hörverstehens in den Erstsprachen Russisch und Türkisch für Schüler:innen mit entsprechendem Sprachhintergrund ergänzt wird (vgl. Edele et al. 2016). Der L1-Test des NEPS beschränkt sich auf die mündlichen Kompetenzen, weil die meisten Kinder mit Migrationshintergrund die Erstsprache informell innerhalb ihrer Familie lernen und dabei oftmals keine schriftsprachlichen Kompetenzen erwerben (Kristen et al. 2011). Die verfügbaren Informationen zur L1-Kompetenz in den Daten ermöglichen es,

die Fragestellung an zentralen Stellen ebenfalls auf die Erstsprache auszurichten und diesbezüglich theoretisch bedeutsame Zusammenhänge auch empirisch – anhand einer vergleichsweise umfangreichen Stichprobe – zu untersuchen (vgl. Kapitel 5 und 6).

Darüber hinaus wurden im Rahmen der Befragung der Schüler:innen aber auch Selbsteinschätzungen zu den eigenen Sprachkompetenzen, sowohl im Deutschen als auch in anderen erlernten Erstsprachen, erhoben. Die *subjektive* Messung der jeweiligen Sprachkompetenz folgt dem verbreiteten Schema und erfasst die globalen Fähigkeiten in den vier Dimensionen Verstehen, Sprechen, Lesen und Schreiben (vgl. hierzu schon Kapitel 2.2.1). Für jede Dimension sollten die Jugendlichen angeben, wie gut sie die entsprechende Sprache beherrschen, wobei sie ihre Antwort anhand der Kategorien „überhaupt nicht“, „sehr schlecht“, „eher schlecht“, „eher gut“ und „sehr gut“ einstufen konnten. Somit liegen in den Daten sowohl subjektive als auch objektive Messungen der Sprachkompetenzen vor, was einen Abgleich beider Messungen ermöglicht und damit auch eine systematische empirische Prüfung der Validität subjektiver Einschätzungen sowie der speziellen Faktoren, die möglichen Antwortverzerrungen zugrunde liegen, zulässt (vgl. Kapitel 6 und 7).

Zusätzlich enthalten die Daten umfassende Informationen zum sozioökonomischen und familiären Hintergrund der Schüler:innen, zu deren Lernumfeld, sowie zu persönlichkeitsbezogenen, motivationalen und kognitiven Eigenschaften, Orientierungen und Einstellungen der Jugendlichen. Eine Auswahl dieser Charakteristika wird je nach Fragestellung zur Kontrolle oder als erklärende Variablen in den Analysen berücksichtigt. Ein besonderes Analysepotential bieten die verfügbaren Indikatoren zum Sprachgebrauch in verschiedenen alltagsrelevanten Kontexten (mit Eltern, Geschwistern, Freund:innen, Mitschüler:innen, sowie bei der Mediennutzung), welche eine vergleichende Beschreibung der Muster des Gebrauchs der Herkunftssprache innerhalb und außerhalb der Familie ermöglichen (vgl. Kapitel 4). Gerade auch anhand der Information zum Sprachgebrauch mit Mitschüler:innen lässt sich eine in der Forschung weit verbreitete Annahme zur Wirkungsweise der ethnischen Klassenkomposition auf die schulischen Leistungen relativ direkt empirisch abbilden und untersuchen (vgl. Kapitel 5). Auch die verschiedenen etablierten Analysemodelle, die zur Prüfung der Validität subjektiver Sprachkompetenzen herangezogen werden, lassen sich mit den in den Daten zur Verfügung stehenden Informationen auf kongruente Weise rekonstruieren (vgl. Kapitel 6). Für die daran anknüpfende Analyse zu den Bedingungen für verzerrte Selbsteinschätzungen lassen sich die drei theoretisch bedeutsamen Konstrukte (allgemeine kognitive Fähigkeiten, Selbstbild und sprachlicher Referenzrahmen) ebenfalls stringent mit den Daten operationalisieren (vgl. Kapitel 7).

Methodisches Vorgehen

Aufgrund des Designs der verwendeten Daten bleibt die empirische Untersuchung der Arbeit auf den Querschnitt und entsprechend dafür geeignete statistische Verfahren beschränkt. Die Fragestellungen der Dissertation lassen sich in weiten Teilen aber auch mittels Querschnittsanalysen umfänglich untersuchen, etwa die Replikation etablierter Analysemodelle zur Prüfung der Validität subjektiver Sprachmessung. Dennoch

bildet der querschnittliche Charakter der Auswertungen eine zentrale Einschränkung, weswegen keine kausalen Aussagen getroffen werden können und dort, wo theoretische Kausalbeziehungen von Interesse wären, muss sich die Arbeit – streng genommen – auf die Interpretation von empirischen Korrelationen beschränken. Um sich dennoch den bedeutsamen Wirkzusammenhängen zu nähern, verfolgt die Arbeit einen theoriegeleiteten Ansatz und versucht die theoretischen Argumente analytisch so zu separieren, dass daraus spezifische Prognosen abgeleitet werden können, die sich theoretisch nur auf eine Weise begründen lassen und dadurch zumindest einen indirekten empirischen Test ermöglichen.

Für die meisten Analysen werden lineare Regressionsmodelle (mittels „Ordinary Least Square“ Verfahren) geschätzt, um die multivariaten Zusammenhänge zwischen der jeweiligen abhängigen Variable und den relevanten unabhängigen Variablen zu bestimmen. Dabei werden die Standardfehler nach Schulklassen geclustert, um die genestete Stichproben- und Datenstruktur (Schüler:innen in Klassen bzw. Schulen) auch in den Analysen zu berücksichtigen (vgl. Abadie et al. 2023, Moulton 1986). Dies ist insofern notwendig, als dass Schüler:innen einer Klasse speziellen Einflüssen ausgesetzt sein können, die innerhalb der Klasse alle gleichermaßen betreffen und sich auf die Beziehungen zwischen den interessierenden individuellen Variablen auswirken. Einflüsse der Klassenzusammensetzung oder der Lehrer:innen wären Beispiele für derartige Prozesse. Ohne weitere Berücksichtigung potentieller Einflüsse auf Schul- bzw. Klassenebene wäre – wenn diese denn vorlägen – die zentrale Modellannahme der Unabhängigkeit der Residuen verletzt, was eine Unterschätzung der Standardfehler und damit eine Überbewertung der statistischen Signifikanz der betroffenen Zusammenhänge zur Folge hätte (vgl. Moulton 1986). Die Clusterung der Standardfehler wirkt diesen verzerrenden Tendenzen entgegen.

Eine Ausnahme bilden die Analysen zur Bedeutung der ethnischen bzw. ethnolinguale Klassenkomposition (Kapitel 5), in welchen der Schulklassenkontext im Rahmen hierarchischer linearer Regressionsmodelle (HLM) direkt als zentrale Analyseebene mit modelliert wird. Dabei lassen sich Einflüsse der unterschiedlichen Ebenen analytisch voneinander trennen und dadurch beispielsweise Einflüsse der Klassenzusammensetzung (Level 2), die über die Gesamtheit der individuellen Einflüsse (Level 1) hinausgehen, statistisch bestimmen (vgl. Hox 2010; Raudenbush und Bryk 2002). Die vergleichsweise umfangreiche Stichprobe auch an Schulklassen ($N=1.006$) ermöglicht eine aussagekräftige Analyse von Einflüssen auf Klassenebene. Gleichzeitig ist für eine zuverlässige Schätzung von Kompositionseffekten insbesondere aber die Stichprobenausschöpfung innerhalb der Klassen entscheidend, gerade wenn die Kompositionsmerkmale auf einer Aggregation der individuellen Information basieren (vgl. Schunck 2016). Bei geringer Ausschöpfung auf Individualebene sinkt die Reliabilität der aggregierten Kompositionsmerkmale, was eine Unterschätzung der Kompositionseffekte zur Folge hat (ebd.). Derartige Tendenzen lassen sich für die vorliegende Untersuchung nicht vollständig ausschließen, gerade auch vor dem Hintergrund einer freiwilligen Teilnahme der Schüler:innen. Allerdings liegen in den Daten nur wenige Klassen mit sehr geringer Ausschöpfung vor, welche bei der Modellierung ohnehin – den Fallzahlen entsprechend – mit geringerem Gewicht in die Schätzung einfließen, weswegen das Ausmaß potentieller

Verzerrungen als eher gering einzustufen ist.

In den multivariaten Analysen zu allen Beiträgen werden fehlende Werte auf den Modellvariablen (*Item-Nonresponse*) mittels multipler Imputation berücksichtigt (vgl. Allison 2001). Dazu wurde das in der Software Stata (StataCorp 2015, 2021) implementierte iterative Verfahren „chained equations“ (MICE; White et al. 2011) verwendet, um – je nach Anwendungsfall und Anteil fehlender Information – bis zu 65 imputierte Datensätze mit vollständigen Informationen zu generieren. Die Analysenmodelle werden dann an jedem imputierten Datensatz separat geschätzt und die Ergebnisse anschließend nach der von Rubin (1987) vorgeschlagenen Regel kombiniert. Dabei wird bei der Schätzung der Standardfehler zusätzlich zur herkömmlichen Varianz zwischen den Beobachtungen auch die Varianz zwischen den Ergebnissen der einzelnen imputierten Datensätze einberechnet, um so möglichen Unsicherheiten und Zufallsschwankungen bei der Imputation Rechnung zu tragen. Mit steigender bzw. bei ausreichender Anzahl an Imputationen lässt sich der Einfluss derartiger Schwankungen auf ein Minimum reduzieren (vgl. von Hippel 2020), weswegen in den meisten Beiträgen der vorliegenden Arbeit die Anzahl an Imputationen relativ hoch angesetzt wurde.

4 *Artikel 1*: Spracherhalt oder Sprachverlagerung? Erstsprachgebrauch und Zweitsprachkompetenzen bei Jugendlichen mit Migrationshintergrund

Dieser Beitrag entstand in Kooperation mit Bernadette Strobel und ist in der *Kölner Zeitschrift für Soziologie und Sozialpsychologie* (©2016, Springer Fachmedien Wiesbaden) erschienen. Der nachfolgende Text und Inhalt wurde unverändert aus dem veröffentlichten Originalbeitrag übernommen und an das Format der Dissertation angepasst.* Die Referenz lautet:

Strobel, B., und Seuring, J. (2016). Spracherhalt oder Sprachverlagerung? Erstsprachgebrauch und Zweitsprachkompetenzen bei Jugendlichen mit Migrationshintergrund. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(2):309–339. <https://doi.org/10.1007/s11577-016-0360-y>

* Ergänzend zu dem Originalbeitrag wurden im Rahmen der Dissertation Zusatzauswertungen angefertigt, die der besseren Einordnung der Befunde dienen sollen. Die Befunde dazu sind im Anhang 4.8 dieses Kapitels ausgewiesen. Der gesamte Anhang 4.8 ist nicht Teil des in der *Kölner Zeitschrift für Soziologie und Sozialpsychologie* veröffentlichten Originalbeitrags. Im Text wird an relevanten Stellen mittels Fußnote („0“) auf diese Zusatzauswertungen verwiesen. Alle Stellen sind eindeutig als Ergänzung gekennzeichnet.

Zusammenfassung

Geringere Zweitsprachkompetenzen bei Jugendlichen mit Migrationshintergrund werden häufig mit einer herkunftssprachlichen Kommunikation innerhalb der Familie in Verbindung gebracht. Der vorliegende Beitrag erweitert den bisherigen Fokus. Er beschreibt den Sprachgebrauch in weiteren alltagsrelevanten Kontexten über drei Zuwanderergenerationen hinweg und untersucht, wie dieser mit den Zweitsprachkompetenzen zusammenhängt. Die Verwendung der Zweitsprache wird über den damit einhergehenden Zugang und eine erhöhte Lernmotivation mit den Deutschkompetenzen verknüpft. Anhand der Daten des Nationalen Bildungspanels lässt sich zeigen, dass die Herkunftssprache mit den Eltern häufiger verwendet wird als mit Geschwistern. Mit Peers wird überwiegend Deutsch gesprochen, während sich beim Medienkonsum eine komplementäre Nutzung beider Sprachen abzeichnet. Türkischstämmige Jugendliche verwenden in allen Kontexten häufiger die Herkunftssprache als diejenigen aus Polen oder der ehemaligen Sowjetunion. Wird mit den Eltern ausschließlich Deutsch gesprochen, dann auch meist in allen anderen Kontexten. Jugendliche mit einer nicht-deutschen Elternsprache variieren hingegen stark hinsichtlich des Sprachgebrauchs in weiteren Kontexten. Die mit der Familiensprache assoziierten Sprachdefizite können durch den Deutschgebrauch mit Peers reduziert werden.

4.1 Einleitung

Kompetenzen in der Sprache des Aufnahmelandes gelten als Schlüsselqualifikation für den Bildungserfolg von Jugendlichen mit Migrationshintergrund. In diesem Zusammenhang wird zunehmend die Bedeutung des Sprachgebrauchs für den Erwerb von Kompetenzen in der Unterrichtssprache diskutiert (Chiswick und Miller 2001; Konsortium Bildungsberichterstattung 2006, S. 175; Kristen et al. 2011). Empirische Studien, die vorwiegend den familiären Sprachgebrauch fokussieren, deuten wiederholt darauf hin, dass Schülerinnen und Schüler, die ihre Herkunftssprache (L1) in der Familie verwenden, oft sprachlich bedingte Nachteile im Bildungssystem aufweisen (z. B. Azzolini et al. 2012; Baumert und Schümer 2001, S. 376 ff.; Esser 2006, S. 311 ff.; Kristen 2008; Müller und Stanat 2006; Verwiebe und Riederer 2013; Walter 2009).¹ Es wird angenommen, dass durch die Verwendung der Herkunftssprache weniger Gelegenheiten bestehen, die Sprache des Aufnahmelandes (L2) zu sprechen und dabei entsprechende Kenntnisse zu erwerben oder auszubauen (Chiswick und Miller 2001; Esser 2006). Darüber hinaus werden allerdings auch Vorteile der herkunftssprachlichen Kommunikation innerhalb der Familie für den Bildungserwerb thematisiert. Derartige Effekte sollten insbesondere dann zu erwarten sein, wenn die Eltern über geringe L2-Kompetenzen verfügen oder bildungsförderliche Orientierungen über die L1 nachdrücklicher kommuniziert werden können (Mouw und Xie 1999; Portes und Hao 2002).

Bei der empirischen Prüfung dieser Argumente wird der Sprachgebrauch meist über eine einzelne Frage nach der zuhause gesprochenen Sprache erhoben. Bei den so erhaltenen Angaben lässt sich jedoch nicht differenzieren, ob die L1 ausschließlich mit den Eltern, mit einem Elternteil oder zusätzlich auch mit den Geschwistern und sonstigen Haushaltsmitgliedern verwendet wird. Darüber hinaus wird nicht berücksichtigt, welche Sprache die Befragten außerhalb des familiären Kontexts, beispielsweise mit gleichaltrigen Bezugspersonen, den sogenannten Peers, oder beim Medienkonsum verwenden. Wenn angenommen wird, dass Sprachkompetenzen durch die entsprechenden Gelegenheiten zum Sprachgebrauch gefördert werden, ergeben sich in Abhängigkeit der Ausprägung des familiären und außerfamiliären Sprachgebrauchs unterschiedliche Prognosen für die L2-Kompetenzen. So könnte der L2-Gebrauch in anderen Kontexten den Nachteilen, die oftmals mit dem L1-Erhalt in der Familie assoziiert werden, entgegenwirken. Daher ist fraglich, ob die Familiensprache die mit heterogenen Sprachgewohnheiten verknüpften Zusammenhänge ausreichend abbildet.

Bisher liegen für den deutschen Kontext nur eingeschränkt Informationen darüber vor, wie sich der Sprachgebrauch von Jugendlichen mit Migrationshintergrund in verschiedenen Bereichen gestaltet und ob diesbezüglich Unterschiede zwischen Herkunftsgruppen und im Generationenverlauf bestehen. Während bezüglich der zuhause verwendeten Sprache durchaus belastbare Ergebnisse verfügbar sind, erweist sich die Befundlage zu weiteren Kontexten als weniger ergiebig. Da ein Forschungsüberblick über den kontextspezifischen Sprachgebrauch bislang aussteht, werden im Folgenden die zentralen

¹Im Folgenden wird die Herkunftssprache als L1 oder Erstsprache, die im Aufnahmeland gesprochene Sprache als L2 oder Zweitsprache bezeichnet, unabhängig davon, in welcher Reihenfolge die beiden Sprachen erlernt wurden (Kristen et al. 2011).

4 *Spracherhalt oder Sprachverlagerung?*

Befunde systematisch zusammengetragen.

Aus den vorhandenen Studien lässt sich entnehmen, dass der Erhalt der Herkunftssprache innerhalb von Zuwandererfamilien insgesamt sehr verbreitet ist (Ramm et al. 2005; Strobel und Kristen 2015); von der ersten hin zur zweiten Generation allerdings stark abnimmt (Hans 2010; Ramm et al. 2005; Segeritz et al. 2010). Gleichzeitig zeigen sich innerhalb der Familie generationale Unterschiede. So wird mit den Eltern deutlich häufiger die Herkunftssprache gesprochen als unter Geschwistern (Haug 2008, S. 41; Nauck 2007, 2009).

Im außerfamiliären Kontext zeigt sich ein gegenläufiges Muster. Personen mit Migrationshintergrund verwenden in der Kommunikation im Freundes- und Bekanntenkreis mehrheitlich die Sprache des Aufnahmelandes (Gärtig et al. 2011; Nauck 2007; Salentin und Wilkening 2003, S. 294). Ähnliches zeigt sich auch in Bezug auf den Medienkonsum. Allerdings ist hier zusätzlich der komplementäre Gebrauch beider Sprachen relativ stark ausgeprägt (Diehl und Schnell 2006; Simon und Neuwöhner 2011; Worbs 2010). Mit Blick auf Unterschiede im Sprachgebrauch zwischen Herkunftsgruppen weist der Forschungsstand übereinstimmend darauf hin, dass türkischstämmige Personen über alle Kontexte hinweg verstärkt die L1 erhalten, wohingegen dies bei anderen Zuwanderungsgruppen, insbesondere bei polnischen Migranten, vergleichsweise selten der Fall ist (Gostomski 2010; Hans 2010; Haug 2008; Nauck 2007, 2009; Noll und Weick 2011; Strobel und Kristen 2015; Worbs 2010).

Die bisherige Befundlage liefert erste Hinweise darauf, dass sich der Sprachgebrauch in Abhängigkeit der verschiedenen Kontexte äußerst heterogen gestaltet. Allerdings basieren die genannten Befunde auf zum Teil sehr unterschiedlichen Alterskohorten, Herkunftsgruppen und Zuwanderergenerationen, was einen direkten Vergleich der Ergebnisse erschwert. Daher lässt sich nicht eindeutig abgrenzen, ob der Sprachgebrauch tatsächlich nach Kontexten oder lediglich in Abhängigkeit der jeweils zugrunde liegenden Analysepopulation variiert. Darüber hinaus ist anzumerken, dass bisherige Studien auf Informationen zum Sprachgebrauch in der ersten und zweiten Zuwanderergeneration beschränkt sind. In Bezug auf den außerfamiliären Sprachgebrauch sowie den Medienkonsum bleibt eine Differenzierung nach Generationen weitestgehend vollständig aus. Dies ist aber insofern bedeutsam, als dass sich Zuwanderungsgruppen stark in ihrer generationalen Zusammensetzung unterscheiden und deshalb ein Vergleich spezifischer Merkmale, wie beispielsweise auch der Sprachgewohnheiten, zwischen Herkunftsgruppen nur unter zusätzlicher Berücksichtigung des Generationenstatus' aussagekräftige Befunde liefert (Gresch und Kristen 2011). Besonders auch vor dem Hintergrund, dass in den späteren Generationen teilweise Tendenzen einer „reactive ethnicity“ (Portes und Rumbaut 2001, S. 148), d. h. einer verstärkten Orientierung an der Herkunftskultur, vermutet werden, ist eine möglichst weitreichende Differenzierung des Generationenstatus' unerlässlich (z. B. Diehl und Schnell 2006; Jacob und Kalter 2013).

Des Weiteren erlauben die von der bisherigen Forschung gebotenen, aggregierten Verteilungen keine Rückschlüsse auf die individuelle Sprachnutzung über verschiedene Kontexte hinweg. Dies bedeutet, dass es unklar ist, ob eine Person, die beispielsweise mit den Eltern hauptsächlich die L1 spricht, diese auch in anderen Kontexten verwendet oder nicht. Eine Betrachtung individueller Muster ist aber zentral, wenn Unterschiede

im Sprachgebrauch mit den Deutschkompetenzen in Verbindung gebracht werden sollen. Obwohl die Bedeutung weiterer Kontexte in der Forschung immer wieder diskutiert wird, dominiert der Einsatz der Familiensprache als alleiniger Indikator für den individuellen Sprachgebrauch.

Der vorliegende Beitrag trägt zur Schließung der Forschungslücken bei, indem er, basierend auf ein und derselben Analysepopulation, zunächst einen differenzierten Überblick darüber bietet, welche Sprache Jugendliche mit Migrationshintergrund mit dem Vater, der Mutter, den Geschwistern, der besten Freundin oder dem besten Freund, Mitschülerinnen und Mitschülern sowie beim Medienkonsum verwenden. Dadurch lassen sich die bisher wenig vergleichbaren Ergebnisse früherer Studien anhand einer einheitlichen Datengrundlage validieren. Gleichzeitig geht der Beitrag über die bisherige Befundlage hinaus und beschreibt die kontextspezifischen Muster des Sprachgebrauchs in verschiedenen Herkunftsgruppen über drei Generationen hinweg.

Weiterhin werden im Anschluss an die aggregierte Betrachtung die Angaben zu den unterschiedlichen Kontexten in individuellen Sprachgebrauchsprofilen kombiniert. Aus diesen lässt sich ablesen, in welchen Kontexten eine Person die Herkunftssprache erhält und in welchen verstärkt die L2 verwendet wird. Ausgehend davon wird schließlich der Frage nachgegangen, ob unter Berücksichtigung verschiedener Kontexte des Sprachgebrauchs auch differenziertere Aussagen über die Zusammenhänge mit den Zweitsprachkompetenzen der Zugewanderten möglich werden, als durch die alleinige Betrachtung der Familiensprache. Die Befunde liefern letztendlich Hinweise darauf, inwiefern zusätzliche Kontexte für die Deutschkompetenzen und damit für den Bildungserfolg bedeutsam sind. Für die Analysen werden Daten des Nationalen Bildungspanels (NEPS) zu Schülerinnen und Schülern der neunten Klasse herangezogen.

Abschnitt 4.2 schildert zunächst in allgemeiner Form die theoretischen Annahmen zur Wirkungsweise des Sprachgebrauchs für den Zweitspracherwerb. In Abschn. 4.3 werden diese Argumente auf unterschiedliche alltagsrelevante Sprachkontexte übertragen. Auf die Beschreibung der Datengrundlage in Abschn. 4.4, der verwendeten Instrumente sowie des methodischen Vorgehens folgt die Präsentation der deskriptiven und multivariaten Befunde in Abschn. 4.5. Der Beitrag schließt mit einer Diskussion der zentralen Ergebnisse sowie deren Implikationen für die zukünftige Forschung.

4.2 Die Bedeutung des Sprachgebrauchs für die Zweitsprachkompetenz

Die Zweitsprachkompetenz von Zuwanderern ist das Ergebnis eines Lernprozesses, der als eine Investition in aufnahmelandsspezifische Ressourcen betrachtet werden kann.² Der Spracherwerb wird von der Lernmotivation, dem Lernvermögen und dem Zugang zu dieser Sprache determiniert (Chiswick und Miller 2001; Esser 2006). Die Motivation

²Aus Gründen der Lesbarkeit werden unter dem Begriff Zuwanderer geschlechtsneutral alle männlichen und weiblichen Personen mit Migrationshintergrund zusammengefasst. Die Formulierung schließt sowohl Personen mit eigener Wanderungserfahrung als auch deren Nachkommen in der zweiten und dritten Generation mit ein.

4 Spracherhalt oder Sprachverlagerung?

ergibt sich aus den zusätzlichen Erträgen, welche die Beherrschung der L2 im Vergleich zur L1 verspricht. Das Lernvermögen umfasst die jeweilige Lerneffizienz, also die Fähigkeit, bestimmte Inhalte in einem gegebenen Zeitraum zu verarbeiten. Der Zugang beschreibt den Kontakt zur jeweiligen Sprache, der in einem bestimmten Zeitumfang und in unterschiedlicher Intensität besteht (Chiswick und Miller 2001, S. 393).

Im Rahmen des Zweitspracherwerbs kommt dem Sprachgebrauch eine besondere Bedeutung zu, da er sich maßgeblich auf zwei der beschriebenen Parameter auswirkt. Einerseits bestimmt er den Zugang zur L2 in Form einer aktiven Verwendung dieser Sprache. Andererseits spiegelt der Sprachgebrauch aber auch die Verfügbarkeit potenzieller Gelegenheiten zur Verwendung der L1 sowie der L2 wider. Je mehr Möglichkeiten bestehen, eine Sprache zu verwenden, desto höher ist die Motivation, diese zu erlernen (Strobel und Kristen 2015, S. 129). Beispielsweise bietet der familiäre Kontext keinen oder nur einen sehr eingeschränkten Zugang zur L2, wenn die Eltern über unzureichende L2-Kompetenzen verfügen und eine Kommunikation ausschließlich in der Herkunftssprache möglich ist. Gleichzeitig besteht dadurch seitens des Kindes ein hoher Anreiz, die L1 zu erhalten (oder zu erlernen), um die Interaktion mit den Eltern zu gewährleisten. Bezogen auf den familiären Kontext wäre demzufolge mit einer eingeschränkten Investitionsneigung in den Zweitspracherwerb zu rechnen.

Sowohl die Motivation als auch der Zugang stellen notwendige Bedingungen für eine Investition in die Zweitsprache dar: Selbst wenn ausreichend Motivation vorhanden ist, werden keine Sprachkompetenzen erworben, sofern keinerlei Zugang zur L2 besteht, und *vice versa* (Esser 2006, S. 73). Nennenswerte Zuwächse in den Sprachkompetenzen durch den Sprachgebrauch sind folglich nur dann zu erwarten, wenn sprachliche Kontaktmöglichkeiten bestehen und diese gleichzeitig auch entsprechend genutzt werden.

4.3 Kontexte des Sprachgebrauchs

Den allgemeinen theoretischen Zusammenhängen zufolge determiniert der Sprachgebrauch sowohl die Motivation als auch den Zugang zur L2. Daraus lässt sich ableiten, dass mit zunehmender Verwendung der Zweitsprache auch höhere Kompetenzen zu erwarten sind. Es ist davon auszugehen, dass der Sprachgebrauch in Abhängigkeit von Bezugskontext und Personen variiert (Fishman 1965; Strobel und Kristen 2015). Folglich deckt die bisher im Fokus der Forschung stehende Familiensprache nur einen Teil des tatsächlichen Sprachgebrauchs ab. Um differenziertere Prognosen über den Kompetenzerwerb aufzustellen, gilt es daher, die individuelle Gesamtnutzung zu berücksichtigen, welche aus dem Umfang und der Intensität des L2-Gebrauchs in verschiedenen alltagsrelevanten Kontexten resultiert (Biedinger et al. 2015, S. 1523; Chiswick und Miller 2001). Je häufiger und in je mehr Kontexten die Zweitsprache verwendet wird, desto höhere L2-Kompetenzen sind zu erwarten.

Umgekehrt können die Sprachkompetenzen aber auch die Nutzung der jeweiligen Sprache bedingen (Strobel und Kristen 2015; Tubergen und Kalmijn 2009). Mit steigenden Kompetenzen eröffnen sich zusätzliche Gelegenheiten, die Sprache zu verwenden. Diese Interdependenz impliziert, dass für eine potenzielle Verwendung der Sprache

zunächst ausreichende Sprachkompetenzen vorliegen müssen. Mit Blick auf den Sprachgebrauch von Personen mit Migrationshintergrund ist demzufolge ausschlaggebend, ob überhaupt ein Mindestmaß an herkunftssprachlichen Kompetenzen gegeben ist und sich dadurch Möglichkeiten zu deren Verwendung ergeben. Hier spielt der familiäre Kontext eine entscheidende Rolle, da die Nachkommen von Zuwanderern die L1 hauptsächlich innerhalb der Familie erlernen (Alba et al. 2002, S. 469; Biedinger et al. 2015).

Wird mit den Eltern ausschließlich die L2 gesprochen, so ist eher davon auszugehen, dass die Kinder die L1 nicht erlernen und die aus der Zeit vor der Migration vorhandenen Fähigkeiten nicht weiter ausgebaut werden. Darüber hinaus kann die Elternsprache ethnische Orientierungen ausdrücken (Strobel und Kristen 2015), die sich wiederum auf die sprachlichen Gelegenheiten der Kinder auswirken können. Beispielsweise wird durch die Wahl der Wohnumgebung auch die ethnische und damit sprachliche Zusammensetzung des sozialen Umfelds der Kinder vorstrukturiert (Bauer et al. 2005). Nutzen die Eltern selbst ausschließlich die L2, legen sie womöglich wenig Wert darauf, herkunftssprachliche Medien im Haushalt verfügbar zu machen (Piga 2007, S. 222). Des Weiteren kann angenommen werden, dass Eltern alle Kinder in der gleichen Sprache erziehen und dadurch auch die Kommunikation unter den Geschwistern beeinflusst wird (Nauck 2009). Demzufolge determiniert die Elternsprache die Möglichkeiten des Sprachgebrauchs in allen weiteren Kontexten. Daher wird, wenn mit den Eltern ausschließlich die L2 gesprochen wird, ein kontextübergreifender L2-Gebrauch erwartet.

Die Argumentation lässt sich in ähnlicher Weise auch auf den Gebrauch der Herkunftssprache mit den Eltern übertragen. Wird die L1 innerhalb der Familie erhalten, eröffnen sich potenzielle Gelegenheiten, diese auch in anderen Kontexten zu verwenden. Zum einen erwerben die Kinder dadurch Kompetenzen in der L1, zum anderen könnte der elterliche L1-Erhalt auf eigenethnische Orientierungen hindeuten, die sich wiederum in der ethnischen Komposition des sozialen Umfelds niederschlagen. Allerdings wirkt sich der L1-Gebrauch mit den Eltern weniger deterministisch auf den Sprachgebrauch in anderen Kontexten aus, als dies bei der Familiensprache L2 der Fall ist. Unabhängig davon, ob Kinder innerhalb der Familie die Herkunftssprache erhalten oder erlernen, sollten sie durch den unausweichlichen Sprachkontakt im Aufnahmekontext (beispielsweise in der Schule) Kenntnisse in der Zweitsprache erwerben. Somit besteht in diesem Fall grundsätzlich die Möglichkeit, in anderen Kontexten sowohl die L1 als auch die L2 zu verwenden. Es stellt sich folglich die Frage, ob Kinder, die mit ihren Eltern in der Herkunftssprache kommunizieren, die L1 dann auch verstärkt in weiteren Kontexten verwenden.

Da der Aufnahmekontext nur eingeschränkt Gelegenheiten und Anreize zur Verwendung der Herkunftssprache außerhalb der Familie bietet, ist anzunehmen, dass der L1-Gebrauch dort insgesamt weniger verbreitet ist (z.B. Alba et al. 2002, S. 469). Bisherige Befunde deuten darauf hin, dass die L1 innerhalb der Familie, im Besonderen bei der Eltern-Kind-Interaktion, vergleichsweise häufig aufrechterhalten wird (Haug 2008, S. 41; Nauck 2007, 2009), während beim Medienkonsum (Diehl und Schnell 2006; Simon und Neuwöhner 2011; Worbs 2010) und in der außerfamiliären Kommunikation zunehmend die L2 Verwendung findet (Gärtig et al. 2011; Nauck 2007). Dies lässt vermuten,

dass sich Kinder, die mit den Eltern die L1 sprechen, stark darin unterscheiden, ob sie die L1 auch in zusätzlichen Kontexten, wie beispielsweise mit Personen außerhalb der Familie oder beim Medienkonsum, gebrauchen oder nicht. Belastbare Studien liegen hierfür nach unserem Wissensstand bislang nicht vor. Allerdings weisen Biedinger et al. (2015) darauf hin, dass sich in Abhängigkeit des kontextübergreifenden Sprachgebrauchs auch Unterschiede bei den herkunftssprachlichen Kompetenzen ergeben. Gleichzeitig betonen sie die besondere Relevanz der Elternsprache für den Spracherwerb. Entsprechend den theoretischen Annahmen lassen sich diese Zusammenhänge analog auch auf den Zweitspracherwerb übertragen. Ein zunehmender L2-Gebrauch steigert die Motivation sowie den Zugang in dieser Sprache. Demzufolge sollten Kinder, die mit den Eltern die Herkunftssprache verwenden, mit jedem zusätzlichen Kontext, in dem die Zweitsprache genutzt wird, zunehmende L2-Kompetenzen aufweisen.

Zusammenfassend wird deutlich, dass die Familiensprache nur einen, wenn auch bedeutenden Teilaspekt des Sprachgebrauchs umfasst. Der Annahme folgend, dass die Elternsprache den Sprachgebrauch in anderen Kontexten bedingt, werden höhere L2-Kompetenzen erwartet, wenn in diesem Zusammenhang die Zweitsprache anstatt der Herkunftssprache verwendet wird. Wird die L1 mit den Eltern erhalten, könnte jedoch der L2-Gebrauch mit anderen Bezugspersonen oder beim Medienkonsum den Nachteilen, welche häufig mit einer herkunftssprachlichen Eltern-Kind-Interaktion in Verbindung gebracht werden, entgegenwirken. Folglich kann angenommen werden, dass eine ausschließliche Betrachtung der Familiensprache die Heterogenität insbesondere innerhalb der Gruppe der Zuwanderer, die mit den Eltern die Herkunftssprache verwenden, nur unzureichend berücksichtigt.

Im Folgenden wird daher untersucht, inwiefern sich für Schülerinnen und Schüler mit Migrationshintergrund unterschiedliche Muster hinsichtlich ihres Sprachgebrauchs innerhalb und außerhalb der Familie zeigen und wie diese mit den Zweitsprachkompetenzen zusammenhängen. Dabei wird der relative Stellenwert der Familiensprache im Vergleich zu weiteren Kontexten untersucht und geprüft, ob sich durch die zusätzliche Berücksichtigung des Sprachgebrauchs in verschiedenen Kontexten differenzielle Befunde im Vergleich zur Familiensprache ergeben.

4.4 Daten und Methode

Als Grundlage für die Analysen werden Daten der Startkohorte 4 des Nationalen Bildungspanels (NEPS; Blossfeld et al. 2011) herangezogen.³ Zum Zeitpunkt der Befragung besuchten die Zielpersonen die neunte Klasse. Die Auswahl der Untersuchungspopulation erfolgte mittels einer geschichteten Klumpenstichprobe, bei der Schulen verschiedener Schultypen zufällig ausgewählt wurden. Innerhalb dieser Schulen wurden

³doi:10.5157/NEPS:SC4:4.0.0. Die Daten des NEPS wurden von 2008 bis 2013 als Teil des Rahmenprogramms zur Förderung der empirischen Bildungsforschung erhoben, welches vom Bundesministerium für Bildung und Forschung (BMBF) finanziert wurde. Seit 2014 wird das NEPS vom Leibniz-Institut für Bildungsverläufe e. V. (IIfBi) an der Otto-Friedrich-Universität Bamberg in Kooperation mit einem deutschlandweiten Netzwerk weitergeführt.

jeweils zwei bis drei Klassen gezogen, deren Schülerinnen und Schüler alle zur Teilnahme eingeladen wurden (Skopek et al. 2013). Die Analysen beschränken sich auf Personen mit Migrationshintergrund ($N=4721$), wobei zwischen Personen polnischer und türkischer Abstammung sowie Personen aus der ehemaligen Sowjetunion unterschieden wird. Schülerinnen und Schüler aus weiteren Herkunftsländern werden unter der Kategorie „Sonstige“ zusammengefasst.

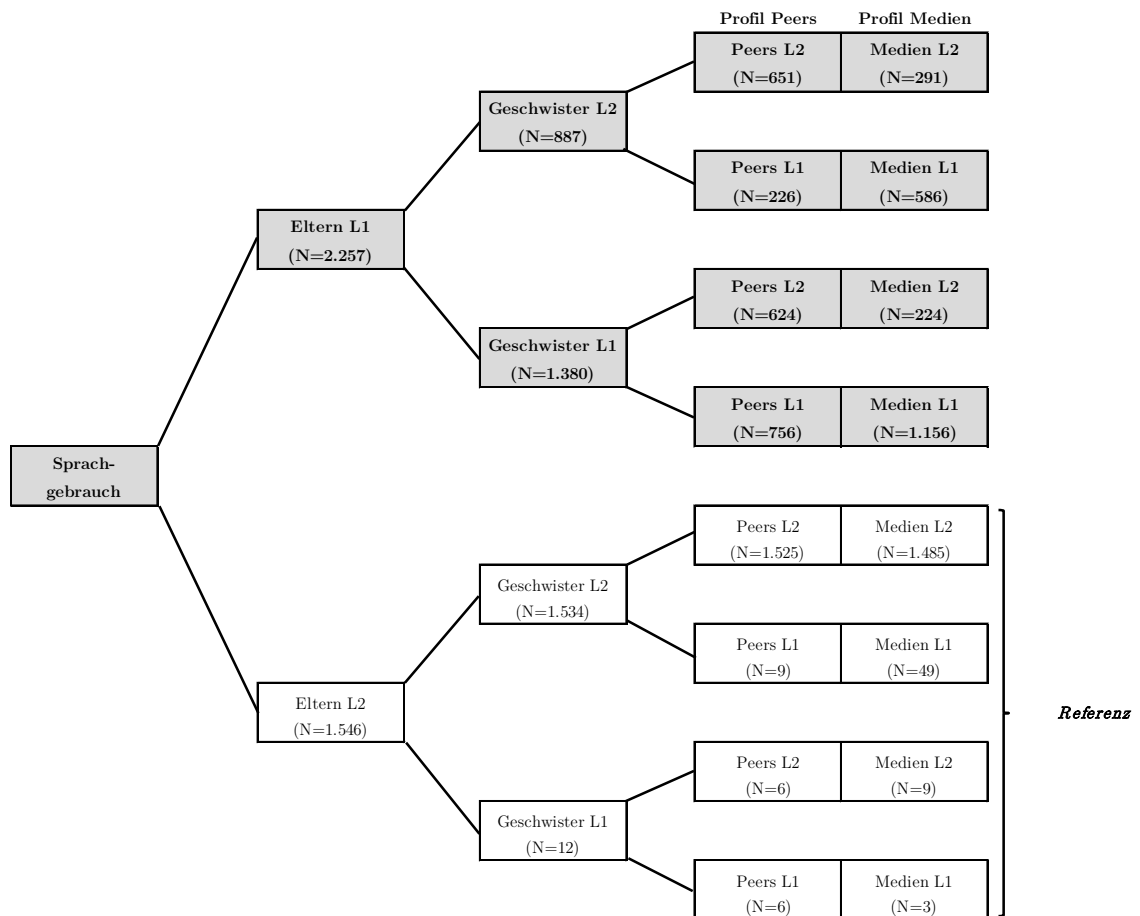
Zur Abbildung des *Sprachgebrauchs* liegen Indikatoren zu sechs unterschiedlichen Kontexten vor. Zielpersonen, die neben Deutsch noch eine weitere Sprache in der Familie gelernt haben ($N=3302$), wurden gefragt, ob sie mit Vater, Mutter, Geschwistern, der besten Freundin oder dem besten Freund, Mitschülerinnen oder Mitschülern und bei der Nutzung bestimmter Medien nur Deutsch, meistens Deutsch, meist die andere Sprache oder nur die andere Sprache verwenden. Als Indikator für die Sprache der Mediennutzung werden sechs Items herangezogen (Videos und DVDs, Sendungen im TV, Bücher, Zeitungen, Nachrichten im Internet, Surfen im Internet; Cronbach's Alpha=0,87), welche in einem Mittelwert zusammengefasst und anschließend den Ausgangskategorien zugeordnet werden. Schülerinnen und Schülern, die zwar einen Migrationshintergrund, jedoch Deutsch als einzige Muttersprache angegeben hatten, wurden keine weiteren Fragen zum Sprachgebrauch gestellt. Hier wird angenommen, dass diese in allen Kontexten nur Deutsch verwenden.⁰

Um den individuellen Sprachgebrauch kontextübergreifend zu analysieren, werden die Angaben zu den einzelnen Bezugspersonen oder zum Medienkonsum in sogenannten *Sprachgebrauchsprofilen* kombiniert. Diese geben an, in welchen Kontexten eine Person die Herkunftssprache erhält und in welchen sie hauptsächlich die Zweitsprache verwendet. Es ist anzunehmen, dass der Sprachgebrauch mit Vater und Mutter nur geringfügig variiert (Nauck 2009). Gleiches kann für den Sprachgebrauch im Peer-Kontext vermutet werden, da sich die Angaben zu der besten Freundin oder dem besten Freund und den Mitschülerinnen oder Mitschülern möglicherweise auf denselben Personenkreis beziehen. Daher werden diese Angaben zu jeweils einem Indikator zusammengefasst. Ausgehend davon lassen sich vier zentrale Kontexte des Sprachgebrauchs kontrastieren: Eltern, Geschwister, Peers und Medien. Für jeden dieser Kontexte wird unterschieden, ob die Befragten dort nur Deutsch oder auch die Herkunftssprache verwenden.

Abbildung 4.1 veranschaulicht das zugrundeliegende Prinzip zur Konstruktion der Sprachgebrauchsprofile. In einem ersten Schritt lässt sich unterscheiden, ob mit den Eltern die Herkunftssprache erhalten wird oder nicht. Innerhalb dieser beiden Gruppen kann weiterhin nach dem Sprachgebrauch mit den Geschwistern differenziert werden. Im letzten Schritt können diese vier Konstellationen zusätzlich um den Sprachgebrauch mit den Peers oder beim Medienkonsum erweitert werden. Um die spätere Interpretation der Ergebnisse zu vereinfachen, wird in einem Sprachgebrauchsprofil nur jeweils einer der beiden Kontexte, Peers oder Medien, hinzugenommen. Daraus ergeben sich letztendlich zwei Sprachgebrauchsprofile mit jeweils acht möglichen Ausprägungen, die

⁰ERGÄNZUNG ZUM ORIGINALBEITRAG: Die Analysen wurden zusätzlich für die Substrichprobe multilingualer Schülerinnen rekonstruiert. Die Ergebnisse und deren Diskussion finden sich im Anhang 4.8 dieses Kapitels.

Abbildung 4.1: Individuelle Profile des Sprachgebrauchs



Anmerkung: Daten: NEPS, Startkohorte 4. Fallzahlen eingegrenzt auf Personen mit gültigen Angaben in allen gezeigten Kontexten.

in Abb. 4.1 zusammengefasst abgebildet werden. Bei Betrachtung der entsprechenden Fallzahlen zeigt sich, in Übereinstimmung mit den theoretischen Erwartungen, ein kontextübergreifender L2-Gebrauch, wenn mit den Eltern ausschließlich die Zweitsprache verwendet wird. Dementsprechend kann in diesem Fall eine weitere Differenzierung nach dem Sprachgebrauch mit Geschwistern, Peers und beim Medienkonsum vernachlässigt werden. Die vier Ausprägungen werden bei beiden Profilen in der Ausprägung *Eltern L2* zusammengefasst, welche in den multivariaten Analysen als Referenzkategorie herangezogen wird.

Das Vorgehen zur Konstruktion der Sprachgebrauchsprofile spiegelt gleichzeitig den schrittweisen Modellaufbau der darauf folgenden Regressionsanalysen wider. Um zu testen, ob neben der Familiensprache auch weitere Kontexte mit den Zweitsprachkompetenzen in Zusammenhang stehen, wird der Sprachgebrauch schrittweise zunehmend differenziert aufgenommen. Analog zum Aufbau der Sprachgebrauchsprofile stellt die Unterscheidung nach dem Sprachgebrauch mit den Eltern das Ausgangsmodell dar

(Modell 1). Anschließend wird zusätzlich der Sprachgebrauch mit den Geschwistern (Modell 2) und mit Peers (Modell 3) sowie beim Medienkonsum (Modell 4) ergänzt. Durch den Vergleich der Modelle lässt sich somit prüfen, ob sich durch eine differenzierte Betrachtung des Sprachgebrauchs Unterschiede mit Blick auf die Zweitsprachkompetenzen ergeben. Aufgrund des querschnittlichen Studiendesigns können die theoretischen Zusammenhänge zwischen Sprachgebrauch und Sprachkompetenzen zwar nicht kausalanalytisch geprüft werden, jedoch lassen sich Aussagen über die relative Bedeutung weiterer Kontexte in Abgrenzung zur Familiensprache treffen.

Die *Zweitsprachkompetenzen* der Zielpersonen werden anhand der Wortschatz- und Lesefähigkeiten in Deutsch abgebildet. Die beiden Maße ermöglichen es, unterschiedliche Einflüsse des Sprachgebrauchs auf bestimmte Fähigkeitsbereiche des Sprachvermögens zu kontrollieren. Zur Messung der Wortschatzkompetenz wird eine deutsche Version des „Peabody Picture Vocabulary Tests“ (PPVT; Dunn und Dunn 2004) herangezogen. In diesem Hörverstehenstest werden den Teilnehmerinnen und Teilnehmern nacheinander 89 Begriffe vorgelesen, denen sie jeweils eines aus vier vorgegebenen Bildern passend zuordnen sollen. Die Testergebnisse liegen als Summe richtig gelöster Aufgaben vor. Die Lesekompetenz wird durch einen im NEPS entwickelten Leseverständnistest erhoben, bei dem die Zielpersonen aus fünf verschiedenen Texten Informationen ermitteln, reflektieren und textbezogene Schlussfolgerungen ziehen sollen (Gehrer et al. 2012). Der Test setzt sich aus insgesamt 31 Aufgaben zusammen, deren Ergebnisse als gewichtete Maximum-Likelihood-Schätzer vorliegen (WLE; Haberkorn et al. 2012). Zur besseren Vergleichbarkeit der beiden Testinstrumente werden diese anhand der Analysestichprobe standardisiert ($M=0$, $sd=1$).

Der *Generationenstatus* lässt sich durch das Geburtsland der Jugendlichen, ihrer Eltern und Großeltern bestimmen (Olczyk et al. 2014). Zielpersonen mit eigener Migrationserfahrung werden der ersten Generation zugeordnet. Zur zweiten Generation zählen Befragte, die selbst in Deutschland geboren sind, aber mindestens ein im Ausland geborenes Elternteil haben. Dabei werden Fälle mit nur einem zugewanderten Elternteil als 2,5. Generation definiert, sofern zusätzlich mindestens ein Großelternanteil in Deutschland geboren ist. Zur dritten Zuwanderergeneration zählen schließlich Befragte, die selbst ebenso wie beide Eltern in Deutschland geboren sind, aber mindestens zwei der Großeltern zugewandert sind. Für die multivariaten Analysen werden alle in Deutschland geborenen Zielpersonen zusammengefasst und der ersten Generation gegenübergestellt, um Einflüsse einer unterschiedlichen Aufenthaltsdauer auf die Deutschkompetenz zu berücksichtigen.

In den multivariaten Analysen werden zusätzlich das Geschlecht, das Alter (in Monaten), die besuchte Schulform, die kognitiven Fähigkeiten sowie der sozioökonomische Hintergrund kontrolliert. Die kognitiven Fähigkeiten der Schülerinnen und Schüler werden durch einen Test ihrer Schlussfolgerungskompetenz erhoben und in Form eines Summenscores (0–12) berücksichtigt (Haberkorn und Pohl 2013). Der sozioökonomische Hintergrund wird über die höchste Bildung (in Jahren) und berufliche Stellung (ISEI-08; Ganzeboom 2010) der Eltern sowie durch die Anzahl der im Haushalt vorhandenen Bücher erfasst. Tabelle 4.1 gibt einen Überblick über die Randverteilungen der Modellvariablen.

Tabelle 4.1: Randverteilungen der in den multivariaten Analysen verwendeten Variablen (Prozent bzw. Mittelwerte)

	Polen (N=414)	Türkei (N=787)	Ehem. SU (N=668)	Sonstige (N=2213)	Insgesamt (N=4082)	% fehlende Werte
Wortschatz	M (sd)	-0,7 (0,9)	0,0 (0,9)	0,2 (1,0)	0,0 (1,0)	2,9
Leseverständnis	M (sd)	-0,5 (0,9)	0,0 (0,9)	0,1 (1,0)	0,0 (1,0)	8,9
%						10,1
Elternsprache						
Mit Eltern L2 (Ref.)	59,3	11,3	25,8	50,1	40,2	
Mit Eltern L1	40,7	88,7	74,2	49,9	59,8	
%						11,8
Familiensprache						
Mit Eltern L2 (Ref.)	59,3	11,6	26,7	51,0	41,1	
Mit Eltern L1 +						
Geschwister L2	16,9	27,0	29,3	21,4	23,1	
Geschwister L1	23,3	61,4	44,0	27,6	35,8	
%						13,9
Profil Peers						
Mit Eltern L2 (Ref.)	59,9	11,7	26,8	51,2	41,2	
Mit Eltern L1 +						
Geschwister L2/Peers L2	14,3	15,4	20,5	18,0	17,5	
Geschwister L2/Peers L1	2,5	11,5	9,0	3,4	5,6	
Geschwister L1/Peers L2	12,4	16,3	16,2	16,5	16,0	
Geschwister L1/Peers L1	10,9	45,0	27,5	10,9	19,6	
%						12,2
Profil Medien						
Mit Eltern L2 (Ref.)	60,3	12,0	27,6	51,9	42,0	
Mit Eltern L1 +						
Geschwister L2/Medien L2	5,2	6,6	10,0	7,3	7,4	
Geschwister L2/Medien L1	11,1	20,3	18,5	13,6	15,3	
Geschwister L1/Medien L2	3,9	5,7	10,5	4,8	5,8	
Geschwister L1/Medien L1	19,3	55,4	33,4	22,3	29,6	
Selbst eingewandert	15,4	11,2	46,9	17,6	21,1	
%						0,0
Schulform						0,0
Hauptschule (Ref.)	21,4	49,8	39,6	33,6	36,4	
Realschule	28,4	20,2	29,8	22,7	24,0	
Gymnasium	28,0	16,4	18,0	30,2	25,4	
Sonstige Schulform	22,1	13,6	12,6	13,5	14,3	
Männlich	44,5	49,9	48,3	49,0	48,6	
Alter (in Monaten)	M (sd)	183,8 (8,2)	184,7 (8,9)	182,5 (8,6)	183,0 (8,5)	0,0
Kognitive Fähigkeit	M (sd)	7,2 (2,6)	8,3 (2,5)	8,2 (2,6)	8,1 (2,6)	8,8
Bücher im Haushalt (kategorisiert)	M (sd)	2,8 (1,3)	3,2 (1,3)	3,5 (1,6)	3,3 (1,5)	5,2
Berufliche Stellung Eltern (ISEI-08)	M (sd)	34,0 (17,1)	35,5 (17,9)	44,4 (21,6)	41,0 (20,7)	28,3
Bildung Eltern (in Jahren)	M (sd)	10,0 (2,1)	11,7 (2,3)	11,5 (2,5)	11,3 (2,5)	29,8

Anmerkung: Daten: NEPS, Startkohorte 4. Werte werden gewichtet, Fallzahlen ungewichtet angegeben. Durch Rundung können sich Abweichungen von 100% ergeben.

Zur Beantwortung der Forschungsfrage, wie die unterschiedlich stark differenzierten Indikatoren des Sprachgebrauchs mit den Zweitsprachkompetenzen in Verbindung stehen, werden lineare Regressionsmodelle geschätzt. Das geschichtete Stichprobendesign wird durch die Schätzung robuster Standardfehler berücksichtigt. Im Gegensatz zum deskriptiven Überblick beschränkt sich die multivariate Analyse auf Zielpersonen an Regelschulen ($N=4358$), da für die Erhebung an Förderschulen spezielle Sprachtests eingesetzt wurden, welche keinen direkten Vergleich erlauben. Zudem werden Einzelkinder ($N=276$) ausgeschlossen, da für diese keine Angaben zum Sprachgebrauch mit Geschwistern vorliegen. Für die verbleibenden 4082 Fälle werden fehlende Werte auf einzelnen Variablen mittels multipler Imputation berücksichtigt (Allison 2000; White et al. 2011). Dabei werden zunächst 65 Datensätze mit vollständigen Informationen generiert. Die Analysen werden an jedem dieser Datensätze separat durchgeführt und die Ergebnisse anschließend nach der von Rubin (2004) vorgegebenen Regel kombiniert. Die deskriptiven Befunde basieren ausschließlich auf den Originaldaten. Dabei werden gewichtete Prozentwerte berichtet (Aßmann et al. 2012). Die multivariaten Analysen haben sich im Abgleich der nicht gewichteten und gewichteten Ergebnisse als robust erwiesen und werden deshalb ohne Gewichtung angeführt.

4.5 Ergebnisse

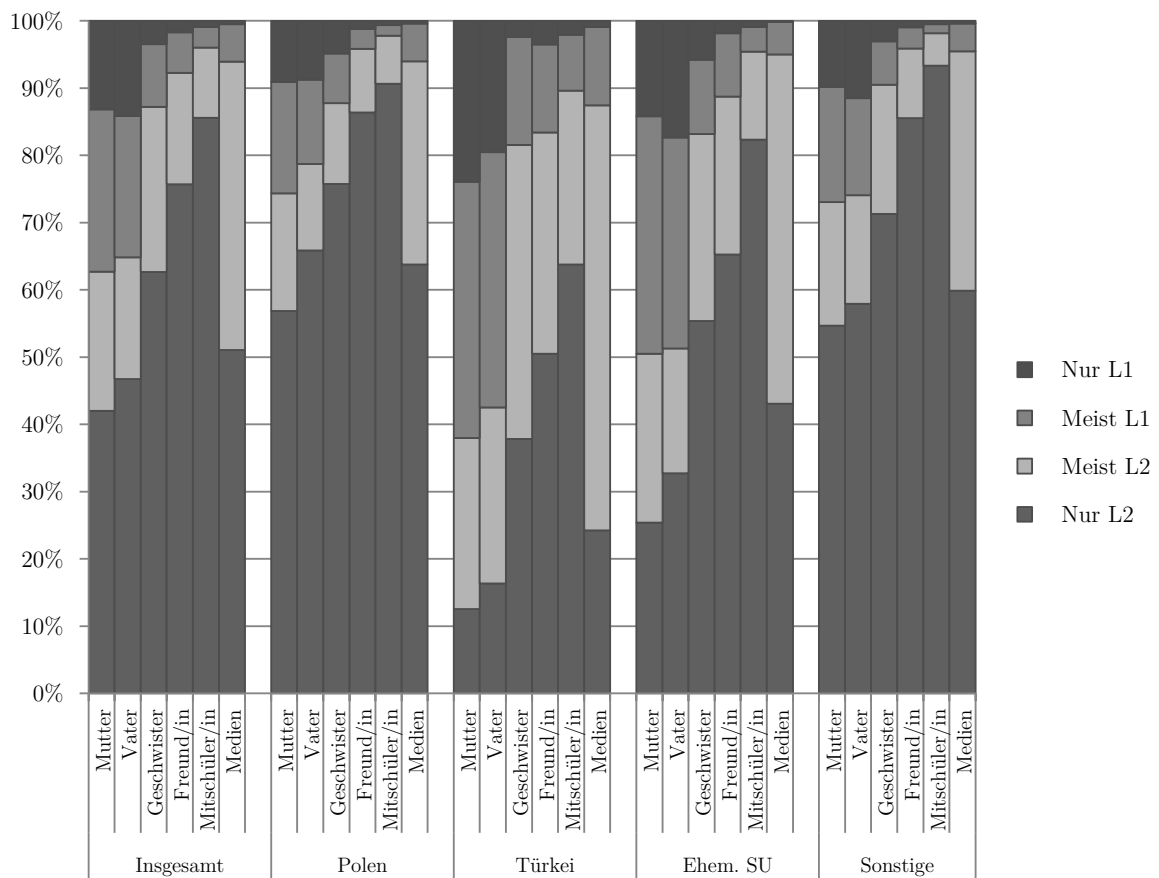
Im Folgenden wird zunächst ein aggregierter Überblick über den Sprachgebrauch in verschiedenen Kontexten nach Herkunftsgruppen (Abschn. 4.5.1) und zusätzlich im Generationenverlauf (Abschn. 4.5.2) gegeben. Anschließend werden die Muster des kontextübergreifenden Sprachgebrauchs anhand der individuellen Sprachgebrauchsprofile veranschaulicht (Abschn. 4.5.3). Darauf aufbauend wird in Abschn. 4.5.4 der Zusammenhang zwischen den Sprachgebrauchsprofilen und dem Zweitspracherwerb multivariat untersucht.

4.5.1 Sprachgebrauch verschiedener Herkunftsgruppen

Abbildung 4.2 zeigt insgesamt und getrennt nach Herkunftsgruppen, welche Sprache(n) die Befragten in den verschiedenen Kontexten verwenden. Insgesamt spricht die Mehrzahl der Jugendlichen über alle betrachteten Kontexte hinweg überwiegend Deutsch. Zwischen den Kontexten lassen sich allerdings klare Unterschiede ausmachen. So wird mit der Mutter oder dem Vater deutlich seltener meist oder nur deutsch verwendet (62,7% bzw. 64,9%) als in der Kommunikation mit Geschwistern (87,2%), der besten Freundin oder dem besten Freund (92,3%), Mitschülerinnen und Mitschülern (96,0%) oder als beim Medienkonsum (94,0%). Besonders ausgeprägt sind die kontextspezifischen Unterschiede, wenn danach differenziert wird, ob Personen ausschließlich die L2 oder auch die L1 verwenden. Dabei fällt auf, dass über die Hälfte der Schülerinnen und Schüler mit den Eltern die Herkunftssprache erhalten. Unterschiede zwischen dem Sprachgebrauch mit dem Vater und der Mutter gibt es kaum: 13,1% sprechen mit der Mutter, 14,2% mit dem Vater nur die L1. Zwischen Geschwistern wird hingegen nur

4 Spracherhalt oder Sprachverlagerung?

Abbildung 4.2: Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland



Daten: NEPS, Startkohorte 4.

bei 3,5% ausschließlich und bei 9,3% der Befragten meist in der L1 kommuniziert.

Die Beobachtung, dass Jugendliche mit Migrationshintergrund häufiger die Sprache des Aufnahmelandes übernehmen als ihre Eltern, deckt sich mit den Ergebnissen bisheriger Forschung (z. B. Alba et al. 2002; Haug 2008, S. 41; Nauck 2007, 2009) und wird in Bezug auf die Kommunikation mit Gleichaltrigen außerhalb der Familie verstärkt deutlich. Hier sprechen 75,7% mit der besten Freundin oder dem besten Freund sowie 85,6% mit den Mitschülerinnen und Mitschülern ausschließlich Deutsch. Dies sollte auch damit zusammenhängen, dass viele Peers keinen Migrationshintergrund haben oder aus anderen Herkunftsländern stammen und die L1 der Zielperson nicht beherrschen. Im Gegensatz zum Peer-Kontext werden beim Medienkonsum oftmals beide Sprachen komplementär verwendet (48,5%), mit einer stärkeren Tendenz zum L2-Gebrauch. Darin spiegeln sich einerseits der durch die fortschreitende Digitalisierung vereinfachte Zugang zu fremdsprachigen Medien und andererseits die Unausweichlichkeit deutscher Medien im Aufnahmekontext wider (Trebbe 2007).

Der im Vergleich zum Familienkontext verstärkte Deutschgebrauch in der Kommunikation mit Gleichaltrigen und beim Medienkonsum ist in allen Herkunftsgruppen

zu beobachten (Abb. 4.2). In Übereinstimmung mit früheren Befunden wird deutlich, dass türkischstämmige Schülerinnen und Schüler über alle Kontexte hinweg häufiger ihre Herkunftssprache erhalten (z. B. Noll und Weick 2011; Strobel und Kristen 2015). Auffällig ist vor allem, dass etwa zwei Drittel der türkischen Zuwanderer mit den Eltern überwiegend die L1 spricht, während bei Befragten aus Gebieten der ehemaligen Sowjetunion jeder Zweite und bei polnischen Zuwanderern nur jeder Fünfte Entsprechendes angibt. Auch beim Medienkonsum, vor allem aber im außerhäuslichen Kontext, nutzen Zugewanderte aus der Türkei verstärkt beide Sprachen komplementär. Beispielsweise verwendet die Hälfte der türkischstämmigen Befragten mit der besten Freundin oder dem besten Freund auch oder nur die Herkunftssprache und 36,2% mit Mitschülerinnen oder Mitschülern. Bei Personen aus der ehemaligen Sowjetunion (34,8% bzw. 17,7%) und besonders bei polnischen Zuwanderern (13,6% bzw. 9,4%) ist der L1-Gebrauch im Peer-Kontext deutlich geringer verbreitet. Der erhöhte Erhalt der Herkunftssprache außerhalb der Familie lässt auf eine vergleichsweise stark ausgeprägte ethnische Segregation türkischer Zuwanderer in den sozialen Netzwerken sowie im Klassenkontext und den damit verbundenen Möglichkeiten zur Verwendung der L1 schließen (Bellin 2009; Friedrichs 2008; Schacht et al. 2014).

4.5.2 Sprachgebrauch im Generationenverlauf

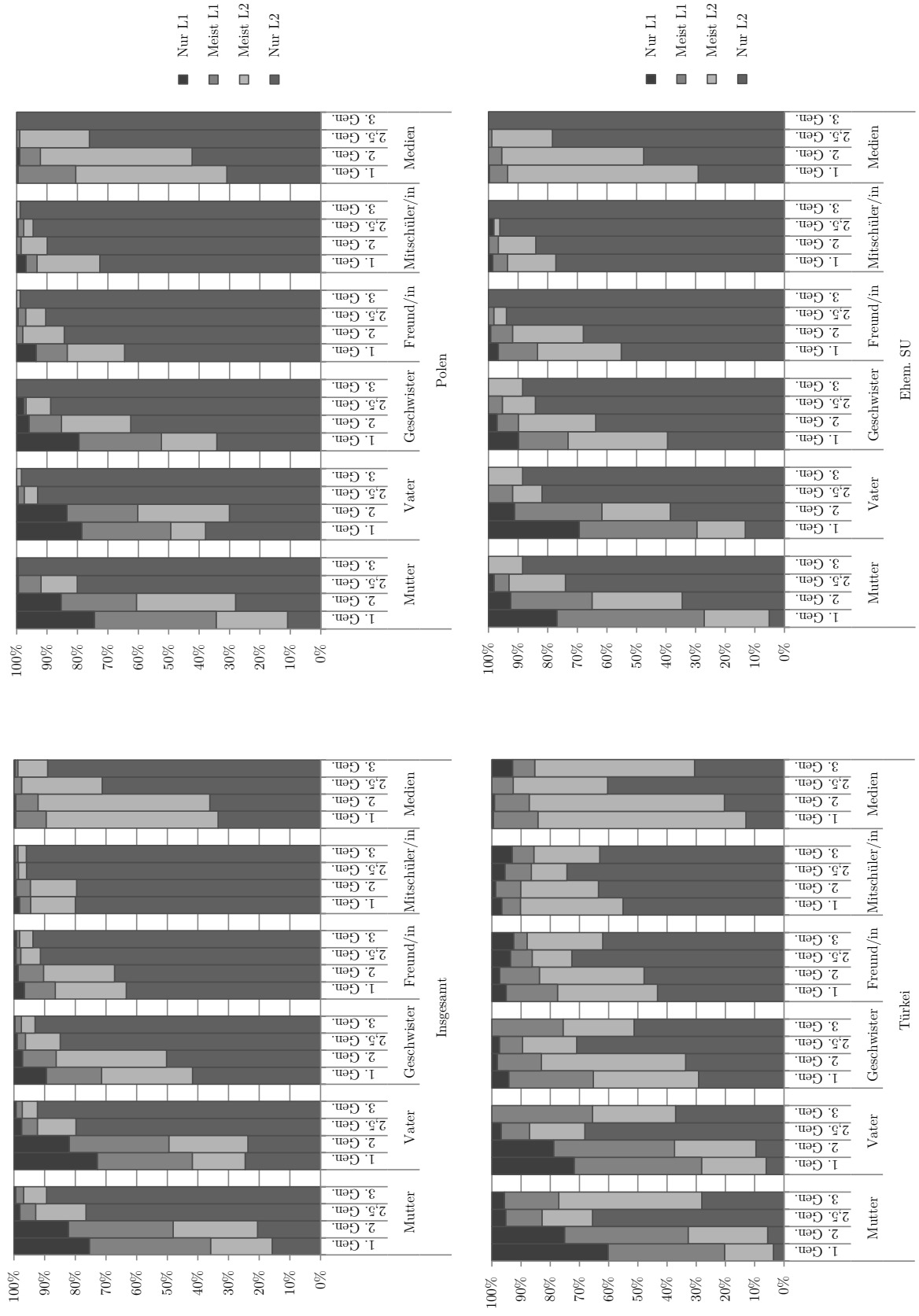
Weitere Einblicke ergeben sich, wenn innerhalb einzelner Sprachkontexte nach der Zuwanderungsgeneration unterschieden wird (Abb. 4.3).⁰ Werden zunächst alle Herkunftsgruppen gemeinsam betrachtet, lässt sich im Generationenverlauf kontextübergreifend eine zunehmende Sprachverlagerung von der L1 zur L2 feststellen. Dieser Befund entspricht den Erwartungen der Assimilationstheorie (Alba et al. 2002; Gordon 1964; Lopéz 1982), konnte jedoch bisher im deutschen Kontext nicht über drei Generationen hinweg belegt werden. Hinweise einer im Generationenverlauf zunehmenden (Rück-)Orientierung an der Herkunftskultur (Portes und Rumbaut 2001, S. 148) lassen sich in Bezug auf den L1-Gebrauch entsprechend nicht bestätigen.

Im Freundeskreis sowie im schulischen Kontext wird bereits in der ersten Generation mehrheitlich ausschließlich Deutsch gesprochen (63,3% bzw. 80,0%). Mit den Eltern hingegen wird sowohl in der ersten als auch in der zweiten Generation überwiegend die Herkunftssprache genutzt. Erst ab der 2,5. Generation, also bei denjenigen, die ein Elternteil ohne Migrationshintergrund haben, verliert die L1 an Bedeutung und wird nur noch von etwa 20% anstatt vorher von rund 80% der Befragten genutzt. Sobald also mindestens ein Elternteil keinen Migrationshintergrund aufweist, findet die Kommunikation mit beiden Eltern nahezu ausschließlich auf Deutsch statt. Ein vergleichbares Muster zeigt sich auch in allen anderen Kontexten, besonders stark jedoch bei der Kommunikation mit Geschwistern und beim Medienkonsum. Diese Zusammenhänge deuten auf einen ausgeprägten Einfluss der Elternsprache auf den Sprachgebrauch in weiteren Kontexten hin.

⁰ERGÄNZUNG ZUM ORIGINALBEITRAG: Eine Übersicht zu den Fallzahlen nach Generation und Herkunftsgruppe befindet sich im Anhang 4.8, Tabelle A4.2.

4 Spracherhalt oder Sprachverlagerung?

Abbildung 4.3: Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland im Generationenverlauf



Daten: NEPS, Startkohorte 4.

Im Vergleich zur Kommunikation mit den Geschwistern und Gleichaltrigen im außerfamiliären Kontext ist der Medienkonsum auch in späteren Generationen noch stark von einem komplementären Gebrauch beider Sprachen geprägt. Zwar konsumieren bereits 89,5% der selbst zugewanderten Jugendlichen überwiegend deutschsprachige Medien, gleichzeitig nutzen jedoch 66,6% zumindest manchmal Medien in der L1. Auch in der 2., 5. und 3. Generation ist der L1-Gebrauch im Vergleich zu anderen Kontexten stärker verbreitet (28,8% bzw. 11,0%).

Werden die Generationenverläufe in einzelnen Herkunftsgruppen betrachtet (Abb. 4.3), fällt auf, dass unter Personen osteuropäischer Abstammung der Anteil derer, die nur in der L2 kommunizieren, kontextübergreifend bereits in der ersten Generation höher ist als bei der türkischen Gruppe. Das Muster des weit verbreiteten L1-Erhalts türkischer Zuwanderer bleibt über alle Generationen hinweg bestehen.⁴ Beispielsweise sprechen Zuwanderer der zweiten Generation aus Polen zu 28,0% und aus der ehemaligen Sowjetunion zu 34,6% ausschließlich die L2 mit ihrer Mutter, während dies nur auf 5,6% der türkischen Zuwanderer zutrifft. Die gleiche Tendenz zeigt sich auch beim Sprachgebrauch mit den anderen Familienmitgliedern. Dies könnte damit zusammenhängen, dass es sich bei vielen Eltern der Befragten aus Polen und den Gebieten der ehemaligen Sowjetunion um (Spät-)Aussiedler oder deren Nachkommen handelt, die teilweise im Herkunftsland die deutsche Sprache erhalten hatten (Dietz 2006; Worbs 2010). Des Weiteren wird Deutsch in diesen Ländern häufig als Fremdsprache an den Schulen unterrichtet (Földes 2003). Folglich ist bei diesen Gruppen eher damit zu rechnen, dass zum Zeitpunkt der Migration bei allen Familienmitgliedern bereits Deutschkompetenzen vorliegen, welche den Gebrauch der L2 ermöglichen.

Weiterhin zeigt sich, dass Zuwanderer der ersten Generation häufiger mit dem Vater als mit ihrer Mutter ausschließlich auf Deutsch kommunizieren. Bei der polnischen Gruppe sind diese Unterschiede besonders stark ausgeprägt. Dies könnte mit geschlechtsspezifischen Unterschieden in der Aufenthaltsdauer und den damit verbundenen Akkulturationsunterschieden zusammenhängen. Viele Väter aus polnischen Familien leben und arbeiten schon länger in Deutschland als die Mütter, welche oft später mit den Kindern nachgeholt wurden (Gostomski 2010, S. 64 ff.).

Zusammenfassend belegen die aggregierten Verteilungen, dass sich der Sprachgebrauch in Abhängigkeit des betrachteten Kontexts äußerst heterogen gestaltet und darüber hinaus auch zwischen Herkunftsgruppen und Generationen stark variiert. Gleichzeitig wird deutlich, dass sich die Kommunikation mit Vater und Mutter teils stark von der mit Geschwistern, mit Peers und beim Medienkonsum abhebt. Dies liefert erste Hinweise darauf, dass Personen mit nicht-deutscher Elternsprache in anderen Kontexten häufiger die L2 verwenden als mit ihren Eltern. In welchen Kontexten diese dann tatsächlich Deutsch nutzen, wird durch die im Folgenden abgebildeten Sprachgebrauchsprofile sichtbar.

⁴Eine Interpretation der Verteilungen in der dritten Generation der Türkischstämmigen ist aufgrund zu geringer Fallzahlen ($N < 10$) nicht sinnvoll.

4.5.3 Individuelle Muster des Sprachgebrauchs: Sprachgebrauchsprofile

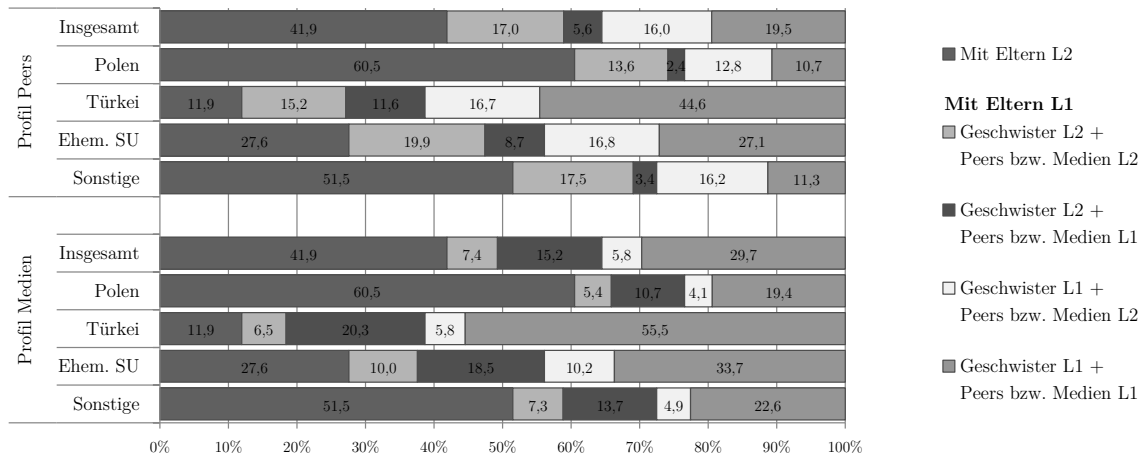
In Abb. 4.4 lässt sich ablesen, dass 41,9% der Zuwanderer mit den Eltern, und damit weitgehend auch in allen anderen Kontexten, ausschließlich Deutsch sprechen. Jugendliche, die mit den Eltern die Herkunftssprache sprechen, nutzen auch in der Kommunikation mit den Geschwistern häufig die L1 (Profil Peers: 16,0%+19,5% bzw. Profil Medien: 5,8%+29,7%=35,5%). 22,6% (17,0%+5,6%, bzw. 7,4%+15,2%) verwenden hier hingegen die L2. Unter denjenigen, die mit Geschwistern die L2 verwenden und denen, die eher L1 sprechen, kann zusätzlich nach dem Sprachgebrauch mit Peers und beim Medienkonsum differenziert werden. In beiden Gruppen findet die Kommunikation mit den Peers überwiegend auf Deutsch statt (17,0%+16,0%=33%), allerdings verwendet auch ein nicht unbeträchtlicher Anteil die L1 (5,6%+19,5%=25,1%). Die mit Geschwistern genutzte Sprache schlägt sich auch im außerfamiliären Sprachgebrauch nieder: Wird in diesem Kontext Deutsch gesprochen, findet diese Sprache auch im Peer-Kontext überproportional häufig Anwendung; Personen, die mit den Geschwistern die L1 sprechen, tun dies auch mehrheitlich im Peer-Kontext. Insgesamt sprechen 19,5% der Schülerinnen und Schüler in allen Kontexten, das heißt sowohl mit den Eltern als auch mit Geschwistern und den Peers, die L1. Bei knapp 40% der Befragten stimmt die Elternsprache jedoch nicht mit dem Sprachgebrauch in den anderen Kontexten überein.

Diese Heterogenität fällt mit Blick auf den Medienkonsum deutlich geringer aus. Jugendliche, die mit ihren Eltern L1 sprechen, konsumieren mehrheitlich auch Medien in ihrer Herkunftssprache (15,2%+29,7%=44,9%). Dies könnte damit zusammenhängen, dass Eltern verstärkt Einfluss auf den Medienkonsum der Kinder nehmen und beispielsweise regelmäßig mit diesen gemeinsam in der L1 Fernsehen (Piga 2007; Worbs 2010).

Vergleicht man die Sprachprofile einzelner Herkunftsgruppen, bestätigt sich der bereits auf Aggregatsebene vermutete starke L1-Erhalt unter türkischen Zuwanderern. Der Anteil an Jugendlichen, die kontextübergreifend nur Deutsch verwenden, fällt hier mit 11,9% um mehr als die Hälfte geringer aus als bei Befragten aus der ehemaligen Sowjetunion und ist etwa fünfmal geringer als bei polnischen Zuwanderern. Gleichzeitig verwenden 44,6% aller türkischstämmigen Jugendlichen die Herkunftssprache neben dem familiären Kontext auch mit den Peers und 55,5% beim Medienkonsum. Diese Anteile fallen bei Zuwanderern aus Polen (10,7% bzw. 19,4%) und der ehemaligen Sowjetunion (27,1% bzw. 33,7%) deutlich geringer aus.

Insgesamt bestätigen die individuellen Sprachgebrauchsprofile, dass sich Zuwanderer, die mit den Eltern die Herkunftssprache sprechen, besonders hinsichtlich der Kommunikation mit den Geschwistern und Peers stark unterscheiden. Dies legt die Vermutung nahe, dass die alleinige Betrachtung der innerhalb der Familie gesprochenen Sprache die tatsächlichen Muster des Sprachgebrauchs nicht vollständig abdeckt.

Abbildung 4.4: Kombinierte Profile des Sprachgebrauchs mit Eltern, Geschwistern und Peers oder Medien nach Herkunftsland (Prozentwerte)



Anmerkungen: Daten: NEPS, Startkohorte 4. Fallzahlen eingegrenzt auf Personen mit gültigen Angaben in allen gezeigten Kontexten. Durch Rundung können sich geringfügige Abweichungen von 100% ergeben.

4.5.4 Zusammenhänge des Sprachgebrauchs mit den Sprachkompetenzen

Um letztlich der Frage nachzugehen, ob die heterogenen Muster des Sprachgebrauchs auch mit Unterschieden in den Zweitsprachkompetenzen einhergehen, werden die Zusammenhänge mit den Wortschatz- und Lesekompetenzen der Schülerinnen und Schüler untersucht (Tab. 4.2). Im ersten Modell werden nur Informationen zur Elternsprache berücksichtigt. In Übereinstimmung mit früheren Studien, die den Sprachgebrauch ausschließlich über die Familiensprache abbilden, weisen Jugendliche, die mit den Eltern die Herkunftssprache nutzen, signifikante Nachteile in der Wortschatzkompetenz (Modell 1a) sowie in reduziertem Umfang auch in der Lesekompetenz auf (Modell 1b).

In Modell 2a und 2b werden Informationen zum Sprachgebrauch mit Geschwistern hinzugenommen. Dadurch wird eine nach Bezugspersonen differenzierte Betrachtung der Familiensprache möglich. Der zusätzliche L2-Gebrauch mit den Geschwistern ist lediglich mit einer marginalen Verbesserung der Wortschatz- und Lesekompetenzen verbunden. Demzufolge macht es unter Jugendlichen mit nichtdeutscher Familiensprache kaum einen Unterschied, in welcher Sprache sie mit den Geschwistern kommunizieren. Gleichzeitig deutet das Ergebnis darauf hin, dass die in früheren Studien mit einer nicht-deutschen Familiensprache assoziierten Nachteile in den Deutschkompetenzen vermutlich vorrangig mit der Eltern-Kind-Kommunikation in Zusammenhang stehen und Geschwister hierbei eine untergeordnete Rolle spielen.

Wird zusätzlich der Sprachgebrauch mit Gleichaltrigen berücksichtigt, ergeben sich differenzielle Muster im Hinblick auf die Zweitsprachkompetenzen (Modell 3a und 3b). Der L2-Gebrauch mit den Peers ist mit Vorteilen im Spracherwerb verknüpft. Dabei profitieren besonders Jugendliche, die mit den Geschwistern die Herkunftssprache ver-

Tabelle 4.2: Lineare Regression der Lese- und Wortschatzkompetenz

	Wortschatz				Leseverständnis			
	Modell 1a	Modell 2a	Modell 3a	Modell 4a	Modell 1b	Modell 2b	Modell 3b	Modell 4b
Selbst eingewandert	-0,15 (0,03) ***	-0,15 (0,03) ***	-0,14 (0,03) ***	-0,15 (0,03) ***	-0,07 (0,03) ***	-0,06 (0,03) ***	-0,06 (0,03) ***	-0,06 (0,03) ***
<i>Herkunftsgruppe (Ref. Türkei)</i>								
Polen	0,51 (0,05) ***	0,51 (0,05) ***	0,49 (0,05) ***	0,51 (0,05) ***	0,25 (0,06) ***	0,25 (0,05) ***	0,23 (0,05) ***	0,25 (0,05) ***
Ehem. SU	0,50 (0,05) ***	0,50 (0,05) ***	0,48 (0,05) ***	0,50 (0,05) ***	0,22 (0,05) ***	0,21 (0,05) ***	0,20 (0,04) ***	0,21 (0,05) ***
Sonstiges Ausland	0,46 (0,04) ***	0,46 (0,04) ***	0,43 (0,04) ***	0,46 (0,04) ***	0,21 (0,04) ***	0,21 (0,04) ***	0,18 (0,04) ***	0,21 (0,04) ***
Kogn. Fähigkeiten	0,06 (0,01) ***	0,06 (0,01) ***	0,06 (0,01) ***	0,06 (0,01) ***	0,10 (0,01) ***	0,10 (0,01) ***	0,10 (0,01) ***	0,10 (0,01) ***
Bücher im Haushalt	0,12 (0,01) ***	0,12 (0,01) ***	0,12 (0,01) ***	0,12 (0,01) ***	0,08 (0,01) ***	0,08 (0,01) ***	0,08 (0,01) ***	0,08 (0,01) ***
Berufliche Stellung Eltern	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
Bildung Eltern	0,01 (0,01)	0,01 (0,01)	0,01 (0,01)	0,01 (0,01)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)	0,00 (0,00)
<i>Sprachgebrauch (Ref. Eltern L2)</i>								
Eltern L1	-0,29 (0,03) ***				-0,16 (0,03) ***			
Eltern L1/Geschwister L2		-0,26 (0,04) ***				-0,14 (0,04) ***		
Eltern L1/Geschwister L1		-0,31 (0,03) ***				-0,17 (0,03) ***		
Eltern L1/Geschwister L2/Peers L2 ^a			-0,25 (0,04) ***				-0,15 (0,04) ***	
Eltern L1/Geschwister L2/Peers L1			-0,31 (0,06) ***				-0,13 (0,06) *	
Eltern L1/Geschwister L1/Peers L2 ^a			-0,23 (0,04) ***				-0,08 (0,04) ***	
Eltern L1/Geschwister L1/Peers L1			-0,38 (0,04) ***				-0,26 (0,04) ***	

Tabelle 4.2: Lineare Regression der Lese- und Wortschatzkompetenz (Fortsetzung)

	Wortschatz				Leseverständnis			
	Modell 1a	Modell 2a	Modell 3a	Modell 4a	Modell 1b	Modell 2b	Modell 3b	Modell 4b
Eltern L1/Geschwister L2/Medien L2				-0,22 (0,06) ***				-0,14 (0,05) **
Eltern L1/Geschwister L2/Medien L1				-0,28 (0,04) ***				-0,14 (0,04) ***
Eltern L1/Geschwister L1/Medien L2				-0,25 (0,06) ***				-0,14 (0,06) *
Eltern L1/Geschwister L1/Medien L1				-0,31 (0,04) ***				-0,17 (0,04) ***
Konstante	-0,67 (0,30) 0,47 *	-0,69 (0,30) 0,47 *	-0,67 (0,30) 0,47 *	-0,69 (0,30) 0,47 *	-0,36 (0,30) 0,42	-0,37 (0,30) 0,43	-0,35 (0,30) 0,42	-0,38 (0,30) 0,42
R ²								

Anmerkung: Daten: NEPS, Startkohorte 4. Robuste Standardfehler in Klammern. N=4082. * p<0,05, ** p<0,01, *** p<0,001. ^a signifikanter Unterschied zu Eltern L1/Geschwister L1/Medien L1 bzw. Peers L1 bei Wortschatz und bei Lesekompetenz. Kompetenzunterschiede zwischen anderen Ausprägungen der Sprachgebrauchsprofile sind nicht signifikant.

4 *Spracherhalt oder Sprachverlagerung?*

wenden, von einem außerfamiliären L2-Gebrauch und schneiden signifikant besser ab als diejenigen, die auch im außerfamiliären Kontext teilweise in der L1 kommunizieren (Modell 3a). Die Nachteile einer nicht-deutschen Familiensprache werden durch den L2-Gebrauch im Peer-Kontext zwar nicht vollständig kompensiert, aber zumindest reduziert. Die stärksten Nachteile zeigen sich erwartungsgemäß bei den Jugendlichen, die in keinem der untersuchten Kontexte ausschließlich die Sprache des Aufnahmelandes nutzen.

Dies zeigt sich auch, wenn der Medienkonsum der Befragten berücksichtigt wird. Im Gegensatz zum Peer-Kontext trägt der Konsum deutschsprachiger Medien nur in sehr geringem Umfang dazu bei, die mit dem familiären L1-Erhalt assoziierten Nachteile zu kompensieren. Die Regressionskoeffizienten weisen zwar tendenziell auf höhere Wortschatzkompetenzen (Modell 4a) durch den Gebrauch deutscher Medien hin, jedoch sind die Unterschiede zu Personen mit einem herkunftssprachlichen Medienkonsum nicht signifikant. Für den Erwerb von Lesekompetenzen (Modell 4b) scheint die Sprache der Mediennutzung keine über die Effekte der Elternsprache hinausgehende Bedeutung zu haben.

Ein Vergleich der Modellgüte (R^2) in den Modellen 1 bis 4 deutet zudem darauf hin, dass eine über die Familiensprache hinausgehende Differenzierung des Sprachgebrauchs nur wenig zur Erklärung von sprachlichen Kompetenzunterschieden beiträgt.

4.6 **Fazit und Diskussion**

Die vorliegende Studie geht der Frage nach, welche Sprache Schülerinnen und Schüler mit Migrationshintergrund in verschiedenen alltagsrelevanten Kontexten verwenden und wie dies mit den Zweitsprachkompetenzen zusammenhängt. Dabei werden, basierend auf einer einheitlichen Datengrundlage, nicht nur herkunftsgruppenspezifische Muster des Sprachgebrauchs in verschiedenen Kontexten herausgearbeitet, sondern auch deren Veränderung über drei Generationen hinweg nachgezeichnet. Als Erweiterung zur bisherigen Forschung, welche den Fokus primär auf die Familiensprache richtet, wird eine differenziertere Sichtweise geboten, die innerfamiliäre Sprachgewohnheiten mit Vater, Mutter und Geschwistern, die Kommunikation mit Gleichaltrigen sowie den Medienkonsum berücksichtigt. Damit greift der Beitrag Kritik an früheren Studien auf, welche durch die Fokussierung auf die Familiensprache andere, ebenfalls wichtige Kontexte vernachlässigen. Dieser Aspekt ist vor allem relevant, wenn es darum geht, den individuellen Sprachgebrauch mit dem Bildungserfolg, insbesondere mit den Zweitsprachkompetenzen, in Verbindung zu bringen. Um die relative Bedeutung einzelner Kontexte im Hinblick auf die Sprachkompetenzen abschätzen zu können, werden sogenannte Sprachgebrauchsprofile erstellt. Diese kombinieren den individuellen Sprachgebrauch über verschiedene Kontexte hinweg und stellen damit einen Indikator für die Gesamtnutzung einer Sprache dar. Ausgehend von einem allgemeinen Modell des Spracherwerbs werden die verschiedenen Ausprägungen dieser Sprachgebrauchsprofile anschließend mit Unterschieden in den Sprachkompetenzen in Verbindung gebracht.

Für die Analysen werden Daten des Nationalen Bildungspanels (NEPS) zu Schüle-

rinnen und Schülern der neunten Klasse herangezogen. Deskriptiv lässt sich bestätigen, dass Jugendliche mit Migrationshintergrund innerhalb der Familie vor allem mit den Eltern die Herkunftssprache erhalten, mit Geschwistern jedoch mehrheitlich Deutsch sprechen. Beim Vergleich des Sprachgebrauchs der Elternteile fällt auf, dass insbesondere in der ersten Generation deutlich häufiger mit dem Vater Deutsch gesprochen wird als mit der Mutter. Dies hängt vermutlich mit der im Schnitt längeren Aufenthaltsdauer der Väter in Deutschland zusammen. Im außerfamiliären Kontext mit Freundinnen und Freunden sowie mit Mitschülerinnen und Mitschülern ist die Herkunftssprache weniger kommunikationsrelevant. Beim Medienkonsum zeichnet sich darüber hinaus eine häufig komplementäre Nutzung beider Sprachen ab, wobei auch hier die Verwendung der deutschen Sprache dominiert. Im Vergleich der Herkunftsgruppen wird deutlich, dass der Erhalt der Herkunftssprache unter Jugendlichen türkischer Herkunft in allen betrachteten Kontexten deutlich verbreiteter ist als bei Zuwanderern aus der ehemaligen Sowjetunion und insbesondere als bei denjenigen aus Polen. Wie auch von der klassischen Assimilationstheorie erwartet wird, kann kontextübergreifend für alle Herkunftsgruppen von der ersten hin zur dritten Generation eine zunehmende Verlagerung des Sprachgebrauchs von der Herkunftssprache zur deutschen Sprache belegt werden. Insbesondere ab der 2,5. Generation, d. h. wenn ein Elternteil ohne Migrationshintergrund vorhanden ist, verliert die Herkunftssprache über alle Kontexte hinweg massiv an Bedeutung. Tendenzen einer „reactive ethnicity“, d. h. einer verstärkten Orientierung an der Herkunftskultur unter Zuwanderern späterer Generationen (Portes und Rumbaut 2001, S. 148), sind damit im Hinblick auf den Sprachgebrauch nicht zu beobachten.

Die Betrachtung der individuellen, kontextübergreifenden Sprachmuster in den Sprachgebrauchsprofilen betont die zentrale Bedeutung der Elternsprache. Wird mit den Eltern nur Deutsch gesprochen, überträgt sich dies auch weitestgehend auf alle anderen Kontexte. Bei Jugendlichen, die mit ihren Eltern die Herkunftssprache sprechen, gestaltet sich der Sprachgebrauch mit Geschwistern, mit Peers und beim Medienkonsum jedoch deutlich heterogener. Zwar stimmt auch in diesem Fall die Elternsprache nicht selten mit dem Sprachgebrauch in anderen Kontexten überein. Besonders die Kommunikation unter Geschwistern und der Medienkonsum orientieren sich häufig an der Elternsprache. Gleichzeitig verwendet jedoch ein erheblicher Anteil dieser Schülerinnen und Schüler mit den Geschwistern und/oder Peers oder beim Medienkonsum ausschließlich die L2. Bei einer nach Herkunftsgruppen differenzierten Betrachtung der Sprachgebrauchsprofile lässt sich der bereits auf Aggregatsebene vermutete erhöhte L1-Erhalt bei Jugendlichen mit türkischem Migrationshintergrund bestätigen. Der Anteil derjenigen, die kontextübergreifend die Herkunftssprache verwenden, ist in dieser Gruppe teilweise mehr als doppelt so hoch wie in den Vergleichsgruppen.

Die heterogenen Muster im Sprachgebrauch stehen allerdings nur bedingt mit den Zweitsprachkompetenzen in Zusammenhang. Die multivariaten Analysen belegen signifikante Defizite in den Wortschatz- und Lesekompetenzen bei Jugendlichen, die mit ihren Eltern die Herkunftssprache verwenden. Diese Nachteile fallen allerdings etwas geringer aus, wenn im Peer-Kontext auf Deutsch kommuniziert wird. Entgegen den Erwartungen lassen sich hinsichtlich des Sprachgebrauchs mit den Geschwistern und

4 *Spracherhalt oder Sprachverlagerung?*

beim Medienkonsum nur marginale Kompetenzunterschiede finden.

Die unterschiedliche Relevanz der betrachteten Kontexte für die Zweitsprachkompetenzen könnte mit Unterschieden im zeitlichen Umfang und der Qualität des sprachlichen Zugangs zusammenhängen. So wäre beispielsweise denkbar, dass Jugendliche mehr Zeit mit ihren Eltern und Peers verbringen und dementsprechend stärker vom Sprachgebrauch in diesen Kontexten beeinflusst werden. Darüber hinaus ist zu bedenken, dass Befragte, die in einem Kontext beide Sprachen verwenden, zwar relativ gesehen, weniger deutsch sprechen als Personen, die einen ausschließlich deutschen Sprachgebrauch aufweisen, die beiden Gruppen sich aber möglicherweise hinsichtlich der absoluten Nutzungsdauer der Zweitsprache kaum unterscheiden.

Im Hinblick auf die Qualität des Zugangs lassen sich in Abhängigkeit der Sprachkompetenzen der Bezugspersonen unterschiedliche Einflüsse auf die sprachlichen Fähigkeiten der Schülerinnen und Schüler vermuten (Place und Hoff 2011). Beispielsweise könnte das sprachliche Niveau in der Kommunikation unter Geschwistern häufiger durch sprachliche Fehler beeinträchtigt sein als der Sprachgebrauch mit oftmals muttersprachlichen Peers und daher nur marginale Vorteile für die Zweitsprachkompetenzen mit sich bringen. Beim Medienkonsum hingegen kann von einer weitestgehend fehlerfreien Vermittlung der deutschen Sprache ausgegangen werden. Die vergleichsweise geringe Bedeutung der Nutzung deutschsprachiger Medien könnte jedoch mit dem intellektuellen Anspruch der vermittelten sprachlichen Inhalte zusammenhängen (vgl. Linebarger und Walker 2005). Ebenso deuten Befunde darauf hin, dass Medien, wie beispielsweise Fernsehen, vergleichsweise wenig ausschlaggebend für den Zweitspracherwerb sind (z. B. Patterson 2002). Tatsächlich zeigen weiterführende Analysen, dass sich auch bei denjenigen Personen, die nicht ausschließlich deutschsprachige Medien konsumieren, der Gebrauch der Herkunftssprache vorwiegend auf das Fernsehen oder das Sehen von Filmen beschränkt, während sie beim Lesen von Büchern oder Zeitungen vorrangig die L2 nutzen (hier nicht angeführt). Darüber hinaus liegen in den verwendeten Daten keine Informationen zum zeitlichen Umfang sowie zur Qualität des sprachlichen Zugangs vor, weswegen diese Annahmen im Rahmen dieser Arbeit nicht näher geprüft werden konnten.

Als weiteres Ergebnis zeigt sich, dass der Sprachgebrauch stärker mit den Wortschatzkenntnissen verknüpft ist als mit den Lesekompetenzen. Diese Beobachtung deckt sich mit Befunden anderer Studien, die darauf hinweisen, dass Jugendliche mit nicht-deutscher Muttersprache häufig Nachteile im Wortschatz, jedoch nicht in anderen sprachlichen Kompetenzbereichen aufweisen (z. B. Oller et al. 2007; Portocarrero et al. 2007). Dies könnte damit zusammenhängen, dass schriftsprachliche Fähigkeiten vorwiegend im Schulunterricht gefördert werden und daher weniger durch die Kommunikation in Familie und Freizeit beeinflusst werden (Söhn 2005, S. 7 f.).

Insgesamt bleibt festzuhalten, dass vor allem die mit den Eltern gesprochene Sprache von zentraler Bedeutung für den Zweitspracherwerb ist. Obwohl der Sprachgebrauch in verschiedenen Kontexten zum Teil sehr heterogen ausfällt, kann der Einfluss der Familiensprache bei Berücksichtigung zusätzlicher Kontexte nur geringfügig relativiert werden. Aus methodischer Sicht stellt die Familiensprache einen verlässlichen Indikator für den kontextübergreifenden Sprachgebrauch dar.

Bei der Interpretation der multivariaten Befunde gilt es einschränkend zu beachten, dass sich die Wirkungsweise des Sprachgebrauchs auf den Spracherwerb sowie etwaige Interdependenzen zwischen beiden Konstrukten auf Basis der herangezogenen Querschnittsdaten nicht kausalanalytisch prüfen lassen. Für einen konkreten Test der theoretischen Argumente sind Wiederholungsmessungen des Sprachgebrauchs sowie der Sprachkompetenzen erforderlich, welche in nachfolgenden Wellen der NEPS-Daten verfügbar sein werden und somit eine Grundlage für weiterführende Forschungsarbeiten bieten. Nichtsdestotrotz lässt sich auch im Querschnitt der relative Stellenwert weiterer Sprachkontexte im Vergleich zur Familiensprache abschätzen, worin das zentrale Anliegen dieses Beitrags besteht.

Ferner ist zu berücksichtigen, dass beim Sprachgebrauch nur unterschieden wird, ob eine Person ausschließlich Deutsch oder auch die Herkunftssprache verwendet. Unterschiede im Ausmaß des herkunftssprachlichen Gebrauchs werden vernachlässigt. Dies ist vor allem dahingehend kritisch zu sehen, dass Jugendliche, die überwiegend Deutsch sprechen, mit jenen gleichgesetzt werden, die mehrheitlich oder ausschließlich die Herkunftssprache verwenden. Um zu prüfen, ob sich in Abhängigkeit der Operationalisierung unterschiedliche Befunde ergeben, wurden die Analysen zusätzlich mit einer anderen Kategorisierung des Sprachgebrauchs (nur/meist Deutsch versus meist/nur L1) repliziert. Die Ergebnisse (hier nicht angeführt) erwiesen sich im Kern als robust. Weiterhin lassen sich die Befunde nur eingeschränkt verallgemeinern, beispielsweise in Bezug auf andere Altersgruppen oder auf den Erwerb oder den Erhalt von Kompetenzen in der Herkunftssprache. Erkenntnissen aus der Bildungsforschung zufolge, ist der familiäre Kontext in der frühen Bildungsphase besonders prägend, verliert aber mit steigendem Alter zunehmend an Bedeutung (z. B. Hillmert und Jacob 2005). Übertragen auf den Anwendungsfall wird demzufolge vermutet, dass der Sprachgebrauch mit den Eltern und Geschwistern vor allem in jungen Jahren ausschlaggebend für den Zweitspracherwerb ist, später jedoch der relative Einfluss der außerfamiliären Kommunikation (z. B. im Freundeskreis) zunimmt.

Aus theoretischer Sicht sollte der familiäre Sprachgebrauch beim Erwerb herkunftssprachlicher Kompetenzen einen höheren Stellenwert einnehmen als beim Zweitspracherwerb, da die Herkunftssprache im Aufnahmekontext in der Regel weder institutionell gefördert wird, noch außerhalb der Familie viele Möglichkeiten bestehen, diese zu erlernen und zu verwenden (Alba et al. 2002, S. 469). Aktuelle Befunde belegen, dass neben der Familiensprache auch die Sprache der Mediennutzung maßgeblich zur Kompetenzentwicklung in der Herkunftssprache von türkischstämmigen Kindergartenkindern beiträgt (Biedinger et al. 2015). Mit Blick auf die Deutschkompetenzen konnte in der vorliegenden Studie hingegen kein Zusammenhang mit dem Medienkonsum festgestellt werden. Wenngleich diese Befunde aufgrund ihrer unterschiedlichen Untersuchungsgrundlage nur bedingt vergleichbar sind, könnte dies darauf hinweisen, dass sich herkunftssprachlicher Medienkonsum nicht nachteilig auf den Zweitspracherwerb auswirkt, gleichzeitig aber Kompetenzen in der Herkunftssprache begünstigt. Vor dem Hintergrund der Debatte um die Vor- und Nachteile des Erhalts der Herkunftssprache sollten diese Zusammenhänge in zukünftigen Studien näher untersucht werden.

4.7 Literatur

- Alba, R., Logan, J., Lutz, A., und Stults, B. (2002). Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants. *Demography*, 39(3):467–484.
- Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. *Sociological Methods and Research*, 28(3):301–309.
- Aßmann, C., Steinhauer, H. W., und Zinn, S. (2012). *Starting cohorts 3 and 4: 5th/9th grade (SC3/SC4). Scientific use file, version 1.0.0. Data manual [supplement]: Weighing*. Bamberg: University of Bamberg, National Educational Panel Study.
- Bauer, T., Epstein, G. S., und Gang, I. N. (2005). Enclaves, language, and the location choice of migrants. *Journal of Population Economics*, 18(4):649–662.
- Baumert, J. und Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Baumert, J., Artelt, C., und Klieme, E., Herausgeber, *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*, Seiten 323–407. Opladen: Leske + Budrich.
- Bellin, N. (2009). *Klassenkomposition, Migrationshintergrund und Leistung: Mehrebenenanalysen zum Sprach- und Leseverständnis von Grundschulern*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Biedinger, N., Becker, B., und Klein, O. (2015). Turkish-language ability of children of immigrants in Germany: Which contexts of exposure influence preschool children’s acquisition of their heritage language? *Ethnic and Racial Studies*, 38(9):1520–1538.
- Blossfeld, H.-P., Roßbach, H.-G., und von Maurice, J. (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS)*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Chiswick, B. R. und Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, 38(3):391–409.
- Diehl, C. und Schnell, R. (2006). “Reactive ethnicity” or “assimilation”? Statements, arguments, and first empirical evidence for labor migrants in Germany. *International Migration Review*, 40(4):786–816.
- Dietz, B. (2006). Aussiedler in Germany: From smooth adaption to tough integration. In Lucassen, L., Feldman, D., und Oltmer, J., Herausgeber, *Paths of integration: Migrants in Western Europe (1880–2004)*, Seiten 116–138. Amsterdam: Amsterdam University Press.
- Dunn, L. M. und Dunn, D. M. (2004). *Peabody Picture Vocabulary Test (PPVT) (Deutsche Version)*. Göttingen: Hogrefe.

- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt am Main.: Campus.
- Fishman, J. A. (1965). Who speaks what language to whom and when? *La linguistique*, 1:67–88.
- Földes, C. (2003). Deutschunterricht und Deutschdidaktik in Ostmitteleuropa. In Bredel, U., Günther, H., Klotz, P., Ossner, J., und Siebert-Ott, G., Herausgeber, *Didaktik der deutschen Sprache*, Band 8236, Seiten 969–982. Paderborn: Schöningh.
- Friedrichs, J. (2008). Ethnische Segregation. In Kalter, F., Herausgeber, *Migration und Integration*, Seiten 380–411. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Ganzeboom, H. (2010). *A new Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007, with an analysis of quality of occupational measurement in ISSP*. Lisbon: Annual Conference of International Social Survey Programme.
- Gärtig, A.-K., Plewnia, A., und Rothe, A. (2011). *Wie Menschen in Deutschland über Sprache denken: Ergebnisse einer bundesweiten Repräsentativerhebung zu aktuellen Spracheinstellungen (IDS Arbeitspapiere und Materialien zur deutschen Sprache, Bd. 40)*. Mannheim: Institut für Deutsche Sprache.
- Gehrer, K., Zimmermann, S., Artelt, C., und Weinert, S. (2012). *The assessment of reading competence (including sample items for grade 5 and 9). Scientific use file 2012, version 1.0.0*. Bamberg: University of Bamberg, National Educational Panel Study.
- Gordon, M. M. (1964). *Assimilation in American life: The role of race, religion, and national origins*. New York: Oxford University Press.
- Gostomski von, C. B. (2010). *Fortschritte der Integration: Zur Situation der fünf größten in Deutschland lebenden Ausländergruppen (Forschungsbericht 8)*. Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Gresch, C. und Kristen, C. (2011). Staatsbürgerschaft oder Migrationshintergrund. *Zeitschrift für Soziologie*, 40(3):208–227.
- Haberkorn, K. und Pohl, S. (2013). *Cognitive basic skills – Data in the scientific use file*. Bamberg: University of Bamberg, National Educational Panel Study.
- Haberkorn, K., Pohl, S., Hardt, K., und Wiegand, E. (2012). *NEPS technical report for reading – Scaling results of starting cohort 4 in ninth grade (NEPS Working Paper 16.)*. Bamberg: University of Bamberg, National Educational Panel Study.
- Hans, S. (2010). *Assimilation oder Segregation? Anpassungsprozesse von Einwanderern in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.

4 Spracherhalt oder Sprachverlagerung?

- Haug, S. (2008). *Sprachliche Integration von Migranten in Deutschland (Working Paper 14)*. Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Hillmert, S. und Jacob, M. (2005). Institutionelle Strukturierung und inter-individuelle Variation. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 57(3):414–442.
- Jacob, K. und Kalter, F. (2013). Intergenerational change in religious salience among immigrant families in four European countries. *International Migration*, 51(3):38–56.
- Konsortium Bildungsberichterstattung. (2006). *Bildung in Deutschland: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung und Migration*. Bielefeld: Bundesministerium für Bildung und Forschung.
- Kristen, C. (2008). Schulische Leistungen von Kindern aus türkischen Familien am Ende der Grundschulzeit. In Kalter, F., Herausgeber, *Migration und Integration*, Seiten 230–251. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kristen, C., Edele, A., Kalter, F., Kogen, I., Schulz, B., Stanat, P., und Will, G. (2011). The education of migrants and their children across the life course. In Blossfeld, H.-P., Roßbach, H.-G., und von Maurice, J., Herausgeber, *Education as a lifelong process: The German National Educational Panel Study (NEPS)*, Seiten 121–138. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Linebarger, D. L. und Walker, D. (2005). Infants' and toddlers' television viewing and language outcomes. *American Behavioral Scientist*, 48(5):624–645.
- Lopéz, D. E. (1982). *The maintenance of Spanish over three generations in the United States*. Los Alamitos: National Center for Bilingual Research.
- Mouw, T. und Xie, Y. (1999). Bilingualism and the academic achievement of first- and second-generation Asian Americans: Accommodation with or without assimilation? *American Sociological Review*, 64(2):232–252.
- Müller, A. G. und Stanat, P. (2006). Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Zuwanderern aus der ehemaligen Sowjetunion und aus der Türkei. In Baumert, J., Stanat, P., und Watermann, R., Herausgeber, *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*, Seiten 221–255. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Nauck, B. (2007). Immigrant families in Germany: Family change between situational adaptation, acculturation, segregation and remigration. *Zeitschrift für Familienforschung*, 19(1):34–54.

- Nauk, B. (2009). Intergenerational transmission, social capital, and interethnic contact in immigrant families. In Schönpflug, U., Herausgeber, *Cultural transmission: Psychological, developmental, social, and methodological aspects*, Seiten 161–184. Cambridge: Cambridge University Press.
- Noll, H.-H. und Weick, S. (2011). Zuwanderer mit türkischem Migrationshintergrund schlechter integriert: Indikatoren und Analysen zur Integration von Migranten in Deutschland. *Informationsdienst Soziale Indikatoren*, 46:1–6.
- Olczyk, M., Will, G., und Kristen, C. (2014). *Immigrants in the NEPS: Identifying generation status and group of origin (NEPS Working Paper 41a.)*. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.
- Oller, D. K., Pearson, B. Z., und Cobo-Lewis, A. B. (2007). Profile effects in early bilingual language and literacy. *Applied Psycholinguistics*, 28(02):191–230.
- Patterson, J. L. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Applied Psycholinguistics*, 23(04):493–508.
- Piga, A. (2007). Mediennutzung von Migranten: Ein Forschungsüberblick. In Bonfadelli, H. und Moser, H., Herausgeber, *Medien und Migration: Europa als multikultureller Raum?*, Seiten 95–116. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Place, S. und Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, 82(6):1834–1849.
- Portes, A. und Hao, L. (2002). The price of uniformity: Language, family and personality adjustment in the immigrant second generation. *Ethnic and Racial Studies*, 25(6):889–912.
- Portes, A. und Rumbaut, R. G. (2001). *Legacies: The story of the immigrant second generation*. Berkeley: University of California Press.
- Portocarrero, J. S., Burright, R. G., und Donovan, P. J. (2007). Vocabulary and verbal fluency of bilingual and monolingual college students. *Archives of Clinical Neuropsychology*, 22(3):415–422.
- Ramm, G., Walter, O., Heidemeier, H., und Prenze, M. (2005). Soziokulturelle Herkunft und Migration im Ländervergleich. In Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., und Schiefele, U., Herausgeber, *PISA 2003: Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?*, Seiten 269–297. Münster: Waxmann.
- Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys*. Hoboken: Wiley-Interscience.

4 Spracherhalt oder Sprachverlagerung?

- Salentin, K. und Wilkening, F. (2003). Ausländer, Eingebürgerte und das Problem einer realistischen Zuwanderer-Integrationsbilanz. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55(2):278–298.
- Schacht, D., Kristen, C., und Tucci, I. (2014). Interethnische Freundschaften in Deutschland. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(3):445–458.
- Segeritz, M., Walter, O., und Stanat, P. (2010). Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62(1):113–138.
- Simon, E. und Neuwöhner, U. (2011). Medien und Migranten 2011: Zielsetzung, Konzeption und Basisdaten einer repräsentativen Untersuchung der ARD/ZDF-Medienkommission. *Media Perspektiven*, 10:458–470.
- Skopek, J., Pink, S., und Bela, D. (2013). *Starting cohort 4: Grade 9 (SC4). Scientific use file, version 1.1.0. Data manual*. Bamberg: University of Bamberg, National Educational Panel Study.
- Söhn, J. (2005). *Zweisprachiger Schulunterricht für Migrantenkinder: Ergebnisse der Evaluationsforschung zu seinen Auswirkungen auf Zweitspracherwerb und Schulerfolg (AKI-Forschungsbilanz 2)*. Berlin: Wissenschaftszentrum für Sozialforschung.
- Strobel, B. und Kristen, C. (2015). Erhalt der Herkunftssprache? – Muster des Sprachgebrauchs in Migrantenfamilien. *Zeitschrift für Erziehungswissenschaft*, 18(1):125–142.
- Trebbe, J. (2007). Akkulturation und Mediennutzung von türkischen Jugendlichen in Deutschland. In Bonfadelli, H. und Moser, H., Herausgeber, *Medien und Migration: Europa als multikultureller Raum?*, Seiten 183–208. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Van Tubergen, F. und Kalmijn, M. (2009). Language proficiency and usage among immigrants in the Netherlands: Incentives or opportunities? *European Sociological Review*, 25(2):169–182.
- Verwiebe, R. und Riederer, B. (2013). Die Lesekompetenzen von Jugendlichen mit Migrationshintergrund in westlichen Gesellschaften. *Zeitschrift für Soziologie*, 42(3):201–221.
- Walter, O. (2009). Herkunftsassoziierte Disparitäten im Lesen, der Mathematik und den Naturwissenschaften: Ein Vergleich zwischen PISA 2000, PISA 2003 und PISA 2006. In Prenzel, M. und Baumert, J., Herausgeber, *Vertiefende Analysen zu PISA 2006*, Seiten 149–168. Wiesbaden: VS Verlag für Sozialwissenschaften.
- White, I. R., Royston, P., und Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399.
- Worbs, S. (2010). *Mediennutzung von Migranten in Deutschland (Integrationsreport 8)*. Nürnberg: Bundesamt für Migration und Flüchtlinge.

4.8 Anhang

Der Anhang und die daran enthaltenen Zusatzauswertungen sind nicht Teil des in der Kölner Zeitschrift für Soziologie und Sozialpsychologie veröffentlichten Originalbeitrags.

Rekonstruktion der Analysen für die Population von multilingualen Schüler:innen

Im Rahmen der vorliegenden Untersuchung wurden auch Fälle berücksichtigt, die neben Deutsch keine weitere Sprache gelernt haben. Das Ziel des Beitrags war es ein möglichst umfassendes Bild des Sprachgebrauchs in der Population von *Schüler:innen mit Migrationshintergrund* zu erhalten. Für eine akkurate Beschreibung des Anwendungsfalls ist die Information von zentraler Bedeutung, welche Schüler:innen überhaupt eine andere Sprache als Kind gelernt haben. Auch für die Analysen zu Unterschieden in den Deutschkompetenzen in Abhängigkeit des Sprachgebrauchs, bildet die Gruppe der Schüler:innen mit Migrationshintergrund, die nur Deutsch gelernt haben, eine wichtige Referenz und knüpft an die bisherige Forschung zu ethnischen Ungleichheiten an.

Gleichzeitig stellt sich die Frage, ob die Berücksichtigung dieser Population nicht das Gesamtbild verzerrt, insbesondere mit Blick auf den relativen Stellenwert des kontextübergreifenden L2-Gebrauchs und/oder bei einem spezifischen Interesse an der Population von *multilingualen Schüler:innen*, die Deutsch als Zweitsprache lernen. Um dieser Frage nachzugehen und um die Ergebnisse der vorliegenden Arbeit zusätzlich abzusichern, wurden die Analysen für die Substichprobe multilingualer Schüler:innen ($N=3302$) rekonstruiert.

Der Anteil der Schüler:innen, die mit den Eltern bzw. kontextübergreifend nur die L2 verwenden, fällt in dieser Population erwartungsgemäß deutlich geringer aus ($N=161$; 5,8%). Infolgedessen verschiebt sich in den deskriptiven Analysen der relative Anteil der verschiedenen Ausprägungen des Sprachgebrauchs innerhalb einzelner Kontexte bzw. Profile und Gruppen. Dies bedeutet, dass die Abstufungen des L1-Gebrauchs („nur L1“, „meist L1“ und „meist L2“) gegenüber „nur L2“ relativ an Bedeutung gewinnen (vgl. Abb. A4.1–A4.3). Die Verhältnisse zwischen den Abstufungen des L1-Gebrauchs, sowie zwischen den Kontexten bzw. Profilen und Gruppen, welche im Vordergrund der vorliegenden Analysen stehen, bleiben davon hingegen unberührt. Die Befundmuster hierzu sind identisch.

Es gilt jedoch zu beachten, dass sich die generationale Zusammensetzung zwischen den zwei Populationen unterscheidet: Multilinguale Schüler:innen gehören anteilig deutlich häufiger der 1. und 2. Generation und sehr selten der 3. Generation an (vgl. Tab. A4.2). Darin spiegelt sich der allgemeine Trend der Sprachverlagerung von L1 zu L2 im Generationenverlauf wider. Aufgrund der unterschiedlichen Zusammensetzung fallen die Ergebnisse zum Vergleich der Generationen (Abb. A4.2) teilweise unterschiedlich aus, insbesondere hinsichtlich der 2,5. und 3. Generation. Die maßgeblichen Tendenzen, auch im Vergleich der Herkunftsgruppen, lassen sich aber bei multilingualen Schüler:innen ebenso finden.

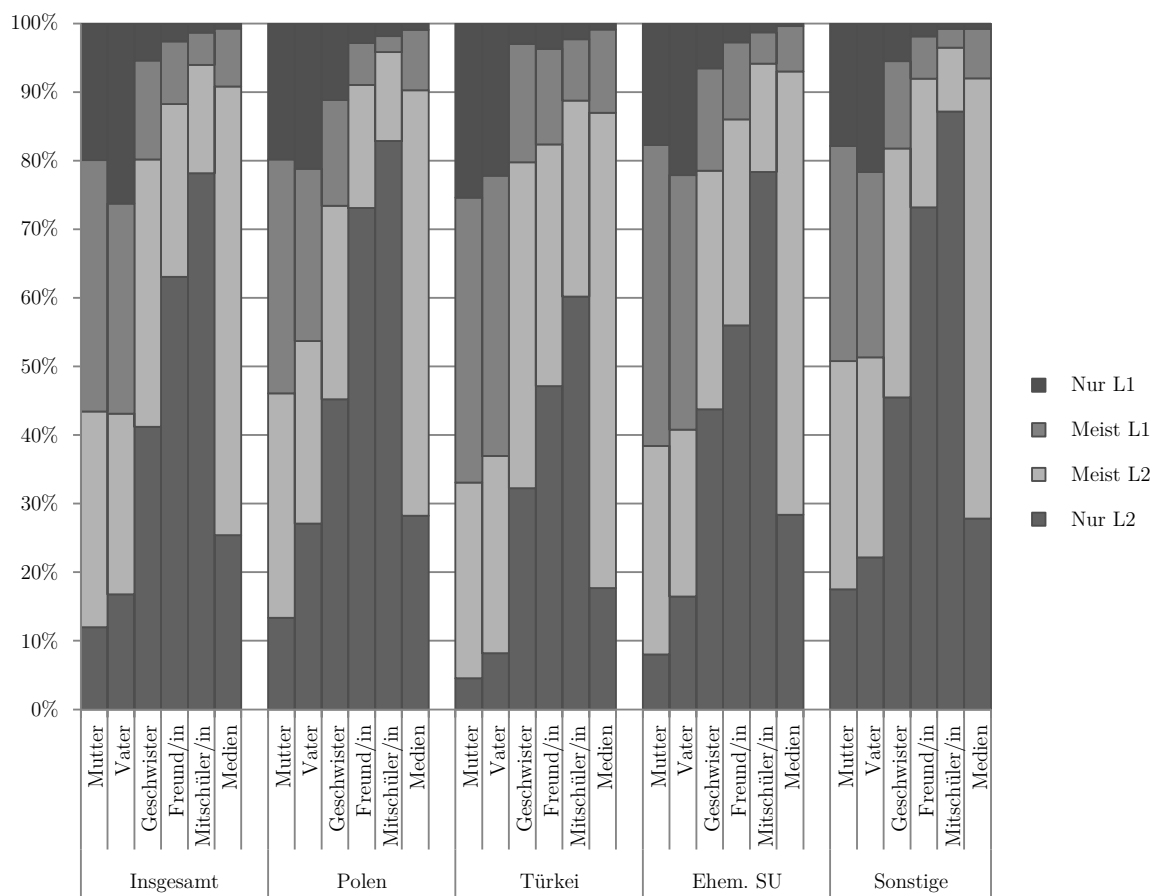
Ein ähnliches Bild zeigt sich bei den Regressionsanalysen zum Zusammenhang zwischen Sprachgebrauch und Deutschkompetenzen (vgl. Tab. A4.1). Infolge der veränderten Referenzkategorie – nur L2 unter multilingualen Schüler:innen versus nur L2 aller Schüler:innen mit Migrationshintergrund – fallen die Unterschiede innerhalb der Gruppe multilingualer Schüler:innen im Vergleich zur Referenz „nur L2“ zwar deutlich geringer aus. Aber auch hier bleiben die relativen Verhältnisse zwischen den Ausprä-

gungen der Profile – um die es im Kern geht – unverändert.

Insgesamt ergeben sich im Vergleich der Populationen keine substantiellen Unterschiede hinsichtlich der zentralen Befundmuster und dementsprechend auch der Schlussfolgerungen, die aus der Untersuchung gezogen werden können. Die Befunde zu multilingualen Schüler:innen bieten darüber hinaus eine nützliche Ergänzung zu den Untersuchungen in Kapitel 5 und 7, in denen speziell diese Population betrachtet wird.

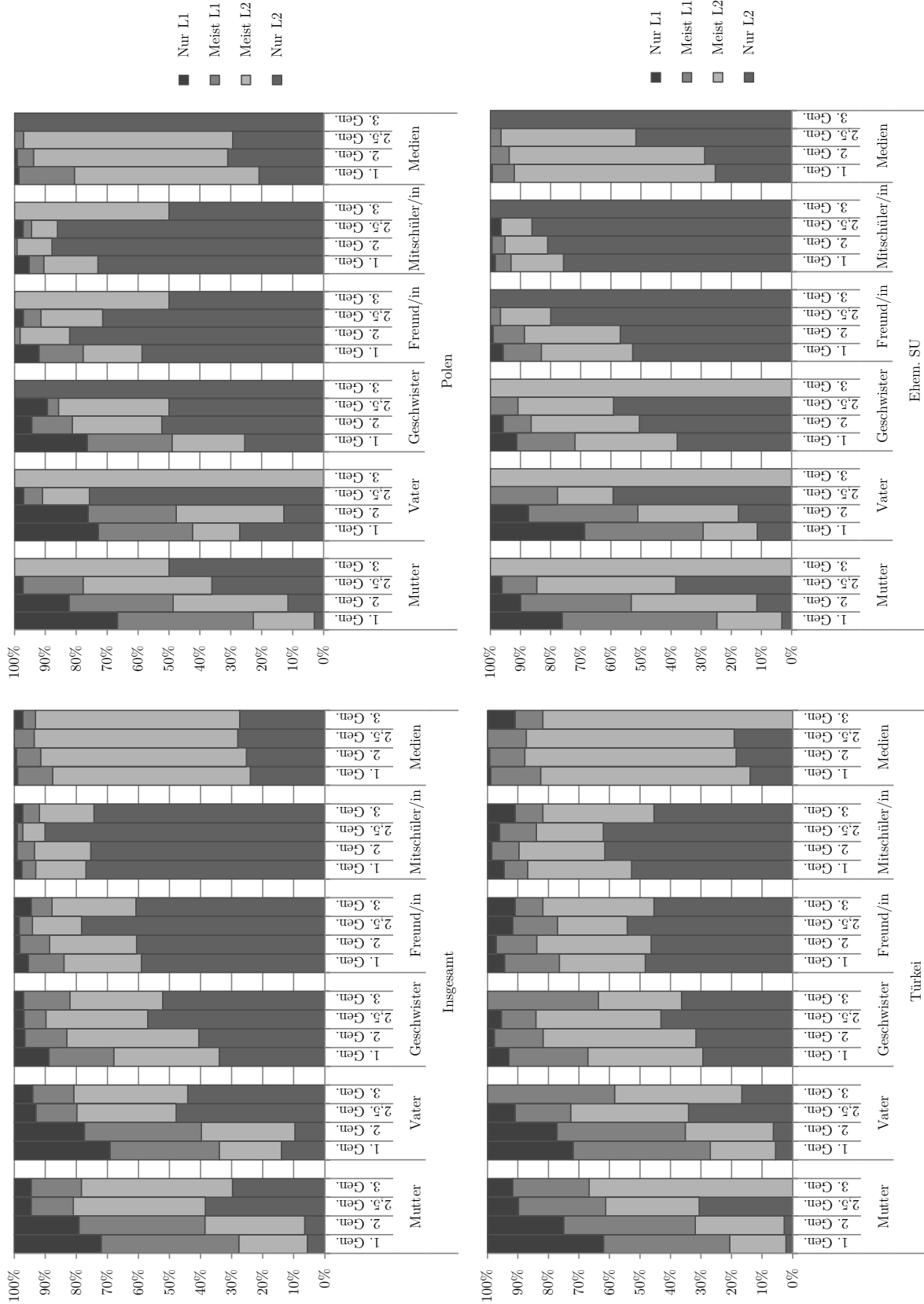
4 Spracherhalt oder Sprachverlagerung?

Abbildung A4.1: Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland (nur multilinguale Schüler:innen)



Daten: NEPS, Startkohorte 4.

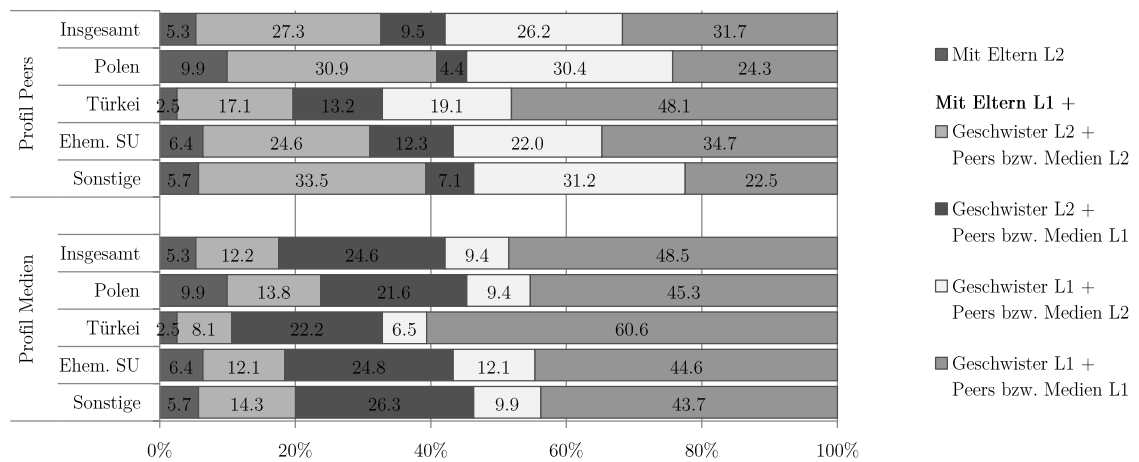
Abbildung A4.2: Sprachgebrauch in verschiedenen Kontexten nach Herkunftsland im Generationsverlauf (nur multilinguale Schüler:innen)



Daten: NEPS, Startkohorte 4. Die Fallzahlen zu den Herkunftsgruppen und Generationen sind in Tab. A4.2 enthalten.

4 Spracherhalt oder Sprachverlagerung?

Abbildung A4.3: Kombinierte Profile des Sprachgebrauchs mit Eltern, Geschwistern und Peers oder Medien nach Herkunftsland (nur multilinguale Schüler:innen)



Anmerkungen: Daten: NEPS, Startkohorte 4. Fallzahlen eingegrenzt auf Personen mit gültigen Angaben in allen gezeigten Kontexten ($N=2384$). Durch Rundung können sich geringfügige Abweichungen von 100% ergeben.

Tabelle A4.1: Lineare Regression der Lese- und Wortschatzkompetenz (nur multilinguale Schüler:innen)

	Wortschatz				Leseverständnis			
	Modell 1a	Modell 2a	Modell 3a	Modell 4a	Modell 1b	Modell 2b	Modell 3b	Modell 4b
Selbst eingewandert	-0,14 (0,04)	*** -0,13 (0,04)	*** -0,13 (0,04)	*** -0,13 (0,04)	*** -0,04 (0,04)	*** -0,03 (0,04)	*** -0,03 (0,04)	*** -0,03 (0,04)
<i>Herkunftsgruppe (Ref. Türkei)</i>								
Polen	0,46 (0,07)	*** 0,45 (0,07)	*** 0,42 (0,07)	*** 0,45 (0,07)	*** 0,31 (0,07)	*** 0,31 (0,07)	*** 0,28 (0,07)	*** 0,31 (0,07)
Ehem. SU	0,53 (0,05)	*** 0,52 (0,05)	*** 0,51 (0,05)	*** 0,52 (0,05)	*** 0,22 (0,05)	*** 0,21 (0,05)	*** 0,20 (0,05)	*** 0,21 (0,05)
Sonstiges Ausland	0,47 (0,04)	*** 0,47 (0,04)	*** 0,44 (0,04)	*** 0,47 (0,04)	*** 0,20 (0,05)	*** 0,20 (0,05)	*** 0,16 (0,04)	*** 0,20 (0,05)
<i>Sprachgebrauch (Ref. Eltern L2)</i>								
Eltern L1	-0,13 (0,06)	*			-0,09 (0,07)			
Eltern L1/Geschwister L2		-0,12 (0,06)			-0,08 (0,07)			
Eltern L1/Geschwister L1		-0,16 (0,06)	*		-0,12 (0,07)			
Eltern L1/Geschwister L2/Peers L2 ^a			-0,10 (0,06)			-0,09 (0,07)		
Eltern L1/Geschwister L2/Peers L1			-0,17 (0,07)	*		-0,08 (0,08)		
Eltern L1/Geschwister L1/Peers L2 ^a			-0,08 (0,06)			-0,02 (0,07)		
Eltern L1/Geschwister L1/Peers L1			-0,22 (0,06)	***		-0,22 (0,07)	***	
Eltern L1/Geschwister L2/Medien L2							-0,09 (0,07)	
Eltern L1/Geschwister L2/Medien L1							-0,09 (0,07)	
Eltern L1/Geschwister L1/Medien L2							-0,09 (0,07)	
Eltern L1/Geschwister L1/Medien L1							-0,12 (0,08)	
Konstante	-0,25 (0,37)	-0,25 (0,37)	-0,24 (0,36)	-0,24 (0,36)	-0,07 (0,37)	-0,08 (0,37)	-0,05 (0,37)	-0,07 (0,37)
R ²	0,40	0,40	0,41	0,40	0,38	0,38	0,39	0,38

Anmerkung: Daten: NEPS, Startkohorte 4. Robuste Standardfehler in Klammern. $N=2744$. Zusätzliche Variablen analog zu Tab. 4.2 kontrolliert.

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

^a signifikanter Unterschied zu ElternL1/Geschwister L1/Medien L1 bzw. Peers L1 bei Wortschatz und bei Lesekompetenz. Kompetenzunterschiede zwischen anderen Ausprägungen der Sprachgebrauchsprofile sind nicht signifikant.

4 Spracherhalt oder Sprachverlagerung?

Tabelle A4.2: Verteilung Generationenstatus nach Herkunftsgruppe (Fallzahlen und Prozent)

	Polen		Türkei		Ehem. SU		Sonstige		Insgesamt	
Schüler:innen mit Migrationshintergrund										
1. Generation	83	(17,3%)	106	(11,7%)	387	(51,0%)	459	(17,8%)	1035	(21,9%)
2. Generation	155	(32,3%)	673	(74,3%)	289	(38,1%)	848	(32,9%)	1965	(41,6%)
2,5. Generation	131	(27,3%)	105	(11,6%)	73	(9,6%)	926	(36,0%)	1235	(26,2%)
3. Generation	111	(23,1%)	22	(2,4%)	10	(1,3%)	343	(13,3%)	486	(10,3%)
<i>Total</i>	480	(100,0%)	906	(100,0%)	759	(100,0%)	2576	(100,0%)	4721	(100,0%)
Multilinguale Schüler:innen										
1. Generation	78	(31,3%)	105	(12,5%)	379	(58,7%)	377	(24,0%)	939	(28,4%)
2. Generation	125	(50,2%)	657	(78,4%)	229	(35,5%)	712	(45,4%)	1723	(52,2%)
2,5. Generation	41	(16,5%)	60	(7,2%)	36	(5,6%)	405	(25,8%)	542	(16,4%)
3. Generation	5	(2,0%)	16	(1,9%)	2	(0,3%)	75	(4,8%)	98	(3,0%)
<i>Total</i>	249	(100,0%)	838	(100,0%)	646	(100,0%)	1569	(100,0%)	3302	(100,0%)

Anmerkung: Daten: NEPS, Startkohorte 4. Durch Rundung können sich Abweichungen von 100% ergeben. Die Verteilung wird für die in den deskriptiven Analysen des Beitrags verwendete Stichprobe aller Schüler:innen mit Migrationshintergrund ($N=4721$) sowie für die Substichprobe von multilingualen Schüler:innen ($N=3302$) aus den Zusatzanalysen im Anhang 4.8 ausgewiesen.

5 *Artikel 2: Ethnic classroom composition and minority language use among classmates: Do peers matter for students' language achievement?*

Dieser Beitrag entstand in Kooperation mit Camilla Rjosk und Petra Stanat und ist in *European Sociological Review* (©The Author(s) 2020. Published by Oxford University Press. All rights reserved.) erschienen. Der nachfolgende Text und Inhalt wurde unverändert aus dem veröffentlichten Originalbeitrag übernommen und an das Format der Dissertation angepasst. Die Referenz lautet:

Seuring, J., Rjosk, C. und Stanat, P. (2020). Ethnic classroom composition and minority language use among classmates: Do peers matter for students' language achievement? *European Sociological Review*, 36(6):920–936. <https://doi.org/10.1093/esr/jcaa022>

Abstract

This article examines the relationship between ethnic classroom composition and students' language-related achievement. We investigate whether minority language use among classmates accounts for effects of ethnic composition on minority students' German reading comprehension and their proficiency in the minority languages Russian and Turkish. We employ multilevel models using cross-sectional data from a sample of ninth-grade students participating in the German National Educational Panel Study. Our findings indicate that students' minority language background rather than their ethnic origin accounts for ethnic composition effects. We find a negative relationship between the ethno-lingual classroom composition and students' German reading comprehension, but the association is small and limited to minority students. Moreover, the ethno-lingual classroom composition is positively correlated with minority language proficiency, specifically among Turkish-speaking students. These associations are substantially reduced after controlling for students' minority language use with their classmates, indicating that a higher proportion of minority language students in a classroom provides additional opportunities for acquiring or maintaining higher levels of proficiency in the minority language. Overall, the ethnic classroom composition does not appear to substantially reinforce existing inequalities between minority and majority students.

5.1 Introduction

International large-scale assessment studies commonly report an achievement gap between ethnic minority¹ and majority students. This gap is particularly pronounced in language-related achievement such as reading comprehension (OECD 2010). These ethnic disadvantages are typically related to differences between minority and majority students' resources such as resources associated with socioeconomic status (SES) and proficiency in the language of instruction (Heath and Brinbaum 2007). In addition to these individual characteristics, recent research on educational achievement has analysed the role of the learning context such as the ethnic composition of schools and classrooms.

The average socioeconomic and prior achievement composition of schools and classrooms has repeatedly been shown to affect individual students' educational achievement (for an overview, see Van Ewijk and Sleegers 2010a). For example, students in classrooms with a high average SES achieve better educational outcomes than comparable students in classrooms with a lower average SES. However, whether the ethnic classroom composition is also related to students' educational achievement over and above socioeconomic composition remains unclear, as the evidence is mixed. This issue is highly relevant to the question of whether ethnically segregated classrooms reinforce existing inequalities between minority and majority students.

Moreover, previous research has not elucidated the potential factors that may account for ethnic composition effects (Driessen 2002; Van Ewijk and Sleegers 2010b). Although several theoretical assumptions have been developed over the last decades, only a few studies have attempted to address them empirically (Agirdag, Van Houtte and Van Avermaet 2012; Rjosk et al. 2014). Identifying these factors is important for understanding the processes underlying the associations between the learning context and individuals' educational achievement, which could also help guide interventions aiming to reduce achievement gaps.

One hypothesis is that minority language use in the school context may affect ethnic minority students' achievement (Van Ewijk and Sleegers 2010b). A high proportion of minority language students in a classroom may increase the likelihood of minority language use among ethnic minority students. The resulting lack of exposure to the instructional language may subsequently impede its development and, thus, more generally, educational achievement (Driessen 2002; Peetsma et al. 2006). At the same time, minority language use should reinforce the maintenance and further development of proficiency in the minority language (Van Tubergen and Mentjox 2014).

The present study addresses this hypothesized mechanism by investigating whether minority language use with classmates mediates the association between ethnic classroom composition and students' language-related achievement. Using data of ninth-grade students from the German National Educational Panel Study, we examine the interrelations among ethnic classroom composition, minority language use with class-

¹When referring to ethnic minority students, we include both students who were born in a different country and students who were born in Germany but whose parents or grandparents were born abroad.

mates, and students' German reading comprehension as well as minority students' language proficiency in Russian and Turkish, which are the languages spoken by the two largest minority groups in Germany.

Our study contributes to the research on ethnic composition effects in several ways. Some previous findings have been based on small or selective samples, especially within the German context. We replicate and evaluate these findings by using large-scale data from a nationwide school sample, which can yield more robust estimates of composition effects. Moreover, the bulk of existing studies has solely examined the association between ethnic classroom composition and educational achievement but has not empirically addressed the underlying mechanisms. We extend the scope of previous analyses by introducing students' self-reported minority language use with classmates as a covariate and attempt to disentangle the potential mechanisms that could transmit the effects. Furthermore, very few studies have analysed the relationship between ethnic classroom composition and minority language proficiency. Investigating minority language proficiency serves as an additional evaluation of the assumption that minority language use may mediate ethnic classroom composition effects. Because students' minority languages are rarely part of the school curriculum, other factors of the classroom context cannot account for this composition effect. Accordingly, our results provide new insight into the potential mechanisms underlying the effects of ethnic classroom composition.

5.2 The German context

In Germany, secondary education is differentiated into three basic school tracks that differ in curricula, duration, and certificate: vocational track ('Hauptschule'), intermediate track ('Realschule'), and academic track ('Gymnasium'). After completing primary education, usually at the end of grade four, students attend one of these school tracks or a comprehensive school type for at least 5 years. Given the stratified nature of the system, the student composition strongly varies between school tracks and schools, while the variation within schools is relatively low compared with other countries (Baumert, Stanat and Watermann 2006). Although track mobility is basically possible, the majority of students stay on the initial school track until the end of compulsory education (i.e., end of ninth grade). Thus, students are exposed to a rather stable school or classroom environment during secondary education. Consequently, influences of the classroom composition on achievement in ninth grade, which we address empirically, represent the cumulated effects of the learning environment over several years and should be particularly pronounced.

The ethnic minority composition in the German school system has increased in magnitude and diversity over the last decades. Students of Turkish origin and from the former Soviet Union represent the two largest minority groups (Olczyk et al. 2016b). Most of these students are so-called second- and third-generation immigrants, i.e., they were born in Germany, and their parents or grandparents were born abroad. Ethnic minority students more frequently attend the lower school tracks and, on average, show

lower (reading) achievement in secondary school than the majority; students of Turkish origin appear to be particularly disadvantaged (Heath and Brinbaum 2007; Stanat, Rauch and Segeritz 2010). One reason for this disadvantage could be that despite having grown up in Germany, a large proportion of ethnic minority students use a minority language in their everyday communication with parents and, to a lesser extent, with friends and classmates (Strobel and Seuring 2016). Based on the numerical relevance of Turkish- and Russian-speaking students combined with the highly stratified school system and residential segregation, these groups are most likely to attend schools and classrooms with a high proportion of students who belong to the same ethnic minority and who speak the same minority language.

5.3 Ethnic classroom composition effects

Research on the role of ethnic composition in schools and classrooms for students' educational achievement has been performed in various disciplines. In some countries, such as Germany, the findings consistently challenge the relevance of ethnic composition effects (Eksner and Stanat 2012). However, on a larger, international scale, research investigating the relationship between ethnic classroom composition and educational achievement has yielded inconclusive results.

Several previous studies have found a negative association between ethnic composition in classrooms or schools and educational achievement (Bankston and Caldas 1996; Roscigno 1998; Driessen 2002; Van Ewijk and Slegers 2010b; Mickelson, Bottia and Lambert 2013), suggesting that on average, students in classrooms with a high proportion of ethnic minority classmates reach lower achievement levels than students with fewer minority classmates. However, findings from other studies suggest that this negative association is reduced or even disappears after controlling for the socioeconomic composition and/or average prior achievement in the classroom (Peetsma et al. 2006; Stanat, Schwippert and Gröhlich 2010; Van der Slik, Driessen and De Bot 2010; Rjosk et al. 2014). Some studies have even found a positive effect of the ethnic classroom composition on educational achievement after controlling for socioeconomic composition (Portes and Hao 2004; Fekjær and Birkelund 2007). Accordingly, whether and how the ethnic classroom composition affects students' educational achievement over and above the socioeconomic composition and average prior achievement remains unclear.

Moreover, previous studies have used a broad range of characteristics to operationalize ethnic classroom composition, including students' race, citizenship, country of birth, and minority language (Eksner and Stanat 2012). However, these characteristics may be associated with different processes that could potentially underlie the effects of the ethnic classroom composition—an issue that has rarely been addressed in previous research. For example, one prominent hypothesis is that language-related factors such as minority language use among classmates mediate the negative effects of ethnic classroom composition (Driessen 2002; Esser 2006; Peetsma et al. 2006; Van Ewijk and Slegers 2010b). Consequently, it may not be the proportion of ethnic minority

students (*ethnic minority composition*) but rather the proportion of minority language students in the classroom (*ethno-lingual composition*) that influences students' academic achievement.

This reasoning is supported by studies that have simultaneously examined the effects of the proportion of ethnic minority students and the proportion of minority language students in a classroom. For example, Stanat, Schwippert and Gröhlich (2010) found that, before controlling for socioeconomic and prior achievement composition, only the latter was significantly correlated with students' reading achievement. This finding suggests that ethnic composition effects may be associated with minority students' language backgrounds rather than with their ethnic backgrounds.

We address this notion in the following section by outlining the main arguments linking language-related factors in classroom composition to educational achievement while focusing on minority language use among classmates.

5.3.1 Ethno-lingual classroom composition and educational achievement

In general, factors that may account for the relationship between student composition in classrooms and educational achievement can be broadly assigned to three interrelated dimensions: (i) peer interaction processes, (ii) instructional quality, and (iii) school resources (Thrupp 1999, Thrupp Lauder and Robinson 2002).

The *interaction among peers* (*i*) may directly affect individual achievement through an exchange of resources. Such resources include learning support among classmates (e.g., helping with homework) and the transmission and reinforcement of shared values, aspirations, and behaviours that favour education (Driessen 2002; Rumberger and Palardy 2005; Fekjær and Birkelund 2007; Hermansen and Birkelund 2015). Several studies indicate that educational aspirations and behaviours of peers are significantly related to students' achievement (Roscigno 1998, Stewart 2008; Agirdag Van Houtte and Van Avermaet 2012). Moreover, ethnic minority students seem to benefit more from interactions with majority peers than with other minority peers (Baysu, Phalet and Brown 2014; Wölfer, Caro and Hewstone 2019). The ethno-lingual classroom composition could determine minority language students' opportunities to interact and form relationships with majority peers and, thus, the range of potential resources provided within the classroom context.

In classrooms with a high proportion of students who share the same minority language, it is likely that they will use that language with each other rather than the language of instruction, for example, during breaks (Driessen 2002; Peetsma et al. 2006; Van Ewijk and Slegers 2010b). Students who speak a minority language with peers inevitably spend less time using the language of instruction, which should impair their acquisition of the language of instruction and, consequently, their educational achievement (Esser 2006; Peetsma et al. 2006; Scheele, Leseman and Mayo 2010). This 'time on task' hypothesis focuses on the quantity of exposure to a given language, which is assumed to determine the acquisition of that language (Chiswick and Miller

2001). Attending classrooms with many minority language students may impair the educational achievement of minority students in these classrooms, which, in turn, should reinforce existing ethnic disadvantages. At the same time, minority language use should have positive effects on students' proficiency in the minority language. That is, students who are exposed to the minority language not only at home but also in school are likely to maintain or develop higher levels of proficiency in their minority language (Esser 2006; Van Tubergen and Mentjox 2014).

In addition to reducing exposure to the instructional language, minority language use can indicate conditions of ethnic closure within the classroom. Fewer interactions between minority and majority students could hamper learning through reduced learning support and cohesion among classmates (Hoxby 2000; Van Ewijk and Slegers 2010b; Hermansen and Birkelund 2015). This reasoning is based on the assumption that minority (language) students generally possess fewer or less favourable resources and have lower educational aspirations than majority students (Driessen 2002; Cebolla-Boado and Medina 2011). Thus, minority students who attend segregated classrooms should benefit less from interactions with their peers than students who have more majority classmates.

However, another branch of research challenges this negative view and emphasizes the additional benefits of intra-ethnic contacts. According to this perspective, ethnic minority communities could reinforce high educational aspirations, prevent deviant behaviour, and provide learning support through a strong cohesion and social control among its members (Bankston and Zhou 1995; Portes and Hao 2002). Speaking the minority language may reflect a strong attachment to the ethnic group and may provide access to resources of the ethnic network (Mouw and Xie 1999). Minority language students who attend more ethnically segregated classrooms might motivate and support one another in learning, and thereby reinforce high educational aspirations within the ethnic peer-group. For instance, students can use the minority language to help each other to master school-related tasks.

A prerequisite for such positive effects of the ethno-lingual classroom composition would be that the ethnic network exhibits attitudes and aspirations that favour education (Kroneberg 2008). In contrast, low educational aspirations should adversely affect achievement. Previous findings indicate that despite performing lower on average, ethnic minority students often show similar or even higher educational aspirations than the majority (Salikutluk 2016), supporting a potential positive mechanism of intra-ethnic contacts.

The *instructional quality* (ii) in classrooms is an additional factor that may mediate composition effects. The primary components of instructional quality include cognitively challenging learning opportunities, efficient classroom management, and individualized learning support (Klieme 2006). Previous findings indicate that aspects of learning support and cognitively challenging instruction appear to be most important (Seidel and Shavelson 2007; Baumert et al. 2010; Rjosk et al. 2014). Based on their perceptions of students' average achievement and specific needs, teachers adapt their teaching levels and standards to create a stimulating learning environment tailored to the demands of the student body in classroom (Hoxby 2000; Harker and Tymms

2004; Cebolla-Boado and Medina 2011). However, teachers may have unjustifiably low perceptions of ethnic minority and language minority students' educational capabilities (Driessen 2002; Van Ewijk and Slegers 2010b; Hermansen and Birkelund 2015). Previous research indicates that teachers evaluate minority language use and difficulties with the language of instruction as particularly detrimental factors for students' achievement (Agirdag, Van Avermaet and Van Houtte 2013). Thus, teachers may be prone to underestimating the true potential of minority language students. In classrooms with many of these students, teachers may adapt their instruction to their lowered expectations (Fekjær and Birkelund 2007), affecting further educational achievement of all students in these classrooms.

Moreover, teachers may communicate their low expectation to the students, which could cause minority language students to lower their educational aspirations, resulting in lower achievement (Rosenthal and Jacobson 1968). This tendency of a 'self-fulfilling prophecy' might be further amplified through peer interaction processes if minority students who share the same experiences in classroom adapt their lowered aspirations to one another. Conversely, a strong cohesion among minority students in classroom could also have a protective function that counteracts discriminatory influences by reinforcing high aspirations within this group.

Finally, an alternative explanation for the relationship between classroom composition and educational achievement pertains to *school resources (iii)*. Organizational and structural features of schools, such as teacher qualifications, teaching materials, and classroom equipment, can contribute to a favourable learning environment (Rumberger and Palardy 2005; Cebolla-Boado and Medina 2011). Although empirical research often demonstrates positive associations between school resources and students' educational achievement (Roscigno 1998; Cebolla-Boado and Medina 2011), school resources do not seem to substantially account for classroom composition effects (Hanushek 1997; Rumberger and Palardy 2005). However, in countries with tracked school systems, minority language students frequently attend lower, often vocationally oriented school tracks (Heath and Brinbaum 2007). These tracks usually have a less challenging curriculum than academic school tracks (Hattie 2002; Rumberger and Palardy 2005). Consequently, the relationship between the proportion of minority language students in the classroom and educational achievement does not necessarily represent a composition effect of the student body and may primarily reflect selection processes for specific school tracks that account for the composition of students. To rule out this alternative explanation, it is important to account for differences among school tracks when examining effects of the ethno-lingual composition.

5.3.2 Disentangling the mechanisms associated with ethno-lingual classroom composition

From a theoretical perspective, the mechanisms associated with ethno-lingual classroom composition and minority language use could inversely affect educational achievement. Restricted exposure to the instructional language, a less challenging instructional

quality, and lower teacher expectations should hamper (minority) students' achievement, whereas positive effects may arise from high educational aspirations within ethnic minority networks. Empirical evidence regarding the processes underlying the effects of ethno-lingual classroom composition is scarce, and it is as-yet unclear which of the potential mechanisms are significant and whether the negative and positive influences counteract each other.

Although previous studies examining the effects of ethnic and ethno-lingual classroom composition have yielded heterogeneous findings, on a larger scale, meta-analyses have revealed a negative overall effect on educational achievement (Van Ewijk and Sleegers 2010b; Mickelson, Bottia and Lambert 2013), suggesting that negative influences outweigh potential positive influences. Similarly, findings on the individual level indicate that favourable aspirations and resources within the family do not compensate for the negative effects of minority language use at home (Strobel 2016). In light of these findings, we expect a negative association between the ethno-lingual classroom composition and educational achievement.

The processes underlying the effects of ethno-lingual classroom composition should be particularly relevant for students' language achievement. In the following, we therefore focus on the association between ethno-lingual classroom composition and students' reading achievement in the language of instruction—as well as minority students' proficiency in the minority language. Reading comprehension not only constitutes an indicator of students' proficiency in the instructional language but also forms an important basis for general educational achievement. Limited reading proficiency hampers learning in almost all other school subjects. Consequently, research investigating classroom composition effects has frequently focused on reading achievement (Entwisle and Alexander 1994; Roscigno 1998; Stanat, Schwippert and Gröhlich 2010; Maestri 2011; Rjosk et al. 2014).

Influences associated with speaking a minority language, such as minority language use in school, require the presence of classmates who also speak this language. Thus, if the effects of ethnic classroom composition on reading achievement primarily result from students' language background, the classroom's ethnic minority composition should not be associated with achievement over and above the ethno-lingual composition (*Hypothesis 1*). Of course, students' ethnic and language backgrounds inherently correlate with each other. Virtually all students who speak a minority language also have an immigration background. In contrast, not all ethnic minority students have also learned a minority language at home, especially students of the second or third immigrant generation. Consequently, in principle, the ethnic minority classroom composition could affect students' achievement, independent of the ethno-lingual composition; however, this should not be the case if students' language background is the driving factor underlying ethnic composition effects.

Students who speak the minority language with their peers may lack exposure to the instructional language. This should impair minority students' acquisition of the instructional language whereas majority students—who do not speak a minority language—should be largely unaffected. Similarly, conditions of ethnic closure in classroom—both positive and negative conditions—and biased teacher expectations

should mainly determine minority language students' achievement without affecting the majority. The only exception pertains to the mechanism of lowered instructional quality in classrooms with many minority language students, which should equally affect achievement of both minority and majority students. Consistent with previous findings indicating zero effect for majority students (Van Ewijk and Slegers 2010b), we expect that the disadvantages associated with ethno-lingual classroom composition exclusively apply to minority language students (*Hypothesis 2*).

Furthermore, the processes of support and shared attitudes towards education among minority classmates and lowered ambitions based on teacher expectations should affect students' educational aspirations. The time spent using the minority language with classmates reduces exposure to the instructional language and should hamper educational achievement, over and above students' educational aspirations. Thus, if minority students in classrooms with many minority language students have restricted exposure to the instructional language, the negative association between ethno-lingual classroom composition and reading achievement for minority students should remain, even after accounting for students' educational aspirations (*Hypothesis 3*). In addition, we expect these remaining disadvantages to substantially decrease after controlling for students' minority language use with classmates (*Hypothesis 4*).

Finally, if classrooms with a higher proportion of minority language students cause these students to use their minority language more frequently to communicate with each other, they are likely to maintain or develop high levels of proficiency in that language. Because minority languages are usually not explicitly taught in school, other factors related to the classroom context, such as teacher expectations and motivational aspects of peer interactions, cannot drive this relationship. Accordingly, this situation allows for an additional evaluation of the proposition that minority language use mediates ethnic classroom composition effects by determining the amount of exposure to the respective language. We expect that the proportion of classmates who speak the same minority language is positively related to minority students' proficiency in their minority language (*Hypothesis 5*). Again, this positive association should substantially decrease after controlling for students' minority language use with classmates (*Hypothesis 6*).

5.4 Data, variables, and methods

5.4.1 Data

We used data from the 'German National Educational Panel Study' (NEPS) for our analyses (Blossfeld, Roßbach and Maurice 2011). The NEPS collects data for nationwide samples of various age cohorts in Germany and provides detailed information on respondents' characteristics, including competence tests in different domains. We focused on the first and second waves of data collection in Starting Cohort 4 (ninth grade).² Students were sampled based on a stratified multistage sampling design. In

²This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Grade 9, doi:10.5157/NEPS: SC4:7.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework

the first stage, schools of different types were randomly selected. In the second stage, all students from one or two classrooms per school were invited to participate in the survey (Skopek, Pink and Bela 2013). The paper-and-pencil questionnaires and tests were administered in fall/winter 2010/2011 (wave 1) and spring 2011 (wave 2). Because the two waves collected information on different characteristics of students' backgrounds and test domains, we integrated both waves into one cross-sectional dataset.

The initial sample included $N=16,425$ students. Excluding students who did not participate in at least one wave reduced the sample to $N=16,254$. Furthermore, we focused on students who attended secondary school types regarded as general schooling, i.e., vocational track ('Hauptschule'), intermediate track ('Realschule'), academic track ('Gymnasium'), and comprehensive school types ('Gesamtschule' or integrated school types with several tracks). Students attending schools for children with special educational needs ($N=1,155$) and Waldorf education schools ($N=118$), which are both unique in several ways, were excluded from the analysis. Finally, we excluded students who did not answer questions on their ethnic and language origin ($N=166$). The final analysis sample consisted of 14,815 students in 1,006 classrooms in 539 schools, with an average of 15 students per classroom. Table 5.1 shows the distributions of selected sample characteristics and all model variables.

5.4.2 Variables

The primary dependent variable is students' score on the *reading comprehension test in German* that was administered in the NEPS. This test is based on five texts (approximately 200–550 words each) of different types, including literary and factual texts (Gehrer et al. 2012). After reading each text, students were asked to complete five to seven tasks (mostly multiple-choice questions) that required them to select information from the text, draw text-related conclusions, or reflect on the content provided in the text. In total, the test consists of 31 items. A generalized Rasch model was used to estimate student achievement scores on a common scale. The test scores are weighted maximum likelihood estimates (WLEs) of student achievement, which are constrained to have a mean of $M=0$. The variance of the WLEs is not restricted.

To test the hypothesized relationship between minority language use and proficiency, we used scores from the *listening comprehension tests in Russian and Turkish* languages. In its assessment of minority language skills, the NEPS focuses on oral comprehension because minority students born in Germany typically learn the minority language in their families and do not generally learn to read or write in that language (Kristen et al. 2011). Ethnic minority students of Turkish origin and from the Former Soviet Union were asked to participate in a screening test. Only students who showed a minimum level of minority language proficiency (i.e., students who correctly completed at least three of eight screening tasks) subsequently participated in the full test (Edele

Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, NEPS is carried out by the Leibniz Institute for Educational Trajectories (LifBi) at the University of Bamberg in cooperation with a nationwide network.

et al. 2012). The test instrument consists of seven texts (approximately 100-150 words each), including dialogues, expository texts, and narratives. After listening to each text, students were asked to answer three to four multiple-choice questions about the content for a total of 31 items. Test scores (WLEs) are available for 499 Russian-speaking and 662 Turkish-speaking students. The analyses of minority language proficiency are restricted to these sub-samples (also see columns 3 and 4 in Table 5.1).³

Respondents were defined as ethnic minority students if they, at least one of their parents or at least two of their grandparents were born in a country other than Germany (Olczyk, Will and Kristen 2016a). Furthermore, students who indicated that they had learned a language other than German at home during childhood were considered *minority language students*. Ethnic minority students make up approximately 29 percent of the total sample, and many of these students have also learned a minority language in their families (70 per cent). Of all minority language students, Turkish- (27 per cent) and Russian- (21 per cent) speaking students represent the numerically most dominant groups (see Appendix 5.9, Table A5.1).

General cognitive ability captures students' broad learning ability and has been shown to be an adequate proxy for *prior achievement* (Baumert, Stanat and Waterman, 2006). The measure of cognitive ability employed in the NEPS is a matrices test with 12 items assessing reasoning (Haberhorn and Pohl 2013). The test asks students to complement sequences of geometric elements. We calculated a sum score representing the number of tasks that each student solved correctly.

We operationalized students' *socioeconomic background* based on the highest 'International Socio-Economic Index of Occupational Status' (ISEI-08; Ganzeboom 2010) of the mother or father. Furthermore, we considered parents' highest level of educational qualification and distinguished among 'no educational qualification or low level secondary education' (up to 9 years of schooling), 'middle level secondary education' (10–11 years of schooling), and 'higher level secondary education' (12 years or more of schooling).

We used students' *idealistic aspirations* as an indicator of their general attitude towards education. Students could indicate the type of school-leaving certificate that they would like to have, irrespective of their actual performance and attended school track. We differentiated between students who aspire to the highest possible certificate and all others (1 'high' versus 0 'low' educational aspirations).

Finally, we used an indicator of *minority language use* in the school context provided by the NEPS data. Minority language students were asked to name the language(s) that they speak with their classmates in the schoolyard (from 1 'only German' to 4 'only the minority language'). We converted this information into a dummy variable for the analyses (0 'only German' versus 1 'at least sometimes the minority language') and coded all majority language students as speaking only German.⁴ Approximately

³The sub-samples include 76/79 per cent of all Russian-/Turkish-speaking students in the sample because not all of these students were present when minority language testing occurred in wave 2.

⁴Additional analyses using the original four-point scale yielded analogous findings (Appendix 5.9, Table A5.2). Nevertheless, we used dummy variables for the main analyses because only a few students mostly or only speak the minority language with classmates (Appendix 5.9, Figure A5.1).

one-fifth of all minority language students at least sometimes use the minority language with classmates, whereas almost all of them speak the minority language at home on a regular basis. To rule out the possibility that minority language use in school merely reflects influences of family language, we additionally controlled for language use with parents.

At the classroom level, we aggregated the information provided by students in each classroom to indicate the percentage of ethnic minority students (ethnic minority composition), the percentage of minority language students (ethno-lingual composition), the average cognitive ability (prior achievement composition), and the average HISEI (socioeconomic composition).

The average proportion of ethnic minority and minority language students in classroom reflects the distribution of these students on the individual level (Table 5.1).⁵ However, minority language students, and particularly Turkish-speaking students, are more inclined to attend classrooms with a higher proportion of minority students. Further analyses show that the proportion of minority language students in classroom is highly correlated with the number of classmates who speak the *same* language (see Appendix 5.9, Table A5.1), indicating tendencies of ethnic segregation between classrooms.

To analyse the relationship between ethno-lingual classroom composition and minority language proficiency, we calculated the percentage of students who speak Russian or Turkish, respectively, within each classroom. To ensure that these students possess at least a minimum level of minority language proficiency, which is necessary for communicating in that language with classmates, we included only students who passed the screening test in Russian or Turkish. Consequently, the analysis was restricted to classrooms with at least one student who passed the respective screening test, resulting in $N=298$ classrooms in the Russian sub-sample and $N=305$ classrooms in the Turkish sub-sample (also see columns 3 and 4 in Table 5.1).

In addition, in all analyses, we controlled for gender, age, country of birth (Germany versus abroad), and attended school track.

5.4.3 Methods

We employed random intercept hierarchical linear models with students at level 1 and classrooms at level 2. To test for differential effects of the ethno-lingual composition for language minority and majority students, we further specified random slopes with cross-level interactions. All continuous variables were standardized ($M = 0$, $SD = 1$) at the individual and classroom levels using grand mean centering to facilitate comparative interpretations of the coefficients (Enders and Tofghi 2007). All indicators of the ethnic and minority language classroom composition were divided by 10 so that the regression coefficients reflect changes in the dependent variable associated with an increase in the

⁵Although both indicators are highly correlated ($r=0.89$, $p<0.001$), the overlap between the two variables is limited (Appendix 5.9, Figure A5.2). In approximately 30 per cent of all classrooms, ethnic minority and minority language composition coincide, whereas in most classrooms, there are higher proportions of ethnic minority than language minority students.

proportion of minority (language) students by 10 per cent.

We accounted for item nonresponse at the individual level with multiple imputation by estimating 50 datasets with complete information (Allison 2001). The aggregated indicators at the classroom level were recalculated within each classroom for the complete data sets. The multilevel models were estimated separately for each of the 50 imputed data sets and combined following Rubin's (1987) rule. Analyses of imputed and original data showed comparable results. The descriptive results shown in Table 5.1 are based on the original data.

We report unweighted values for all of the descriptive and multivariate results. The estimates of the multilevel models proved robust across weighted and unweighted analyses.

Table 5.1: Distribution of model variables and sample characteristics

	Total		MiL students		Russian		Turkish		Miss. (%) ^a
	M	SD	M	SD	M	SD	M	SD	
Level 1: Individual characteristics	N=14,815		N=3,123		N=499		N=662		
Reading competence (German)	0.0	(1.0)	-0.5	(1.0)	-0.4	(0.9)	-0.8	(0.9)	7.6
Minority language proficiency	–	–	–	–	0.0	(1.0)	0.0	(1.0)	0.0
Ethnic minority (%)	28.8		95.5		100.0		100.0		0.0
Minority language (%)	21.1		100.0		100.0		100.0		0.0
Idealistic aspirations (%)	58.5		53.3		49.7		49.2		6.5
MiL use with classmates (%)	3.9		22.0		22.2		41.1		4.0
MiL use with parents (%)	16.7		92.8		92.4		97.1		3.8
Born abroad (%)	6.2		27.2		54.4		11.0		0.0
Female (%)	49.8		50.9		50.5		48.3		0.1
Age	15.2	(0.6)	15.4	(0.7)	15.5	(0.7)	15.4	(0.7)	0.2
General cognitive ability	8.7	(2.5)	7.8	(2.7)	8.3	(2.4)	7.3	(2.5)	7.6
SES: HISEI	51.1	(20.6)	42.5	(20.4)	39.4	(18.8)	36.4	(17.6)	21.7
Parents' highest education (%)									21.5
No/low level	14.8		25.2		7.2		47.4		
Middle level	40.2		34.1		49.2		32.6		
Higher level	45.0		40.7		43.7		20.0		
Level 2: Classroom characteristics	N=1,006		N=798		N=298		N=305		
Proportion of ethnic min. students	30.0	(23.6)	35.2	(22.7)	38.9	(20.6)	46.5	(22.9)	0.0
Proportion of MiL students	22.3	(22.7)	28.1	(22.1)	31.6	(19.9)	39.1	(22.8)	0.0
Proportion of same MiL students	–	–	12.8	(11.1)	11.0	(7.3)	14.9	(11.9)	0.0
General cognitive ability (mean)	8.4	(1.5)	8.4	(1.5)	8.4	(1.4)	8.0	(1.5)	1.7
SES (mean)	49.0	(11.7)	48.8	(11.6)	46.6	(10.6)	46.0	(10.7)	0.2
Class size	14.9	(6.8)	15.9	(6.5)	16.9	(6.0)	16.6	(6.2)	0.0
School track (%)									0.0
Vocational	33.4		35.3		40.6		46.2		
Intermediate	28.5		27.1		29.2		23.6		
Academic	30.9		30.7		25.2		20.7		
Comprehensive	7.2		6.9		5.0		9.5		

Notes: Due to rounding, the sum of the percentages may slightly deviate from 100%. The columns 'Russian' and 'Turkish' refer to sub-samples of students with the respective minority language background (level 1) and classrooms with at least one of those students (level 2). HISEI, highest International Socio-Economic Index of Occupational Status; MiL, minority language; SES, socioeconomic status. ^a Proportion of missings refers to total sample.

5.5 Results

5.5.1 Ethnic minority and ethno-lingual classroom composition

Table 5.2 shows the results of the multilevel regressions of German reading comprehension on the ethnic minority and ethno-lingual classroom composition. Students in classrooms with higher proportions of ethnic minority students show significantly lower levels of reading proficiency than students with fewer minority classmates (Model 1). Specifically, students' reading scores decrease by approximately 0.03 standard deviations with each 10-per cent increase in the proportion of ethnic minority students. The effect size for the ethno-lingual classroom composition is practically the same (Model 2).

Furthermore, the negative association between the proportion of ethnic minority students in the classroom and students' reading achievement disappears after including the ethno-lingual classroom composition in the model (Model 3). Consistent with *Hypothesis 1*, the correlation between ethnic classroom composition and reading achievement appears to be associated with language-related factors rather than with students' ethnic origin per se.

5.5.2 Ethno-lingual classroom composition and German reading comprehension

Model 4 in Table 5.3 shows the association between ethno-lingual classroom composition and reading comprehension, net of mean cognitive ability and the socioeconomic composition. Both of these classroom composition characteristics are positively related to students' reading comprehension. On average, students in classrooms with a high mean cognitive ability and high average SES are significantly more proficient in reading than students in classrooms with less privileged peers. Moreover, the results indicate that the negative association between ethno-lingual composition and reading achievement (Table 5.2, Model 2) can be partially attributed to a lower mean cognitive ability and lower socioeconomic composition in these classrooms; yet, the coefficient remains statistically significant even after controlling for these classroom characteristics.

Students' average reading proficiency decreases by 0.015 standard deviations with each 10-per cent increase in the proportion of minority language students in the classroom. To illustrate, the effect size corresponds to a difference in average reading achievement of 0.1 standard deviations between classrooms with relatively few minority language students (10 per cent) and highly segregated classrooms (80 per cent). Thus, the association between ethno-lingual composition and reading comprehension is small in magnitude compared to that of other factors such as prior achievement composition or individual characteristics. Nevertheless, this finding indicates that factors associated with ethno-lingual classroom composition, such as minority language use among classmates, might influence students' language achievement over and above the socioeconomic and prior achievement composition of the classroom.

Next, we examine whether the ethno-lingual classroom composition is differentially

Table 5.2: Predicting German reading comprehension by ethnic minority and ethno-lingual classroom composition (hierarchical linear models)

	Model 0	Model 1	Model 2	Model 3	
Fixed Part					
<i>Level 2</i>					
Ethnic minority students (%)		-0.029 (0.005)	***		0.005 (0.011)
Minority language students (%)				-0.030 (0.005)	*** -0.034 (0.011) **
School track (<i>ref. vocational</i>)					
Intermediate		0.355 (0.027)	***	0.347 (0.027)	*** 0.348 (0.027) ***
Academic		0.852 (0.028)	***	0.836 (0.029)	*** 0.836 (0.029) ***
Comprehensive		0.305 (0.042)	***	0.299 (0.042)	*** 0.299 (0.042) ***
<i>Level 1</i>					
Ethnic minority (<i>ref. majority</i>)		-0.117 (0.017)	***		-0.005 (0.023)
Minority language (<i>ref. German</i>)				-0.204 (0.020)	*** -0.200 (0.027) ***
Born abroad		-0.070 (0.031)	*	-0.013 (0.031)	-0.012 (0.031)
Female		0.193 (0.014)	***	0.195 (0.014)	*** 0.195 (0.014) ***
Age		-0.051 (0.007)	***	-0.048 (0.007)	*** -0.048 (0.007) ***
Cognitive ability		0.250 (0.008)	***	0.248 (0.008)	*** 0.248 (0.008) ***
SES: HISEI		0.061 (0.009)	***	0.055 (0.009)	*** 0.055 (0.009) ***
Parents' education (<i>ref. no/low level</i>)					
Middle level		0.007 (0.022)		0.004 (0.022)	0.004 (0.022)
High level		0.067 (0.025)	**	0.066 (0.025)	** 0.066 (0.026) *
Intercept		-0.447 (0.033)	***	-0.452 (0.031)	*** -0.456 (0.032) ***
Random Part					
Within-classroom variance	0.619	0.564		0.561	0.561
Between-classroom variance	0.391	0.048		0.047	0.047
R ² within classrooms		0.09		0.09	0.09
R ² between classrooms		0.88		0.88	0.88

Notes: Coefficients represent standard deviation changes in the dependent variable, with standard errors in parentheses. N students=14,815, N classrooms=1,006. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

related to reading achievement between minority language students and majority students (Table 5.3, Model 5). Because the potential mechanisms associated with the ethno-lingual classroom composition should primarily affect minority language students, we expect lower German reading scores among these students (*Hypothesis 2*). The negative and statistically significant cross-level interaction between language background and ethno-lingual composition supports this assumption. However, the overall association for minority students is very small in magnitude (-0.008-0.016=-0.024; $p < 0.01$). Moreover, the non-significant conditional main effect in Model 5 indicates that majority students are not disadvantaged in classrooms with a high proportion of minority language students. Thus, the negative association between ethno-lingual composition and reading achievement for all students in the classrooms (Model 4) ap-

Table 5.3: Predicting German reading comprehension by ethno-lingual classroom composition and minority language use (hierarchical linear models)

	Model 4	Model 5	Model 6	Model 7	Model 8
Fixed Part					
<i>Level 2</i>					
Minority language students (%)	-0.015 (0.005)	** -0.008 (0.006)	-0.011 (0.006)	-0.011 (0.006)	-0.011 (0.006)
Cognitive ability (mean)	0.133 (0.016)	*** 0.132 (0.016)	*** 0.129 (0.016)	*** 0.128 (0.016)	*** 0.128 (0.016)
SES (mean)	0.060 (0.016)	*** 0.059 (0.016)	*** 0.048 (0.016)	** 0.048 (0.016)	** 0.048 (0.016)
<i>Cross-level interaction</i>					
MiL x MiLS (%)		-0.016 (0.008)	* -0.017 (0.008)	-0.014 (0.008)	-0.013 (0.008)
<i>Level 1</i>					
Minority language (<i>ref. German</i>)	-0.210 (0.020)	*** -0.159 (0.033)	*** -0.173 (0.032)	*** -0.170 (0.033)	*** -0.055 (0.062)
Born abroad	-0.014 (0.031)	-0.013 (0.031)	-0.020 (0.031)	-0.018 (0.031)	-0.012 (0.031)
Cognitive ability	0.230 (0.008)	*** 0.229 (0.008)	*** 0.221 (0.008)	*** 0.221 (0.008)	*** 0.221 (0.008)
SES: HISEI	0.047 (0.009)	*** 0.047 (0.009)	*** 0.041 (0.010)	*** 0.041 (0.009)	*** 0.041 (0.009)
Idealistic aspirations (<i>ref. low</i>)			0.221 (0.019)	0.221 (0.019)	0.222 (0.019)
MiL use with classmates (<i>ref. only German</i>)				-0.088 (0.039)	-0.079 (0.039)
MiL use with parents (<i>ref. only German</i>)					-0.132 (0.060)
Intercept	-0.309 (0.034)	*** -0.323 (0.034)	*** -0.362 (0.034)	*** -0.359 (0.034)	*** -0.358 (0.034)
Random Part					
Within-classroom variance	0.561	0.561	0.555	0.555	0.555
Between-classroom variance	0.039	0.039	0.037	0.037	0.037
R ² within classrooms	0.09	0.09	0.10	0.10	0.10
R ² between classrooms	0.90	0.90	0.91	0.91	0.91

Note: Coefficients represent standard deviation changes in the dependent variable, with standard errors in parentheses. Analyses are controlled for school track (level 2) and female, age, parents' education (level 1). N students=14,815, N classrooms=1,006. MiL(S)=minority language (students). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

pears to reflect relationships that only apply to minority language students.

The negative cross-level interaction remains stable after controlling for students' idealistic aspirations (Model 6). Although aspirations are positively correlated with reading achievement, the findings do not indicate that the ethno-lingual classroom composition might be associated with students' aspirations (*Hypothesis 3*). However, consistent with *Hypothesis 4*, the negative cross-level interaction is considerably reduced and no longer significant after controlling for minority language use among classmates (Model 7). Students who at least sometimes use the minority language within the school context have significantly lower German reading scores than students who speak only German with their peers. This association also remains robust when additionally controlling for language use with parents (Model 8). The findings might tentatively support the proposition that minority language use among classmates mediates ethno-lingual classroom composition effects through lowered exposure to the instructional language.

5.5.3 Ethno-lingual classroom composition and minority language proficiency

Finally, we examine the role of the ethno-lingual classroom composition for minority language proficiency. If the language-related composition provides opportunities for using the minority language, higher proportions of students speaking the same minority language in the classroom should foster proficiency in that language (*Hypothesis 5*). Therefore, we regressed minority students' listening comprehension scores in Russian and Turkish on the proportion of Russian- or Turkish-speaking students in the classroom (see Table 5.4).

Consistent with the theoretical assumptions, the proportion of Turkish-speaking students in the classroom is positively related to language proficiency in Turkish (Model 1_T). For example, a 10-per cent increase in the proportion of Turkish-speaking students in the classroom is associated with an increase of 0.08 standard deviations in students' test scores. The effect size is approximately 3.5 times higher than the effect size of the ethno-lingual classroom composition on minority students' German reading proficiency (see Table 5.3, Model 5). Moreover, the association between ethno-lingual classroom composition and students' Turkish proficiency is reduced and no longer significant after controlling for language use with classmates in Model 2_T (*Hypothesis 6*). Turkish students who use the minority language with classmates show significantly higher levels of Turkish listening comprehension than Turkish students who speak only German in school. Again, this association remains stable, even after controlling for language use with parents (Model 3_T).

For the Russian sub-sample, there are no significant associations between students' Russian listening comprehension and ethno-lingual classroom composition (Model 1_R) or minority language use with classmates (Model 2_R). The coefficients are similar in size to the coefficients of the Turkish-speaking students, but they do not reach statistical significance. One explanation for this discrepancy could be that approximately half of

Table 5.4: Predicting minority language proficiency by the proportion of Russian-/Turkish-speaking classmates (hierarchical linear models)

	Russian			Turkish			
	Model 0 _R	Model 1 _R	Model 2 _R	Model 0 _T	Model 1 _T	Model 2 _T	Model 3 _T
Fixed Part							
<i>Level 2</i>							
Russian-speaking students (%)	0.108 (0.062)		0.102 (0.063)	0.101 (0.062)	0.081 (0.040)	0.067 (0.041)	0.066 (0.040)
Turkish-speaking students (%)							
Cognitive ability (mean)	0.099 (0.080)		0.100 (0.080)	0.090 (0.079)	0.204 (0.077)	0.210 (0.077)	0.217 (0.077)
SES (mean)	-0.002 (0.088)		-0.001 (0.088)	0.013 (0.087)	0.005 (0.088)	0.014 (0.088)	0.016 (0.088)
<i>Level 1</i>							
Born abroad	0.603 (0.084)	***	0.601 (0.084)	0.570 (0.084)	0.110 (0.111)	0.107 (0.111)	0.094 (0.111)
Cognitive ability	0.061 (0.045)		0.061 (0.045)	0.063 (0.045)	0.157 (0.040)	0.161 (0.040)	0.156 (0.040)
SES: HISEI	0.051 (0.056)		0.052 (0.056)	0.043 (0.057)	-0.034 (0.053)	-0.037 (0.053)	-0.031 (0.053)
Idealistic aspirations (<i>ref. low</i>)	0.064 (0.093)		0.069 (0.094)	0.066 (0.093)	0.170 (0.084)	0.168 (0.083)	0.162 (0.083)
MiL use with classmates (<i>ref. only German</i>)			0.061 (0.100)	0.034 (0.100)		0.177 (0.079)	0.157 (0.079)
MiL use with parents (<i>ref. only German</i>)			0.343 (0.152)	0.343 (0.152)			0.515 (0.224)
Intercept	-0.718 (0.205)	**	-0.732 (0.206)	-1.031 (0.243)	-0.141 (0.150)	-0.173 (0.150)	-0.651 (0.224)
Random Part							
Within-classroom variance	0.525		0.485	0.481	0.617	0.609	0.611
Between-classroom variance	0.513		0.273	0.267	0.260	0.262	0.245
R ² within classrooms	0.08		0.08	0.08	0.07	0.08	0.08
R ² between classrooms	0.46		0.47	0.48	0.28	0.27	0.32

Note: Coefficients represent standard deviation changes in the dependent variable, with standard errors in parentheses. Analyses controlled for school track (level 2) and female, age, parents' education (level 1). N students=499(RU)/662(TR), N classrooms=298(RU)/305(TR). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the Russian-speaking students were born abroad. Those students may have acquired high levels of minority language proficiency in their country of origin; therefore, whether they use this language with their peers at school may be less relevant.

5.6 Limitations and robustness of findings

The aim of the present study was to analyse the proposition that minority language use among classmates may mediate ethno-lingual classroom composition effects, and to elucidate the mechanisms underlying this association. We followed a more indirect approach of disentangling the distinct processes by hypothesizing associations that could specifically be assigned to certain mechanisms. Consistent with our hypotheses, the findings indicated an interrelation among ethno-lingual classroom composition, minority language use with classmates, and language achievement. However, the applied analytic approach is limited in several regards and does not allow us to unequivocally test the mediation of classroom composition effects through minority language use with classmates or the processes underlying these associations. Selection bias, interdependencies between variables, and omitted variables may affect the reliability of our findings.

The correlation between ethno-lingual classroom composition and language proficiency could reflect selection processes of students with lower achievement attending classrooms with many minority students, and may thus be spurious. Moreover, minority students who have many co-ethnic classmates might be more likely to live in ethnically segregated areas and more frequently use the minority language in their everyday communication with parents and peers outside of school (neighbourhood context effects).

In addition, the causal relationship between language use and language proficiency is interdependent. We argue that language use influences students' language proficiency through increased exposure to the respective language (Chiswick and Miller 2001; Esser 2006). However, students' proficiency in a given language also determines their opportunities to use that language (Van Tubergen and Kalmijn 2009). For example, students must have at least basic proficiency in their minority language to use that language with classmates. These interdependencies may cause endogeneity problems in cross-sectional analyses, which could lead to overestimating the 'effects' of minority language use on language proficiency.

Moreover, the applied analytical approach does not provide a reliable test of the potential mechanisms that underlie the associations found in our analyses. The omission of relevant covariates may limit the reliability of our results. For example, we argue that minority language use with classmates hampers achievement through lowered exposure to the language of instruction. However, we do not have directly assessed information on students' language exposure and cannot rule out that other mechanisms might transmit the effects of minority language use. Examining these mechanisms more directly would require reliable indicators of the underlying factors, which are difficult to assess in surveys and are usually not available in large-scale data sets. Furthermore, the

proposed mechanisms associated with peer interaction processes should only apply to students who speak the *same* minority language; however, when examining associations with German reading comprehension, we focused on the proportion of *all* minority students in classroom.

In our analyses, we accounted for issues related to selection processes and endogeneity by controlling for the selection of students to different school tracks and general cognitive ability as a proxy for students' prior achievement. In addition, we controlled for alternative accounts by including country of origin, educational aspirations, and language use with parents as covariates. To evaluate the robustness of our findings, we replicated the analysis on German reading comprehension only for minority language students and additionally controlled for the proportion of co-ethnics in the neighbourhood and the proportion of classmates who speak the same minority language (Appendix 5.9, Table A5.3). The results proved to be robust against potential selection bias associated with the co-ethnic neighbourhood composition. Furthermore, the correlation between ethno-lingual classroom composition and reading achievement appears to be associated with the presence of classmates who share the same minority language rather than the proportion of minority language students per se. This finding might further suggest that minority students in segregated classrooms are more inclined to use their minority language among themselves.

However, the applied controls and robustness analyses may not capture all confounding influences and we cannot entirely rule out that alternative accounts might cause the interrelations among ethno-lingual classroom composition, minority language use with classmates, and achievement. However, our findings still provide first indications of the potential mechanisms that might be relevant in transmitting the effects of ethno-lingual classroom composition. This insight also contributes to previous research, as only very few studies have attempted to empirically address the mechanisms underlying (ethnic) classroom composition effects.

5.7 Summary and conclusion

The primary question addressed in the present study was whether and how the ethnic classroom composition is associated with students' language-related achievement. It has frequently been suggested that a high proportion of ethnic minority students in a classroom could enhance opportunities for minority language use among classmates, which may hamper the acquisition of the language of instruction. We empirically addressed this theoretical assumption and analysed the interrelations among ethnic classroom composition, minority language use, and students' German reading comprehension, as well as minority students' language proficiency in Russian and Turkish.

The findings indicated that the association between ethnic classroom composition and language achievement could be attributed to classmates' foreign language backgrounds rather than their ethnic origin. We found that the ethno-lingual classroom composition is negatively correlated with minority language students' German reading proficiency and positively with Turkish-speaking students' minority language profi-

ciency. These associations were considerably reduced and statistically no longer significant after controlling for students' reported language use with classmates. The findings provide tentative support for the proposition that a high proportion of minority language students in a classroom is associated with more frequent minority language use among classmates, which could affect language achievement through de-/increased exposure to that language.

Overall, the relationships between the proportion of minority language students in a classroom and minority students' language achievement were small in magnitude. Hence, the effects of ethnic classroom composition do not appear to substantially reinforce the achievement gap between ethnic minority and majority students. Other classroom characteristics, such as socioeconomic and prior achievement compositions, appear to be more closely related to students' educational achievement. Future studies should continue to explore the mechanisms underlying the effects of these factors as the available empirical evidence is far from being conclusive.

5.8 References

- Agirdag, O., Van Avermaet, P., and Van Houtte, M. (2013). School segregation and math achievement: A mixed-method study on the role of self-fulfilling prophecies. *Teachers College Record*, 115(3):1–50.
- Agirdag, O., Van Houtte, M., and Van Avermaet, P. (2012). Why does the ethnic and socio-economic composition of schools influence math achievement? The role of sense of futility and futility culture. *European Sociological Review*, 28(3):366–378.
- Allison, P. D. (2001). *Missing data*. Thousand Oaks: Sage Publications.
- Bankston III, C. L. and Caldas, S. (1996). Majority African American schools and social injustice: The influence of de facto segregation on academic achievement. *Social Forces*, 75(2):535–555.
- Bankston III, C. L. and Zhou, M. (1995). Effects of minority-language literacy on the academic achievement of Vietnamese youths in New Orleans. *Sociology of Education*, 68(1):1–17.
- Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., Klusmann, U., Krauss, S., Neubrand, M., and Tsai, Y.-M. (2010). Teachers' mathematical knowledge, cognitive activation in the classroom, and student progress. *American Educational Research Journal*, 47(1):133–180.
- Baumert, J., Stanat, P., and Watermann, R. (2006). Schulstruktur und die Entstehung differenzieller Lern- und Entwicklungsmilieus. In Baumert, J., Stanat, P., and Watermann, R., editors, *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit. Vertiefende Analysen im Rahmen von PISA 2000*, pages 95–188. Wiesbaden: VS Verlag für Sozialwissenschaften.

- Baysu, G., Phalet, K., and Brown, R. (2014). Relative group size and minority school success: The role of intergroup friendship and discrimination experiences. *British Journal of Social Psychology*, 53(2):328–349.
- Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J. (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS)*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Cebolla-Boado, H. and Medina, L. G. (2011). The impact of immigrant concentration in Spanish schools: School, class, and composition effects. *European Sociological Review*, 27(5):606–623.
- Chiswick, B. R. and Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, 38(3):391–409.
- Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Studies in Educational Evaluation*, 28(4):347–368.
- Edele, A., Schotte, K., Hecht, M., and Stanat, P. (2012). *Listening comprehension tests of immigrant students' first languages (L1) Russian and Turkish in grade 9: Scaling procedure and results (NEPS Working Paper 13)*. Bamberg: University of Bamberg, National Educational Panel Study.
- Eksner, J. and Stanat, P. (2012). Effects of ethnically segregated learning settings in the United States and Germany. In Gallagher, K., Goodyear, R., Brewer, D., and Rueda, R., editors, *Urban education: A model for leadership and policy*, pages 826–878. New York/London: Routledge.
- Enders, C. and Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods*, 12(2):121–138.
- Entwisle, D. and Alexander, K. (1994). The racial composition of schools and learning to read. *American Sociological Review*, 59(3):446–460.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt am Main: Campus.
- Fekjær, S. and Birkelund, G. (2007). Does the ethnic composition of upper secondary schools influence educational achievement and attainment? A multilevel analysis of the Norwegian case. *European Sociological Review*, 23(3):309–323.
- Ganzeboom, H. (2010). *A new Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007, with an analysis of quality of occupational measurement in ISSP*. Lisbon: Annual Conference of International Social Survey Programme.

- Gehrer, K., Zimmermann, S., Artelt, C., and Weinert, S. (2012). *The assessment of reading competence (including sample items for grade 5 and 9). Scientific use file 2012, version 1.0.0.* Bamberg: University of Bamberg, National Educational Panel Study.
- Haberkorn, K. and Pohl, S. (2013). *Cognitive basic skills – Data in the scientific use file.* Bamberg: University of Bamberg, National Educational Panel Study.
- Hanushek, E. (1997). Assessing the effects of school resources on student performance: An update. *Educational Evaluation and Policy Analysis*, 19(2):141–164.
- Harker, R. and Tymms, P. (2004). The effects of student composition on school outcomes. *School Effectiveness and School Improvement*, 15(2):177–199.
- Hattie, J. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, 37(5):449–481.
- Heath, A. F. and Brinbaum, Y. (2007). Guest editorial: Explaining ethnic inequalities in educational attainment. *Ethnicities*, 7(3):291–304.
- Hermansen, A. and Birkelund, G. (2015). The impact of immigrant classmates on educational outcomes. *Social Forces*, 94(2):615–646.
- Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation (NBER Working Series 7876).* Cambridge: National Bureau of Economic Research.
- Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. *Zeitschrift für Pädagogik*, 52(6):765–773.
- Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., and Will, G. (2011). The education of migrants and their children across the life course. In Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J., editors, *Education as a lifelong process: The German National Educational Panel Study (NEPS)*, pages 121–137. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kristen, C., Olczyk, M., and Will, G. (2016). Identifying immigrants and their descendants in the National Educational Panel Study. In Blossfeld, H.-P., von Maurice, J., Bayer, M., and Skopek, J., editors, *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*, pages 195–211. Wiesbaden: Springer VS.
- Kroneberg, C. (2008). Ethnic communities and school performance among the new second generation in the United States: Testing the theory of segmented assimilation. *The Annals of the American Academy of Political and Social Science*, 620(1):138–160.

- Maestri, V. (2011). *A deeper insight into the ethnic make-up of school cohorts: Diversity and school achievement (AIAS Working Paper 11-108)*. Amsterdam: Amsterdam Institute for Advanced Labour Studies, University of Amsterdam.
- Mickelson, R., Bottia, M., and Lambert, R. (2013). Effects of school racial composition on K-12 mathematics outcomes: A metaregression analysis. *Review of Educational Research*, 83(1):121–158.
- Mouw, T. and Xie, Y. (1999). Bilingualism and the academic achievement of first- and second-generation Asian Americans: Accommodation with or without assimilation? *American Sociological Review*, 64(2):232–252.
- OECD (2010). *PISA 2009 results: Overcoming social background—Equity in learning opportunities and outcomes (volume II)*. Paris: OECD.
- Olczyk, M., Seuring, J., Will, G., and Zinn, S. (2016). Migranten und ihre Nachkommen im deutschen Bildungssystem: Ein aktueller Überblick. In Diehl, C., Hunkler, C., and Kristen, C., editors, *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*, pages 33–70. Wiesbaden: Springer VS.
- Peetsma, T., van der Veen, I., Koopman, P., and van Schooten, E. (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement*, 17(3):275–302.
- Portes, A. and Hao, L. (2002). The price of uniformity: Language, family and personality adjustment in the immigrant second generation. *Ethnic and Racial Studies*, 25(6):889–912.
- Portes, A. and Hao, L. (2004). The schooling of children of immigrants: Contextual effects on the educational attainment of the second generation. *Proceedings of the National Academy of Sciences of the United States of America*, 101(33):11920–11927.
- Rjosk, C., Richter, D., Hochweber, J., Lüdtke, O., Klieme, E., and Stanat, P. (2014). Socioeconomic and language minority classroom composition and individual reading achievement: The mediating role of instructional quality. *Learning and Instruction*, 32:63–72.
- Roscigno, V. (1998). Race and the reproduction of educational disadvantage. *Social Forces*, 76(3):1033–1060.
- Rosenthal, R. and Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review*, 3(1):16–20.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Rumberger, R. and Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers College Record*, 107(9):1999–2045.

- Salikutluk, Z. (2016). Why do immigrant students aim high? Explaining the aspiration-achievement paradox of immigrants in Germany. *European Sociological Review*, 32(5):581–592.
- Scheele, A. F., Leseman, P. P. M., and Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1):117–140.
- Seidel, T. and Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4):454–499.
- Skopek, J., Pink, S., and Bela, D. (2013). *Starting cohort 4: Grade 9 (SC4). Scientific use file, version 1.1.0. Data manual*. Bamberg: University of Bamberg, National Educational Panel Study.
- Stanat, P., Rauch, D. P., and Segeritz, M. (2010a). Schülerinnen und Schüler mit Migrationshintergrund. In Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., and Stanat, P., editors, *PISA 2009: Bilanz nach einem Jahrzehnt*, pages 200–230. Münster: Waxmann.
- Stanat, P., Schwippert, K., and Gröhlich, C. (2010b). Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb: Längsschnittliche Überprüfung eines umstrittenen Effekts. In Allemann-Ghionda, C., Stanat, P., Göbel, K., and Röhner, C., editors, *Migration, Identität, Sprache und Bildungserfolg*, pages 147–164. Weinheim: Beltz.
- Stewart, E. (2008). School structural characteristics, student effort, peer associations, and parental involvement: The influence of school- and individual-level factors on academic achievement. *Education and Urban Society*, 40(2):179–204.
- Strobel, B. (2016). Does family language matter? The role of foreign language use and family social capital in the educational achievement of immigrant students in Germany. *Ethnic and Racial Studies*, 39(14):2641–2663.
- Strobel, B. and Seuring, J. (2016). Spracherhalt oder Sprachverlagerung? Erstsprachgebrauch und Zweitsprachkompetenzen bei Jugendlichen mit Migrationshintergrund. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(2):309–339.
- Thrupp, M. (1999). *Schools making a difference: Let's be realistic!* Buckingham: Open University Press.
- Thrupp, M., Lauder, H., and Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37(5):483–504.
- Van der Slik, F., Driessen, G., and De Bot, K. (2006). Ethnic and socioeconomic class composition and language proficiency: A longitudinal multilevel examination in Dutch elementary schools. *European Sociological Review*, 22(3):293–308.

- Van Ewijk, R. and Sleegers, P. (2010a). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review*, 5(2):134–150.
- Van Ewijk, R. and Sleegers, P. (2010b). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21(3):237–265.
- Van Tubergen, F. and Kalmijn, M. (2009). Language proficiency and usage among immigrants in the Netherlands: Incentives or opportunities? *European Sociological Review*, 25(2):169–182.
- Van Tubergen, F. and Mentjox, T. (2014). Minority language proficiency of adolescent immigrant children in England, Germany, the Netherlands, and Sweden. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(1):241–262.
- Wölfer, R., Caro, D. H., and Hewstone, M. (2019). Academic benefit of outgroup contact for immigrant and nonimmigrant students. *Group Processes & Intergroup Relations*, 22(3):419–433.

5.9 Appendix

Table A5.1: Distribution and description of minority language population

	<i>N</i>	%	Corr. MiL(%) and sMiL(%) ^a	
Turkish	836	26.8	.75	*
Russian	655	21.0	.54	*
Polish	246	7.9	.21	*
Italian	176	5.6	.39	*
Arabic	161	5.2	.49	*
Albanian	109	3.5	.35	*
English	95	3.0	-.10	
Spanish	76	2.4	.23	*
Kurdish	72	2.3	.15	
Greek	72	2.3	.49	*
Bosnian	63	2.0	.08	
Croatian	62	2.0	.23	
Portuguese	54	1.7	-.13	
Serbian	51	1.6	.32	*
Other	395	12.7	.06	
Total	3,123	100.0	.51	*

Notes: 'Other' combines all languages spoken by less than 50 students in the sample.

^a MiL(%)=proportion of minority language students in classroom; sMiL(%)=proportion of students speaking the same minority language in classroom (Pearson's r ; * $p < 0.05$).

Table A5.2: Robustness analysis of German reading comprehension using 4-point scales of minority language use

	Model 1		Model 2	
Fixed Part				
<i>Level 2</i>				
Minority language students (%)	-0.011 (0.006)		-0.011 (0.006)	
<i>Cross-level interaction</i>				
MiL*MiLS (%)	-0.013 (0.008)		-0.010 (0.008)	
<i>Level 1</i>				
Minority language (<i>ref. German</i>)	-0.172 (0.033)	***	-0.068 (0.062)	
MiL use with classmates (<i>ref. only German</i>)				
mostly German	-0.047 (0.044)		-0.032 (0.045)	
mostly MiL	-0.175 (0.074)	*	-0.146 (0.074)	*
only MiL	-0.250 (0.134)		-0.195 (0.135)	
MiL use with parents (<i>ref. only German</i>)				
mostly German			-0.093 (0.061)	
mostly MiL			-0.167 (0.064)	**
only MiL			-0.220 (0.074)	**
Intercept	-0.358 (0.034)	***	-0.355 (0.034)	***
Random Part				
Within-classroom variance	0.555		0.554	
Between-classroom variance	0.037		0.037	
R ² within classrooms	0.10		0.11	
R ² between classrooms	0.91		0.91	

Notes: Coefficients represent standard deviation changes in the dependent variable, with standard errors in parentheses. Analyses are controlled for school track, mean cognitive ability, mean SES (Level 2) and cognitive ability, SES, parents' education, idealistic aspirations, born abroad, female, age (Level 1). *N* students=14,815, *N* classrooms=1,006. MiL(S)=Minority language (students).

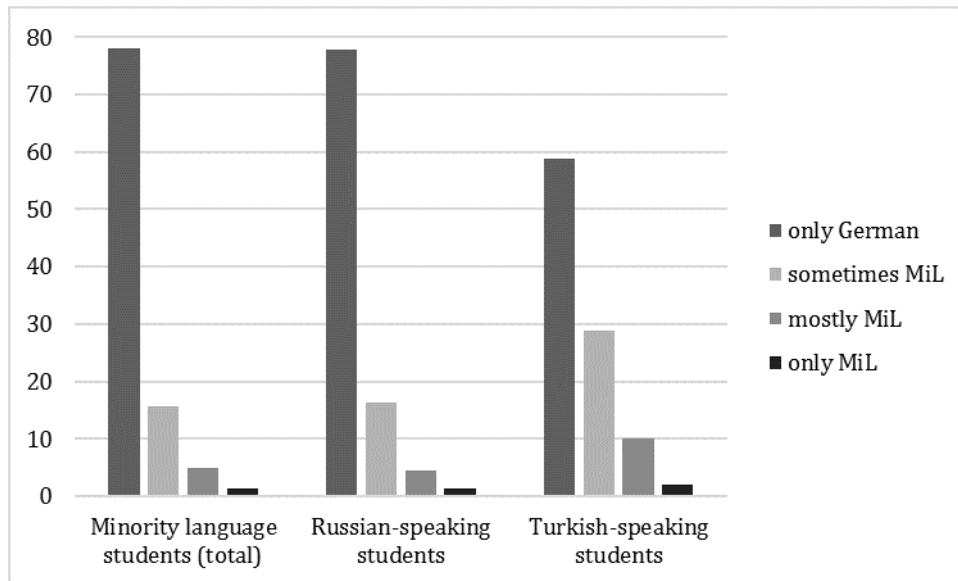
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A5.3: Robustness analysis of German reading comprehension with additional controls (minority language students only)

	Model 0	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Fixed Part							
<i>Level 2</i>							
Minority language students (%)	-0.027 (0.008)	**	-0.021 (0.008)	*	-0.014 (0.009)	-0.015 (0.009)	-0.014 (0.009)
<i>Level 1</i>							
Born abroad	-0.010 (0.033)		-0.015 (0.033)		-0.019 (0.033)	-0.018 (0.033)	-0.011 (0.033)
Proportion of co-ethnics in neighbourhood ^a			-0.053 (0.009)	***	-0.049 (0.009)	-0.048 (0.009)	-0.047 (0.009)
Proportion of same MiL students ^b			-0.031 (0.014)	*	-0.032 (0.014)	-0.022 (0.014)	-0.023 (0.014)
Idealistic aspirations (<i>ref. low</i>)					0.131 (0.035)	0.128 (0.035)	0.131 (0.035)
MiL use with classmates (<i>ref. only German</i>)							
MiL use with parents (<i>ref. only German</i>)							
Intercept	-0.463 (0.060)	***	-0.336 (0.064)	***	-0.349 (0.064)	-0.388 (0.065)	-0.277 (0.081)
Random Part							
Within-classroom variance	0.553		0.495		0.493	0.490	0.489
Between-classroom variance	0.378		0.046		0.046	0.046	0.046
R ² within classrooms	0.09		0.10		0.11	0.11	0.12
R ² between classrooms	0.88		0.88		0.88	0.88	0.88

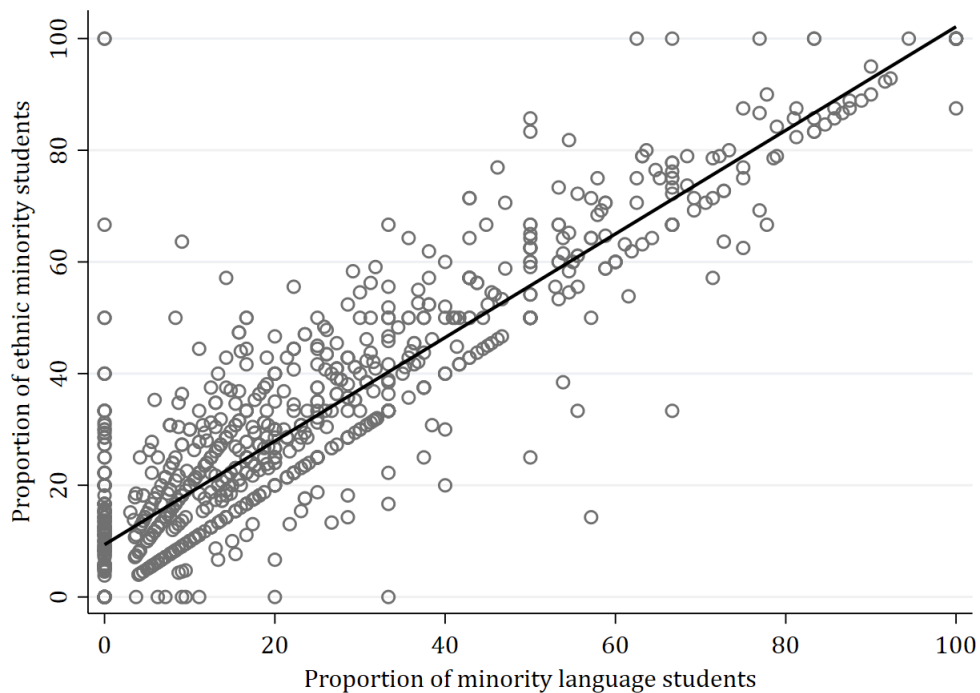
Note: Coefficients represent standard deviation changes in the dependent variable, with standard errors in parentheses. Analyses are controlled for school track, mean cognitive ability, mean SES (Level 2) and cognitive ability, SES, parents' education, female, age (Level 1). *N* students=3,123, *N* classrooms=798. MiL=Minority language. ^a Based on minority students' self-ratings of the proportion of people living in their neighbourhood or their parents that migrated from the same country of origin as the students themselves or their parents (1 "up to 10 percent" to 6 "more than 50 percent"). ^b Proportion of classmates who have learned the same minority language in family as individual student; aggregated from individual information for each student. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Figure A5.1: Distribution of language use with classmates (in percent; minority language students only)



Notes: MiL = minority language.

Figure A5.2: Correlation between proportion of ethnic minority students and proportion of minority language students on classroom level



Notes: $N=1,006$ classrooms. Pearson's $r=.89$ ($p<0.001$).

6 *Artikel 3: Why bother with testing?* The validity of immigrants' self- assessed language proficiency

Dieser Beitrag entstand in Kooperation mit Aileen Edele, Cornelia Kristen und Petra Stanat und ist in *Social Science Research* (©2015 Elsevier Inc. All rights reserved.) erschienen. Der nachfolgende Text und Inhalt wurde unverändert aus dem veröffentlichten Originalbeitrag übernommen und an das Format der Dissertation angepasst. Die Referenz lautet:

Edele, A., Seuring, J., Kristen, C. und Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, 52:99–123. <https://doi.org/10.1016/j.ssresearch.2014.12.017>

Abstract

Due to its central role in social integration, immigrants' language proficiency is a matter of considerable societal concern and scientific interest. This study examines whether commonly applied self-assessments of linguistic skills yield results that are similar to those of competence tests and thus whether these self-assessments are valid measures of language proficiency. Analyses of data for immigrant youth reveal moderate correlations between language test scores and two types of self-assessments (general ability estimates and concrete performance estimates) for the participants' first and second languages. More importantly, multiple regression models using self-assessments and models using test scores yield different results. This finding holds true for a variety of analyses and for both types of self-assessments. Our findings further suggest that self-assessed language skills are systematically biased in certain groups. Subjective measures thus seem to be inadequate estimates of language skills, and future research should use them with caution when research questions pertain to actual language skills rather than self-perceptions.

6.1 Introduction

Across disciplines, research on immigrants and their descendants¹ views language skills as a key to their educational, economic and social integration into receiving countries (Chiswick and Miller 2001; Jimenez 2011; Portes and Rumbaut 2006). A substantial body of research has examined the conditions for acquiring the language of the country of residence (second language, L2) and for the maintenance or loss of the language spoken in the country of origin (first language, L1) (e.g., Alba et al. 2002; Carliner 2000; Chiswick and Miller 2001; Van Tubergen and Kalmijn 2009). In addition, various studies have aimed to determine the effects of immigrants' language skills in both L1 and L2 on core aspects of their integration, such as educational attainment (e.g., Bankston and Zhou 1995; Lutz and Crist 2009; Mouw and Xie 1999; Nielsen and Lerner 1986) or economic success (e.g., Chiswick and Miller 1995; Dustmann and Fabbri 2003; Hwang et al. 2010; Kaida 2013).

Whereas studies of smaller samples sometimes apply psychometric tests to assess immigrants' linguistic skills, population and household censuses as well as large-scale studies typically measure participants' language proficiency with self-assessments. A substantial proportion of research on immigrants' language proficiency, particularly in sociology and economics but also in other fields, relies on data from these studies and thus on self-reports (e.g., Berry et al. 2006; Carliner 2000; Chiswick and Miller 1995; Chiswick et al. 2004; Mouw and Xie 1999; Pendakur and Pendakur 2002; Van Tubergen and Kalmijn 2009).

Self-assessments (or self-estimates) of ability (or skill) reflect a person's self-concept, which can be broadly defined as "a person's perception of himself" (Shavelson et al. 1976, p. 411) in a particular domain. Whereas earlier views of this construct focused on the global self-concept, domain-specific conceptualizations now prevail (e.g., Bong and Skaalvik 2003; Marsh 1990, 1992). As psychological research shows, self-concepts of ability correlate substantially with actual levels of ability in a given domain (Marsh 1992). Whether self-assessed skills present valid indicators of actual skills, however, is questionable, and the extent to which self-assessments of language skills measure the same (or at least very similar) skills as language tests is unclear. In other words, we cannot presume that, for example, investigations using self-assessed language skills as predictors of immigrants' educational outcomes will arrive at the same conclusions as analyses using language tests.

The main purpose of the present study is to determine whether language self-assessments of the type that is frequently used in research on immigrants yield results similar to those of competence tests. Our analyses draw on a large sample of immigrant youth participating in the German National Educational Panel Study (NEPS), for which both self-assessed language skills and tested language abilities in L1 and L2 are available. The first set of analyses focuses on L2 abilities and includes all students of immigrant descent assessed in the 9th-grade sample of the NEPS (4078 students in

¹The term "immigrant", as it is used in this paper, refers to immigrants (first generation) and their children (second and third generations).

total). The families of the two largest groups of immigrant students immigrated to Germany from the Former Soviet Union or from Turkey, and the number of students of Polish descent is also substantial. Reflecting the heterogeneity of the immigrant population in Germany, however, most students are of other origins. The second set of analyses focuses on immigrant students' L1 ability and is based on a subsample of Russian- or Turkish-speaking students (1164 students in total).

In a first step, we correlate language self-assessments with performance on language tests, assuming that the language tests provide valid estimates of participants' actual linguistic skills. We then move beyond this bivariate approach and use self-assessed language skills vs. language test scores in a number of regression models pertaining to substantive research questions to compare the findings. The validity of self-assessments is supported if these models yield similar findings, whereas substantial differences in the results indicate that self-assessments are of limited validity.

Determining the validity of self-assessed language skills is important for at least two reasons. First, it provides an indication of the extent to which the findings from the large number of previous studies that have relied on self-assessments are valid (e.g., Berry et al. 2006; Carliner 2000; Chiswick and Miller 1995; Mouw and Xie 1999; Pendakur and Pendakur 2002; Van Tubergen and Kalmijn 2009). Second, our study provides important guidance for designing future research. Because testing is time consuming and expensive, researchers typically prefer to use self-assessments. However, this route should only be taken if the measures applied are valid.

6.2 Evidence for the validity of ability self-assessments

In general, the accuracy of ability self-assessments is questionable. It is not just an amusing bon mot but an empirical finding that eight out of ten drivers report that they have above-average driving skills (McCormick et al. 1986). Driving is certainly not the only domain in which people have difficulty adequately estimating their own abilities. Meta-analyses examining the correlation between ability self-assessments and test scores or teacher ratings in various domains indicate that self-assessments are often inaccurate (Falchikov and Boud 1989; Freund and Kasten 2012; Hansford and Hattie 1982; Mabe and West 1982). These studies find mean correlations that are positive, yet moderate (typically ranging between .3 and .4) with considerable variability in the correlation coefficients across studies. Although some researchers argue that ability self-assessments are accurate (e.g., Ackerman et al. 2002), most scientists agree that their validity is limited (e.g., Carter and Dunning 2008; Dunning et al. 2004; Kruger and Dunning 1999).

The self-assessment measures that have been implemented in empirical studies vary substantially. An important distinguishing feature is their degree of specificity. Some self-assessments refer to broad domains and underlying abilities that affect behavior across situations and over time, such as scientific or reading ability, and are thus quite

general. Other self-assessments focus on more specific skills, such as understanding a science article, or refer to performance on a concrete task that participants just completed in the same study or typically encounter in everyday situations. Many researchers assume that self-assessment measures with high degrees of specificity are more likely to evoke accurate responses because they are less ambiguous than more general measures (Ackerman et al. 2002; Freund and Kasten 2012; Mabe and West 1982; Marsh 1990). Empirical findings support this assumption. Kruger and Dunning (1999), for instance, found a substantial correlation between performance on a grammar test and a self-estimated test score, whereas participants' self-assessments of their general grammar ability were not significantly related to their test performance. Integrating the findings from a larger number of studies, a meta-analysis confirmed that self-assessed performance on specific tasks relates more strongly to the validation criteria than to self-assessed underlying abilities (Mabe and West 1982). Thus, self-assessments with high degrees of specificity tend to be more accurate and may therefore produce results more similar to those of tests.

6.2.1 Correlations between self-assessed and tested language proficiency

Some studies on the validity of self-assessments have focused on measures of linguistic abilities. A meta-analysis of the relationship between self-assessed language ability and language tests or teacher ratings revealed a mean correlation of $r=.63$ (Ross 1998). However, this investigation included only 10 studies, which primarily focused on foreign language learners and used very small, selective samples; therefore, its generalizability is limited.

We reviewed a larger number of studies, including those in Ross' (1998) meta-analysis (see Appendix 6.8, Table A6.1). To be considered, the studies had to contain self-assessments of linguistic skills as well as performance on psychometric tests or teacher ratings. The majority of studies focused on students acquiring a foreign language in college. Fewer investigations included second language learners for whom the language of the residence country was an L2 or focused on participants' first language abilities. The number of studies including more heterogeneous adult samples or younger participants was also limited (see column 2 for details on the samples). The majority of these studies reported positive correlations. The most striking finding is the large variation in the correlation coefficients across studies, ranging from zero to very high correlations (see column 5). One reason for this is most likely the large variability in the quality of the instruments the studies used. Whereas some studies utilized well-established tests as criteria, others relied on criteria with questionable validity, such as "language tests" of dubious quality or teacher ratings (see column 4). Similarly, the studies included very different types of self-assessments (see column 3). Some analyses utilized the type of language self-assessment that is typically applied in large-scale studies and censuses to measure participants' language proficiency (e.g., Brantmeier 2006; Charette and Meng 1994; Finnie and Meng 2005). These questionnaire items ask participants to indicate

their language ability with respect to one or several global linguistic dimensions (e.g., understanding, speaking, reading, writing) on four- or five-point rating scales and thus represent a general type of self-assessment. The majority of studies, however, used more specific types of self-assessments that asked participants to estimate their ability with regard to concrete criteria. These criteria were often linguistic tasks participants typically encounter in everyday situations (labeled “real-life tasks” in Appendix 6.8, Table A6.1), such as understanding a letter describing the events, feelings and wishes of the writer (see Brantmeier et al. 2012). Another type of specific self-assessment refers to linguistic tasks completed by participants in the same study (labeled “study tasks” in Appendix 6.8, Table A6.1), such as the self-rating of how well participants’ understand a text they have read (Brantmeier and Vanderplank 2008) or the estimated number of correct answers in a test (see Lin et al. 2001).

The few studies that have examined the validity of the general question type have yielded inconsistent findings. A study drawing on a large sample of adult Canadians reported a correlation of .42 between self-assessed literacy (a composite measure for reading and writing skills) and a test of functional literacy (Finnie and Meng 2005). In contrast, Brantmeier (2006) did not identify a relationship between this type of self-assessment and test scores. It is thus unclear whether the general type of self-assessment on which most empirical evidence on immigrants’ language skills relies is reliably related to tested language proficiency.

The majority of studies utilizing more specific self-assessments, in contrast, have found substantial correlations between participants’ self-estimated language abilities and their test scores. Brantmeier et al. (2012), who focused on advanced foreign language learners in college, found a medium correlation ($r=.34$) between a self-assessment scale including 20 “can-do” items reflecting the content of participants’ language course and a computer-based multiple choice test. Similarly, the authors reported that three types of specific self-assessment indicators (an item describing different levels of reading ability in great detail; a scale including 15 “can-do-statements”; and 4 items assessing the performance in the just-completed reading tests) accounted for participants’ performance in a computer-based reading test as well as in a paper-based reading test (Brantmeier and Vanderplank 2008). These findings are in line with a study based on a very large sample of European foreign language learners that correlated specific self-assessment indicators of reading, listening and writing with tests of these abilities and found medium correlations (Alderson 2005). Some studies revealed even higher correlations between specific self-assessments and test scores. LeBlanc and Painchaud (1985) asked foreign language learners in college to estimate their listening and reading skills based on extensive self-assessment scales (30 and 20 items, respectively) that were tailored to the university context and were thus highly context-specific. They found correlations up to .80 between the self-assessments and test scores. A study of adult learners of English as a foreign language that used self-assessments describing levels of speaking and listening ability in great detail reported similarly high correlations with participants’ test performance (Wilson 1999).

Although some studies have reported smaller correlations between specific self-assessments and test scores (e.g., Peirce et al. 1993; Rauch et al. 2010), the current state

of research overall indicates that self-assessment indicators referring to specific criteria rather than abstract abilities are substantially related to participants' performance in language tests (see Brantmeier et al. 2012; Ross 1998). However, these measures are typically quite extensive and take a long time to complete. Moreover, positive correlations between self-assessments and test scores alone do not ensure unbiased findings. Provided that a language test has been properly constructed and meets general psychometric requirements, a less than perfect correlation between test scores and self-assessments suggests that the latter lacks reliability. If measurement error is random and falls within a reasonable range, however, such self-reports may nevertheless be useful. They would typically yield smaller effects in analyses, yet the conclusions would not necessarily be incorrect. The situation is different when the error is systematic and especially when it is associated with other factors of interest. Such biases can be detected by comparing patterns of relationships in analyses using self-reports of language proficiency and language proficiency tests.

6.2.2 Multivariate analyses using self-assessed and tested language proficiency

Research on immigrants' language proficiency typically either seeks to identify factors that influence their linguistic skills (language proficiency as the dependent variable) or to determine the relative effect of linguistic skills on various outcomes, such as academic success or labor-market integration (language proficiency as an independent variable). The question is whether the use of self-assessment measures biases such multivariate analyses. Only a few studies have addressed this issue systematically by comparing results using subjective measures vs. test scores. One early investigation contrasted self-assessments and test scores in models examining the determinants of immigrants' L2 skills and in models analyzing the effects of L2 on various aspects of social integration (Esser 1985). Overall, its findings were similar for both measures (e.g., participants' education was the strongest predictor of L2 in both analyses). Finnie and Meng (2005), in contrast, analyzed data from a larger sample and found substantial differences between the results for self-assessments and test scores. When using test scores instead of self-assessments as the dependent variable, the estimates were more consistent with theoretical considerations and previous findings. In addition, tested L2 ability predicted participants' employment status and income reliably, whereas self-assessed language skills did not.

6.2.3 Limitations of previous research

Although the validity of self-assessed language abilities has been explored in a number of studies, the state of research is still inconclusive for various reasons.

First, previous research has primarily focused on foreign language learners, and it is questionable whether these findings generalize to immigrants. Whereas foreign language learners typically can draw on explicit feedback regarding their linguistic performance (e.g., teacher responses, grades, exams), immigrants often lack such feedback.

In addition, foreign language learners have a relatively clear frame of reference for estimating their language skills, namely, curricular expectations and the performance of their classmates. The reference for immigrants, in contrast, is more ambiguous. For example, they may compare their L2 skills to those of other immigrants or to those of non-immigrants and their L1 skills to those of co-ethnics with good, basic or no skills in the language. This ambiguity may decrease the accuracy of immigrants' self-assessments. Second, most past research has focused on L2; thus, we know very little about the validity of immigrants' self-assessed L1 abilities. A third limitation of previous research is that most studies have used small, selective samples, and the extent to which their findings can be generalized is unclear. Fourth, previous research on the validity issue has often applied criterion measures of limited quality, such as teacher ratings or "language tests" of questionable validity rather than sound psychometrical instruments. A fifth limitation is that only a few analyses have addressed the validity of the different types of self-assessments that are typically implemented in large-scale studies, yet there is evidence that the accuracy of self-assessments may depend on the characteristics of the measure, particularly its degree of specificity. Finally, only a few studies have moved beyond the common bivariate approach and compared the effects of using self-assessments vs. test results in multivariate analyses.

6.3 Analytical approach

This study addresses the validity of self-assessed language skills in L1 and L2 using data from a large sample of immigrant students. We compare two different self-assessment measures to sound psychometrical tests. We focus on types of self-assessments that are typically used in research on immigrants—that is, very efficient measures—rather than developing potentially more valid instruments. In addition to bivariate correlational analyses, we examine the validity of self-assessments with multivariate models using self-assessments and test scores as dependent and independent variables. If self-assessed language skills are valid indicators of immigrants' language proficiency, then the findings for self-assessments and test scores should be nearly identical.

Our analyses build upon current research on immigrants' language proficiency. One set of questions raised in the literature addresses the conditions under which immigrant students develop high levels of language skills. Analyses on this issue treat language proficiency as the dependent variable. Another set of questions concerns language skills as an independent variable that affects further language learning and educational achievement. Below, we briefly outline central theoretical arguments related to both sets of questions. Because the main purpose of the present study is to determine the validity of self-reports, we limit this description to the main points.

6.3.1 Determinants of L1 and L2 skills

Models of language acquisition frequently refer to three types of factors relevant to learning: motivation, exposure and learning efficiency (e.g., Chiswick et al. 2004; Chiswick

and Miller 1995, 2001; Esser 2006). *Motivation* typically refers to economic incentives; in a broader view, it encompasses a variety of incentives associated with the command of a particular language (Esser 2006), such as the benefits of being able to speak to friends from the majority population. *Exposure* refers to the time units during which immigrants are exposed to a language and to the intensity of that exposure (Chiswick and Miller 1995). *Efficiency* relates to “the extent of improvement in destination-language skills per unit of exposure” (Chiswick and Miller 2001, p. 393), recognizing, for instance, that the capacity for language learning is greater in younger immigrants (Long 1990).

The three types of factors can be linked to a variety of determinants that influence the development of skills in L1 and L2. For example, strong identification with the country of residence can increase the *motivation* to learn and improve L2, whereas strong identification with the country of origin should increase the motivation to maintain L1. Similarly, one would expect that co-ethnic networks offer an incentive to retain and further develop L1 skills, whereas contacts and friendships with members of the majority population should increase the motivation for L2 acquisition. At the same time, depending on the composition of an immigrant’s networks, *exposure* to either L1 or L2 may increase or decrease. Another important condition relevant to exposure includes the language students use in their everyday lives—in particular, with their family and peers. Moreover, the first immigrant generation is likely to have experienced more exposure to L1 (for example, in the country of origin prior to moving) than the second generation. Finally, the *efficiency* of language acquisition relates to the capacity for learning, which varies with such factors as age and cognitive ability.

6.3.2 L1 skill as a determinant of L2 proficiency

Some authors argue that L1 proficiency has positive effects on L2 acquisition, whereas others consider L1 skills detrimental or irrelevant to L2 learning (Kristen et al. 2011). These arguments typically relate to efficiency and exposure. The linguistic interdependence hypothesis, for example, assumes that knowledge in L1 transfers to L2 and therefore increases a student’s efficiency in acquiring the second language (Cummins 1979, 2000). According to this view, the linguistic level attained in L1 partially accounts for the skill level second language learners develop in L2. An opposing perspective argues that L1 skills are irrelevant or detrimental to the development of high levels of L2 (e.g., Esser 2006). A core argument for this position is the language competition hypothesis, which states that using L1 detracts time that could otherwise be invested in improving the L2. This reasoning pertains to language exposure.

6.3.3 L2 skill as a determinant of educational achievement

Linguistic abilities are an important tool for learning, especially for acquiring academic competencies. According to the learning model outlined above, higher proficiency in a school’s language of instruction, which is often an L2 in the case of immigrant students, increases the efficiency of learning and thus of academic achievement, for instance in mathematics. In line with this assumption, there is considerable evidence that L2 skills

are an important condition for academic achievement, including mathematics (e.g., Duncan et al. 2007; Grimm 2008; LeFevre et al. 2010). The central role of L2 skills in mathematics achievement is also apparent in immigrant students (e.g., Barrett et al. 2012; Kleemans et al. 2011; Vukovic and Lesaux 2013). For instance, Beal et al. (2010) found that the L2 reading skills of adolescent language minority students predicted their math performance.

6.3.4 Two types of self-assessment

In our multivariate analyses of the validity of self-assessed language proficiency, we addressed these issues with different models using two types of self-assessments. The first type is typically used in large-scale studies and censuses. Participants estimate their linguistic ability with regard to general linguistic dimensions (i.e., understanding, speaking, reading and writing, see Appendix 6.8, Figures A6.1 and A6.2). This measure—hereinafter referred to as *ability self-assessment* or *ability SA*—is rather broad. Because the bulk of findings on immigrants' language proficiency relies on this general type of self-assessment, it is particularly important to explore its validity.

The second measure—labeled *performance self-assessment* or *performance SA*—refers to self-estimated performance immediately after completing a test of linguistic ability (see Appendix 6.8, Figures A6.3 and A6.4 and Instruments section for a detailed description of the measure). Because it refers to a concrete task and situation this indicator is more specific than the *ability self-assessment*. Relating to a linguistic task the participants encounter within the study, it is more specifically a study-task (see Section 6.2.1). Previous findings suggest that specific self-assessments correlate more highly with test scores in various domains (e.g., Ackerman et al. 2002; Mabe and West 1982) including language (Brantmeier and Vanderplank 2008; Brantmeier et al. 2012; Ross 1998). Although our *performance SA* measure differs in several respects from the specific measures that show particularly high correlations in other studies (these instruments, for instance, often refer to real-life situations rather than study tasks and typically have a larger number of items or describe the criteria in great detail; see LeBlanc and Painchaud 1985; Wilson 1999), it is clearly more specific than general *ability SAs*. Therefore, we expected *performance SAs* to correlate more closely with test scores and to show a higher congruence with regression models using test scores compared with *ability SAs*.

6.4 Method

6.4.1 Data

We used data from the National Educational Panel Study (NEPS), a nationwide study conducted in Germany that draws upon a multi-cohort sequence design (Blossfeld et

al. 2011). The analyses focus on the starting cohort 4 sample of 9th-grade students.² For this cohort, the NEPS provides test scores in L2 for all students and in L1 for students who speak Russian or Turkish.³ These are the first languages of the two largest immigrant groups in Germany. In addition to test scores, the data include *ability* and *performance self-assessments*.

The first subsample used in the analyses of L1 included 502 9th graders who completed the Russian L1 test and 662 students who completed the Turkish L1 test (see Table 6.1, sample 1). Our analyses of L2 abilities included all students of immigrant origin. Within this second subsample, we distinguished four groups: 708 students whose families migrated from the Former Soviet Union, 878 students of Turkish origin, 275 students of Polish descent, and 2217 students of other origins (see Table 6.1, sample 2).

6.4.2 Instruments

The NEPS includes both test-based and self-estimated indicators of students' language skills in L1 and L2. In addition, it provides two measures of math achievement, one of the central outcome variables in our analyses. Moreover, the NEPS provides several indicators of motivation, language exposure and learning efficiency, as well as a number of control variables.

L1 tests: The L1 tests assessed students' listening comprehension at the text level in their first language: Russian or Turkish. Listening comprehension constitutes a complex, integrated process that requires a multitude of top-down and bottom-up processes (Andringa et al. 2012; Flowerdew and Miller 2005; Rost 2011). Unlike instruments that assess linguistic component skills (e.g., vocabulary tests), tests of listening comprehension thus cover the mastery of a language broadly and circumvent an overly narrow representation of the construct (see, e.g., Messick 1995). The Russian and the Turkish tests were constructed analogously and consisted of seven short texts with approximately 100–150 words each. Three different text types were represented in the tests: two dialogues, three narrative texts, and three expository texts. Every text was followed by three to six multiple-choice questions with 4 or 5 response options each that asked students about the content of what they had heard, resulting in a total of 31 test items.⁴ Cognitive interviews ensured that participants answered items based on information from the text rather than based on their prior knowledge (Edele et al., 2012, 2016). The internal consistency (Cronbach's alpha) was .87 for the Russian test and .84 for the Turkish test in our sample (sample 1).

²doi:10.5157/NEPS:SC4:1.0.1. From 2008 to 2013, NEPS data were collected as part of the Framework Programme for the Promotion of Empirical Educational Research, which was funded by the German Federal Ministry of Education and Research and supported by the federal states. As of 2014, the NEPS survey is conducted by the Leibniz Institute for Educational Trajectories (LifBi).

³Students who demonstrated a basic level of competence in Russian or Turkish in a short screening test participated in the L1 assessment (Edele et al. 2016).

⁴One item out of the 32 items originally included in the test was removed because of low model fit.

Table 6.1: Sample characteristics and distribution of model variables in terms of absolute and relative frequencies or means and standard deviations

	Sample 1						Sample 2																	
	Former SU			Turkey			Total			Former SU			Turkey			Poland			Other			Total		
	N	%	sd	N	%	sd	N	%	sd	N	%	sd	N	%	sd	N	%	sd	N	%	sd	N	%	sd
Overall	502	44.8		662	55.2		1,164	100.0		708	17.6		878	20.5		275	7.2		2,217	54.7		4,078	100.0	
M		sd	M		sd	M		sd	M		sd	M		sd	M		sd	M		sd	M		sd	
Language tests																								
L1 test ^a (WLE)	0.04	1.19	0.00	1.02																				
L2 test (WLE)	-0.49	1.10	-1.00	1.09	-0.78	1.12				-0.47	1.16	-0.97	1.13	-0.29	1.19				-0.44	1.27	-0.55	1.23		
Self-assessed language skills																								
Ability SA L1	2.87	0.72	3.32	0.57	3.13	0.68																		
Self-report L1: Underst. ^b	4		4		4																			
Self-report L1: Speaking ^b	4		4		4																			
Self-report L1: Reading ^b	3		3		3																			
Self-report L1: Writing ^b	2		3		3																			
Performance SA L1	0.62	0.23	0.68	0.23	0.65	0.23																		
Ability SA L2	3.59	0.52	3.52	0.50	3.55	0.51				3.60	0.53	3.51	0.53	3.64	0.49				3.60	0.55	3.58	0.54		
Self-report L2: Underst. ^b	4		4		4					4		4		4					4		4			
Self-report L2: Speaking ^b	4		4		4					4		4		4					4		4			
Self-report L2: Reading ^b	4		4		4					4		4		4					4		4			
Self-report L2: Writing ^b	4		3		3					4		3		4					4		4			
Performance SA L2	0.69	0.18	0.66	0.18	0.67	0.18				0.69	0.18	0.66	0.18	0.70	0.17				0.69	0.19	0.69	0.18		
Competencies																								
Mathematics test (WLE)	-		-		-					-0.39	1.05	-0.74	0.88	-0.24	1.12				-0.36	1.16	-0.44	1.09		
Grades in mathematics	-		-		-					3.04	1.02	3.21	1.02	3.06	1.04				3.19	1.05	3.16	1.04		
General cognitive ability	8.31	2.43	7.25	2.54	7.71	2.54				8.30	2.46	7.23	2.60	8.15	2.70				8.02	2.73	7.91	2.68		
N		%	N		%	N		%	N		%	N		%	N		%	N		%	N		%	
Generation status																								
Born abroad	272	52.6	78	11.2	350	29.7				368	50.3	105	11.6	76	25.8				501	22.3	1,050	25.3		
Born in Germany	230	47.4	584	88.8	814	70.3				340	49.8	773	88.5	199	74.2				1,716	77.7	3,028	74.7		
Gender																								
Female	256	52.7	327	50.7	583	51.6				375	54.6	430	50.0	151	57.0				1,113	51.0	2,069	51.8		
Male	246	47.3	335	49.3	581	48.4				333	45.5	448	50.0	124	43.0				1,104	49.0	2,009	48.2		
School type																								
Hauptschule	206	31.4	330	38.9	536	35.6				289	31.8	423	38.4	70	19.5				815	28.5	1,597	30.5		
Realschule	124	30.8	130	25.0	254	27.6				170	29.4	158	22.7	61	25.9				403	22.5	792	24.0		
Gymnasium	98	24.3	94	17.9	192	20.7				139	24.5	149	20.3	70	29.9				583	31.0	941	27.6		
Gesamtschule	47	9.1	97	16.3	144	13.1				68	9.4	131	16.5	49	15.6				279	12.3	527	12.9		
MBG	27	4.3	11	2.0	38	3.0				42	5.0	17	2.2	25	9.2				137	5.8	221	5.1		

The test data were scaled according to a one-parameter logistic model (Rasch model; e.g., Embretson and Reise 2000), and the mean of the person estimates was restricted to zero. A test score close to zero indicates average L1 skills. The Russian and Turkish tests were scaled separately, resulting in *weighted maximum-likelihood estimates* (WLEs) as indicators of students' L1 abilities.

The Russian and Turkish tests were translations of a German listening comprehension test. The translation process aimed to maintain the comparability of the text content and questions, but it did not strive for complete equivalence. Investigations of measurement equivalence with multigroup confirmatory factor analyses for categorical data (e.g., Meredith 1993; Schroeders and Wilhelm 2011) demonstrated configural equivalence, corroborating that the tests assessed the same construct. More restrictive models of equivalence, however, were not confirmed (see Edele et al. 2016). Thus, neither the test scores nor the coefficients in the analyses for the two groups are directly comparable.

Analyses of data from a pilot study with 224 Russian-speaking 9th-graders and 310 Turkish-speaking 9th-graders that included several additional validation criteria and data for the first sample of the current study support the construct validity of the L1 tests by demonstrating their convergent and discriminant validity (see Edele et al. 2016). For example, the analyses indicated medium to high correlations between the L1 test scores and another test-based indicator of students' skill level in Russian or Turkish based on two scales of the Bilingual Verbal Ability Test, BVAT (Muñoz-Sandoval et al. 1998). The L1 tests also related substantially to four self-assessment indicators and to parents' appraisals of the students' L1 proficiency. In addition, analyses showed the expected positive associations between L1 tests and students' age at migration as well as frequency of L1 use.

L2 test: As an indicator of proficiency in German, the NEPS provides scores on a reading comprehension test (Gehrer et al. 2012, 2013), which was constructed based on the literacy concept (OECD 2009). Like listening comprehension, reading comprehension at the text level constitutes a multiply determined, integrated process requiring a multitude of top-down and bottom-up skills (e.g., Anderson 1995; Snow et al. 1998). Tests of reading comprehension thus present a broad indicator of language ability. The L2 test consisted of five texts with approximately 200–550 words each and represented five different text types: an information text, a commenting text, an advertising text, an instruction text, and a literary text. To ensure that the test was independent of students' prior knowledge, the topics of the texts pertained to general world knowledge and did not require particular knowledge, such as discipline-specific vocabulary or familiarity with specific text types. The five to seven tasks following each text required students to find specific information in the text, to draw text-related conclusions, or to reflect on the text more generally in terms of content or form. Overall, the test consisted of 31 items.⁵ Most of the items had a multiple-choice format (4 response options), but the test also included items that required participants to decide whether

⁵Two items out of the 33 items originally included in the test were removed because of low model fit.

a statement was correct or incorrect or to assign titles to text sections. Cronbach's alpha of the L2 test in sample 2 was .81.

These tests were scaled according to a partial-credit model that constrains the mean of the person parameters in the calibration sample to zero. The calibration sample included all students of starting cohort 4 (i.e., both monolingual and language-minority students). The mean of the achievement scores on the L2 test was quite low for the Turkish group and almost identical to the average of the calibration sample for the Polish group (see Table 6.1).

The pattern of relationships of the L2 test with external criteria in the current sample (sample 2) supports the construct validity of the instrument. For instance, the L2 test showed convergent validity because it correlated highly with a L2 vocabulary test based on the German version of the *Peabody Picture Vocabulary Test*, PPVT (Dunn and Dunn 2004) and correlated moderately with two indicators of self-assessed L2 ability. In addition, higher age at migration was negatively related to students' L2 test scores, whereas frequent use of the German language was positively related to the test score.

Ability self-assessment in L1/L2: Students with a first language other than German were asked to estimate their abilities in both L1 and L2 (“How good is your command of the other language/German?”) on the basic linguistic dimensions of understanding, speaking, reading and writing (see Appendix 6.8, Figures A6.1 and A6.2). The 5-point response scale ranged from “not at all” (0) to “very good” (4). We computed the mean value of students' ratings across these four dimensions and used it as a composite measure for *ability SA*. The reliability (Cronbach's alpha) of the L1 scale was .73 in the Russian sample and .76 in the Turkish sample, whereas the reliability of the L2 scale was .82 across all immigrant groups. All immigrant groups indicated average levels of L2 skills ranging between “quite good” (3) and “very good” (4). The ratings of L1 skills were somewhat lower, particularly in the Russian sample (see Table 6.1).

Performance self-assessment in L1/L2: After completing the L1 and L2 tests, the students were asked to estimate the number of items that they had answered correctly (see Appendix 6.8, Figures A6.3 and A6.4). The estimated number of correctly solved items divided by the total number of items in the test served as an indicator of students' *performance SA*.

Mathematics achievement: Students' mathematics achievement constitutes an important indicator of their academic success and was assessed with a test of mathematical literacy in the NEPS (Neumann et al. 2013). WLEs serve as estimates of students' mathematical abilities, with a reliability of .79 (Duchhardt and Gerdes 2013). Another indicator of mathematical achievement is a student's grade from his or her most recent report card. The self-reported grades ranged from 1 (best grade) to 6 (poorest grade).

Cultural identity: Cultural identity is assumed to affect immigrants' motivation for acquiring L2 and maintaining L1. The student questionnaire included two dimensions of cultural identity: identification with the culture of the country of residence (e.g., "I feel that I am part of the German culture") and identification with the culture of the country of origin (e.g., "I feel closely attached to this culture"). The 4-point response scale ranged from "does not apply" (1) to "fully applies" (4). Each dimension was measured with 4 items, for which we computed the respective mean values.

Co-ethnic neighborhood: A high proportion of co-ethnics in an immigrant's neighborhood is expected to positively affect both motivation to maintain L1 and exposure to L1 while also decreasing both motivation to better learn L2 and exposure to L2. Immigrant students estimated how many persons in their neighborhood belong to their own immigrant group. The 6-point response scale ranged from "up to 10 percent" (1) to "more than 50 percent" (6).

Language use: The extent to which students use a language is another core indicator of exposure. All of the language use items applied a 4-point response scale ranging from "only the other language" (1) to "only German" (4). Items about the mother, the father and siblings were aggregated to form an indicator of family language (e.g., "What language do you speak with your mother?"). Ratings for best friends and classmates were aggregated to indicate language use with peers. The mean value of seven items on media use that, for example, captured the language in which students read books or watch television, served as an indicator of the language of media use.

Generation status: Whether students are born in the country of origin or in the country of residence typically affects their exposure to L1 and L2, respectively. We therefore distinguished between students who were born abroad (first generation) and those who were born in Germany (including the second and third generations).

General cognitive ability: Students' nonverbal general cognitive ability serves as an indicator of learning efficiency and is used as a control variable in several regression analyses. The measure is based on a 12-item reasoning test consisting of matrices similar to those of the traditional Raven's test (e.g., Raven 2000). We used the sum of correct answers for the analyses, ranging from 0 to 12; in the second sample, the reliability of the scale was .70.

Control variables: We included several indicators of students' socioeconomic and sociocultural background as control variables. The highest educational qualification of the mother or the father was computed, distinguishing three categories: "no educational degree or low-level secondary education" (a maximum of 9 years of schooling), "middle-level secondary education" (at least 10 years of schooling) and "higher-level secondary education" (12 years or more of schooling). The highest International Socio-Economic Index of Occupational Status of the mother or the father served as an indicator of the

family's socioeconomic status (HISEI-08; Ganzeboom 2010). Families' cultural capital was captured by the number of books available in students' homes. The 6-point scale ranged from "none or very few (0 to 10 books)" to "enough to fill a large set of shelves (201 to 500 books)".

As additional control variables, we accounted for students' gender and whether they attended a *Hauptschule* (the lowest track of the German secondary school system), a *Realschule* (the intermediate track), a *Gymnasium* (the highest track leading to a university entrance degree), a *Gesamtschule* (a comprehensive school type) or a *Schule mit mehreren Bildungsgängen* (a school with several educational tracks).

6.4.3 Data analyses

To address item nonresponse, we used multiple imputation and estimated five datasets with complete information (e.g., Allison, 2001). The regression analyses were based on these imputed datasets. Following Rubin's (1987) approach, we combined the results of the single regression analyses performed on each dataset. Descriptive results are based on the original dataset.

In the ordinary least squares regression analyses, estimating robust standard errors allowed us to account for the multilevel data structure with students clustered in schools (Huber 1967; White 1982). In addition, we used design weights to adjust for information stemming from different strata (i.e., school types; Aßmann et al. 2012). In the table with descriptive statistics, we report weighted percentages (Table 6.1); for the multivariate models, we report unweighted coefficients. The estimates are robust across analyses based on the weighted and the unweighted data sets.

6.5 Results

We first examined the covariation of *ability SAs* and *performance SAs* with test scores. Next, we explored the exemplary research questions delineated above and alternately include the language test scores and the two types of self-assessments in a sequence of multiple regression models.

6.5.1 Correlations of self-assessed language skills with language test scores

First, we correlated the self-assessment indicators of linguistic ability with the scores in the respective language test for both L1 and L2 (see Table 6.2). In addition to the *performance SAs* and the composite *ability SAs*, we separately examined the self-assessment dimensions of understanding, reading, speaking and writing.

The upper half of Table 6.2, which shows the results for L1, indicates only moderate correlations, with substantially larger coefficients in the Russian sample than in the Turkish sample. In both groups, the composite indicator of *ability SAs* correlates more strongly with the L1 test scores than with the individual dimensions. As expected,

the correlations between test scores and *performance SAs* are notably higher than the correlations between test scores and *ability SAs*.

Table 6.2: Correlations between the self-assessment indicators and test scores for L1 and L2

	Under-stand	Speak	Read	Write	Ability SA	Perform. SA
<i>Self-assessment L1</i>						
L1 test						
Former SU	.24***	.33***	.42***	.34***	.44***	.58***
Turkey	.10*	.19***	.24***	.24***	.25***	.49***
<i>Self-assessment L2</i>						
L2 test						
Former SU	.23***	.23***	.22***	.14**	.22***	.28***
Turkey	.21***	.19***	.22***	.08*	.20***	.32***
Poland	.24***	.28***	.31***	.26***	.34***	.46***
Other	.22***	.23***	.25***	.19***	.25***	.39***

Note: Spearman's rho. * p<0.05, ** p<0.01, *** p<0.001.

The lower half of Table 6.2 shows the L2 results for students whose families immigrated to Germany from the Former Soviet Union, Turkey, Poland and other countries. The correlation coefficients between L2 test scores and *ability SAs* were small to medium in magnitude. Although the correlation between the composite *ability SA* and the L2 test did not consistently exceed the correlations with the individual linguistic dimensions, in general, it was slightly larger. Overall, differences among the correlations for the four immigrant groups were marginal and unsystematic, with the exception of the Polish group for which the coefficients tended to be higher. Again, the correlations between the test scores and *performance SAs* were substantially higher than for the *ability SAs*.

The relatively low correlations for L1 and L2 may be partly due to ceiling effects: the majority of students indicated “quite good” or “very good” language proficiency (see Table 6.1). Thus, the variability of the self-assessment is lower than that of the test scores. Taken together, *ability SAs* in particular do not seem to be very accurate, and the correlation analyses clearly challenges the assumption that self-assessments are valid indicators of language proficiency.

6.5.2 Multivariate analyses of immigrant students' L1 and L2 skills

6.5.2.1 Determinants of L1 skills

The first group of models analyzes determinants of L1 maintenance. Based on the considerations outlined above and consistent with previous research, we included factors assumed to drive the motivation for language learning (i.e., identification with the country of origin and the proportion of co-ethnics in the neighborhood), aspects of language exposure (i.e., generational status, the proportion of co-ethnics in the neighborhood and aspects of language use) and aspects of learning efficiency (i.e., general cognitive

ability). In addition, we controlled for indicators of students' social background, school type and gender.

For each language group, Table 6.3 shows three models. The first models (1_R and 1_T) predict students' scores on the L1 tests, the second models (2_R and 2_T) predict students' L1 *ability SAs*, and the third models (3_R and 3_T) predict students' L1 *performance SAs*. We standardized the dependent variables such that within the Russian and Turkish samples, it is possible to compare regression coefficients across models. Our findings support the validity of self-assessments if the effects of the independent variables on the self-assessed language skills are similar to their effects on the test scores.

Several predictors showed consistent effects regardless of the L1 indicator used. In the Russian sample, for instance, students born abroad consistently demonstrated higher L1 abilities than students born in Germany; in the Turkish sample, these two groups did not differ on any of the language indicators. Similarly, in both samples, frequent use of German in the family had a negative effect on L1 proficiency, regardless of how L1 was measured.

Other variables, in contrast, showed differential effects depending on the L1 indicator used. As expected, general cognitive ability predicted proficiency on the L1 tests in both samples (models 1_R and 1_T), whereas the models using self-assessments (models 2_R, 3_R, 2_T and 3_T) failed to identify this effect. In addition, male students achieved lower scores on the L1 tests, whereas there were no or notably smaller gender differences in models based on self-assessments. Thus, even though boys and students with lower cognitive ability displayed lower levels of L1, they did not rate their L1 ability lower than did girls and students with higher cognitive ability. Another differential finding emerged for identification. This variable was only associated with self-assessed L1 proficiency, not with test scores. Students who strongly identified with their country of origin tended to perceive themselves as more proficient in L1 than students with lower levels of identification, yet these differences in orientation were not reflected in tested L1 skills.

Taken together, the discrepancies suggest that the use of self-assessments produces inaccurate findings. On the one hand, the models using self-assessments failed to detect effects that emerged when test scores were used as L1 indicators. On the other hand, analyses of self-assessments identified effects for some variables on L1 that were not supported in models using test scores. In addition, the analyses were not in line with the assumption that *performance SAs* yield more valid results than *ability SAs*. When the effects that emerged in the self-assessment models deviated from those in the test score models, this occurred, for the most part, for both types of self-assessments. In the few instances in which different results emerged only in one self-assessment model, this was not the case more often for *ability SAs* than for *performance SAs*.

Table 6.3: Predicting L1 skills (ordinary least squares regression models)

	<i>L1: Russian</i>			<i>L1: Turkish</i>		
	Model 1 _R : L1 test ^a	Model 2 _R : Ability SA L1 ^a	Model 3 _R : Perform. SA L1 ^a	Model 1 _T : L1 test ^a	Model 2 _T : Ability SA L1 ^a	Model 3 _T : Perform. SA L1 ^a
Born abroad	0.52 (0.09)	*** (0.09)	*** (0.35)	0.10 (0.12)	0.00 (0.11)	0.04 (0.13)
Male	-0.26 (0.08)	** (0.09)	-0.00 (0.09)	-0.22 (0.08)	-0.18 (0.08)	* (0.08)
General cognitive ability	0.05 (0.02)	** (0.02)	-0.02 (0.02)	0.09 (0.02)	0.00 (0.02)	0.03 (0.02)
Parent's secondary education ^b						
Middle-level	-0.09 (0.22)	-0.07 (0.19)	0.05 (0.21)	-0.08 (0.11)	0.03 (0.08)	0.08 (0.09)
Higher-level	0.07 (0.24)	-0.03 (0.19)	0.06 (0.23)	0.02 (0.14)	0.26 (0.12)	0.19 (0.12)
HISEI	0.08 (0.06)	0.09 (0.04)	0.11 (0.06)	-0.02 (0.07)	-0.05 (0.09)	-0.06 (0.08)
Cultural capital	0.00 (0.03)	0.05 (0.04)	0.03 (0.04)	0.05 (0.04)	0.05 (0.03)	0.04 (0.04)
Identification with CO	0.08 (0.05)	0.23 (0.06)	0.15 (0.06)	0.08 (0.06)	0.32 (0.07)	0.13 (0.06)
Co-ethnic neighborhood	0.01 (0.02)	0.01 (0.03)	0.04 (0.03)	0.02 (0.02)	0.03 (0.03)	0.07 (0.02)
German language use in family	-0.23 (0.07)	*** (0.07)	-0.21 (0.06)	-0.20 (0.08)	-0.36 (0.08)	-0.19 (0.08)
German language use with peers	0.03 (0.10)	0.11 (0.08)	-0.05 (0.09)	0.04 (0.08)	0.13 (0.05)	* (0.08)
German language use in media	-0.34 (0.13)	** (0.11)	-0.16 (0.12)	-0.29 (0.10)	-0.36 (0.07)	*** (0.11)
R ²	0.35	0.31	0.19	0.20	0.25	0.14

Note: Robust standard errors in parentheses; additionally controlled for school track; $N=502(\text{Russian})/662(\text{Turkish})$.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

^a Standardized variables. ^b Reference group: No degree/low-level of secondary school.

6.5.2.2 Determinants of L2 skills

The next set of regression models investigated the accuracy of the L2 self-assessments (see Table 6.4). These models aimed to predict L2 skills, as measured with the German language test (model 1), the L2 *ability SAs* (model 2) and the L2 *performance SAs* (model 3). Similar to the models of L1 maintenance, we included indicators for the motivation to acquire high levels of German proficiency (i.e., identification with the country of residence and the ratio of co-ethnics in the neighborhood), measures of exposure to the German language (i.e., generation status, the proportion of co-ethnics in the neighborhood and use of German) and an indicator of learning efficiency (i.e., cognitive ability). Because previous research has shown that students from different countries of origin differ considerably in their academic achievement, we further distinguished between different immigrant groups (i.e., Turkey, the former Soviet Union, Poland, and other). Students of Turkish origin born in Germany served as the reference group. We also controlled for students' social background and gender.

Again, the findings reveal significant differences between self-assessment measures and L2 tests. In model 1, all immigrant groups showed a level of L2 reading ability that was equal to or higher than that of the Turkish reference group. This result is in line with previous research that has repeatedly found particularly low L2 reading skills for students of Turkish descent (e.g., Edele and Stanat 2011; Stanat and Christensen 2006). Model 2, however, showed a very different pattern: compared to several of the other immigrant groups, Turkish students of second or higher generations tended to rate their L2 abilities higher; even though their actual levels of proficiency were lower. Model 3, moreover, did not identify any disparities among the immigrant groups. Another important difference among the models appeared for general cognitive ability. Model 1 demonstrated that, as expected, general cognitive ability was associated with students' L2 proficiency, whereas model 2 did not show this effect, and model 3 revealed only a weak association. Differential findings for self-assessments and L2 tests also emerged for other variables, such as the ratio of co-ethnics in the neighborhood.

In addition, the models differed remarkably with respect to R^2 . Although the various predictors accounted for a considerable proportion of the variance in L2 test scores (41%), they explained much less variance in the *ability SAs* and *performance SAs* (15% and 10%, respectively). Thus, the variables included in our models were notably more likely to account for variations in the L2 test scores than for variations in self-assessed L2 abilities. Analyses drawing on self-assessments were likely to underestimate the effect of important conditions relevant to L2 acquisition.

The results shown in Table 6.4 also provide limited support for the assumption that *performance SAs* are more valid than *ability SAs*. The differences between models using self-assessments and the models using test scores were sometimes less pronounced when *performance SAs* were used instead of *ability SAs* (e.g., with respect to differences among immigrant groups, general cognitive ability or the language of media use). However, consistent with the results of the test score models, the *ability SAs* indicated gender effects, whereas the *performance SAs* did not, suggesting that *performance SAs* do not generally yield more valid findings.

6 Why bother with testing?

In sum, different variables accounted for the tested L2 abilities and for self-assessed L2 skills, implying that the validity of self-assessments as L2 indicators is limited. The findings suggest that using self-assessments instead of test scores as outcomes carries the risks of overlooking existing effects and of identifying effects that are not present when using test-based measures.

Table 6.4: Predicting L2 skills (ordinary least squares regression models)

	Model 1: L2 test ^a	Model 2: Ability SA L2 ^a	Model 3: Performance SA L2 ^a	
Immigrant group^b				
Former SU, born abroad	0.18 (0.05)	*** (0.06)	-0.03 (0.07)	0.06 (0.06)
Former SU, born in Germany	0.18 (0.05)	*** (0.07)	-0.15 (0.07)	* (0.07)
Turkey, born abroad	0.04 (0.07)	-0.31 (0.11)	** (0.10)	-0.02 (0.10)
Poland, born abroad	0.24 (0.07)	*** (0.07)	0.12 (0.07)	0.11 (0.07)
Poland, born in Germany	0.11 (0.09)	-0.24 (0.12)	-0.10 (0.12)	-0.10 (0.12)
Other, born abroad	0.14 (0.04)	*** (0.05)	-0.06 (0.05)	0.06 (0.05)
Other, born in Germany	0.08 (0.05)	-0.35 (0.07)	*** (0.07)	-0.08 (0.07)
Male	-0.19 (0.03)	*** (0.03)	-0.16 (0.03)	*** (0.03)
General cognitive ability	0.10 (0.01)	*** (0.01)	-0.00 (0.01)	0.02 (0.01)
Parent's secondary education^c				
Middle-level	-0.02 (0.03)	0.02 (0.05)	0.05 (0.06)	0.05 (0.06)
Higher-level	-0.08 (0.05)	0.10 (0.07)	0.08 (0.07)	0.08 (0.07)
HISEI	0.00 (0.02)	0.03 (0.03)	0.04 (0.02)	* (0.02)
Cultural capital	0.08 (0.01)	*** (0.02)	0.06 (0.02)	*** (0.01)
Identification with CR	0.06 (0.02)	** (0.03)	0.13 (0.03)	*** (0.03)
Co-ethnic neighborhood	-0.05 (0.01)	*** (0.01)	0.02 (0.01)	-0.01 (0.01)
German language use in family	0.04 (0.02)	-0.05 (0.03)	-0.05 (0.03)	* (0.03)
German language use with peers	0.13 (0.02)	*** (0.03)	0.21 (0.03)	*** (0.05)
German language use in media	-0.03 (0.04)	0.25 (0.04)	*** (0.05)	0.04 (0.05)
R ²	0.41	0.15	0.10	

Note: Robust standard errors in parentheses; additionally controlled for school track; $N=4,078$. * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

^a Standardized variables. ^b Reference group: Turkey, born in Germany.

^c Reference group: No degree/low-level of secondary school.

6.5.2.3 L1 as determinant of L2 proficiency

We next address the accuracy of language self-assessments as independent variables. First, we explored the role of immigrant students' L1 abilities in their German reading skills (see Table 6.5), alternately including the three indicators of L1 as independent variables. In these models, we accounted for a number of relevant variables, including students' social background, their general cognitive ability, and their language use.

In line with the linguistic interdependence hypothesis, L1 test scores significantly correlated with reading skills in L2 for both the Russian and Turkish samples (models 1_R and 1_T). The self-assessment indicators, in contrast, did not relate to L2 reading skills (models 2_R, 3_R, 2_T and 3_T). Therefore, using self-assessment measures again yielded results that were different from those obtained with test scores. In this case, self-assessments failed to identify an effect, the existence of which continues to be a hotly debated issue in the field.

6.5.2.4 L2 as determinant of mathematical achievement

The final set of analyses assesses the role of L2 in immigrant students' mathematical skills (see Table 6.6). Again, we alternately used the three indicators of L2. In addition, we controlled for several variables, including students' social background, general cognitive ability, country of origin and generation status. Math test scores (models 1, 2 and 3) and math grades (models 4, 5 and 6) served as indicators of mathematical achievement.

As expected, the L2 test results showed a strong effect on mathematics test scores (model 1). This effect also appeared when using self-assessments as L2 indicators (model 2 and 3), although it was less pronounced. The models for math grades revealed even larger deviations between self-assessment measures and the L2 test. Although the test scores significantly related to students' math grades (see model 4; the negative sign indicates that students with higher scores on the L2 test were more likely to achieve good grades), this effect did not emerge for the two self-assessment indicators (models 5 and 6). Thus, when using self-assessments instead of test scores, we underestimated the association between L2 proficiency and mathematical achievement. This finding applies to both self-assessment indicators; *performance SAs* do not seem to be more accurate than *ability SAs*.

6.6 Discussion and conclusion

The main purpose of the current study was to determine whether self-assessments of language proficiency, as measured with the general indicator of language proficiency typically used in large-scale assessments (*ability SAs*) and a more specific indicator referring to a concrete task participants just completed (*performance SAs*), provide accurate indicators of language skills. The *ability SAs* correlated only moderately with tested linguistic ability. The *performance SAs* showed higher correlations with the test

Table 6.5: Predicting L2 reading test scores based on L1 skills (ordinary least squares regression models)

	Former SU			Turkey		
	Model 1 _R	Model 2 _R	Model 3 _R	Model 1 _T	Model 2 _T	Model 3 _T
L1 test ^a	0.22 (0.05)	***		0.20 (0.04)	***	
Ability SA L1 ^a		-0.02 (0.05)			0.02 (0.04)	
Performance SA L1 ^a			0.03 (0.05)			0.01 (0.04)
Born abroad	-0.03 (0.10)	0.09 (0.10)	0.08 (0.09)	-0.10 (0.10)	-0.08 (0.10)	-0.08 (0.10)
Male	-0.11 (0.09)	-0.16 (0.09)	-0.16 (0.09)	-0.22 (0.07)	*** (0.26)	*** (0.26)
General cognitive ability	0.12 (0.02)	*** (0.13)	*** (0.14)	*** (0.11)	*** (0.13)	*** (0.13)
Parent's secondary education ^b						
Middle-level	0.04 (0.17)	0.02 (0.19)	0.02 (0.18)	-0.09 (0.09)	-0.10 (0.10)	-0.11 (0.10)
Higher-level	-0.30 (0.20)	-0.28 (0.22)	-0.28 (0.21)	-0.22 (0.12)	-0.22 (0.12)	-0.21 (0.12)
HISEI	-0.05 (0.05)	-0.03 (0.05)	-0.03 (0.05)	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)
Cultural capital	0.06 (0.03)	0.06 (0.03)	0.05 (0.03)	0.04 (0.03)	0.05 (0.03)	0.05 (0.03)
Identification with CO	0.04 (0.06)	0.05 (0.07)	0.05 (0.07)	0.12 (0.05)	0.11 (0.05)	0.11 (0.05)
Co-ethnic neighborhood	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.08 (0.02)	-0.08 (0.02)	-0.08 (0.02)
German language use in family	0.17 (0.07)	* (0.11)	0.12 (0.07)	0.18 (0.07)	** (0.14)	0.14 (0.07)
German language use with peers	0.07 (0.08)	0.08 (0.08)	0.08 (0.08)	0.13 (0.07)	0.13 (0.07)	0.14 (0.07)
German language use in media	0.11 (0.09)	0.02 (0.09)	0.03 (0.08)	0.03 (0.07)	-0.02 (0.07)	-0.03 (0.07)
R ²	0.43	0.40	0.40	0.39	0.36	0.36

Note: Robust standard errors in parentheses; additionally controlled for school track; N=502(Russian)/662(Turkish).

* p<0.05, ** p<0.01, *** p<0.001.

^a Standardized variables. ^b Reference group: No degree/low-level of secondary school.

Table 6.6: Predicting mathematics achievement based on L2 skills (ordinary least square models)

	Performance in mathematics test			Grades in mathematics		
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
L2 test ^a	0.29 (0.02)	***		-0.05 (0.02)	*	
Ability SA L2 ^a		0.05 (0.01)	**		0.02 (0.02)	
Performance SA L2 ^a			0.03 (0.01)	*		-0.02 (0.02)
Immigrant group ^b						
Former SU, born abroad	0.16 (0.06)	**	0.24 (0.06)	***	-0.09 (0.07)	-0.09 (0.07)
Former SU, born in Ger- many	0.09 (0.05)		0.15 (0.05)	**	-0.04 (0.07)	-0.04 (0.07)
Turkey, born abroad	0.03 (0.08)		0.03 (0.08)		-0.11 (0.11)	-0.11 (0.11)
Poland, born abroad	0.15 (0.08)	*	0.25 (0.08)	**	0.04 (0.09)	0.02 (0.09)
Poland, born in Ger- many	0.33 (0.09)	***	0.40 (0.09)	***	0.09 (0.09)	0.09 (0.10)
Other, born abroad	0.12 (0.04)	**	0.19 (0.04)	***	0.12 (0.05)	0.11 (0.05)
Other, born in Ger- many	0.02 (0.05)		0.07 (0.05)		0.04 (0.06)	0.03 (0.06)
Male	0.45 (0.03)	***	0.40 (0.03)	***	-0.33 (0.04)	-0.32 (0.04)
General cognitive ability	0.08 (0.01)	***	0.11 (0.01)	***	-0.07 (0.01)	-0.08 (0.01)
Parent's secondary edu- cation ^c						
Middle-level	-0.00 (0.04)		-0.01 (0.04)		0.01 (0.06)	0.01 (0.06)
Higher-level	0.03 (0.05)		0.01 (0.05)		-0.19 (0.07)	-0.18 (0.07)
HISEI	0.00 (0.02)		0.01 (0.02)		-0.04 (0.03)	-0.04 (0.03)
Cultural capital	0.01 (0.01)	**	0.04 (0.01)	**	-0.02 (0.01)	-0.02 (0.01)
R ²	0.47		0.43		0.10	0.10

Note: Robust standard errors in parentheses; additionally controlled for school track; $N=4,078$.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

^a Standardized variables. ^b Reference group: Turkey, born in Germany. ^c Reference group: No degree/low-level of secondary school.

scores comparable to those found in other studies using specific self-assessments (see Alderson 2005; Brantmeier et al. 2012). Because the overlap was far from perfect, this higher correlation did, however, not necessarily ensure that these self-assessments provide unbiased findings when used to address substantive research questions.

Our study therefore additionally explored the validity issue with multivariate analyses pertaining to substantive research questions. We conducted a variety of analyses that either predicted language skills based on a set of factors assumed to be relevant for language acquisition (language as dependent variable) or examined the effects of language on other outcomes (language as independent variable). In all analyses, both types of self-assessments yielded different effects than test scores. Our findings thus indicate that the types of self-assessments used in our study do not provide accurate estimates of language skills.

In contrast to our expectations, the results for the specific self-assessment indicators (i.e., the *performance SAs*) were not more similar to the test score models than the results for the *ability SAs*. Thus, employing instruments in which participants estimate their performance on a concrete linguistic task did not improve validity in the multivariate analyses. This finding seems to contradict previous research which suggests that a higher degree of specificity increases the accuracy of self-assessments (Brantmeier et al. 2012; Ross 1998). Yet, the *performance SAs* correlated substantially with the test scores in our study as well, suggesting that they were reasonably accurate. However, the multivariate results indicate that even types of self-assessments that relate more closely to test scores do not ensure unbiased findings in substantive analyses and that they should thus be used with caution.

The findings from the multivariate analyses further suggest that self-assessments of both types are systematically biased in some groups. For instance, even though female students scored higher on the language tests, they estimated their L1 skills on a similar level as their male peers. This result indicates that, in relative terms, boys tend to overestimate their abilities. We observed similar biases in comparing different immigrant groups. Whereas the students of Turkish origin who were born in Germany exhibited lower skills on the L2 tests than did most other groups, they rated their L2 skills on a similar or even higher level than students from other immigrant groups did. Students of Turkish origin thus tend to overestimate their language abilities. Similarly, students with higher cognitive ability scored higher on the L1 and L2 tests, yet this association did not emerge for the self-assessment indicators. The self-assessments thus failed to capture the linguistic advantages of students with higher cognitive ability. These examples illustrate that self-assessments measures may lead to systematically biased results because they are not comparable across groups. Previous findings of group differences in language proficiency that rely on self-assessments may thus be erroneous.

Moreover, we found deviations in both directions. On the one hand, some models using self-assessments exhibited effects that were not revealed by analyses using language tests. This type of deviation, however, was rare. Interestingly, one of these cases involved the relationship between identification with the country of origin and L1 proficiency, which was stronger for the self-assessments. Identification with the country

of origin—like self-estimated L1 proficiency—reflects a person’s self-perception and orientation toward his or her own group (e.g., Phinney and Ong 2007). It seems likely that the observed covariation between self-assessed L1 skills and identification with the country of origin is based on this self-perceived affiliation rather than on an actual relationship between language competence and cultural identity. On the other hand, regression models using self-assessments frequently failed to reveal group differences, such as those between immigrant groups, that were identified in models using test results. In addition, tested language abilities clearly related to math achievement and reading performance in L2, whereas the self-assessment measures either showed much less pronounced or no effects. Taken together, these findings indicate that the use of self-assessments increases the risk of overlooking relevant effects.

Our findings are in line with previous research revealing that self-assessment measures often underestimate existing effects. In Finnie and Meng’s (2005) study, for example, self-assessed language abilities failed to detect the effect of language on employment status and were less reliable in predicting income differences. Because previous investigations, particularly analyses based on data from large-scale studies, have typically relied on self-assessed language proficiency, it seems likely that the current state of research systematically underestimates the role of language abilities in the integration of immigrants into the country of residence.

At the same time, it is possible that under different conditions self-assessments may yield more accurate findings than in our study. In this context studies that found particularly high correlations between tests and self-assessments are of special interest (e.g., LeBlanc and Painchaud 1985; Wilson 1999). These studies examined foreign-language learners instead of immigrants. Self-estimates for foreign-language learners are likely to be more accurate due to the feedback they receive on their language skills and their less ambiguous frame of reference (see Section 6.2). In addition, these studies used specific self-assessments referring to real-life situations that participants most likely had encountered repeatedly in their daily lives. Our *performance SAs*, in contrast, required students to estimate their success on a psychometric test (i.e., a study-task), a type of task with which students likely had little experience and which, in its concrete form, they encountered for the first time. Because individual behavior varies largely across situations and time, it may be easier to adequately assess one’s own performance when it is possible to aggregate information on repeatedly encountered tasks (e.g., Ajzen 1987; Epstein 1983). Previous research further indicates that to arrive at accurate self-assessments it is important to tailor items exactly to participants’ experience, for instance by matching them to the curriculum of the students’ language class (e.g., Brantmeier et al. 2012; Ross 1998). Note, however, that it would rarely be possible to design instruments to match participants’ experience this closely in the case of immigrant populations. Immigrants typically acquire their language skills not only (if at all) in language classes, but also in informal contexts and consequently have very heterogeneous prior experiences and language skills.

The types of self-assessments that showed particularly high correlations with test scores (e.g., LeBlanc and Painchaud 1985; Wilson 1999) were also more extensive than the self-assessments used in our study; they either used long item batteries or extensive

verbal descriptions. Such instruments would not be noticeably more efficient than tests. However, it is precisely because of their efficiency that many studies, particularly large-scale studies, have thus far preferred self-assessments over tests. Because there is no evidence that self-assessments are superior to tests, at least when the goal is to assess participants' actual language ability rather than their self-perceptions, the use of elaborate self-assessments would thus have no significant advantage over the use of tests.

Our study is limited in several respects. First, our sample is restricted to adolescents, and it is unclear whether the findings generalize to other age groups. Previous research has indicated that children's conceptualization of their abilities is not yet clearly differentiated (Stipek and Mac Iver 1989) and that their ability self-estimates are often inaccurate (e.g., Blatchford 1992; Butler and Lee 2006; Demetriou and Kazi 2006). Adolescents, in contrast, have relatively accurate estimations of their own abilities (Demetriou and Kazi 2006). Self-assessments of younger participants are thus likely to be even more biased than those in our sample.

In addition, our study tested language skills in L1 and L2 in only one domain each (i.e., listening in L1 and reading in L2). Depending on the research topic, other linguistic domains may be of interest. For instance, students' listening skills in L2 may be central to the acquisition of academic competencies because they are necessary for following classroom instructions. However, both listening comprehension and reading comprehension are useful indicators of participants' language skills. Unlike components skills, such as vocabulary knowledge, they involve a multitude of linguistic processes and therefore ensure a broad assessment of the language ability construct.

Another possible concern with our study is that the observed correlation of the language tests with other tested competencies may (partly) rely on a third factor rather than on an actual relationship between the competencies. For instance, general test-taking ability may have accounted for the relationship between the L2 test and the mathematics test. We cannot entirely rule out this possibility, yet, we also found self-assessments to be less clearly related to other variables that were not test-based (i.e., math grades or gender). It therefore seems unlikely that our findings were affected by unobserved heterogeneity in the test scores rather than indicating higher accuracy of language tests compared to language self-assessments.

Future studies should replicate our findings for other types of self-assessments and language tests, for other countries and immigrant groups, for different age groups and with regard to additional outcomes. Because competence testing is costly and time-consuming, not all empirical investigations will be able to implement comprehensive language tests. It would be a promising venue for future research to develop valid language tests that are efficient enough to be used in studies with limited testing time, such as large-scale studies and censuses. It would also be important to develop self-assessment measures that are both more valid than the ones commonly used in large-scale assessments and efficient. As described above, previous research suggests that elaborate, specific self-estimates that are tailored to foreign language learners' experience proved to be quite accurate. Although this proposal seems challenging, an attempt should be made to condense such measures and adapt them to the situation

of immigrant populations that have more diverse experiences and language levels (see above).

In addition, it would be helpful to identify factors that enhance the validity of self-assessments. This would permit researchers to specify the conditions under which the use of self-assessed language proficiency is reasonable, meaningful and adequate, and the conditions under which self-assessments are particularly prone to bias. As outlined above, it seems to be more difficult to obtain accurate self-assessments in heterogeneous populations such as immigrants compared to foreign language learners. Previous findings further indicate that the accuracy of self-assessments varies, for example, with respondents' level of competence (e.g., Brantmeier and Vanderplank 2008; Kruger and Dunning 1999) and with the frame of reference to which respondents refer (Marsh and Hau 2004; Shajek et al. 2006). Our findings suggest that additional conditions associated with gender, immigrants' country of origin or participants' cognitive ability may also moderate the accuracy of self-assessments. However, research has only begun to explore such distorting factors and the underlying processes. It is therefore important for future research to gain further insight into the determinants that affect the accuracy of self-assessments. This type of knowledge may also help to develop statistical techniques that can reduce the bias in data based on self-assessments (see Dustmann and Van Soest 2001).

It is important to note that our findings do not imply that the use of language tests is always superior to the use of self-assessments. Our study examines issues pertaining to immigrants' *actual* language proficiency. For other research questions, participants' self-perceptions may be of central interest. For instance, individual decisions, such as whether one's linguistic skills are sufficient to apply for a job or whether further investments in language learning are necessary, depend on subjective perceptions. Furthermore, self-assessments seem to be valuable tools for advancing self-regulated learning. They may raise students' awareness of learning goals, help them to monitor their progress, increase their learning attitudes and motivation, and consequently improve subsequent language learning and classroom performance (Butler and Lee 2010; Chen 2008; Oscarson 1989). In addition, self-assessments may be useful for the placement of advanced language learners, particularly when used in combination with language tests (Brantmeier and Vanderplank 2008; Brantmeier et al. 2012; LeBlanc and Painchaud 1985).

The findings of this study indicate that test-based indicators and self-assessments of linguistic ability cannot be used interchangeably. A clear distinction between the two concepts at the empirical and conceptual levels is indispensable to obtain valid information on immigrants' language skills.

6.7 References

Ackerman, P. L., Beier, M. E., and Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, 33(4):587–605.

- Ajzen, I. (1987). Attitudes, traits, and actions: Dispositional prediction of behavior in personality and social psychology. *Advances in Experimental Social Psychology*, 20:1–63.
- Alavi, S. M. and Akbarian, I. (2008). Validating a self-assessment questionnaire on vocabulary knowledge. *TELL*, 2:125–154.
- Alba, R., Logan, J., Lutz, A., and Stults, B. (2002). Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants. *Demography*, 39(3):467–484.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Allison, P. D. (2001). *Missing data*. Thousand Oaks: Sage Publications.
- Anderson, J. R. (1995). *Cognitive psychology and its implications*. New York: Freeman.
- Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., and Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, 62:49–78.
- Aßmann, C., Steinhauer, H. W., and Zinn, S. (2012). *Starting cohorts 3 and 4: 5th/9th grade (SC3/SC4). Scientific use file, version 1.0.0. Data manual [supplement]: Weighting*. Bamberg: University of Bamberg, National Educational Panel Study.
- Bachman, L. F. and Palmer, A. S. (1981). The construct validation of the FSI oral interview. *Language Learning*, 31(1):67–86.
- Bachman, L. F. and Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16(4):449–465.
- Bankston III, C. L. and Zhou, M. (1995). Effects of minority-language literacy on the academic achievement of Vietnamese youths in New Orleans. *Sociology of Education*, 68(1):1–17.
- Barrett, A. N., Barile, J. P., Malm, E. K., and Weaver, S. R. (2012). English proficiency and peer interethnic relations as predictors of math achievement among Latino and Asian immigrant students. *Journal of Adolescence*, 35(6):1619–1628.
- Beal, C. R., Adams, N. M., and Cohen, P. R. (2010). Reading proficiency and mathematics problem solving by high school English language learners. *Urban Education*, 45(1):58–74.
- Berry, J. W., Phinney, J. S., Sam, D. L., and Vedder, P. E. (2006). *Immigrant youth in cultural transition: Acculturation, identity, and adaptation across national contexts*. Mahwah: Lawrence Erlbaum Associates.

- Blanche, P. and Merino, B. J. (1989). Self-assessment of foreign-language skills: Implications for teachers and researchers. *Language Learning*, 39(3):313–338.
- Blatchford, P. (1992). Academic self assessment at 7 and 11 years: Its accuracy and association with ethnic group and sex. *British Journal of Educational Psychology*, 62(1):35–44.
- Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J. (2011). *Education as a lifelong process: The German National Educational Panel Study (NEPS)*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bong, M. and Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1):1–40.
- Brantmeier, C. (2005). Nonlinguistic variables in advanced second language reading: Learners' self-assessment and enjoyment. *Foreign Language Annals*, 38(4):494–504.
- Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. *System*, 34(1):15–35.
- Brantmeier, C. and Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 readers. *System*, 36(3):456–477.
- Brantmeier, C., Vanderplank, R., and Strube, M. (2012). What about me?: Individual self-assessment by skill and level of language instruction. *System*, 40(1):144–160.
- Buck, G. (1992). Listening comprehension: Construct validity and trait characteristics. *Language Learning*, 42(3):313–357.
- Butler, Y. G. and Lee, J. (2006). On-task versus off-task self-assessments among Korean elementary school students studying English. *The Modern Language Journal*, 90(4):506–518.
- Butler, Y. G. and Lee, J. (2010). The effects of self-assessment among young learners of English. *Language Testing*, 27(1):5–31.
- Carliner, G. (2000). The language ability of US immigrants: Assimilation and cohort effects. *International Migration Review*, 34(1):158–182.
- Carter, T. J. and Dunning, D. (2008). Faulty self-assessment: Why evaluating one's own competence is an intrinsically difficult task. *Social and Personality Psychology Compass*, 2(1):346–360.
- Charette, M. and Meng, R. (1994). Explaining language proficiency: Objective versus self-assessed measures of literacy. *Economics Letters*, 44(3):313–321.
- Chen, Y.-M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, 12(2):235–262.

- Chiswick, B. R., Lee, Y. L., and Miller, P. W. (2004). Immigrants' language skills: The Australian experience in a longitudinal survey. *International Migration Review*, 38(2):611–654.
- Chiswick, B. R. and Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economics*, 13(2):246–288.
- Chiswick, B. R. and Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, 38(3):391–409.
- Clark, J. L. and Swinton, S. S. (1979). *An exploration of speaking proficiency measures in the TOEFL context (ETS Research Report Series 4)*. Princeton: Educational Testing Service.
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2):222–251.
- Cummins, J. (2000). Language, power, and pedagogy: Bilingual children in the crossfire. In Barker, C. and Hornberger, N., editors, *Bilingual education and bilingualism*. Clevedon: Multilingual Matters Ltd, UTP.
- Demetriou, A. and Kazi, S. (2006). Self-awareness in g (with processing efficiency and reasoning). *Intelligence*, 34(3):297–317.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., et al. (2007). School readiness and later achievement. *Developmental Psychology*, 43(6):1428–1446.
- Dunn, L. M. and Dunn, D. M. (2004). *Peabody Picture Vocabulary Test (PPVT) (German version)*. Göttingen, Hogrefe.
- Dunning, D., Heath, C., and Suls, J. M. (2004). Flawed self-assessment: Implications for health, education, and the workplace. *Psychological Science in the Public Interest*, 5(3):69–106.
- Dustmann, C. and Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal*, 113(489):695–717.
- Dustmann, C. and Van Soest, A. (2001). Language fluency and earnings: Estimation with misclassified language indicators. *The Review of Economics and Statistics*, 83(4):663–674.
- Edele, A., Schotte, K., Hecht, M., and Stanat, P. (2012). *Listening comprehension tests of immigrant students' first languages (L1) Russian and Turkish in grade 9: Scaling procedure and results (NEPS Working Paper 13)*. Bamberg: Universität Bamberg, Nationales Bildungspanel.

- Edele, A., Schotte, K., and Stanat, P. (2016). Assessment of immigrant students' listening comprehension in their first languages (L1) Russian and Turkish in grade 9: Test construction and validation. In Blossfeld, H.-P., von Maurice, J., Bayer, M., and Skopek, J., editors, *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*, pages 441–463. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Edele, A. and Stanat, P. (2011). PISA's potential for analyses of immigrant students' educational success: The German case. In Pereya, M. A., Kotthoff, H.-G., and Cowen, R., editors, *PISA under examination: Changing knowledge, changing tests, and changing schools*, pages 185–206. Rotterdam: Sense Publishers.
- Embretson, S. E. and Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah: Lawrence Erlbaum Associates.
- Epstein, S. (1983). Aggregation and beyond: Some basic issues on the prediction of behavior. *Journal of Personality*, 51(3):360–392.
- Esser, H. (1985). Zur Validität subjektiver Sprachkompetenzmessung. In Sievering, U. O., editor, *Arbeitsmigrantenforschung in der Bundesrepublik Deutschland: Methodenprobleme der Datenerhebung*, pages 192–226. Frankfurt am Main: Haag + Herchen.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt am Main: Campus.
- Falchikov, N. and Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4):395–430.
- Ferguson, N. (1978). Notes and discussion: Self-assessment of listening comprehension. *International Review of Applied Linguistics in Language Teaching*, 16(2):149–162.
- Finnie, R. and Meng, R. (2005). Literacy and labour market outcomes: Self-assessment versus test score measures. *Applied Economics*, 37(17):1935–1951.
- Fishman, J. A. and Terry, C. (1969). The validity of census data on bilingualism in a Puerto Rican neighborhood. *American Sociological Review*, 34(5):636–650.
- Flowerdew, J. and Miller, L. (2005). *Second language listening: Theory and practice*. New York: Cambridge University Press.
- Freund, P. A. and Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138(2):296–321.
- Ganzeboom, H. (2010). *A new Socio-Economic Index (ISEI) of occupational status for the International Standard Classification of Occupation 2008 (ISCO-08) constructed with data from the ISSP 2002–2007, with an analysis of quality of occupational*

6 Why bother with testing?

measurement in ISSP. Lisbon: Annual Conference of International Social Survey Programme.

- Gehrer, K., Zimmermann, S., Artelt, C., and Weinert, S. (2012). *The assessment of reading competence (including sample items for grade 5 and 9)*. Scientific use file 2012, version 1.0.0. Bamberg: University of Bamberg, National Educational Panel Study.
- Gehrer, K., Zimmermann, S., Artelt, C., and Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online*, 5(2):50–79.
- Glenberg, A. M., Sanocki, T., Epstein, W., and Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116(2):119–136.
- Glover, J. A. (1989). Reading ability and the calibrator of comprehension. *Educational Research Quarterly*, 13(3):7–11.
- Grimm, K. J. (2008). Longitudinal associations between reading and mathematics achievement. *Developmental Neuropsychology*, 33(3):410–426.
- Hansford, B. C. and Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research*, 52(1):123–142.
- Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. Berkeley: University of California Press.
- Hwang, S.-S., Xi, J., and Cao, Y. (2010). The conditional relationship between English language proficiency and earnings among US immigrants. *Ethnic and Racial Studies*, 33(9):1620–1647.
- Janssen-van Dieten, A.-M. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6(1):30–46.
- Jiménez, T. R. (2011). *Immigrants in the United States: How well are they integrating into society*. Washington D.C.: Migration Policy Institute.
- Kaida, L. (2013). Do host country education and language training help recent immigrants exit poverty? *Social Science Research*, 42(3):726–741.
- Kleemans, T., Segers, E., and Verhoeven, L. (2011). Cognitive and linguistic precursors to numeracy in kindergarten: Evidence from first and second language learners. *Learning and Individual Differences*, 21(5):555–561.
- Kristen, C., Edele, A., Kalter, F., Kogen, I., Schulz, B., Stanat, P., and Will, G. (2011). The education of migrants and their children across the life course. In Blossfeld, H.-P., Roßbach, H.-G., and von Maurice, J., editors, *Education as a lifelong process*:

- The German National Educational Panel Study (NEPS)*, pages 121–138. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6):1121–1134.
- LeBlanc, R. and Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19(4):673–687.
- LeFevre, J.-A., Fast, L., Skwarchuk, S.-L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., and Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, 81(6):1753–1767.
- Lin, L.-M., Moore, D., and Zabrocky, K. M. (2001). An assessment of student's calibration of comprehension and calibration of performance using multiple measures. *Reading Psychology*, 22(2):111–128.
- Long, D. R. (1990). What you don't know can't help you: An exploratory study of background knowledge and second language listening comprehension. *Studies in Second Language Acquisition*, 12(1):65–80.
- Lutz, A. and Crist, S. (2009). Why do bilingual boys get better grades in English-only America? The impacts of gender, language and family interaction on academic achievement of Latino/a children of immigrants. *Ethnic and Racial Studies*, 32(2):346–368.
- Mabe, P. A. and West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3):280–296.
- Maki, R. H. and Berry, S. L. (1984). Metacomprehension of text material. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(4):663–679.
- Marsh, H. W. (1990). A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educational Psychology Review*, 2(2):77–172.
- Marsh, H. W. (1992). Content specificity of relations between academic achievement and academic self-concept. *Journal of Educational Psychology*, 84(1):35–42.
- Marsh, H. W. and Hau, K.-T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, 96(1):56–67.
- McCormick, I. A., Walkey, F. H., and Green, D. E. (1986). Comparative perceptions of driver ability—A confirmation and expansion. *Accident Analysis and Prevention*, 18(3):205–208.

6 Why bother with testing?

- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4):525–543.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9):741–749.
- Milleret, M., Stansfield, C. W., and Kenyon, D. M. (1991). The validity of the Portuguese speaking test for use in a summer study abroad program. *Hispania*, 74(3):778–787.
- Mouw, T. and Xie, Y. (1999). Bilingualism and the academic achievement of first- and second-generation Asian Americans: Accommodation with or without assimilation? *American Sociological Review*, 64(2):232–252.
- Muñoz-Sandoval, A. F., Cummins, J., Alvarado, C. G., and Ruef, M. L. (1998). *Bilingual verbal ability tests*. Chicago: Riverside Publishing Company.
- Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., and Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. *Journal for Educational Research Online*, 5(2):80–109.
- Nielsen, F. and Lerner, S. J. (1986). Language skills and school achievement of bilingual Hispanics. *Social Science Research*, 15(3):209–240.
- OECD (2009). *PISA 2009: Assessment framework—Key competencies in reading, mathematics and science*. Paris: OECD.
- Oscarson, M. (1989). Self-assessment of language proficiency: Rationale and applications. *Language Testing*, 6(1):1–13.
- Peirce, B. N., Swain, M., and Hart, D. (1993). Self-assessment, French immersion, and locus of control. *Applied Linguistics*, 14(1):25–42.
- Pendakur, K. and Pendakur, R. (2002). Language as both human capital and ethnicity. *International Migration Review*, 36(1):147–177.
- Peyer, E. and Kaiser, I. (2010). Subjektive und objektive Messgrößen zur Erforschung des fremdsprachlichen Leseverstehens: Selbsteinschätzung und Testergebnisse im Vergleich. *Zeitschrift für Fremdsprachenforschung*, 21(1):79–105.
- Phinney, J. S. and Ong, A. D. (2007). Conceptualization and measurement of ethnic identity: Current status and future directions. *Journal of Counseling Psychology*, 54(3):271–281.
- Portes, A. and Rumbaut, R. G. (2006). *Immigrant America: A portrait*. Los Angeles: University of California Press.

- Rauch, D. P., Jurecka, A., and Hesse, H.-G. (2010). Für den Drittspracherwerb zählt auch die Lesekompetenz in der Herkunftssprache: Untersuchung der Türkisch-, Deutsch- und Englisch-Lesekompetenz bei Deutsch-Türkisch bilingualen Schülern. In Allemann-Ghionda, C., Stanat, P., Göbel, K., and Röhner, C., editors, *Migration, Sprache, Identität und Bildungserfolg*, pages 78–100. Weinheim: Beltz.
- Raven, J. (2000). The raven’s progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1):1–48.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1):1–20.
- Rost, M. (2011). *Teaching and researching listening*. Harlow: Pearson Education.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schroeders, U. and Wilhelm, O. (2011). Equivalence of reading and listening comprehension across test media. *Educational and Psychological Measurement*, 71(5):849–869.
- Shajek, A., Lüdtke, O., Stanat, P., et al. (2006). Akademische Selbstkonzepte bei Jugendlichen mit Migrationshintergrund. *Unterrichtswissenschaft*, 34(2):125–145.
- Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3):407–441.
- Snow, C. E., Burns, M. S., and Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington D.C.: National Academy Press.
- Stanat, P. and Christensen, G. (2006). *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003*. Paris: OECD.
- Stipek, D. and Iver, D. M. (1989). Developmental change in children’s assessment of intellectual competence. *Child Development*, 60(3):521–538.
- Van Tubergen, F. and Kalmijn, M. (2009). A dynamic approach to the determinants of immigrants’ language proficiency: The United States, 1980–2000. *International Migration Review*, 43(3):519–543.
- Vukovic, R. K. and Lesaux, N. K. (2013). The language of mathematics: Investigating the ways language counts for children’s mathematical development. *Journal of Experimental Child Psychology*, 115(2):227–244.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica: Journal of the Econometric Society*, 50(1):1–25.
- Wilson, K. M. (1989). *Enhancing the interpretation of a norm-referenced second language test through criterion referncing: A research assesment of experience in the TOEIC testing context (ETS Research Report Series 89-39)*. Princeton: Educational Testing Service.

6 Why bother with testing?

Wilson, K. M. (1999). *Validity of global self-ratings of ESL speaking proficiency based on an FSI/ILR-referenced scale (ETS Research Report Series 99-13)*. Princeton: Statistics and Research Division.

6.8 Appendix

Table A6.1: Studies comparing self-assessments and test scores/teacher estimates of linguistic skills

Study	Sample	Self-assessment indicators	Test scores/teacher estimates	Major findings
Alavi and Akbarian (2008)	EFL, Iranian university students ($N=293$)	SA scale: vocabulary knowledge (42 items) <i>Type: specific (not identifiable)</i>	Vocabulary Levels Test (VLT): matching words to definitions (5 levels, 30 items each)	Correlation between SA and test: .50
Alderson (2005)	EFL, European students from 11 different language backgrounds ($N=2060$)	SA scales: listening, reading and writing (18 items each) <i>Type: specific (real-life task)</i>	Listening test (56 items) Reading test (50 items) Writing test (54 items)	Correlations between SA and respective test: .49 (listening) .49 (reading) .55 (writing)
Bachman and Palmer (1981)	EFL, Mandarin Chinese speaking university students and their spouses ($N=75$)	SA scales: speaking and reading (number of items not identifiable) <i>Type: specific (real-life task)</i>	Teacher ratings: (1) Speaking (15-30 minutes structured conversation and recordings of answers) (2) Reading (questions and translation of the text passage)	Correlations between SA and teacher ratings: .46 < r < .56 (speaking) .54 < r < .68 (reading)
Bachman and Palmer (1982)	ESL, university students from 18 different language backgrounds ($N=110$)	SA scales: grammar (6 items), pragmatic skills (9 items) and socio-linguistic competence (9 items) <i>Type: specific (real-life task)</i>	Grammar, pragmatic skills and socio-linguistic competence assessed (119 items): (1) Oral interview (2) Writing tasks (3) Multiple choice test	Correlation between SA and respective test scale: .43 < r < .71
Brantmeier (2005)	FLL, English-speaking university students in advanced level Spanish course ($N=88$)	SA scale: reading (1 item) <i>Type: general</i>	Reading comprehension: (1) Written recall (2) Multiple choice test (10 items)	Significant association between SA and written recall (no correlation coefficients reported) No significant association between SA and multiple choice test
Brantmeier (2006)	FLL, English-speaking college students in advanced level Spanish courses ($N=34$)	SA scale (pre-test): reading (2 items) <i>Type: general</i> SA scale (post-test): performance in just completed tests (2 items) <i>Type: specific (study task)</i>	Reading comprehension: (1) Multiple choice test (8 texts, number of items not identifiable) (2) Written recall, sentence completion and multiple-choice test (number of items not identifiable)	No significant association between SA measures and respective test scale (no correlation coefficients reported)

Table A6.1: Studies comparing self-assessments and test scores/teacher estimates of linguistic skills (*Continued*)

Study	Sample	Self-assessment indicators	Test scores/teacher estimates	Major findings
Brantmeier and Vanderplank (2008)	FLL, English-speaking college students in advanced level Spanish courses ($N=359$)	SA scale (pre-test): describing reading levels at 6 stages (1 item) and reading comprehension (15 items) <i>Type: specific (real-life task)</i> SA scale (post-test): performance in just completed tests (4 items) <i>Type: specific (study task)</i>	Reading comprehension: (1) Multiple choice test (8 texts, number of items not identifiable) (2) Written recall, sentence completion and multiple-choice test (2 texts, number of items not identifiable)	Significant association between pre-test SAs (descriptive/reading) and test 1 ($p<.05/p<.05$) Correlation between post-test SA and test 2 subscales: $.16<r<.49$
Brantmeier, Vanderblank, and Strube (2012)	FLL, English-speaking college students in advanced level Spanish courses ($N=150$)	SA scales: listening, speaking, reading and writing (number of items not identifiable) <i>Type: specific (real-life task)</i>	Listening test: multiple choice (30 items) Vocabulary test: multiple choice (20 items) Reading test: multiple choice (30 items) Grammar test: multiple choice (20 items)	Correlations between SA and test subscales: $.21<r<.33$ Correlation between SA and tests (overall): $.34$
Buck (1992)	EFL, Japanese university students ($N=220$)	SA scales: listening and reading (8 items each) <i>Type: specific (not identifiable)</i>	Listening test: picture recognition (25 items) and cloze-test (24 items) Reading test: picture recognition (25 items) and gap-filling (24 items)	Correlations between SA and respective test scales: $.46<r<.58$
Butler and Lee (2006)	EFL, Korean students in 4th grade or 6th grade ($N=151$)	SA scales (pre-test): speaking and listening (8 items) <i>Type: specific (real-life task)</i> SA scales (post-test): performance in just-completed tasks (8 items) <i>Type: specific (study task)</i>	Teacher estimates: speaking and comprehension (7 items) Listening test (20/25 items)	Correlations between SA scales and teacher estimates (4th grade/6th grade): $.54/.55$ (pre-test) $.59/.61$ (post-test) Correlations between SA scales and listening test (4th grade/6th grade): $.33/.60$ (pre-test) $.40/.75$ (post-test)
Charette and Meng (1994)	Adult immigrants in Canada tested in French/English ($N=656$)	SA scale: reading (1 item) <i>Type: general</i>	Reading test (34 items)	Correlation between SA and test: $.58$
Clark and Swinton (1979)	ESL, undergraduate and graduate students from 26 different language backgrounds ($N=86$)	SA scale: speaking (1 item) <i>Type: specific (real-life task)</i>	Speaking test: face-to-face structured interview rated by 5 trained raters (long/short version)	Correlation between SA and test (long/short): $.48/.43$

Table A6.1: Studies comparing self-assessments and test scores/teacher estimates of linguistic skills (Continued)

Study	Sample	Self-assessment indicators	Test scores/teacher estimates	Major findings
Ferguson (1978)	EFL, first-year college students ($N=89$) and students at an interpreter's school ($N=90$)	SA scale 1: listening and speaking (35 items) <i>Type: specific (real-life-task)</i> SA scale 2: match of one's own speaking ability to recordings at 18 levels (1 item) <i>Type: specific (real-life task)</i>	Listening comprehension test (141/336 items) Speaking test (number of items not identifiable)	Correlation between SA 1/SA 2 and speaking test: .39/.19 Correlation between SA 2 and listening test: .19
Finnie and Meng (2005)	French-/English-speaking adults in Canada tested in French/English ($N=6995$); English first language $N=5062$; foreign born $N=951$)	SA scale: reading and writing (1 item) <i>Type: general</i>	Functional literacy test (37 items)	Correlations between SA and test: .42 (overall) .29 (English first language) .62 (immigrants)
Glenberg et al. (1987; experiment 1)	English-speaking college students tested in English ($N=80$)	SA scale 1: confidence of correctly identifying a verbatim (15 items) <i>Type: specific (study task)</i> SA scale 2: confidence of correctly judging interferences (15 items) <i>Type: specific (study task)</i>	Reading test (15 texts): (1) verbatim recognition (1 item per text) (2) interference verification (2 items per text) (experimental variation: SAs and tests were administered immediately after each text vs. after reading all texts)	Correlations between SA 1 and test 1 (immediate/delay): -.09/-.04 Correlations between SA 2 and test 2: .04/.06
Glenberg et al. (1987; experiment 2)	English-speaking college students tested in English ($N=40$)	SA scale: confidence of correctly identifying a paraphrase of an idea (15 items) <i>Type: specific (study task)</i>	Reading test (15 texts): identifying the paraphrase of an idea (4 items per text) (experimental variation: SAs and tests administered immediately after each text vs. after reading all texts)	Correlations between the SA and test (immediate/delay): -.21/.03
Glenberg et al. (1987; experiments 6-8)	English-speaking college students tested in English ($N=46/38/37$)	SA scale: confidence of correctly identifying an idea in the text (16 items) <i>Type: specific (study task)</i>	Reading test (16 texts): identifying an idea (3 items)	Correlations between SA and test: .08< r <.13 (experiment 6) .04< r <.26 (experiment 7) .11< r <.27 (experiment 8)
Glover (1989)	English-speaking college students ($N=240$) and 11th/12th graders ($N=118$) tested in English	SA scale: confidence of number of items correctly answering in subsequent test (1 item) <i>Type: specific (study task)</i>	Reading test: multiple choice (40 items)	Correlations between SA and test (college/high school): .04/-06

Table A6.1: Studies comparing self-assessments and test scores/teacher estimates of linguistic skills (*Continued*)

Study	Sample	Self-assessment indicators	Test scores/teacher estimates	Major findings
Janssen-van Dieten (1989)	Second-language students from different language backgrounds in the Netherlands ($N=730$)	SA scale: listening, speaking, reading and writing (number of items not identifiable) <i>Type: specific (real-life task)</i>	Test 1: listening, speaking, reading, and writing subscales (100 items each) Test 2: C-test	Correlations between SAs and test 1: .19 < r < .69 Correlations between SAs and test 2: .09 < r < .56
LeBlanc and Painchaud (1985; study 1)	French-/English-speaking college students in Canada learning English/French ($N=200$)	SA scale: listening, speaking, reading and writing (10 items each) <i>Type: specific (real-life task)</i>	Test: listening, reading, and cloze subscales (number of items not identifiable)	Correlation between SA and respective test scale: .52 (listening) .49 (reading) .53 (overall)
LeBlanc and Painchaud (1985; study 2)	English-speaking college students in Canada learning French ($N=1739$)	SA scales: listening (30 items) and reading (20 items) <i>Type: specific (real-life task)</i>	Test: listening, reading, and cloze subscales (number of items not identifiable)	Correlation between SA and respective test scale: .78/.81 (listening) .70 (reading) .80 (overall)
Lin, Moore, and Zabrocky (2001)	English-speaking college students tested in English ($N=60$)	SA scales (pre-test): confidence of correctly judging interferences, text ease, text understanding and text interestingness (1 item each) <i>Type: specific (study task)</i> SA scales (post-test): certainty of having answered the test items correctly and estimated number of correctly answered items (1 item each) <i>Type: specific (study task)</i>	Reading comprehension (12 texts): true-false questions (48 items)	Correlation between SAs and test: .14 (pre-test) .28 (post-test)
Maki and Berry (1984; experiment 1)	English-speaking college students tested in English ($N=28$)	SA scales: confidence of correctly answering questions on texts in subsequent test (items 11-17 per text) <i>Type: specific (study task)</i>	Reading comprehension test (4 texts): multiple choice (18 items per text)	Correlation between SA and test: .15
Maki and Berry (1984; experiment 2)	English-speaking college students tested in English ($N=45$)	SA scales: confidence of correctly answering questions on texts in subsequent test (items 9-16 per text) <i>Type: specific (study task)</i>	Reading comprehension test (4 texts): multiple choice (16 items per text) (experimental variation: SAs and tests administered immediately after each text vs. after 1- or 3-day delay)	Correlation between the SA and the test (immediate/delay): .20/.23

Table A6.1: Studies comparing self-assessments and test scores/teacher estimates of linguistic skills (*Continued*)

Study	Sample	Self-assessment indicators	Test scores/teacher estimates	Major findings
Peirce, Swain, and Hart (1993)	FLL, English-speaking 8th graders in a French immersion school program ($N=500$)	SA scales 1: listening, speaking, reading and writing (1 item each) <i>Type: general</i> SA scale 2: ability to solve tasks (4 items) <i>Type: specific (real-life task)</i>	Listening comprehension: multiple choice (15 items) Open speaking task (number of items not identifiable) Reading comprehension: cloze test (number of items not identifiable) Open writing task (number of items not identifiable)	Correlation between SA 1 and tests: .01 < r < .19 Correlation between SA 2 and tests: .12 < r < .25
Peyer and Kaiser (2010)	FLL, French-speaking university students learning German ($N=506$)	SA scale: reading at 6 levels (1 item) <i>Type: specific (real-life task)</i>	Test 1: filling gaps in sentences, multiple choice (30 items) Test 2 (2 texts): reading (21 items)	Correlation between SA and tests: .68 (test 1) -.39 (number of errors in test 2)
Rauch, Jurecka, and Hesse (2010)	9th-grade Turkish-speaking immigrant students in Germany ($N=128$)	SA scale: listening at 7 levels (1 item) <i>Type: specific (real-life task)</i>	Reading comprehension: Computerized adaptive test of Turkish taking 35–40 minutes (number of items not identifiable)	Correlation between SA and test: .20
Ross (1998)	EFL, participants of a language training programme at a Japanese electronics company ($N=205$)	SA scale: listening (20 items) <i>Type: specific (real-life task)</i>	Listening test: (1) cloze test (14 items) (2) matching tasks (12 items)	Correlations between SA and test scales: .39 (cloze test) .50 (match tasks)
Wilson (1989)	EFL, French-speaking adults ($N=56$)	SA scale: listening and speaking at 10 levels (1 item each) <i>Type: specific (real-life task)</i>	Test of English for International Communication (TOEIC): (1) listening (100 items) (2) reading (100 items) (3) teacher rating (15–25 minute structured interview)	Correlation between SA and test scales: .64 (listening) .50 (reading) .64 (teacher rating) .63 (overall)
Wilson (1999)	EFL, French-speaking and adults of other language backgrounds ($N=937$)	SA scale: listening and speaking at 7 levels (1 item each) <i>Type: specific (real-life task)</i>	Test of English for International Communication (TOEIC): (1) listening (100 items) (2) reading (100 items)	Correlations between SA and test scales: .75 (listening) .70 (reading) .75 (overall)

Note: This overview covers studies that relate self-assessed language skills to language tests or teacher ratings of language ability. Several studies were excluded due to small sample size (Miller et al. 1991) or inadequate analyses (Blanche and Merino 1989; Fishman and Terry 1969). EFL = English as a foreign language; ESL = English as a second language; FLL = foreign language learners; SA = self-assessment measure. *Type: “general”* = SAs referring to linguistic ability at a general level; *“specific (real-life task)”* = SAs referring to concrete linguistic tasks encountered in everyday life; *“specific (study task)”* = SAs referring to concrete tasks within the respective study.

Figure A6.1: Ability self-assessment L1

How good is your command of the other language?					
<i>Please check one box in each line.</i>					
	Very poor	Rather poor	Rather good	Very good	Not at all
	[1]	[2]	[3]	[4]	[0]
Writing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Reading	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Speaking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Understanding	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Note: The questionnaire had previously defined the “other language” as the language other than German that participants’ had learned in their family.

Figure A6.2: Ability self-assessment L2 (German)

How good is your command of German?					
<i>Please check one box in each line.</i>					
	Very poor	Rather poor	Rather good	Very good	Not at all
	[1]	[2]	[3]	[4]	[0]
Writing	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Reading	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Speaking	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Understanding	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Figure A6.3: Performance self-assessment L1

Your own assessment

In the second part of this booklet, you answered several questions pertaining to texts in Russian/Turkish. Now, we would like you to assess your performance on these questions.

*Please do **not** turn back to the questions and texts.*

Question 1:

Altogether, we asked you to answer 32 questions about the texts.

How many of these questions do you think you answered correctly?

Please write a number between 0 and 32 in the box.

Figure A6.4: Performance self-assessment L2 (German)

Your own assessment

In this booklet, you answered several questions pertaining to texts. Now, we would like you to assess your performance on these questions.

*Please do **not** turn back to the questions and texts.*

Question 1:

Altogether, we asked you to answer 33 questions about the texts.

How many of these questions do you think you answered correctly?

Please write a number between 0 and 33 in the box.

7 *Artikel 4*: When can we (not) rely on self-reported language proficiency? An investigation of the processes underlying response bias and their implications for research on ethnic inequalities

Dieser Beitrag entstand in Kooperation mit Aileen Edele und befindet sich aktuell in Überarbeitung. Der nachfolgende Text und Inhalt wurde unverändert aus der bisher unveröffentlichten Fassung, welche zur Begutachtung der Dissertation eingereicht wurde (Stand: 28.11.2024), übernommen.

Abstract

Proficiency in the majority language of the destination country is important for immigrants' integration. Research examining integration processes and ethnic inequalities often draws on self-reports as indicators of respondents' language proficiency. However, previous research has indicated that these self-reports are often inaccurate and can result in biased findings. The primary objective of the present study is to investigate the processes and associated factors that cause response bias, specifically overestimation bias, in self-reported language proficiency. This facilitates a more systematic prediction of the errors that arise in analyses using self-reports. Emanating from a general model of the survey response process, we examine how respondents' cognitive abilities, aspects of self-concept, and frame of reference are related to overestimation bias. For the analysis, we use data from the German National Educational Panel Study on 1,630 secondary school students with German as a second language. We determine bias by comparing students' self-reported language proficiency to their scores on a language test. The linear regression models indicate that students with lower cognitive abilities, a high linguistic self-concept and self-esteem, and a weak frame of reference tend to overestimate their language proficiency. Among these factors, cognitive abilities explain the largest share of variance in overestimation bias and moderate the effects of the other factors: Students with low cognitive abilities are particularly susceptible to biases associated with aspects of self-concept and the frame of reference. We discuss the implications of using self-reports to study ethnic inequalities, and the conditions under which their use is (less) problematic.

7.1 Introduction

Given the large and growing number of immigrants worldwide, determining the factors that support the integration process of immigrants and their children is a pressing issue. Proficiency in the majority language is generally seen as an important factor in immigrants' successful integration into the destination context. However, for immigrant families who typically learn the majority language as a second or additional language (L2), it can act as an integration barrier that reinforces ethnic inequalities: those with restricted proficiency often experience lower educational outcomes (e.g., Azzolini et al. 2012; OECD 2023; Heath et al. 2008; Stanat and Edele 2016) and encounter lower returns as well as potential discrimination in the labor market (e.g., Chiswick and Miller 1995; Dustmann and Fabbri 2003; Schmaus 2020). In light of the pronounced academic and societal interest in integration processes and the pivotal role of majority language proficiency, valid and reliable instruments to assess L2 learners' language proficiency are highly important.

The most commonly used indicators of L2 learners' language proficiency, particularly in large-scale assessments, are respondents' self-reports. However, there is growing evidence that the accuracy of self-reports is limited. They correlate only moderately with objective indicators of language proficiency (for overviews, see Edele et al. 2015; Li and Zhang 2021; Ross 1998). Even more importantly, several studies indicate that they often yield biased findings. In particular, analyses that draw on self-reports tend to overlook effects that can be identified with objective indicators of language proficiency on different outcomes of structural integration, or they indicate spurious correlations between variables that are, in fact, not or oppositely correlated (e.g., Charette and Meng 1994; Edele et al. 2015; Finnie and Meng 2005). In addition, self-reports seem to be systematically biased in certain respondent groups (Edele et al. 2015); this can lead to erroneous conclusions regarding group differences, which are a common focus of research interest.

Despite the growing evidence of problems involved in using self-reports as measures of respondents' actual language proficiency, the processes underlying the biases in self-reports have rarely been considered in previous research. The identification of the underlying processes and related factors allows us to assess the extent to which findings in previous research based on self-reported language proficiency are biased; this not only helps to evaluate the reliability of previous research regarding immigrants' and other L2 learners' language proficiency but also facilitates the development of strategies and research designs that are less susceptible to such biases. The present study therefore focuses on the processes and associated factors that lead to response bias in self-reported language proficiency. More specifically, we examine how respondents' general cognitive abilities, linguistic self-concept, self-esteem, and frame of reference are related to response bias in self-reported language proficiency. We analyze these relationships using data from the German National Educational Panel Study (NEPS) on 1,630 secondary school students with German as a second language.

Previous research has indicated that respondents' cognitive abilities, aspects of self-concept, and frame of reference can affect the accuracy of their ability judgments (e.g.,

Critcher and Dunning 2009; Ehrlinger 2008; Marsh et al. 2008; Schild 2024; Truxillo et al. 2008). However, most studies to date have focused on single aspects, with the relative impact of these factors and their interplay not yet having been investigated. Moreover, most of these factors have not yet been analyzed in relation to L2 learners' self-reported language proficiency. It is thus an open question whether—and to what extent—these factors distort the response process of self-reported language proficiency.

Our study contributes to the current state of knowledge by integrating the discussed factors into an overarching theoretical framework of the response process. Emanating from this framework, we investigate the (relative) role of the explanatory factors in the accuracy of self-reported language proficiency, as well as the interplay between these factors. The results contribute to a more nuanced understanding of the underlying processes that cause response bias and, on this basis, to a more systematic prediction of the conditions under which the use of self-reports—as indicators of respondents' actual language proficiency—is (not) problematic.

7.2 Background

7.2.1 Tendencies toward overestimation bias in self-reported language proficiency

Self-reports of respondents' ability have repeatedly been shown to be of limited accuracy for various competence domains (for reviews, see Dunning 2005; Mabe and West 1982; Zell and Krizan 2014). Previous research has revealed two major findings related to bias in self-reports: First, respondents are prevalently overconfident about their abilities, and second, individuals who perform well in a specific domain tend to rate their abilities in this domain more accurately than low performers (e.g., Dunning et al. 2003; Kruger and Dunning 1999; Ehrlinger et al. 2008). This general pattern of overestimation by low performers has also been observed in self-reported language proficiency, including among refugee youth (Schild 2024) and other L2 learner populations (Ross 1998; Suzuki 2015).

Furthermore, previous research has indicated that response bias can vary according to respondents' background characteristics, including their ethnic background. For example, Edele et al. (2015) reported that students of Turkish origin—one of the largest ethnic minority groups in Germany—rated their German proficiency as similar to or even higher than the other ethnic groups under study despite their, on average, lower tested performance in the domain; this indicates that the findings of studies investigating group differences in language proficiency may be distorted if they rely on self-reports. Similarly, response bias appears to vary depending on other background characteristics, such as gender (e.g., Beyer and Bowden 1997) or educational attainment (e.g., Finnie and Meng 2005).

7.2.2 Processes underlying response biases

Although previous research has shown that several respondent groups systematically tend to overestimate their language proficiency, the processes underlying these response biases remain unclear. However, to account for systematic overestimations, it is pivotal to identify the mechanisms that link such respondent characteristics to biases. The processes that respondents undergo to generate estimates of their ability provide a useful starting point for this purpose.

It is generally assumed that individuals do not typically have an elaborate evaluation of their ability in a specific domain, including their language proficiency, which they can readily communicate in a survey. Respondents are typically confronted with the task during the survey situation for the first time and must generate an estimate on the spot or at least re-evaluate and potentially update a preexisting view in light of its fit to the question context (Sudman et al. 1996; Tourangeau et al. 2000). According to general models of the survey response process, respondents need to accomplish four tasks when responding to a survey question (e.g., Cannel et al. 1981; Strack and Martin 1987; Sudman et al. 1996; Tourangeau et al. 2000). First, they must *comprehend* the wording of the questions and their meaning. Based on this information, they need to *retrieve* relevant information from memory or generate new information and evaluate this information to form a *judgment*. Finally, respondents need to translate their judgment into a reasonable *response*.

Ideally, respondents would elaborate each step of the response process with maximum effort and diligence (e.g., Tourangeau et al. 2000). However, assessing one's ability is a demanding task that requires respondents to put considerable cognitive effort into retrieving adequate information and forming a reliable judgment (e.g., Carter and Dunning 2008). When respondents do not possess cognitive capacity or are not willing to expend the cognitive effort to accomplish these tasks thoroughly, the response process is prone to bias (Krosnick 1991; Krosnick et al. 2002). Furthermore, respondents need to define reliable criteria against which they can evaluate the retrieved information to form a judgment (Sudman et al. 1996). The accuracy of self-reports depends heavily on the criteria on which respondents base their answers and on how adequate those criteria are. This suggests that the magnitude of bias in ability self-reports reflects the respondents' cognitive abilities and the reliability of the criteria used in the response process.

In the following, we discuss factors that are likely to affect the accuracy of the response process and biases in ability self-reports. More precisely, we focus on the role of respondents' cognitive abilities, aspects of their self-concept, and their frame of reference.

7.2.3 Cognitive abilities

General cognitive abilities are crucial to adequately perform retrieval and judgment processes (Krosnick 1991). Students with high cognitive abilities can accomplish the demanding tasks of retrieval and judgment more easily. In contrast, students with lower

cognitive abilities are more likely to face problems accessing and evaluating adequate information. Consequently, they are likely to simplify the retrieval and judgment processes by switching to response heuristics that are less cognitively demanding (Krosnick 1991; Krosnick et al. 2002). For example, respondents reduce their cognitive effort by retrieving not all relevant information but only the most accessible information (e.g., Sudman et al. 1996; Tourangeau and Rasinski 1988). Respondents with lower cognitive abilities are thus more likely to omit relevant cues and to provide erroneous estimates of their abilities.

Another line of reasoning, which links cognitive abilities and response bias in a similar way, highlights the specific mechanism of metacognition. Metacognition is defined as the ability to accurately assess and reflect on one's performance (Flavell 1979). It refers to the ability to evaluate shortcomings in one's own performance (e.g., mistakes in test performance) and to distinguish between correct and incorrect responses (Dunning et al. 2003). Respondents with low metacognitive abilities lack the cognitive strategies to retrieve and judge the relevant information efficiently, which is necessary to accurately evaluate their own performance (Kruger and Dunning 1999). Dunning et al. (2003) refer to a "double curse", as the skills needed to score high on a performance test are similar to those needed to evaluate one's performance. For example, respondents need to know the correct answer to a specific test task to judge whether they have solved the task correctly or not. Respondents with low metacognitive abilities thus have difficulties identifying existing deficits in their own performance (Dunning et al. 2003; Ehrlinger et al. 2008). When evaluating their performance, they selectively retrieve and evaluate information: they unintentionally omit negative experiences and overvalue positive experiences, leading to a systematic overestimation of their abilities.

Only a few studies to date have investigated the associations between respondents' cognitive abilities and the accuracy of their ability self-reports. For example, Truxillo et al. (2008) examined perceived and actual performance in a video-based situational judgment test that assessed respondents' coping strategies for critical incidents in work situations. The authors found considerable convergence between self-reports and test scores for respondents with high general cognitive abilities, whereas no significant associations emerged for respondents with low cognitive abilities. Specifically, with respect to the accuracy of self-reported language proficiency, Schild (2024) reported a negative correlation between general cognitive abilities and self-reported German proficiency among refugee adolescents with comparable levels of tested language proficiency, pointing to less accurate judgments of one's own language proficiency in individuals with lower general cognitive abilities.

7.2.4 Self-concept

Bias in self-reports may also arise from relying on unreliable criteria to evaluate one's ability. It is widely assumed that respondents base their self-reports on preexisting self-concepts of their own abilities (e.g., Ehrlinger 2008; Ehrlinger and Dunning 2003). Such self-concepts represent relatively broad and stable beliefs about one's abilities and general potential and draw on evaluations of individual characteristics and experiences

(e.g., Bong and Skaalvik 2003; Freund and Kasten 2012). An individual's self-concept encompasses multiple dimensions that vary in their level of generality, ranging from a general self-concept to more nuanced perceptions of the self regarding specific domains or situations (Shavelson et al. 1976). The dimensions that can influence ability judgments include both specific domains (e.g., academic self-concept) and global dimensions, such as general self-esteem (Critcher and Dunning 2009).

It seems reasonable to assume that respondents base their judgments on their self-concepts. Specifically, respondents may use their self-concept as a benchmark of their ability, against which they evaluate all other retrieved information (e.g., their performance on an exam) to form a judgment. This strategy may indeed provide accurate self-reports if the self-concept represents an adequate picture of respondents' actual abilities. However, the beliefs individuals hold about their abilities are often inflated and only modestly correlate with their actual performance (e.g., Carter and Dunning 2008).

Respondents who base their judgments on a flawed self-concept are likely to provide biased self-reports. Individuals' tendency to maintain a stable self-concept most likely inflates that bias. According to cognitive dissonance theory (Festinger 1957) and related notions (e.g., Aronson 1968; Steele 1988; Swann 1983), individuals seek to attain outcomes that are in accordance with their self-concept. In contrast, information that contradicts one's self-concept should produce cognitive dissonance and discomfort (e.g., Freund and Kasten 2012; Harmon-Jones and Harmon-Jones 2007). To avoid those psychological costs, respondents should judge their performance in a fashion that suits their preexisting self-concept (e.g., Wells and Sweeney 1986). They consequently ignore or pay less attention to cues that contradict their self-concept when forming their judgments, although these cues might enhance the accuracy of predicting their actual ability (e.g., Farh and Dobbins 1989; Kwang and Swann 2010; Swann and Read 1981). This tendency may reflect a simplified response strategy, whereby respondents with a positive self-concept do not retrieve and evaluate all available information but rather only the most positive information.

Prior research indicates a positive association between individuals' ability-related self-concepts and their self-reported performance (e.g., Ackermann and Wolman 2007; Critcher and Dunning 2009; Ehrlinger and Dunning 2003; Kwon and Linderholm 2014). Such correlations have been identified in relation to different aspects of self-concept. For example, previous studies have shown that respondents with exaggerated self-esteem tend to rate their ability in specific domains, including language proficiency, more favorably than respondents with lower self-esteem, who perform at the same level (Wells and Sweeney 1986; Kim et al. 2016). Moreover, additional studies indicate a correlation between the accuracy of ability self-reports and various personality traits associated with an individual's self-concept, including extraversion, agreeableness, and narcissism (Jacobs et al. 2012; John and Robins 1994; Park et al. 2016).

7.2.5 Frame of reference

Another criterion that respondents are likely to use in evaluating their ability is the frame of reference. According to Marsh's (1986) internal/external frame of reference model, individuals compare their own performance to that of other individuals within their social environment and base their judgments on this information. The frame of reference may consist of generalized others (e.g., the average performance level in a school or class), selected individuals (e.g., the performance of close friends), or both (Marsh et al. 2008). Judgments based on social comparison processes are susceptible to bias, as respondents are exposed to different social environments with varying compositions, which may influence their perceptions and evaluations. Specifically, individuals with comparable performance levels rate their own abilities differently depending on their frame of reference: when they refer to subjects with relatively low performance, they tend to overestimate their own abilities, and conversely, they view themselves less favorably when they refer to subjects with relatively high performance ("big-fish-little-pond-effect"; Marsh and Parker 1984).

Many studies indicate that students' frames of reference affect their ability judgments (for reviews and meta-analyses, see Fang et al. 2018; Loyalka et al. 2018). They suggest that the frame of reference affects students' academic self-concepts. Specifically, students in classrooms with a low performance composition reported higher academic self-concepts than did students with similar individual performance in classrooms with a higher average performance level. Experimental research also indicates that students use their frame of reference as a benchmark for evaluating their ability. For example, Strickhouser and Zell (2015) reported that students systematically respond to manipulated feedback about their test performance in comparison to their classmates. Those who were told that they were above average evaluated their abilities more favorably than students who were told that their scores were average or below. This tendency is further corroborated by the findings of a meta-analysis of 102 experiments investigating social comparison processes in ability judgments across various domains (Gerber et al. 2018). Under experimental conditions that induce a downward comparison, individuals tend to rate their ability more favorably. Moreover, the study demonstrated that individuals exhibit a stronger response to social comparisons when they compare themselves to individuals in their immediate social environment (e.g., their classmates) than to more distant or abstract references (e.g., the population's average).

The performance of their classmates represents an obvious frame of reference, particularly for students evaluating their abilities. A salient indicator of one's peers' abilities is their grades in a specific subject or in general. School grades reflect not only students' abilities in absolute terms but also their relative performance among classmates, as teachers tend to adapt their grading to the ability composition of the class (e.g., Hübner et al. 2020; Neumann et al. 2011). Students with comparable performance receive better grades in classrooms with a lower average ability than in classrooms with higher-ability students. Consequently, students with above-average grades have a relatively weak frame of reference and are likely to be overly optimistic in their ability judgments, even if their abilities are rather low in absolute terms. Thus, the classroom,

as a prominent frame of reference, is a potential source of bias in students' ability self-reports.

Previous research on frame of reference effects has focused mainly on students' academic self-concepts and has largely neglected effects on ability self-reports. It is therefore an open question whether frame of reference effects lead directly to bias in ability self-reports or indirectly induce bias through distortions in self-concept that bias the response process (see section 7.2.4).

7.3 Aims and expectations

The primary aim of the present study is to investigate the processes and associated factors that lead to response bias in self-reported language proficiency. In particular, we aim to explain overestimation bias, the main source of bias identified in previous research.

In line with the theoretical considerations and empirical findings on ability self-reports in general, we expect that lower cognitive abilities, a high self-concept and a weak frame of reference will also bias the response process of self-reported language proficiency. Moreover, we expect general cognitive abilities to moderate the biasing influences of self-concept and frame of reference. Respondents with high cognitive abilities retrieve more reliable information and evaluate this information more reliably during the response process. They are likely to reflect more critically on the cues associated with their self-concept and frame of reference and are therefore less susceptible to biases associated with these factors. In contrast, respondents with low cognitive abilities may not be able to critically reflect on the adequacy of this type of information and may therefore rely heavily on it in forming their judgments. Consequently, the biases associated with a high self-concept and a weak frame of reference should be particularly pronounced among respondents with lower cognitive abilities.

In the following, we concurrently examine the roles of general cognitive abilities, self-concept, and frame of reference, as well as their interaction, in the overestimation bias of self-reported German proficiency among adolescent L2 learners. In addition to global self-esteem, we analyze linguistic self-concept as a particularly relevant aspect of respondents' self-concept in assessing their language proficiency.

7.4 Data and methods

7.4.1 Sample

We analyze data from the *German National Educational Panel Study* (NEPS; Blossfeld and Roßbach 2019). The NEPS is carried out by the Leibniz Institute for Educational Trajectories (LIfBi, Germany) in cooperation with a nationwide network. We use data from the first and second waves of Starting Cohort 4, a nationwide sample of ninth-grade students, which includes both tested and self-reported German proficiency (NEPS Network 2024). Students were sampled with a stratified multistage sampling design.

In the first stage, schools from different school types were randomly selected. In the second stage, all students from one or two classes within each school were invited to participate in the survey (Skopek et al. 2013). The self-administered questionnaire, as well as the testing, was conducted in fall/winter 2010/2011 (wave 1) and in spring 2011 (wave 2). As the two waves assess different constructs, we integrate information on both waves into one cross-sectional dataset.

The initial sample of Starting Cohort 4 includes 16,425 students. We restrict our analyses to regular school tracks, i.e., we exclude students attending schools for children with special needs ($N=1,186$) or private schools ($N=118$). In the first wave, students who had learned a language other than German at home were asked to self-report their level of German proficiency. Our analyses are limited to this specific subsample of 3,004 students with German as a second language who participated in wave 1. This includes students who were born abroad and students who were born in Germany but whose parents or grandparents were born abroad. For 2,350 of the L2 students, valid information on both self-reported and tested German proficiency is available. In our analysis, we focus exclusively on students with accurate or overestimated self-reports (see section 7.4.2), resulting in an analysis sample of 1,630 cases.

7.4.2 Determining bias in self-reports

In our analysis, we define the outcome variable bias as the discrepancy between respondents' self-reports and an objective criterion of their actual ability. Specifically, we contrast the self-reported German proficiency of L2 students with the results they obtained in a test of their proficiency in German.

In the questionnaire, students are asked to indicate how well they understand, speak, read and write the German language. The response categories ranged from 1 "not at all" to 5 "very good". We combine the four items into a composite measure of *self-reported language proficiency* ($M=4.59$, $SD=0.52$), reflecting respondents' self-reported overall German proficiency.

Students' *tested language proficiency* is assessed with a reading comprehension test. The test includes five texts (approximately 200 to 550 words each) of different types, including informational, commentary, literary, instructional, and advertising texts (Gehrer et al. 2012). Following every text, the students are asked to complete five to seven tasks (mostly multiple-choice questions), which requires them to find information in the text, draw text-related conclusions, or reflect on the content of the text. In total, the test consists of 31 items. Test scores are provided as weighted maximum likelihood estimates (WLEs) of students' reading competences (Pohl and Carstensen 2012). The WLE scores are mean-centered, thus representing point estimates of individual scores relative to the full sample mean. In the subsample used for the analysis, the scores range from -4.02 to 3.30 ($M=-0.53$, $SD=1.20$).

The determination of bias in self-reports presents certain methodological challenges that must be overcome to achieve reliable results. To facilitate a meaningful comparison between self-reported and tested language proficiency, it is essential that both indicators measure the same construct. The self-report measure combines information

on students' language proficiency across four domains, whereas the test is designed primarily to assess reading comprehension. However, reading comprehension requires the use of a multitude of other language-related skills and is typically seen as an integrated measure of language abilities. Consequently, reading comprehension tests can serve as indicators of respondents' language proficiency in a more general sense and are frequently employed in research as such (see Edele et al. 2015). Nevertheless, to test the robustness of our findings, we conduct sensitivity analyses using alternative operationalizations of self-reports that are more closely aligned with reading comprehension. Specifically, we employ the single item on reading—instead of the composite measure—and an additional measure reflecting students' estimates of the proportion of tasks they have solved correctly in the reading comprehension test (see section 7.5.2). The principal focus of the study is, however, on the composite measure of self-reports, as it is predominantly used in integration research as an indicator of language proficiency.

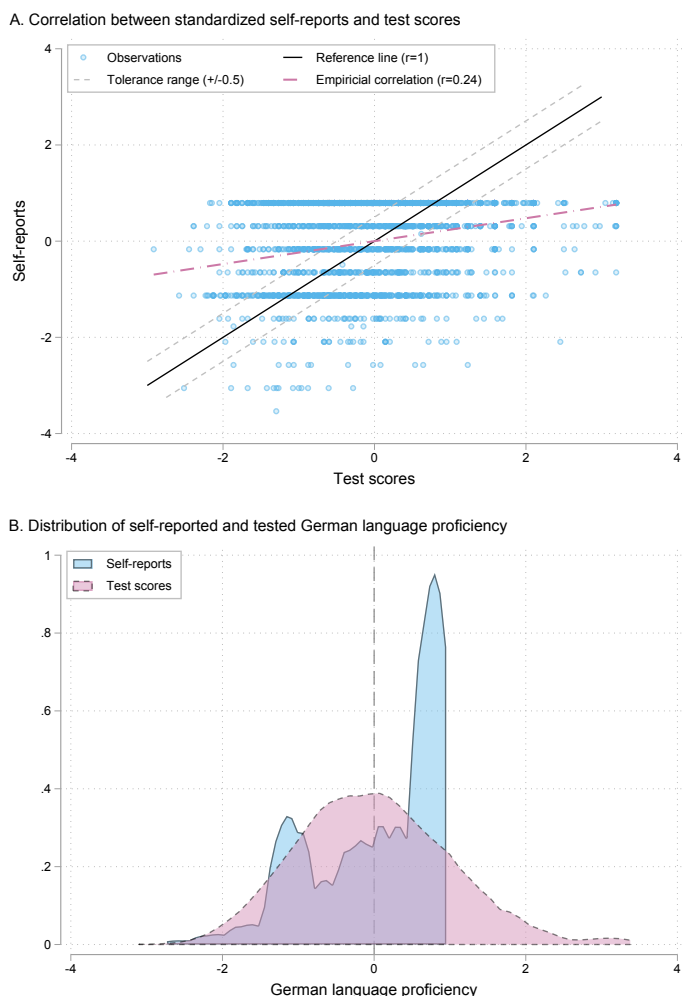
Another challenge in determining bias relates to the method used to compare self-reports with the objective criterion. Two main approaches have been used in previous research: calculating the algebraic difference between the two measures (*difference score*) and estimating the residuals of the regression of self-reports on the objective criterion (*residual score*). The basic difference score requires both measures to be on the same scale, which is often not the case. In addition, general concerns have been raised about the use of difference scores, suggesting that the residual approach is more reliable (e.g., Cronbach and Furby 1970; Edwards 1994; 2001). However, the residual approach removes all variance associated with the objective criterion (e.g., De Los Reyes and Kazdin 2004; Paulhus and John 1998). As a result, the influences of actual language proficiency and related constructs, such as general cognitive abilities, which may also shape the response process, cannot be reliably identified via this approach.

In our study, we therefore follow an alternative strategy employing the *standardized difference* (e.g., De Los Reyes and Kazdin 2004; Whitton et al. 2008) between self-reported and tested language proficiency. We standardize each measure according to the mean and standard deviation of the subsample, yielding z-scores ($M=0$, $SD=1$) on the same scale for both measures. The difference score is calculated by simply subtracting the standardized test scores from the standardized self-reports. The standardized difference represents a relative measure of bias in self-reports. Students who perform on average on the test ($M=0$) should also rate their proficiency at an average level ($M=0$) relative to the population under study. Similarly, students who score one standard deviation below or above the average should rate their proficiency accordingly ($M=-1/1$). In such cases of accurate self-reports, the standardized difference score is close to zero. Negative values indicate a tendency toward underestimation, and positive values indicate a tendency toward overestimation. For example, a score of 1 indicates that a student rated their language proficiency one standard deviation higher than would be expected based on their relative performance in the test.

Figure 7.1 displays the correlation between the standardized self-reports and test scores (A) and the distribution of the two measures (B). The correlation between self-reported and tested language proficiency is only moderate (Pearson's $r=0.24$, $p<0.001$). In terms of R-square, this implies that students' test scores account for only 5.7 percent

7 When can we (not) rely on self-reported language proficiency?

Figure 7.1: Correlation (A) and distribution (B) of standardized self-reports and test scores



Notes: The analysis is based on the subsample of students with German as a second language, and for whom valid data on both self-reported and tested German proficiency are available ($N=2,350$).

of the variance in self-reports. Furthermore, the self-reports show a lower variance and a highly skewed distribution than the test results (see Fig. 7.1B). The majority of the students (over 90 percent) report a high level of German proficiency (rated between 4 “good” and 5 “very good”). In contrast, the results of the test indicate that German proficiency is normally distributed in our subsample.

Consistent with previous studies that have identified overestimation as the predominant pattern of bias in self-reported language proficiency (e.g., Ross 1998; Schild 2024; Suzuki 2015), our findings indicate that a considerable number of students tend to overestimate their German proficiency. In our analysis, we aim to address this form of bias and therefore focus exclusively on overestimations (positive scores) and accurate self-reports (scores around zero). This includes all observations that exceed the reference line illustrated in Figure 7.1A. As the standardized difference scores are

rarely exactly zero, we also classify scores just below zero [-0.5 to 0] as accurate and include these cases in the analysis. To validate our methodological choices, we conduct sensitivity analyses using a stricter cutoff criterion for the analysis sample (only scores of 0 or greater). In addition, we determine bias via a residual approach (see section 7.5.2). In the residual approach, the definition of accurate self-reports is based on the empirical correlation (see the corresponding line illustrated in Fig. 7.1A).

7.4.3 Explanatory and background variables

The various explanatory variables include indicators of the students' cognitive abilities, self-concept, and frame of reference. Students' *general cognitive abilities* are tested based on a matrices test assessing nonverbal reasoning (Haberkorn and Pohl 2013). The test includes 12 items in which students are asked to complement sequences of geometric elements, similar to Raven's matrices test (e.g., Raven 2000). We calculate a sum score containing the number of tasks each student solved correctly.

The self-concepts of the students are assessed via two measures. The *linguistic self-concept* reflects the students' perception of their ability in the specific competence domain (i.e., German). Respondents are asked to answer three questions on their abilities in the school subject German (e.g., "I learn fast in German") from 1 "does not apply" to 4 "applies completely". In addition, we use a scale measuring students' *self-esteem* to detect more general aspects of their self-concept. The NEPS employs a revised version of the Rosenberg self-esteem scale (Rosenberg 1965), which consists of ten questions assessing students' complacency (e.g., "All in all, I am satisfied with myself"), ranging from 1 "does not apply" to 5 "applies completely" (see Wohlkinger et al. 2016). For both indicators, we calculate the mean of the single items.

Frame of reference effects are typically examined by regressing students' academic self-concepts on average class performance while simultaneously controlling for their individual performance (e.g., Marsh et al. 2008; Thijs et al. 2010). As individual students' performance (i.e., tested language proficiency) is part of the difference score used as the outcome variable in our analyses, this approach would result in a highly endogenous regression model and is therefore not suitable for the purpose of this study. We therefore use the students' grades in German class, which serve as an indicator of the student's academic standing in comparison to their classmates, to assess their frame of reference. In Germany, the grading scale ranges from 1 (very good) to 6 (deficient). We reverse-code the scale so that higher values indicate better grades. Furthermore, we center the individual grades on the class average (based on the initial sample, including information on all participating classmates) to account for baseline differences in grades across classrooms. Group-mean centering is a reliable method for indicating students' relative standing in class (e.g., Hox 2010). The grades below the average represent a strong frame of reference, whereas the grades above the average represent a weak frame of reference.

The analyses further include several background characteristics of the students that have emerged in previous studies as being associated with bias: students' *ethnic group*, *migration experience* (born abroad versus born in Germany), *gender*, and attended

school track. Based on the country of birth of the students, their parents and grandparents, and their first language, we distinguish five ethnic groups: students from Turkey, Poland, the former Soviet Union, former Yugoslavia and a mixed category of all other origins. School tracks in the German secondary education system can be broadly grouped into low (“Hauptschulzweig”), intermediate (“Realschulzweig”) and high (“gymnasialer Zweig”). Most schools offer one or more tracks, whereas some integrated school types do not differentiate among tracks and are therefore categorized separately in the analysis.

7.4.4 Analytical approach

We employ linear regression models to examine the hypothesized associations between overestimation bias in self-reported language proficiency and the explanatory variables. To account for the clustered data structure with students nested in school classes, we estimate robust standard errors (Huber 1967; White 1982). To facilitate interpretation of the findings in the regression analyses, we transform all explanatory variables so that they range from zero (low) to one (high). For the regression models, this implies that coefficients represent the mean difference between students with the lowest and the highest empirical values of these variables. Negative coefficients imply more accurate self-reports, whereas positive coefficients indicate tendencies toward overestimation bias. Table 7.1 summarizes the distributions and proportions of missing values for all model variables.

Table 7.1: Distribution of variables used in the analysis

Variables	Min–Max	Mean (SD) / %	Missing in %
Overestimation bias (<i>stand. difference</i>)	-0.50–2.97	0.62 (0.76)	0.0
Ethnic group (in %)			0.0
Turkey		26.7	
Poland		7.6	
Former Soviet Union		20.4	
Former Yugoslavia		9.9	
Other		35.5	
Born abroad (in %)		24.7	0.0
Female (in %)		54.1	0.0
School track (in %)			0.0
Low		46.5	
Intermediate		26.0	
High		20.6	
Integrated		6.9	
General cognitive abilities	0–1	0.62 (0.22)	0.6
Linguistic self-concept	0–1	0.65 (0.20)	2.9
Self-esteem	0–1	0.70 (0.17)	8.4
Frame of reference	0–1	0.51 (0.13)	3.1

Notes: Due to rounding, the sum of percentages may deviate from 100%. The analysis sample includes overestimations and accurate self-reports only ($N=1,630$).

We address item nonresponse at the individual level via multiple imputation and estimate 50 datasets with complete information (e.g., White et al. 2011). In accordance with recommendations set forth by von Hippel (2007), cases exhibiting missing values

for the outcome variable are incorporated into the imputation process and excluded from the analysis models. The regression models are separately estimated on each of the 50 imputed datasets and then combined following Rubin's (1987) rule. The descriptive results are based on the original data. We report unweighted results for the descriptive and regression analyses. Robustness checks in which regression analyses include sample weights yield results analogous to those of unweighted analyses. All analyses are conducted with the statistical software package Stata (Version 17.0; StataCorp 2021).

7.5 Results

7.5.1 Accounting for overestimation bias in self-reported language proficiency

In the regression analysis, we proceed in three steps. First, we examine differences in overestimation bias according to the key background characteristics of the respondents. Second, we add the explanatory variables to the models and analyze their associations with bias. In the final step, we investigate the hypothesized interactions between students' general cognitive abilities and the effects of aspects of their self-concept and their frame of reference (see Table 7.2).

In line with previous findings, Model 1 indicates that the tendency to overestimate their German proficiency significantly varies by the students' ethnic group and attended school track. Students of Turkish origin and from former Yugoslavia show greater tendencies toward overestimation than other ethnic groups, particularly students from the former Soviet Union. Furthermore, students who attend the lowest school track report their language proficiency more optimistically than students in any other track, with the most pronounced differences between the lowest and the highest school track. In addition, students born abroad and female students tend to provide more accurate self-reports, yet the coefficients of the two characteristics in Model 1 are not statistically significant ($p > 0.05$). This pattern is more pronounced in the preceding models, when general cognitive abilities are controlled for, showing statistically significant differences in overestimation bias by migration experience and gender (see Models 2 and 5).

Models 2–4 separately control the explanatory variables, namely, the students' general cognitive abilities (Model 2), their linguistic self-concept and self-esteem (Model 3), and the frame of reference (Model 4). Model 5 includes all explanatory variables, considering possible covariations and their relative importance. The results demonstrate the expected negative association between general cognitive abilities and bias in self-reports. Students with low general cognitive abilities are inclined to overestimate their German proficiency, whereas students with high general cognitive abilities provide more accurate self-reports. Students with the lowest scores on the reasoning test of general cognitive abilities, for example, rate their language proficiency—relative to their language test results—by approximately 0.8 of a standard deviation more favorably than students who performed best in the reasoning test (see Model 2).

Table 7.2: Predicting overestimation bias in self-reported German proficiency (linear regression)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Ethnic group (<i>ref. Turkey</i>)								
Poland	-0.12 (0.08)	-0.07 (0.07)	-0.11 (0.08)	-0.12 (0.08)	-0.06 (0.07)	-0.08 (0.07)	-0.07 (0.07)	-0.07 (0.07)
Former Soviet Union	-0.11 (0.06)	-0.03 (0.05)	-0.10 (0.05)	-0.11 (0.06)	-0.02 (0.05)	-0.03 (0.05)	-0.03 (0.05)	-0.02 (0.05)
Former Yugoslavia	0.03 (0.07)	0.03 (0.07)	0.01 (0.07)	0.03 (0.07)	0.01 (0.07)	0.01 (0.07)	0.01 (0.07)	0.01 (0.07)
Other	-0.11 (0.05)	-0.06 (0.05)	-0.12 (0.05)	-0.11 (0.05)	-0.07 (0.04)	-0.07 (0.04)	-0.07 (0.04)	-0.07 (0.04)
Born abroad	-0.07 (0.04)	-0.10 (0.04)	-0.06 (0.04)	-0.07 (0.04)	-0.09 (0.04)	-0.09 (0.04)	-0.09 (0.04)	-0.09 (0.04)
Female	-0.04 (0.04)	-0.08 (0.04)	-0.05 (0.04)	-0.04 (0.04)	-0.09 (0.04)	-0.08 (0.04)	-0.09 (0.04)	-0.08 (0.04)
School track (<i>ref. low</i>)								
Intermediate	-0.29 (0.05)	-0.20 (0.04)	-0.31 (0.05)	-0.29 (0.05)	-0.22 (0.04)	-0.23 (0.04)	-0.22 (0.04)	-0.23 (0.04)
High	-0.49 (0.05)	-0.32 (0.05)	-0.51 (0.05)	-0.49 (0.05)	-0.36 (0.05)	-0.36 (0.05)	-0.35 (0.05)	-0.37 (0.05)
Integrated	-0.20 (0.08)	-0.12 (0.07)	-0.22 (0.09)	-0.20 (0.08)	-0.15 (0.08)	-0.14 (0.08)	-0.15 (0.08)	-0.13 (0.08)
General cognitive abilities (Cog)								
Linguistic self-concept (LSC)		-0.84 (0.09)			-0.81 (0.09)	0.24 (0.26)	-0.15 (0.33)	0.25 (0.34)
Self-esteem (SE)			0.48 (0.09)		0.55 (0.11)	1.56 (0.26)	0.55 (0.11)	0.58 (0.11)
Frame of reference (FoR)			0.31 (0.12)		0.27 (0.11)	0.27 (0.11)	0.86 (0.33)	0.26 (0.11)
Cog X LSC				0.11 (0.14)	-0.31 (0.16)	-0.27 (0.16)	-0.31 (0.16)	1.00 (0.43)
Cog X SE						-1.60 (0.37)	-0.92 (0.46)	
Cog X FoR								-2.08 (0.62)
Constant	0.92 (0.05)	1.37 (0.07)	0.41 (0.10)	0.86 (0.08)	0.98 (0.11)	0.30 (0.19)	0.56 (0.25)	0.30 (0.23)
R ²	0.08	0.13	0.10	0.08	0.15	0.16	0.16	0.16

Notes: Robust standard errors in parentheses; N=1,630. Cog=general cognitive abilities, LSC=linguistic self-concept, SE=self-esteem, FoR=frame of reference, X indicates the interaction (multiplication) between two variables. * p<0.05, ** p<0.01, *** p<0.001.

Students' linguistic self-concept and self-esteem are both significantly associated with bias in self-reports (Model 3). Students with a high linguistic self-concept and/or a high self-esteem tend to rate their German proficiency more favorably than students with lower expressions of these traits. The results of Model 4 indicate only a weak positive correlation between the frame of reference and overestimation bias, which is not statistically significant. However, the magnitude and direction of the coefficient exhibit a notable change when aspects of the students' self-concept are simultaneously controlled for (Model 5). This finding suggests that the frame of reference may not be directly related to bias in self-reports but may instead shape students' self-concepts in a systematic manner.

The coefficients of general cognitive abilities and the aspects of self-concept in Model 5 are comparable to those observed in Models 2 and 3, indicating that there is no substantial covariation between the two factors. With respect to the relative magnitude of the explanatory variables, the linguistic self-concept ($\beta=0.15$) appears to exert a greater influence than self-esteem ($\beta=0.06$), whereas general cognitive abilities ($\beta=-0.24$) emerge as the most influential factor, explaining the largest share of variance in overestimation bias.¹

In the next step, we examine whether the magnitude of bias associated with aspects of self-concept and the frame of reference differs between students with low versus high general cognitive abilities (Models 6–8). The conditional coefficients in these models represent the correlation between the respective variable and the outcome when the other condition of the interaction term equals zero; in the present case, this pertains to the condition of low general cognitive abilities. The coefficients of the interaction terms indicate the discrepancy in these conditional effects between students with high and those with low general cognitive abilities. The results of the interaction between students' general cognitive abilities and their linguistic self-concept (Model 6), self-esteem (Model 7), and frame of reference (Model 8) reveal a pattern that aligns with the hypothesized associations: all three factors demonstrate a positive conditional coefficient and a negative interaction term, both of which are considerable in magnitude and statistically significant.

Figure 7.2 provides a graphical illustration of the differential influences on students with low versus high general cognitive abilities. Among students with low general cognitive abilities, a positive correlation between the tendency to overestimate their language proficiency and higher expressions of linguistic self-concept (Fig. 7.2A) and self-esteem (Fig. 7.2B) emerged. In contrast, the two aspects of self-concept are fully unrelated to the accuracy of self-reports among students with high general cognitive abilities. Thus, only students with lower general cognitive abilities seem susceptible to the distorting influences associated with an exaggerated self-concept, as reflected in their self-reports.

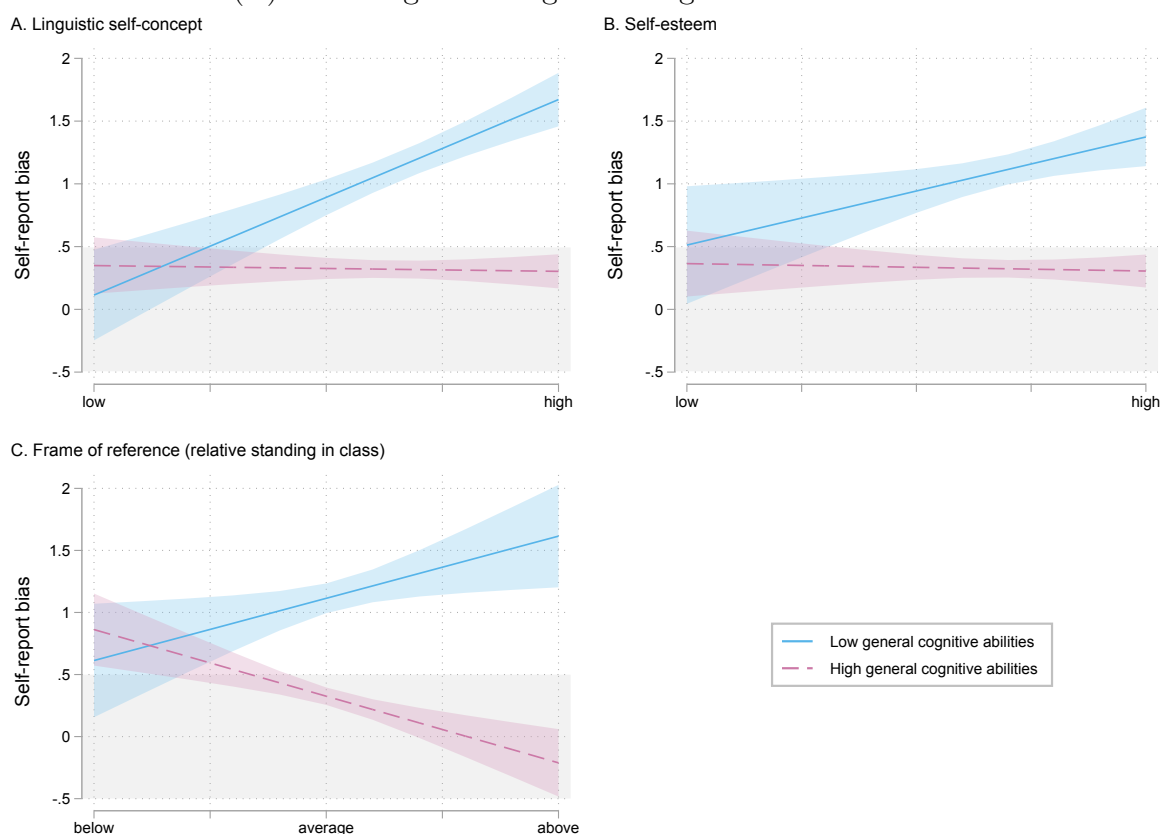
The findings for the influences of the students' frame of reference exhibit a similar

¹The standardized beta coefficients (β) were calculated based on Fisher's z transformation of the coefficients outlined in Model 5 (Table 7.2) using the Stata ado "mibeta" by Yulia Marchenko. The ado was also used to calculate R-square for the imputed data in the regression analyses.

7 When can we (not) rely on self-reported language proficiency?

pattern (Fig. 7.2C). Among students with low general cognitive abilities, those with a weaker frame of reference are more inclined to overestimate their language proficiency than those with a stronger frame of reference. Conversely, students with high general cognitive abilities tend to provide more accurate responses when they attend classrooms with a weaker frame of reference. These differential findings may indicate that students with low general cognitive abilities tend to differentiate themselves from their peers, whereas those with higher cognitive abilities demonstrate an aptitude for aligning their judgments with their frame of reference. The two opposing trends in social comparison processes, when considered in combination, result in an average coefficient that is close to zero and not statistically significant, masking the underlying mechanisms (see Model 4 in Table 7.2).

Figure 7.2: Conditional correlations between overestimation bias in self-reports and students' linguistic self-concept (A), self-esteem (B), and frame of reference (C) according to their general cognitive abilities



Notes: Graphical illustration of the results outlined in Table 7.2, Models 6–8 ($N=1,630$). The estimates represent predictive marginal effects calculated with the Stata ado “mimrgns” by Klein (2014). Categories “below” and “low” (0) represent the lowest empirical values of the variables, “above” and “high” (1) represent the highest values.

7.5.2 Sensitivity analyses

To examine the robustness of our findings, we conduct sensitivity analyses utilizing alternative operationalizations, measures, and approaches to assess self-report bias. Figure 7.3 summarizes the main results of these analyses regarding the background characteristics (A), the explanatory variables (B), and their interactions (C). The full tables containing all the models and coefficients are included in Appendix 7.8 (see Tables A7.1–A7.4).

We conduct four alternative sensitivity analyses. *First*, we calculate the standardized difference using only the self-reported item on reading ability instead of the composite measure (S1). *Second*, we employ the difference between the students' self-reported number of test items they solved correctly—which they were asked to report after completing the reading comprehension test—and their actual test performance (S2). These two alternatives account for the possibility that the composite self-report measure may capture different domains of language proficiency than the test.

Third, we apply a stricter criterion to define the analysis sample of accurate and overestimated self-reports (S3). The analysis essentially resembles the main analysis with a restricted subsample, excluding individuals with a negative difference score close to zero, and presumably more accurate self-reports. With the stricter focus on overestimation, the analysis in S3 specifically addresses variations in the magnitude of the overestimation bias—that is, the absolute discrepancy between self-reports and test results among those individuals who overestimate their proficiency. In contrast, the main analysis also includes (more) individuals with accurate self-reports as a reference, thus combining a measure of both the magnitude and the likelihood of overestimation bias. Differences between the analyses may indicate trends in the magnitude versus the likelihood of overestimation bias.

Fourth, we use a residualized difference score to define accurate and overestimated self-reports (S4). This approach removes all variance associated with the tested language proficiency and its correlates. Therefore, the analysis in S4 addresses only tendencies of bias that are independent of students' actual language proficiency. Unlike the standardized difference approach, the residual approach does not allow for the identification of additional biasing conditions and mechanisms that vary with students' levels of language proficiency. Consequently, discrepancies between the main analysis and the residual approach may provide insight into associations that are ultimately related to differences in actual language proficiency, which may also reflect the general tendency of low-performing individuals to overestimate their abilities.

With respect to students' background characteristics, the regression analyses using alternative specifications S1 and S2 yield coefficients similar to those in the main analysis (see Fig. 7.3A). Many of the differences are even more pronounced, particularly in the analysis using self-reported test performance (S2). The coefficients of the analysis using a restricted sup-sample of overestimations (S3) also point in the same direction, but they are considerably smaller in magnitude than those of the main analysis. For example, analysis S3 reveals no statistically significant differences by ethnic group. Given that the outcome used in S3 specifically measures the magnitude of self-report

bias, the group differences found in the main analysis appear to reflect differences in the likelihood rather than the magnitude of the overestimation bias. Accordingly, some groups tend to overestimate their German proficiency more frequently than others, but the magnitude of the overestimation bias is similar for all groups.

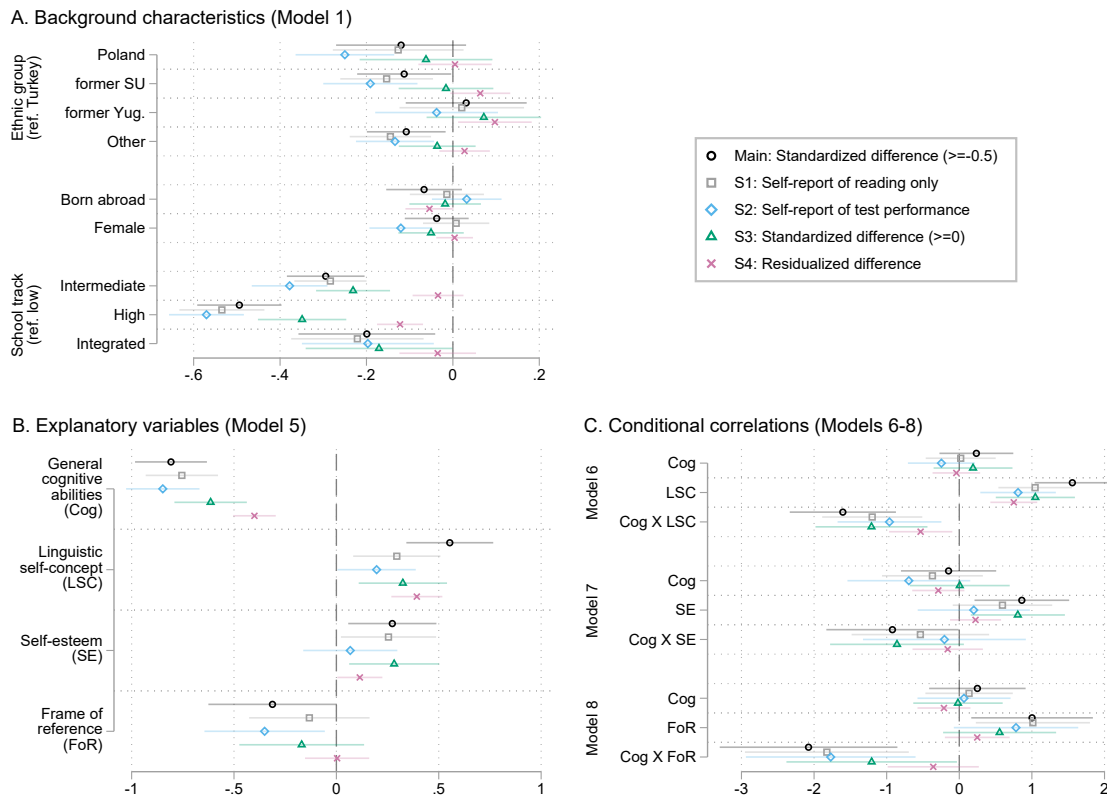
The analysis based on the residual approach (S4) shows hardly any significant differences in overestimation bias by students' background characteristics. This finding indicates that the differences between ethnic groups and school tracks are considerably reduced among students with similar language proficiency according to the test. Consequently, the group differences identified in the main analysis are likely attributable to variations in the average level of actual language proficiency between the groups, as opposed to additional group-specific response patterns. Groups with lower average performance levels are thus more susceptible to overestimating their language proficiency. This assumption is corroborated by the observation that those groups that are most inclined to overestimate their language proficiency also present the lowest average test scores (see Appendix 7.8, Table A7.5). However, the findings of S4 indicate group-specific response biases, e.g., by migration experience. Considering the lower average test performance of foreign-born L2 students, these students tend to report their German proficiency more accurately than those who were born in Germany. This difference is statistically significant after controlling for the explanatory variables (see Appendix 7.8, Table A7.4), which is consistent with the main analysis.

The results for the explanatory variables in Figure 7.3B demonstrate a relatively consistent pattern across the alternative specifications, with minor variations in the magnitude of the coefficients. Irrespective of the outcome measure, we find a negative coefficient for general cognitive abilities, positive coefficients for linguistic self-concept and self-esteem, and a slightly negative coefficient for the frame of reference, which is statistically not significant in most specifications. However, the coefficients for linguistic self-concept and self-esteem are notably smaller in the analysis of bias in self-reported test performance (S2) than in the main analysis and in most other alternatives that use the more abstract type of self-reported language proficiency. This finding may indicate that self-reports of performance in specific situations and tasks may be less susceptible to the influences of individuals' self-concepts and the biases associated with them.

Figure 7.3C displays the results of the conditional correlations of the explanatory variables according to general cognitive abilities and the corresponding interactions. The results consistently indicate positive conditional coefficients for linguistic self-concept, self-esteem, and frame of reference and negative coefficients for the interaction terms across all alternatives. These patterns largely reflect those of the main analysis, with variations in the magnitude and level of statistical significance of specific coefficients for certain alternatives. With respect to self-esteem and frame of reference, several of the sensitivity analyses yield coefficients that are not statistically significant ($p > .05$). These conditional associations are thus of limited robustness. In contrast, the results for the interaction between students' linguistic self-concept and their general cognitive abilities are highly robust across all alternatives.

Notably, the coefficients in analysis S4 are markedly lower than those in the main analysis and most of the other alternatives. The discrepancy between the analyses is

Figure 7.3: Summary of sensitivity analyses of the results regarding background characteristics (A), explanatory variables (B), and conditional correlations (C) using alternative operationalizations of overestimation bias



Notes: The full tables containing all models and coefficients are outlined in Appendix 7.8 (see Tables A7.1–A7.4). The coefficient plots were generated using the Stata ado “coefplot” by Jann (2014). Error bars indicate the 95% confidence intervals. Cog=general cognitive abilities, LSC=linguistic self-concept, SE=self-esteem, FoR=frame of reference, X indicates the interaction (multiplication) between two variables. Main: Results of the main analysis based on the standardized difference score between the composite self-report measure and the reading comprehension test scores (see Table 7.2; $N=1,630$). S1: Standardized difference score using the single self-report item assessing reading ability, rather than the composite measure (see Table A7.1; $N=1,586$). S2: Difference score between self-reported and actual number of items correctly solved on the test (see Table A7.2; $N=1,830$). S3: Standardized difference score analogous to the main analysis using a stricter criterion including only positive values (see Table A7.3; $N=1,215$). S4: Residualized difference score (standardized) based on the regression of self-reports on test scores (see Table A7.4; $N=1,696$).

likely attributable to a substantial covariation between general cognitive abilities and tested language proficiency, which is eliminated in the analysis based on the residual approach. This results in an overall reduced magnitude of the coefficients for general cognitive abilities in S4, which is also evident when the unconditional coefficients are examined (see Fig. 7.3B). Based on these findings, it can be assumed that parts of the biases associated with general cognitive abilities, particularly the interaction with other biasing factors, reflect variations in actual language proficiency and, thus, may account for the general tendency of overestimations among low performers.

Overall, the sensitivity analyses corroborate the findings of the main analysis related to the relevance of the explanatory variables and the interaction between general cognitive abilities and linguistic self-concept. However, the results for the background characteristics and the other interactions are somewhat inconsistent. The inconsistencies primarily pertain to the magnitude and the level of statistical significance of the coefficients rather than their direction. Apart from this, the patterns identified in the various analyses are largely analogous, indicating that the key tendencies are robust toward alternative specifications.

7.6 Discussion and conclusion

This study examines processes associated with bias in L2 students' self-reports of their proficiency in the majority language. Prior research has demonstrated that the use of self-reports as indicators of L2 learners' actual language proficiency can lead to biased findings and erroneous conclusions (e.g., Charette and Meng 1994; Edele et al. 2015; Finnie and Meng 2005). The most prevalent pattern of bias in self-reported language proficiency is characterized by overestimations among individuals with relatively low proficiency levels (e.g., Ross 1998; Schild 2024; Suzuki 2015). Our findings indicate that these tendencies of overestimation bias are the result of distorting influences associated with low general cognitive abilities, an exaggerated linguistic self-concept and self-esteem, and, to some extent, a weak frame of reference. We find that general cognitive abilities exert the greatest influence on overestimation bias and can moderate the distorting influences of self-concept and the frame of reference. This suggests that the task of evaluating one's own language proficiency presents a considerable cognitive challenge, particularly for students with relatively low general cognitive abilities. These students are likely to encounter difficulties in retrieving and judging relevant information during the response process and tend to rely on indicators that are easily accessible but also susceptible to bias, such as aspects of their self-concept and their frame of reference.

The identification of conditions that promote overestimation bias allows us to estimate the potential for bias in analyses that rely on self-reported language proficiency. The finding that respondents with low general cognitive abilities systematically tend to overestimate their language proficiency suggests that the actual correlation between language proficiency and cognitive abilities would be downwardly biased in analyses using self-reports. Similarly, the associations between language proficiency and aspects

of the respondents' self-concept are likely to be upwardly biased in such analyses. Even more importantly, these biases are not limited to correlations with the factors themselves but also extend to other variables that are confounded with them. For example, biases can be expected regarding educational outcomes that are correlated with general cognitive abilities, such as mathematical competences. Analyses based on self-reports are likely to underestimate the relationship between such ability-related variables and language proficiency. Moreover, the biases associated with respondents' self-concept could also manifest, for example, in other personality traits that covary with the linguistic self-concept and/or self-esteem and that have been the subject of prior research on second language learning (e.g., Asfar et al. 2019; Kosyakova and Liable 2024). In such cases, the use of self-reports tends to overestimate the associations between personality traits and language proficiency. Although the findings indicate the direction of biases, it is difficult to determine their extent. The extent may vary considerably, encompassing minor variations in coefficient sizes and more substantial discrepancies. Such errors may include the erroneous indication of the absence or opposition of correlations between variables of interest.

Moreover, the general tendency of overestimation among low performers can blur results regarding group differences, which frequently represent a central focus of research interest. For example, individuals from groups with relatively low levels of actual language proficiency are more inclined to overestimate their abilities; this may result in an aggregated level of self-reported language proficiency that is comparable to or even higher than that of groups exhibiting higher levels of actual proficiency. Previous research has identified such biases in differences between groups of ethnic origin (e.g., Edele et al. 2015). The sensitivity analysis of our study complements these findings by demonstrating that these biases primarily reflect differences in actual language proficiency between ethnic groups rather than group-specific response patterns that imply distinct underlying processes. However, our findings also indicate group-specific response patterns related to other background characteristics, namely, migration experience. Despite exhibiting a lower average L2 proficiency level, foreign-born students tend to provide more accurate self-reports than students from immigrant families who were born in Germany. One potential reason for this finding is that students who start learning German as an additional language at a later age may be more inclined to acknowledge potential deficiencies, as proficiency in that language is less crucial for maintaining a positive self-concept. Consequently, the distorting influences associated with self-concept should be less pronounced among them than among students who have learned German as a second language from an early age. Overall, our findings suggest that the interpretation of results based on self-reports as indicators of actual language proficiency should be approached with great caution, particularly when examining group differences.

In light of the potential biases arising from the use of self-reports as indicators of language proficiency, how future studies should address this issue to generate valid results is a pressing question. The most promising approach is obviously to employ objective test instruments. However, as testing is not always feasible due to restrictions on testing time and budget, it may be worthwhile to develop more reliable self-report

instruments. The literature suggests that self-reports of abstract language competences (e.g., “How well can you write English?”)—the type that is commonly employed in the field of integration research and that is also the focus of our analysis—are more susceptible to response bias than indicators assessing performance in specific tasks, such as “I can write a very simple personal letter, for example, thanking someone for something” (Brantmeier et al. 2012; Li and Zhang 2021; Ross 1998). One possible explanation for this is that evaluating one’s proficiency in a particular task requires less cognitive effort because task-specific questions provide a clear reference point, thus facilitating respondents’ retrieval and judgment processes. Considering the pivotal role that general cognitive abilities play in accurate responses, as demonstrated in our study, questions requiring minimal cognitive effort from respondents are less prone to bias. This assumption is corroborated by the results of our sensitivity analysis, which show that self-reports of test performance are less influenced by aspects of individuals’ self-concept and associated biases than self-reports of abstract language proficiency, even among those with low cognitive abilities. Another potential strategy for reducing bias, particularly frame of reference effects, in self-report measures is to employ comprehensive response categories that set a clear anchor at the endpoints. As an illustration, the range “not at all” to “on a native speaker level” could be used in lieu of the conventional “very poor” to “very well” response scale. However, further development and testing of these concepts is necessary before we can evaluate their potential to reduce biases.

The various sensitivity analyses, which employ different methodologies for assessing bias in self-reports overall, yield comparable findings, emphasizing the robustness of our findings. Nonetheless, the study is limited in several respects. The analyses are based on cross-sectional data, which limits the ability to draw causal conclusions. For example, we suggest that an inflated linguistic self-concept causes respondents to overestimate their language proficiency. However, it is also possible that the relationship is inverse, whereby respondents’ inability to accurately assess their language proficiency may distort their self-concept. While the findings appear to corroborate the theoretical assumptions, the possibility of reversed causality or an influence of third variables cannot be fully ruled out.

Moreover, the scope of our analyses is limited to self-reports of German proficiency among adolescent L2 learners. It is therefore unclear whether the findings can be generalized to other populations, such as adult immigrants or nonimmigrant populations, or to other languages and competence domains. The present study focused on L2 learners because the use of self-reports as an indicator of actual language proficiency is a popular method in research on integration processes. From a theoretical standpoint, however, the examined mechanisms are not specific to adolescents or the majority language. Rather, they should apply to response bias in ability self-reports in general. Consequently, it is likely that the underlying processes hold for other populations and competence domains. Nevertheless, the potential for bias in self-reports may vary depending on the specific population and domain of interest. For example, given the crucial role of cognitive abilities in the accuracy of self-reports, it might be expected that the use of self-reported language proficiency would yield more reliable

results when high-ability populations, such as university students, are studied than when lower-ability samples are studied. Using self-reports may also be less problematic in analyses of populations with personal migration experience, who, based on our findings, tend to provide more accurate self-reports than respondents from immigrant families who were born in the destination country. The potential for bias associated with self-reported language proficiency may be reduced, but not fully eliminated, in samples with such a favorable composition. Consequently, at least minor (systematic) variations in findings based on self-reports are still to be expected.

Finally, while our study indicates that self-reports are not accurate indicators of respondents' actual language proficiency, it is still reasonable to use these indicators in certain cases. Self-reports provide insight into individuals' perceptions of their language proficiency. Depending on the research objective, such subjective beliefs may be of greater interest than individuals' actual language proficiency. However, research on integration processes and ethnic inequalities thus far has typically used self-reports to indicate respondents' actual language proficiency, which is likely to produce biased results.

In conclusion, the study contributes to a more nuanced understanding of the processes that cause response bias in self-reports of language proficiency. The findings offer valuable insights that can help to evaluate the potential for bias in research findings based on such self-reports. Our findings further emphasize the need to improve self-report questions or apply language tests as indicators of language proficiency.

7.7 References

- Ackerman, P. L. and Wolman, S. D. (2007). Determinants and validity of self-estimates of abilities and self-concept measures. *Journal of Experimental Psychology: Applied*, 13(2):57–78.
- Aronson, E. (1968). Dissonance theory: Progress and problems. In Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., and Tannenbaum, P. H., editors, *Theories of cognitive consistency: A sourcebook*, pages 5–27. Chicago: Rand McNally.
- Asfar, D., Born, M. P., Oostrom, J. K., and Van Vugt, M. (2019). Psychological individual differences as predictors of refugees' local language proficiency. *European Journal of Social Psychology*, 49(7):1385–1400.
- Azzolini, D., Schnell, P., and Palmer, J. R. B. (2012). Educational achievement gaps between immigrant and native students in two “new” immigration countries: Italy and Spain in comparison. *The Annals of the American Academy of Political and Social Science*, 643(1):46–77.
- Beyer, S. and Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, 23(2):157–172.

7 When can we (not) rely on self-reported language proficiency?

- Blossfeld, H.-P. and Roßbach, H.-G. (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS)*. Wiesbaden: Springer VS.
- Bong, M. and Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1):1–40.
- Brantmeier, C., Vanderplank, R., and Strube, M. (2012). What about me?: Individual self-assessment by skill and level of language instruction. *System*, 40(1):144–160.
- Cannell, C. F., Miller, P. V., and Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, 12:389–437.
- Carter, T. J. and Dunning, D. (2008). Faulty self-assessment: Why evaluating one’s own competence is an intrinsically difficult task. *Social and Personality Psychology Compass*, 2(1):346–360.
- Charette, M. and Meng, R. (1994). Explaining language proficiency: Objective versus self-assessed measures of literacy. *Economics Letters*, 44(3):313–321.
- Chiswick, B. R. and Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economic*, 13(2):246–288.
- Critcher, C. R. and Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology*, 97(6):931–945.
- Cronbach, L. J. and Furby, L. (1970). How we should measure “change”: Or should we? *Psychological Bulletin*, 74(1):68–80.
- De Los Reyes, A. and Kazdin, A. E. (2004). Measuring informant discrepancies in clinical child research. *Psychological Assessment*, 16(3):330–334.
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.
- Dunning, D., Johnson, K., Ehrlinger, J., and Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3):83–87.
- Dustmann, C. and Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal*, 113(489):695–717.
- Edele, A., Seuring, J., Kristen, C., and Stanat, P. (2015). Why bother with testing? The validity of immigrants’ self-assessed language proficiency. *Social Science Research*, 52:99–123.
- Edwards, J. R. (1994). Regression analysis as an alternative to difference scores. *Journal of Management*, 20(3):683–689.

- Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, 4(3):265–287.
- Ehrlinger, J. (2008). Skill level, self-views and self-theories as sources of error in self-assessment. *Social and Personality Psychology Compass*, 2(1):382–398.
- Ehrlinger, J. and Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84(1):5–17.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., and Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105(1):98–121.
- Fang, J., Huang, X., Zhang, M., Huang, F., Li, Z., and Yuan, Q. (2018). The Big-Fish-Little-Pond effect on academic self-concept: A meta-analysis. *Frontiers in Psychology*, 9:1569.
- Farh, J.-L. and Dobbins, G. H. (1989). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology*, 42(4):835–850.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Finnie, R. and Meng, R. (2005). Literacy and labour market outcomes: Self-assessment versus test score measures. *Applied Economics*, 37(17):1935–1951.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10):906–911.
- Freund, P. A. and Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138(2):296–321.
- Gehrer, K., Zimmermann, S., Artelt, C., and Weinert, S. (2012). *The assessment of reading competence (including sample items for grade 5 and 9)*. *Scientific use file 2012, version 1.0.0*. Bamberg: University of Bamberg, National Educational Panel Study.
- Gerber, J. P., Wheeler, L., and Suls, J. (2018). A social comparison theory meta-analysis 60+ years on. *Psychological Bulletin*, 144(2):177–197.
- Haberkorn, K. and Pohl, S. (2013). *Cognitive basic skills – Data in the scientific use file*. Bamberg: University of Bamberg, National Educational Panel Study.

7 *When can we (not) rely on self-reported language proficiency?*

- Harmon-Jones, E. and Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, 38(1):7–16.
- Heath, A. F., Rothon, C., and Kilpi, E. (2008). The second generation in Western Europe: Education, unemployment, and occupational attainment. *Annual Review of Sociology*, 34(1):211–235.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge, Taylor & Francis.
- Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. University of California Press, Berkeley.
- Hübner, N., Wagner, W., Hochweber, J., Neumann, M., and Nagengast, B. (2020). Comparing apples and oranges: Curricular intensification reforms can change the meaning of students' grades! *Journal of Educational Psychology*, 112(1):204–220.
- Jacobs, K. E., Szer, D., and Roodenburg, J. (2012). The moderating effect of personality on the accuracy of self-estimates of intelligence. *Personality and Individual Differences*, 52(6):744–749.
- Jann, B. (2014). Plotting regression coefficients and other estimates. *The Stata Journal*, 14(4):708–737.
- John, O. P. and Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, 66(1):206–219.
- Kim, Y.-H., Kwon, H., Lee, J., and Chiu, C.-Y. (2016). Why do people overestimate or underestimate their abilities? A cross-culturally valid model of cognitive and motivational processes in self-assessment biases. *Journal of Cross-Cultural Psychology*, 47(9):1201–1216.
- Klein, D. (2022). MIMRGNS: Stata module to run margins after mi estimate. Statistical Software Components S457795, Boston College Department of Economics, revised 25 Jul 2022.
- Kosyakova, Y. and Laible, M.-C. (2024). Importance of personality traits for destination-language acquisition: Evidence for refugees in Germany. *International Migration Review*, 58(1):347–385.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3):213–236.
- Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C., and Conaway, M. (2002). The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66(3):371–403.

- Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6):1121–1134.
- Kwang, T. and Swann, W. B. (2010). Do people embrace praise even when they feel unworthy? A review of critical tests of self-enhancement versus self-verification. *Personality and Social Psychology Review*, 14(3):263–280.
- Kwon, H. and Linderholm, T. (2014). Effects of self-perception of reading skill on absolute accuracy of metacomprehension judgements. *Current Psychology*, 33(1):73–88.
- Li, M. and Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, 38(2):189–218.
- Loyalka, P., Zakharov, A., and Kuzmina, Y. (2018). Catching the big fish in the little pond effect: Evidence from 33 countries and regions. *Comparative Education Review*, 62(4):542–564.
- Mabe, P. A. and West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67(3):280–296.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23(1):129–149.
- Marsh, H. W. and Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47(1):213–231.
- Marsh, H. W., Trautwein, U., Lüdtke, O., and Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology*, 100(3):510–524.
- NEPS Network (2024). National Educational Panel Study, scientific use file of starting cohort grade 9. Bamberg: Leibniz Institute for Educational Trajectories (LIfBi).
- Neumann, M., Trautwein, U., and Nagy, G. (2011). Do central examinations lead to greater grading comparability? A study of frame-of-reference effects on the university entrance qualification in Germany. *Studies in Educational Evaluation*, 37(4):206–217.
- OECD (2023). *PISA 2022 results (volume I): The state of learning and equity in education*. Paris: OECD Publishing.
- Park, S. W., Tignor, S. M., Joo, M. J., and Heo, Y. H. (2016). Accuracy and bias in self-perception of performance: Narcissism matters in Korea as well. *Korean Social Science Journal*, 43(2):29–43.

7 When can we (not) rely on self-reported language proficiency?

- Paulhus, D. L. and John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives. *Journal of Personality*, 66(6):1025–1060.
- Pohl, S. and Carstensen, C. H. (2012). *NEPS technical report – Scaling the data of the competence tests (NEPS Working Paper 14)*. Bamberg: University of Bamberg, National Educational Panel Study.
- Raven, J. (2000). The Raven’s progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41(1):1–48.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton: Princeton University Press.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1):1–20.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Schild, A. (2024). *Selbsteinschätzungen sprachlicher Kompetenzen im Deutschen jugendlicher Flüchtlinge*. Bamberg: Otto-Friedrich-Universität.
- Schmaus, M. (2020). Ethnic differences in labour market outcomes—The role of language-based discrimination. *European Sociological Review*, 36(1):82–103.
- Shavelson, R. J., Hubner, J. J., and Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3):407–441.
- Skopek, J., Pink, S., and Bela, D. (2013). *Starting cohort 4: Grade 9 (SC4). Scientific use file, version 1.1.0. Data manual*. Bamberg: University of Bamberg, National Educational Panel Study.
- Stanat, P. and Edele, A. (2016). Language proficiency and the integration of immigrant students in the education system. In Scott, R. A. and Kosslyn, S. M., editors, *Emerging trends in the social and behavioral sciences*, pages 1–15. Hoboken: Wiley.
- StataCorp (2023). *Stata statistical software: Release 18*. College Station: StataCorp LLC.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, 21:261–302.
- Strack, F. and Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In Hippler, H.-J., Schwarz, N., and Sudman, S., editors, *Social information processing and survey methodology*, pages 123–148. New York: Springer.
- Strickhouser, J. E. and Zell, E. (2015). Self-evaluative effects of dimensional and social comparison. *Journal of Experimental Social Psychology*, 59:60–66.

- Sudman, S., Bradburn, N., and Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Suzuki, Y. (2015). Self-assessment of Japanese as a second language: The role of experiences in the naturalistic acquisition. *Language Testing*, 32(1):63–81.
- Swann, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. *Social Psychological Perspectives on the Self*, 2:33–66.
- Swann, W. B. and Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology*, 17(4):351–372.
- Thijs, J., Verkuyten, M., and Helmond, P. (2010). A further examination of the big-fish–little-pond effect: Perceived position in class, class size, and gender comparisons. *Sociology of Education*, 83(4):333–345.
- Tourangeau, R. and Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103(3):299–314.
- Tourangeau, R., Rips, L. J., and Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Truxillo, D. M., Seitz, R., and Bauer, T. N. (2008). The role of cognitive ability in self-efficacy and self-assessed test performance. *Journal of Applied Social Psychology*, 38(4):903–918.
- Von Hippel, P. T. (2007). 4. Regression with missing Ys: An improved strategy for analyzing multiply imputed data. *Sociological Methodology*, 37(1):83–117.
- Wells, L. E. and Sweeney, P. D. (1986). A test of three models of bias in self-assessment. *Social Psychology Quarterly*, 49(1):1–10.
- White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25.
- White, I. R., Royston, P., and Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399.
- Whitton, S. W., Larson, J. J., and Hauser, S. T. (2008). Depressive symptoms and bias in perceived social competence among young adults. *Journal of Clinical Psychology*, 64(7):791–805.
- Wohlkinger, F., Bayer, M., and Ditton, H. (2016). Measuring self-concept in the NEPS. In Blossfeld, H.-P., Von Maurice, J., Bayer, M., and Skopek, J., editors, *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*, pages 181–193. Wiesbaden: Springer VS.
- Zell, E. and Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9(2):111–125.

7 *When can we (not) rely on self-reported language proficiency?*

7.8 Appendix

Table A7.1: Sensitivity analysis (S1) using the single self-report item on reading ability (linear regression)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
<i>Ethnic group (ref. Turkey)</i>								
Poland	-0.13 (0.08)	-0.09 (0.07)	-0.12 (0.08)	-0.13 (0.08)	+ (0.07)	-0.09 (0.07)	-0.08 (0.07)	-0.08 (0.07)
Former Soviet Union	-0.15 (0.05)	** (0.05)	-0.15 (0.05)	-0.15 (0.05)	** (0.05)	-0.08 (0.05)	-0.08 (0.05)	-0.08 (0.05)
Former Yugoslavia	0.02 (0.07)	0.01 (0.07)	0.00 (0.07)	0.02 (0.07)	-0.00 (0.07)	-0.00 (0.07)	-0.00 (0.07)	-0.00 (0.07)
Other	-0.14 (0.05)	** (0.05)	-0.15 (0.05)	-0.15 (0.05)	** (0.05)	-0.11 (0.05)	-0.11 (0.05)	-0.11 (0.05)
Born abroad	-0.01 (0.04)	-0.04 (0.04)	-0.00 (0.04)	-0.01 (0.04)	-0.04 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)
Female	0.01 (0.04)	-0.03 (0.04)	0.00 (0.04)	0.00 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)
<i>School track (ref. low)</i>								
Intermediate	-0.28 (0.04)	*** (0.04)	-0.29 (0.04)	-0.28 (0.04)	*** (0.04)	-0.22 (0.04)	-0.21 (0.04)	-0.22 (0.04)
High	-0.53 (0.05)	*** (0.05)	-0.55 (0.05)	-0.53 (0.05)	*** (0.05)	-0.41 (0.05)	-0.40 (0.05)	-0.42 (0.05)
Integrated	-0.22 (0.08)	** (0.07)	-0.24 (0.08)	-0.22 (0.08)	** (0.07)	-0.16 (0.07)	-0.17 (0.07)	-0.16 (0.07)
<i>General cognitive abilities (Cog)</i>								
General cognitive abilities (Cog)	-0.78 (0.09)	*** (0.09)	-0.75 (0.09)	-0.75 (0.09)	*** (0.09)	0.02 (0.25)	-0.37 (0.35)	0.13 (0.31)
Linguistic self-concept (LSC)			0.28 (0.09)		** (0.11)	1.04 (0.26)	0.29 (0.11)	0.32 (0.11)
Self-esteem (SE)			0.29 (0.12)	*	0.25 (0.12)	0.25 (0.12)	0.59 (0.35)	0.24 (0.12)
Frame of reference (FoR)				0.12 (0.13)	-0.13 (0.15)	-0.10 (0.15)	-0.13 (0.15)	1.01 (0.40)
Cog X LSC						-1.20 (0.35)		
Cog X SE							-0.54 (0.48)	
Cog X FoR								-1.82 (0.57)
Constant	0.95 (0.05)	*** (0.07)	1.37 (0.10)	*** (0.08)	0.57 (0.11)	*** (0.19)	1.06 (0.25)	0.82 (0.22)
R ²	0.09	0.14	0.11	0.09	0.15	0.15	0.15	0.15

Notes: Robust standard errors in parentheses; $N=1,586$. Cog=general cognitive abilities, LSC=linguistic self-concept, SE=self-esteem, FoR=frame of reference, X indicates the interaction (multiplication) between two variables. + $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A7.2: Sensitivity analysis (S2) using the self-reported test performance (linear regression)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Ethnic group (<i>ref. Turkey</i>)								
Poland	*** (0.06)	*** (0.06)	*** (0.06)	*** (0.06)	*** (0.06)	*** (0.06)	*** (0.06)	*** (0.06)
Former Soviet Union	*** (0.06)	* (0.06)	*** (0.06)	*** (0.06)	* (0.06)	*	*	+
Former Yugoslavia	*** (0.07)	*** (0.07)	*** (0.07)	*** (0.07)	*** (0.07)	*** (0.07)	*** (0.07)	*** (0.07)
Other	** (0.05)	* (0.05)	** (0.05)	** (0.05)	+	*	+	*
Born abroad	0.03 (0.04)	0.01 (0.04)	0.04 (0.04)	0.03 (0.04)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)	0.01 (0.04)
Female	** (0.04)	** (0.04)	** (0.04)	** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)
School track (<i>ref. low</i>)								
Intermediate	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)
High	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)	*** (0.04)
Integrated	* (0.04)	+	*	*	+	+	+	+
General cognitive abilities (Cog)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)
Linguistic self-concept (LSC)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)	(0.23)	(0.43)	(0.33)
Self-esteem (SE)	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.26)	(0.10)	(0.10)
Frame of reference (FoR)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.39)	(0.12)
Cog X LSC	(0.13)	(0.15)	(0.13)	(0.15)	(0.15)	(0.15)	(0.15)	(0.44)
Cog X SE	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)	(0.16)
Cog X FoR	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
Constant	1.28 (0.05)	1.74 (0.07)	1.12 (0.11)	1.37 (0.08)	1.74 (0.13)	1.35 (0.20)	1.64 (0.31)	1.15 (0.25)
R ²	0.12	0.16	0.12	0.12	0.17	0.17	0.17	0.17

Notes: Robust standard errors in parentheses; N=1,830. Cog=general cognitive abilities, LSC=linguistic self-concept, SE=self-esteem, FoR=frame of reference, X indicates the interaction (multiplication) between two variables. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table A7.3: Sensitivity analysis (S3) using a stricter criterion to define the analysis sample (linear regression)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Ethnic group (<i>ref. Turkey</i>)								
Poland	-0.06 (0.08)	-0.04 (0.07)	-0.05 (0.08)	-0.06 (0.08)	-0.03 (0.07)	-0.05 (0.07)	-0.03 (0.07)	-0.04 (0.07)
Former Soviet Union	-0.02 (0.06)	0.04 (0.06)	-0.01 (0.06)	-0.02 (0.06)	0.04 (0.05)	0.04 (0.05)	0.04 (0.05)	0.05 (0.05)
Former Yugoslavia	0.07 (0.07)	0.07 (0.07)	0.05 (0.07)	0.07 (0.07)	0.05 (0.07)	0.05 (0.07)	0.05 (0.07)	0.05 (0.07)
Other	-0.04 (0.05)	-0.01 (0.04)	-0.04 (0.04)	-0.04 (0.05)	-0.02 (0.04)	-0.02 (0.04)	-0.02 (0.04)	-0.02 (0.04)
Born abroad	-0.02 (0.04)	-0.04 (0.04)	-0.01 (0.04)	-0.02 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)
Female	-0.05 (0.04)	-0.09 (0.04)	* (0.04)	-0.05 (0.04)	-0.08 (0.04)	-0.08 (0.04)	-0.09 (0.04)	* (0.04)
School track (<i>ref. low</i>)								
Intermediate	-0.23 (0.04)	-0.16 (0.04)	*** (0.04)	-0.23 (0.04)	-0.18 (0.04)	-0.18 (0.04)	-0.17 (0.04)	*** (0.04)
High	-0.35 (0.05)	-0.23 (0.05)	*** (0.05)	-0.34 (0.05)	-0.26 (0.05)	-0.26 (0.05)	-0.25 (0.05)	*** (0.05)
Integrated	-0.17 (0.09)	-0.12 (0.08)	* (0.09)	-0.17 (0.09)	-0.13 (0.09)	-0.13 (0.09)	-0.14 (0.09)	-0.12 (0.09)
General cognitive abilities (Cog)								
		-0.64 (0.09)	*** (0.09)		-0.61 (0.09)	0.19 (0.28)	0.01 (0.35)	-0.02 (0.31)
Linguistic self-concept (LSC)								
			** (0.09)		0.32 (0.11)	1.05 (0.28)	0.32 (0.11)	0.34 (0.11)
Self-esteem (SE)			** (0.12)		0.28 (0.11)	0.28 (0.11)	0.81 (0.33)	* (0.11)
Frame of reference (FoR)								
				0.12 (0.13)	-0.17 (0.16)	-0.13 (0.15)	-0.17 (0.15)	0.56 (0.40)
Cog X LSC								
						-1.21 (0.39)		
Cog X SE								
							-0.86 (0.47)	+
Cog X FoR								
								-1.21 (0.60)
Constant	1.09 (0.05)	1.43 (0.07)	*** (0.10)	1.04 (0.08)	1.10 (0.11)	*** (0.20)	0.72 (0.25)	*** (0.21)
R ²	0.05	0.09	0.07	0.05	0.11	0.12	0.11	0.11

Notes: Robust standard errors in parentheses; $N=1,215$. Cog=general cognitive abilities, LSC=linguistic self-concept, SE=self-esteem, FoR=frame of reference, X indicates the interaction (multiplication) between two variables. + $p<0.10$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Table A7.4: Sensitivity analysis (S4) using the residualized difference score (linear regression)

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
<i>Ethnic group (ref. Turkey)</i>								
Poland	0.00 (0.04)	0.03 (0.04)	0.01 (0.04)	0.00 (0.04)	0.03 (0.04)	0.03 (0.04)	0.03 (0.04)	0.03 (0.04)
Former Soviet Union	0.06 (0.04)	0.10 (0.04)	** (0.08)	0.06 (0.04)	0.11 (0.03)	0.11 (0.03)	0.11 (0.03)	0.11 (0.03)
Former Yugoslavia	0.10 (0.04)	0.10 (0.04)	* (0.08)	0.10 (0.04)	0.09 (0.04)	0.08 (0.04)	0.08 (0.04)	0.08 (0.04)
Other	0.03 (0.03)	0.05 (0.03)	0.02 (0.03)	0.02 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)
Born abroad	-0.05 (0.03)	-0.07 (0.03)	* (0.05)	-0.05 (0.03)	-0.06 (0.03)	-0.06 (0.03)	-0.06 (0.03)	-0.06 (0.03)
Female	0.00 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)	-0.03 (0.02)
<i>School track (ref. low)</i>								
Intermediate	-0.03 (0.03)	0.01 (0.03)	-0.04 (0.03)	-0.03 (0.03)	0.00 (0.03)	0.00 (0.03)	0.01 (0.03)	0.00 (0.03)
High	-0.12 (0.03)	-0.03 (0.03)	*** (0.13)	-0.11 (0.03)	-0.04 (0.03)	-0.04 (0.03)	-0.04 (0.03)	-0.05 (0.03)
Integrated	-0.04 (0.05)	0.01 (0.04)	-0.05 (0.05)	-0.04 (0.05)	-0.01 (0.05)	-0.01 (0.05)	-0.01 (0.05)	-0.01 (0.05)
<i>General cognitive abilities (Cog)</i>								
Linguistic self-concept (LSC)		-0.42 (0.05)	*** (0.41)		-0.40 (0.05)	-0.04 (0.17)	-0.29 (0.18)	-0.21 (0.19)
Self-esteem (SE)			*** (0.06)		*** (0.39)	*** (0.75)	*** (0.39)	*** (0.40)
Frame of reference (FoR)			* (0.13)		*(0.11)	*(0.16)	*(0.22)	*(0.11)
Cog X LSC								
Cog X SE								
Cog X FoR								
Constant	0.52 (0.03)	*** (0.76)	** (0.18)	*** (0.38)	*** (0.42)	*** (0.17)	*(0.34)	*(0.32)
R ²	0.02	0.05	0.06	0.03	0.09	0.10	0.09	0.09

Notes: Robust standard errors in parentheses; N=1,696. Cog=general cognitive abilities, LSC=linguistic self-concept, SE=self-esteem, FoR=frame of reference, X indicates the interaction (multiplication) between two variables. + p<0.10, * p<0.05, ** p<0.01, *** p<0.001.

Table A7.5: Average tested and self-reported German language proficiency by students' background characteristics (means & standard deviations)

	Test scores	Self-reports
Ethnic group		
Turkey	-0.93 (1.08)	4.52 (0.51)
Poland	-0.25 (1.16)	4.62 (0.52)
Former Soviet Union	-0.44 (1.13)	4.59 (0.54)
Former Yugoslavia	-0.64 (1.21)	4.62 (0.51)
Other	-0.33 (1.25)	4.61 (0.52)
Immigrant generation		
Born abroad	-0.67 (1.13)	4.45 (0.58)
Born in Germany	-0.48 (1.22)	4.64 (0.49)
Gender		
Female	-0.40 (1.12)	4.64 (0.50)
Male	-0.67 (1.22)	4.53 (0.54)
School track		
Low	-1.17 (0.98)	4.48 (0.55)
Intermediate	-0.39 (0.98)	4.59 (0.54)
High	0.40 (1.10)	4.76 (0.37)
Integrated	-0.62 (1.14)	4.56 (0.55)
Total	-0.53 (1.20)	4.59 (0.52)

Notes: The analysis is based on the subsample of students who had learned a foreign language within their family, and for whom valid data on both self-reported and tested German language proficiency are available ($N=2,350$). The differences in the means between the various background characteristics in the analysis sample ($N=1,630$) demonstrate an analogous pattern.

8 Gesamtdiskussion und Fazit

Die kompetente Beherrschung der Sprache des Aufnahmelandes ist eine Schlüsselqualifikation für den Bildungserfolg von Kindern und Jugendlichen mit Migrationshintergrund. Kinder, die mit einer anderen Familiensprache aufwachsen und zuhause kaum Kontakt zur Unterrichtssprache haben, starten ihre Bildungskarriere oftmals mit ungünstigen Voraussetzungen. Dementsprechend wird im Zusammenhang mit sprachlich bedingten Bildungsungleichheiten zwischen Schüler:innen mit und ohne Migrationshintergrund – auch noch in den späteren Bildungsetappen – der Gebrauch der Herkunftssprache innerhalb von zugewanderten Familien immer wieder als limitierende Bedingung diskutiert.

Anknüpfend an diese Debatte wurde in der vorliegenden Dissertation der Frage nachgegangen, inwiefern – zusätzlich zur Kommunikation innerhalb der Familie – der Gebrauch der Herkunftssprache (L1) auch in anderen alltagsrelevanten Kontexten den Kompetenzerwerb in der Sprache des Aufnahmelandes (L2) von Jugendlichen mit Migrationshintergrund bedingt und zur Erklärung ethnischer Leistungsdisparitäten beiträgt. Da die Sprachkompetenzen von Personen mit Migrationshintergrund in zahlreichen Studien ausschließlich über subjektive Einschätzungen der Befragten erfasst werden, wurde in der Dissertation darüber hinaus die Frage nach der Verlässlichkeit von Selbsteinschätzungen der Sprachkompetenz und den darauf basierenden Forschungsergebnissen aufgegriffen. Nachfolgend werden die zentralen Befunde zu diesen beiden Themenkomplexen zusammengefasst und eingeordnet.

Zur Bedeutung des außerfamiliären Sprachgebrauchs für den Zweitspracherwerb

Als Ausgangspunkt für die empirische Analyse wurde zunächst ein deskriptiver Überblick über die Verbreitung des Gebrauchs der Herkunftssprache in verschiedenen Kontexten gegeben. Für die untersuchten Schüler:innen der neunten Klassenstufe zeigen die Ergebnisse, dass im Vergleich zum Familienkontext, in dem die Mehrheit der Jugendlichen mit den Eltern häufig in der Herkunftssprache kommuniziert, der L1-Gebrauch in anderen Kontexten weniger verbreitet ist und beispielsweise mit Freund:innen und Mitschüler:innen meist Deutsch gesprochen wird. Auch Medien (z.B. TV, Bücher, Inhalte im Internet) werden überwiegend auf Deutsch konsumiert. Obwohl die L1 gegenüber der L2 für den alltäglichen Gebrauch quantitativ einen geringeren Stellenwert einnimmt, verwenden je nach Kontext dennoch zwischen 15 und 55 Prozent der Jugendlichen die L1 regelmäßig, wenn auch nur manchmal. Der Alltag vieler Jugendlicher ist folglich durch einen bilingualen Sprachgebrauch geprägt. Dies gilt auch für den familiären Kontext. Zwar gibt es Familien, die ausschließlich die Herkunftssprache zuhause sprechen,

jedoch stellt dies – anders als manche politische und mediale Äußerungen den Anschein erwecken mögen – eher die Ausnahme als die Regel dar. Im Generationenverlauf lässt sich über alle Kontexte hinweg eine zunehmende Verwendung der deutschen Sprache im Vergleich zur Herkunftssprache beobachten. Diese Tendenz entspricht den Erwartungen der Assimilationstheorie (Alba und Nee 1997) und deckt sich mit den Ergebnissen anderer Studien (z.B. Alba et al. 2002; Segeritz et al. 2010). Im Vergleich zwischen den Herkunftsgruppen fallen die Muster des Sprachgebrauchs sehr unterschiedlich aus. Insbesondere unter Jugendlichen türkischer Abstammung sowie unter denjenigen aus der ehemaligen Sowjetunion ist der Gebrauch der Erstsprache innerhalb und außerhalb der Familie verbreitet. Dies könnte darauf zurückzuführen sein, dass diese beiden Gruppen das Zuwanderungsgeschehen nach Deutschland maßgeblich geprägt haben (vgl. Olczyk et al. 2016) und sich dabei vielerorts entsprechende Gelegenheitsstrukturen zur Verwendung der L1 auch außerhalb der Familie etabliert haben.

Mit Blick auf den Zusammenhang zwischen dem Sprachgebrauch und den L2-Kompetenzen stellen sich in den Ergebnissen – neben der Familiensprache – die Kommunikation im Freundeskreis und mit Mitschüler:innen als bedeutsam heraus. Jugendliche, die sowohl zuhause als auch mit Gleichaltrigen die Herkunftssprache verwenden, erzielen im Schnitt geringere Leistungen in den Tests zur deutschen Sprachkompetenz (Wortschatz und Leseverstehen) als diejenigen, die außerhalb der Familie ausschließlich Deutsch sprechen. Vergleichbare Tendenzen können in den Analysen insbesondere auch für den Sprachgebrauch im schulischen Kontext bestätigt werden. Die Befunde deuten darauf hin, dass nachteilige Bedingungen, die mit der ethnischen bzw. ethno-lingualen Klassenkomposition in Zusammenhang stehen, unter anderem durch den L1-Gebrauch zwischen Klassenkamerad:innen vermittelt werden. Damit liefern die Ergebnisse der Dissertation erste empirische Hinweise auf die in der Literatur vielfach diskutierte Annahme, dass eine hohe ethnische bzw. ethno-linguale Konzentration in der Schule für Jugendliche mit Migrationshintergrund Gelegenheitsstrukturen und Anreize schafft untereinander die L1 zu verwenden, was den Zugang zur L2 und damit den Erwerb entsprechender Kompetenzen einschränkt (z.B. Driessen 2002; Eksner und Stanat 2011; Esser 2006; Peetsma et al. 2006; Van Ewijk und Sleegers 2010). In den Ergebnissen wird aber auch deutlich, dass die ethnische bzw. ethno-linguale Klassenzusammensetzung insgesamt nur vergleichsweise schwach mit den L2-Kompetenzen der Jugendlichen zusammenhängt und andere Merkmale des Schulkontextes, wie die soziale und kognitive Komposition, weitaus bedeutender sind. Dies deckt sich weitgehend auch mit der allgemeinen Einschätzung der bisherigen Befundlage (vgl. Eksner und Stanat 2011; Van Ewijk und Sleegers 2010). Demnach sind substantielle Nachteile nur für Schulen bzw. Klassen mit einer außergewöhnlich hohen Konzentration von Schüler:innen mit derselben Herkunftssprache zu erwarten (z.B. Walter 2009a). Die ethnische Klassenkomposition und die damit einhergehenden Gelegenheiten zum L1-Gebrauch unter Schüler:innen mit Migrationshintergrund tragen folglich nur in geringem Umfang zur Genese systematischer ethnischer Bildungsungleichheiten bei.

Im Vergleich zum familiären Kontext kommt dem außerfamiliären Sprachgebrauch insgesamt eine nachgelagerte Bedeutung zu. Die Ergebnisse zeigen zwar auf, dass die mit dem familiären L1-Gebrauch verbundenen Nachteile beim L2-Erwerb durch den

zunehmenden Kontakt zur L2 im Freundeskreis und in der Schule teilweise reduziert werden können, vollständig kompensieren lassen sich diese jedoch nicht. Damit betonen die Befunde der vorliegenden Dissertation einmal mehr den zentralen Stellenwert der Familiensprache. Gerade der Sprachgebrauch zwischen Eltern und Kind ist von entscheidender Bedeutung, auch weil dieser den Sprachgebrauch insgesamt vorstrukturieren kann: Wird mit den Eltern ausschließlich Deutsch gesprochen, dann in der Regel auch in allen anderen Kontexten. Nur wenn Kinder innerhalb ihrer Familie mit der L1 aufwachsen und diese auch regelmäßig nutzen, ergibt sich überhaupt erst die Möglichkeit diese auch mit Geschwistern und in anderen Kontexten außerhalb der Familie – bei entsprechenden Gelegenheitsstrukturen – zu verwenden. Es ist anzunehmen, dass der Sprachgebrauch innerhalb der Familie auch maßgeblich davon abhängt, wie gut die Eltern die L2, gegenüber der L1, beherrschen. Zum Beispiel gibt es aus dem US-amerikanischen Kontext Hinweise darauf, dass der L1-Gebrauch in der Familie besonders relevant ist, solange die Eltern nur über geringe Kompetenzen in der L2 verfügen und eine Verständigung darüber nur sehr eingeschränkt möglich ist (Mouw und Xie 1999). Mit zunehmender Beherrschung der L2 seitens der Eltern lösen sich diese Tendenzen auf. Die Befunde der Dissertation verweisen in Übereinstimmung mit früheren Studien (z.B. Baumert und Schümer 2001; Edele und Stanat 2016; Kristen 2008b; Stanat et al. 2010a; Walter 2009b) ebenfalls auf die zentrale Rolle der Eltern und des Sprachgebrauchs zuhause für den L2-Erwerb der Jugendlichen. Die in der Familie gesprochene Sprache bestimmt dabei maßgeblich den Zugang zur L2 sowie die Motivation, diese zu verwenden und zu lernen. Der Sprachgebrauch in anderen Kontexten hingegen trägt darüber hinaus nur in geringerem Umfang zum L2-Erwerb bei. Demzufolge kann angenommen werden, dass sprachlich bedingte Bildungsungleichheiten zwischen Jugendlichen mit und ohne Migrationshintergrund besonders durch den Sprachgebrauch *innerhalb* der Familie, und weniger durch den *außerfamiliären* Sprachgebrauch, geprägt sind.

Die aus den Ergebnissen der Dissertation gezogenen Schlussfolgerungen beziehen sich vorrangig auf die untersuchte Population von *Jugendlichen mit Migrationshintergrund* sowie auf die Bedeutung des außerfamiliären Sprachgebrauchs für die *Zweitsprachkompetenz* (Deutsch). Darüber hinaus lassen sich ausgehend von den grundlegenden Zusammenhängen aber auch Überlegungen anstellen, inwiefern die Befunde auf andere Populationen und/oder Anwendungsbereiche übertragen werden können:

Nach Erkenntnissen der Bildungsforschung ist der familiäre Hintergrund besonders in der frühen Bildungsphase prägend, verliert aber mit zunehmendem Alter an Bedeutung (z.B. Hillmert und Jacob 2005). Gleichzeitig gewinnen Peers im Jugendalter zunehmend an Bedeutung (Zander et al. 2017). Bezogen auf den Anwendungsfall kann demnach angenommen werden, dass der Sprachgebrauch innerhalb der Familie vor allem in jungen Jahren entscheidend für den Zweitspracherwerb ist, während der relative Einfluss der außerfamiliären Kommunikation (z.B. im Freundeskreis) im späteren Verlauf zunimmt. Befunde aus Untersuchungen zum Zweitspracherwerb von neu nach Deutschland zugewanderten Erwachsenen deuten ebenfalls darauf hin, dass der Sprachgebrauch außerhalb der Familie sowie die Sprache der Mediennutzung ganz besonders ausschlaggebend sind (z.B. Kosyakova et al. 2022; Kristen et al. 2016; Kristen und Seu-

ring 2021). Die unterschiedlichen Befundmuster zum Stellenwert des außerfamiliären Sprachgebrauchs könnten neben dem Alter aber auch mit dem Generationenstatus der jeweiligen Untersuchungspopulationen zusammenhängen. Die Mehrheit der in der Dissertation untersuchten Jugendlichen gehört zur zweiten und dritten Zuwanderungsgeneration, d.h. sie sind in Deutschland geboren und aufgewachsen. Es kann angenommen werden, dass Kinder, die in Deutschland aufwachsen, in der Regel bereits von klein auf ein Mindestmaß an Kontakt zur deutschen Sprache haben. Falls dies nicht innerhalb der Familie der Fall ist, dann zumindest im alltäglichen Umfeld, wie beispielsweise in der Kindertagesstätte. Spätestens jedoch mit Schuleintritt wird der Kontakt zur deutschen Sprache im Rahmen des Unterrichts intensiviert. Im Gegensatz dazu beginnen selbst zugewanderte Kinder und Jugendliche den Zweitspracherwerb im Aufnahmeland meist ohne jegliche Vorkenntnisse und vorherigen Kontakt zur deutschen Sprache. Die entsprechenden Strukturen und Gelegenheiten für den sprachlichen Zugang müssen erst geschaffen werden. Gerade in der frühen Phase des Zweitspracherwerbs könnten die Kommunikation mit Gleichaltrigen sowie die Nutzung von Medien entscheidend sein, um mit der neuen Sprache in Kontakt zu kommen und grundlegende Kompetenzen aufzubauen. Dieser Argumentation folgend sollte der außerfamiliäre L2-Gebrauch beispielsweise auch für die Integrationsverläufe von geflüchteten Kindern und Jugendlichen, die das aktuelle Zuwanderungsgeschehen maßgeblich prägen, in besonderer Weise förderlich sein.

Weiterhin deuten die Ergebnisse zum Sprachgebrauch im schulischen Kontext darauf hin, dass die Verwendung der L1 mit Mitschüler:innen die Kompetenzen in der Herkunftssprache der Jugendlichen begünstigen kann. Der regelmäßige L1-Gebrauch, auch außerhalb der Familie, scheint für den Erhalt und Ausbau von Kompetenzen in der Herkunftssprache eine wichtige Rolle zu spielen (vgl. Kristen et al. 2019). In diesem Zusammenhang werden oft auch potentielle Vorteile einer kompetenten Beherrschung der Herkunftssprache sowie deren aktiven Verwendung für den Bildungserfolg diskutiert (vgl. Edele et al. 2023). Die empirische Befundlage hierzu ist allerdings uneindeutig, wobei sich hinsichtlich des L1-Gebrauchs die Tendenz abzeichnet, dass die mit dem eingeschränkten Zugang zur L2 verbundenen Nachteile – gegenüber potentiell simultan auftretenden Vorteilen – überwiegen (Strobel 2016). Ungeachtet der umstrittenen Frage nach einer bildungsförderlichen Wirkung der Herkunftssprache, kann die Beherrschung der L1 für Jugendliche mit Migrationshintergrund einen Wert an sich darstellen. Zur Unterstützung des Erwerbs und Erhalts der Herkunftssprache ist es förderlich, Gelegenheiten zu schaffen bzw. zuzulassen, in welchen die Jugendlichen auch außerhalb der Familie ihre Herkunftssprache verwenden können, beispielsweise im schulischen Umfeld.

Im Rahmen der Dissertation konnten nicht alle potentiell bedeutsamen Aspekte hinsichtlich der Ausprägungen des Sprachgebrauchs in der nötigen Tiefe behandelt und untersucht werden. In der vorliegenden Arbeit wurde der *Erhalt* der Herkunftssprache und deren Verwendung in unterschiedlichen Kontexten in den Vordergrund gestellt. Dementsprechend wurde im Zuge der Analysen danach unterschieden, ob zumindest manchmal – also überhaupt – die Herkunftssprache oder ausschließlich Deutsch in unterschiedlichen Kontexten verwendet wird. Für den Prozess des Spracherwerbs ist

außerdem wichtig, in welchem zeitlichen Umfang und in welcher Intensität der sprachliche Zugang stattfindet (vgl. Chiswick und Miller 1995, 2001). Demzufolge ist der förderliche Effekt des Zugangs umso stärker, je häufiger und länger ein Individuum der jeweiligen Sprache ausgesetzt ist und je intensiver dieser Kontakt ist. Ausgehend von dieser Argumentation lassen sich weitere, noch spezifischere Prognosen zur Wirkungsweise des Sprachgebrauchs in Abhängigkeit des Umfangs und der Intensität anstellen, auch hinsichtlich des Zusammenspiels der verschiedenen Kontexte. So könnte die relative Bedeutung des außerfamiliären L2-Gebrauchs höher ausfallen, wenn zuhause nur in geringem Umfang oder gar nicht Deutsch gesprochen wird, weil der förderliche Zugang zur L2 dann – wenn überhaupt – nur außerhalb der Familie besteht. Weiterhin ist anzunehmen, dass die Intensität des Inputs von der Sprachkompetenz der Interaktionspartner oder beim Medienkonsum von der sprachlichen Qualität der Inhalte bestimmt wird. Sprechen die Eltern beispielsweise nur sehr schlecht Deutsch, sind selbst bei einem häufigen L2-Gebrauch innerhalb der Familie nur eingeschränkte Vorteile für die L2-Kompetenzen der Kinder zu erwarten. In diesem Fall sollten die Kinder wiederum besonders vom L2-Gebrauch außerhalb der Familie profitieren, und noch mehr, wenn der sprachliche Kontakt dort möglichst anspruchsvoll ist. Hinsichtlich des Sprachgebrauchs zuhause gibt es zwar Hinweise darauf, dass sowohl der Umfang des L2-Inputs (z.B. Scheele et al. 2010) als auch die L2-Kompetenzen der Eltern (z.B. Becker 2011) den L2-Erwerb der Kinder maßgeblich beeinflussen, die Interaktion zwischen Quantität und Qualität des sprachlichen Zugangs sowie zwischen verschiedenen Kontexten hingegen blieb in der bisherigen Forschung weitestgehend unberücksichtigt. Auch in den empirischen Analysen der vorliegenden Dissertation wurden diese Zusammenhänge nicht näher adressiert. Damit bleibt es zukünftigen Forschungsarbeiten vorbehalten diese spezifischen Wirkungsweisen des Sprachgebrauchs genauer zu analysieren und das darin steckende Potential – auch zur Erklärung bestehender Gruppenunterschiede (z.B. zwischen Zuwanderungsgenerationen) – aufzudecken.

Zur Validität subjektiver Einschätzungen von Sprachkompetenzen

Im Rahmen der zweiten Fragestellung der Dissertation wurde die Validität subjektiver Einschätzungen von Sprachkompetenzen – als Indikator für die tatsächlichen Sprachkompetenzen der Befragten – untersucht. Zur Beantwortung der Frage erfolgte zunächst eine systematische Zusammenstellung und Einschätzung des vergleichsweise umfassenden Forschungsstands zum Zusammenhang zwischen subjektiven und objektiven gemessenen Sprachkompetenzen. Daraus geht hervor, dass die bisherigen Studien mit Blick auf die Untersuchungsanlagen sowie die Befundlage sehr heterogen ausfallen und sich in den meisten Fällen eher moderate Korrelation abzeichnen. Diese Einschätzung deckt sich auch mit den Ergebnissen aus einer früheren und einer neueren, zwischenzeitlich erschienenen Meta-Studie dazu (Li und Zhang 2021; Ross 1998). Der zentrale Befund einer moderaten Korrelation zwischen subjektiver und objektiver Sprachkompetenzen kann darüber hinaus auch in der empirischen Analyse der Dissertation bestätigt werden, wobei die Korrelationen bei den untersuchten Jugendlichen mit Migrationshintergrund, je nach Herkunftsgruppe und Sprache (L1

oder L2), mit 0,20 bis 0,44 vergleichsweise gering ausfallen. Der besondere Beitrag der Arbeit besteht aber darin, die Verlässlichkeit von selbsteingeschätzten gegenüber getesteten Sprachkompetenzen auch multivariat in exemplarischen Analysemodellen mit Anwendungsbezug zu untersuchen. Frühere Untersuchungen mit einem ähnlichem Ansatz weisen vereinzelt darauf hin, dass die Verwendung von subjektiven Sprachkompetenzen mit systematischen Verzerrungen einhergehen kann (z.B. Charette und Meng 1994; Finnie und Meng 2005). Die Analysen der vorliegenden Arbeit zeigen ebenfalls, dass bei einer Verwendung von subjektiven Sprachkompetenzen viele Zusammenhänge zwar in der Tendenz korrekt wiedergegeben werden, gleichzeitig aber auch einige zentrale Ergebnisse systematisch verzerrt werden. Als besonders kritisch einzustufen sind unter anderem Abweichungen in den Befunden zu sprachlichen Kompetenzunterschieden zwischen verschiedenen Herkunftsgruppen und Zuwanderungsgenerationen, welche objektiv vorliegen und oft im Zentrum des Forschungsinteresses stehen, aber anhand der subjektiven Messung nicht bzw. teilweise gegenläufig dargestellt werden. Die Ergebnisse aus den exemplarischen Modellen deuten demnach darauf hin, dass Selbsteinschätzungen der Befragten oftmals systematischen Verzerrungen unterliegen, die sich dann auch in den darauf aufbauenden Analysen widerspiegeln und letztendlich zu fehlgeleiteten Rückschlüssen aus den Forschungsbefunden führen können.

In einem weiteren Analyseschritt wurden die Bestimmungsfaktoren, die verzerrten Selbsteinschätzungen zugrunde liegen, näher beleuchtet. Die Ergebnisse zeigen, dass die Antwortverzerrungen mit den kognitiven Fähigkeiten der Befragten, ihrem Selbstbild und ihrem sprachlichen Referenzrahmen zusammenhängen. Jugendliche mit vergleichsweise geringen kognitiven Fähigkeiten, einem hohen akademischen Selbstkonzept, einem hohen allgemeinen Selbstbewusstsein und mit vielen – im Vergleich zur eigenen Leistung – leistungsschwächeren Mitschüler:innen tendieren dazu, ihre Sprachkompetenzen im Deutschen systematisch zu überschätzen. Als besonders relevante Größe erweisen sich in den Analysen die kognitiven Fähigkeiten, da diese die (verzerrenden) Einflüsse der anderen Faktoren moderieren. Jugendliche mit geringen kognitiven Fähigkeiten sind in besonderem Maße anfällig für Verzerrungen durch das Selbstbild und Referenzrahmeneffekte. Mit zunehmenden kognitiven Fähigkeiten nimmt der relative Einfluss dieser Faktoren ab. Bei sehr hoch ausgeprägten kognitiven Fähigkeiten gehen die verzerrenden Einflüsse teilweise vollständig zurück. In Übereinstimmung mit den theoretischen Erwartungen deuten die Ergebnisse demnach darauf hin, dass viele Jugendliche mit der Aufgabe, ihre Sprachkompetenzen selbst einzuschätzen, überfordert sind und im Antwortprozess auf einfachere Heuristiken ausweichen. Dabei ziehen sie verstärkt Indikatoren heran, die ein selektives und teils systematisch überzogenes Bild ihrer Leistung entstehen lassen. Da dies überwiegend auf Personen mit geringeren Fähigkeiten zutrifft, resultiert daraus auch das in der Forschung häufig beobachtete Muster, wonach gerade diejenigen mit geringen Leistungen dazu neigen, ihre Kompetenzen zu überschätzen (vgl. Kruger und Dunning 1999). Auch die oben skizzierten Verzerrungen zu Gruppenunterschieden lassen sich weitestgehend auf diese allgemeine Tendenz zurückführen. In Herkunftsgruppen mit durchschnittlich geringerer *objektiver* Sprachkompetenz neigen die Jugendlichen stärker dazu ihre Kompetenz zu überschätzen, was im Aggregat zu einer Angleichung bzw. Verschiebung der durchschnittlichen

subjektiven Sprachkompetenz zwischen den Gruppen führt.

Insgesamt lassen sich aus diesen Ergebnissen einige Implikationen für die Praxis ableiten, die Forschenden im Umgang mit selbsteingeschätzten Sprachkompetenzen als Hilfestellung dienen können. Die Implikationen gelten in gleicher Weise für Überlegungen zum Einsatz subjektiver Sprachkompetenzen in zukünftigen Studien sowie für die Beurteilung bisheriger Forschungsergebnisse, die darauf beruhen. Die Befunde legen nahe, dass Zusammenhänge mit den kognitiven Fähigkeiten der Befragten unterschätzt und Zusammenhänge mit dem Selbstbild und der sprachlichen Zusammensetzung des sozialen Umfelds überschätzt werden. Die verzerrenden Tendenzen beschränken sich dabei nicht nur auf die Faktoren selbst, sondern können auch auf Kovariate übergreifen, d.h. potentiell auf alle Beziehungen, die mit den Faktoren in Zusammenhang stehen. Zum Beispiel wäre zu erwarten, dass die Leistung in Mathematik, aufgrund ihrer Abhängigkeit von den allgemeinen kognitiven Fähigkeiten, schwächer mit den subjektiven als mit den tatsächlichen L2-Kompetenzen korreliert. Dieses Muster bestätigte sich auch in den Befunden der exemplarischen Analysemodelle zu den Testergebnissen und Noten in Mathematik (vgl. Kapitel 6, Tabelle 6.6). Ein weiteres Beispiel für potentielle Verzerrungen betrifft Einflüsse von Persönlichkeitsmerkmalen, die mit dem Selbstbild zusammenhängen und von denen angenommen wird, dass sie den Spracherwerb bedingen (z.B. Asfar et al. 2019; Kosyakova und Laible 2024). Derartige Merkmale sollten stärker mit den subjektiven als den objektiven L2-Kompetenzen korrelieren. In einer erweiterten Analyse zu dem Beitrag aus Kapitel 6 konnte diese Tendenz exemplarisch für das allgemeine Selbstwertgefühl (vgl. Anhang, Tabelle A.1, S. 248) sowie für vereinzelte Dimensionen der Big Five (vgl. Anhang, Tabelle A.2, S. 249) beobachtet werden.

Das konkrete Ausmaß der potentiellen Verzerrungen, die mit dem Einsatz selbsteingeschätzter Sprachkompetenzen verbunden sind, ist allerdings nur schwer vorhersehbar und hängt jeweils von der Fragestellung und dem Anwendungsbereich ab. Die Schlüsse, die aus den Befunden der Dissertation gezogen werden können, beschränken sich deshalb auf Hinweise darauf, welche Zusammenhänge fehleranfällig sind und in welche Richtung diese potentiell verzerrt werden. Grundsätzlich werden objektive, leistungs-basierte Zusammenhänge in Analysen mit selbsteingeschätzten Sprachkompetenzen tendenziell unterschätzt. Dies betrifft auch Leistungsunterschiede zwischen Gruppen, weswegen der Einsatz von Selbsteinschätzungen als problematisch einzustufen ist, wenn die Analyse bzw. Aufklärung von Gruppenunterschieden im Zentrum des Forschungsinteresses steht. Zielt die Forschungsfrage hingegen darauf ab, relevante Einflussfaktoren zu identifizieren und/oder Mechanismen zu prüfen, können auch subjektive Sprachkompetenzen aufschlussreich sein, da diese zahlreiche bedeutsame Zusammenhänge – zumindest in der Tendenz – korrekt wiedergeben. Beispielsweise konnten für den Sprachgebrauch mit Freund:innen und Mitschüler:innen vergleichbare Zusammenhänge mit den subjektiven wie objektiven L2-Kompetenzen der Jugendlichen festgestellt werden. Völlig unbrauchbar für die Forschung sind Selbsteinschätzungen von Sprachkompetenzen daher nicht, deren Verlässlichkeit sollte jedoch hinsichtlich des jeweiligen Anwendungsfalls kritisch diskutiert und bewertet werden. Die vorliegende Dissertation bietet erste Anhaltspunkte für eine derartige Bewertung.

Vieles deutet darauf hin, dass Befragte oft schlichtweg nicht in der Lage sind, ihre

Leistungen akkurat zu beurteilen (vgl. Dunning et al. 2003; Kruger und Dunning 1999; Ehrlinger et al. 2008). Das bedeutet, dass sie nicht bewusst falsch oder sozial erwünscht antworten, sondern tatsächlich glauben, ihre Kompetenzen seien besser als sie objektiv sind. Vor diesem Hintergrund scheint es nur bedingt möglich zu sein, durch eine Verbesserung der Fragen und Instrumente zur Selbsteinschätzung die Messung zu präzisieren. Es ist eher auszuschließen, dass die vorherrschenden Antwortverzerrungen mittels verbesserter Instrumente komplett ausgeschaltet werden können, aber eine Abschwächung der allgemeinen Tendenzen ließe sich damit unter Umständen schon erzielen. Einige Studien deuten beispielsweise an, dass eine spezifischere Abfrage der Sprachkompetenzen stärker mit objektiven Testleistungen korrelieren als globale Einschätzungsmaße (z.B. Brantmeier und Vanderplank 2008; Brantmeier et al. 2012; Li und Zhang 2021; Ross 1998). Der Annahme nach fällt es den Befragten leichter, ihre Sprachkompetenzen anhand ganz konkreter sprachlicher Anforderungen und Situationen, bestenfalls mit Bezug zu ihrem Alltag und jeweiligem Kompetenzniveau, zu beurteilen. Mit Blick auf die Befunde zu den Bestimmungsfaktoren für Antwortverzerrungen könnte eine solche spezifischere Abfrage durchaus zielführend sein, wenn dadurch die Aufgabenschwierigkeit, und damit die kognitive Leistung, die die Befragten im Antwortprozess aufbringen müssen, gesenkt wird. Mit dem niedrigerem Anforderungsniveau sollten dann auch mögliche Verzerrungen durch das Selbstbild und den Referenzrahmen geringer ausfallen.

Zusätzlich sollten die eingesetzten Skalen möglichst allgemein verständliche und trennscharfe Antwortkategorien umfassen, um einen einheitlichen Bezugsrahmen zu schaffen. Zum Beispiel geben die häufig verwendeten Antwortkategorien „*sehr/eher schlecht*“ und „*eher/sehr gut*“ relativ unscharfe Kompetenzbereiche vor, die keine allgemein definierten Richtwerte repräsentieren und entsprechend von der Befragten sehr unterschiedlich interpretiert werden können. Da die Interpretationsleistung kognitiv sehr voraussetzungsreich ist und die Antwortskala keine klaren Bezugspunkte setzt, könnte dies die Befragten dazu veranlassen, verstärkt auch andere, wenig verlässliche Indikatoren – wie das Selbstkonzept oder soziale Vergleiche – zur Orientierung heranzuziehen. Eine Möglichkeit dem entgegenzuwirken, wäre die Verwendung von klar definierten Endpunkten der Skala (z.B. „*auf muttersprachlichem Niveau*“), um damit einen einheitlichen Beurteilungsmaßstab vorzugeben. Inwiefern eine derartige Strategie die Antwortverzerrungen tatsächlich eindämmen könnte, lässt sich an dieser Stelle nicht beantworten. Bisher liegen keine empirischen Befunde vor, anhand derer eine abschließende Einschätzung dazu getroffen werden könnte.

Bestrebungen zur Entwicklung präziserer Selbsteinschätzungsmaße sind zweifellos wünschenswert, insbesondere für den Einsatz in Erhebungen, die sich keine zeitaufwändigen oder kostenintensiven Kompetenztests leisten können und deswegen auf eine möglichst effiziente Messung angewiesen sind. Allerdings setzt die Verbesserung durch eine Abfrage spezifischer Kompetenzen – so vielversprechend diese auch sein möge – in der Regel den Einsatz umfangreicher Fragebatterien voraus, was dem zeitlichen Aufwand einer Testung nahe kommt. Für einen Großteil der Studien wäre dies vermutlich keine gangbare Alternative, zumal entsprechende Instrumente ohnehin erst entwickelt und erprobt werden müssten. Folglich ist davon auszugehen, dass zukünftig

weiterhin viele Studien die klassischen globalen Indikatoren (z.B. „*Wie gut können Sie auf Deutsch lesen?*“) einsetzen werden – ggf. mit kleineren Modifikationen der Fragestellungen und Antwortskalen entsprechend der hier skizzierten Verbesserungsvorschläge. Umso entscheidender ist es, ein Bewusstsein dafür zu schaffen, dass die Verwendung von Selbsteinschätzungen als Indikator für die tatsächliche Sprachkompetenz mit gewissen Unschärfen verbunden sein kann. Dies sollte bei der Interpretation entsprechender Forschungsbefunde berücksichtigt werden.

In den Analysen der vorliegenden Dissertation wurde die Validität selbsteingeschätzter Sprachkompetenzen anhand einer Stichprobe von Jugendlichen mit Migrationshintergrund untersucht. Es stellt sich folglich die Frage, inwiefern die zentralen Befunde auch über den Geltungsbereich der Untersuchungspopulation hinaus von Bedeutung sind. Die relevanten Bestimmungsfaktoren und Prozesse, die der Entstehung von Antwortverzerrungen zugrunde liegen, sind allgemeiner Natur und sollten demnach in vergleichbarer Weise auch für andere Untersuchungspopulationen und Anwendungsgebiete gelten. Hingegen können sich die jeweiligen empirisch vorherrschenden Randbedingungen zwischen Untersuchungspopulationen durchaus unterscheiden, weswegen die allgemeinen Tendenzen potentieller Verzerrungen unterschiedlich stark ins Gewicht fallen können. Folglich ist bei der Verwendung selbsteingeschätzter Sprachkompetenzen immer mit gewissen (systematischen) Unschärfen zu rechnen, aber das Ausmaß der verzerrenden Tendenzen könnte in Abhängigkeit der Untersuchungspopulation und des Anwendungsfalls variieren. Mit Blick auf die kognitiven Fähigkeiten, denen im Antwortprozess eine tragende Rolle zukommt, sind in der Regel keine substantiellen Unterschiede zwischen verschiedenen Populationen zu erwarten. Lediglich in Ausnahmefällen, bei besonders selektiven Untersuchungspopulationen wie beispielsweise Studierenden-Stichproben, könnten die kognitiven Fähigkeiten im Schnitt höher ausfallen, was verlässlichere Selbsteinschätzungen vermuten ließe. Bisherige Untersuchungen, meist im Kontext von Fremdsprachenkursen an Universitäten, können allerdings auch unter Studierenden nur moderate Zusammenhänge zwischen subjektiven und objektiven Sprachkompetenzen feststellen (z.B. Alavi und Akbarian 2008; Brantmeier und Vanderplank 2008; Brantmeier et al. 2012; Peyer und Kaiser 2010). Die Korrelationen fallen dabei nicht wesentlich höher aus als bei anderen Untersuchungspopulationen, was die Vermutung nahelegt, dass selbst in Stichproben mit eher vorteilhaften Bedingungen ähnliche Verzerrungstendenzen auftreten, wenn auch teilweise in abgeschwächter Form.

Zusätzliche Analysen der in der Dissertation verwendeten Daten deuten an, dass die Selbsteinschätzungen von Jugendlichen *ohne* Migrationshintergrund ebenfalls relativ stark von ihren getesteten Sprachkompetenzen im Deutschen abweichen. Die Korrelation zwischen subjektiven und objektiven Sprachkompetenzen fällt in dieser Gruppe ähnlich gering aus wie in der Untersuchungspopulation mit Migrationshintergrund. Dieser Befund stützt die Annahme, dass die grundlegenden Tendenzen nicht ausschließlich bei Jugendlichen mit Migrationshintergrund vorliegen, sondern allgemeiner Natur sind. Allerdings beruhen die Zusatzanalysen auf einer sehr speziellen Teilstichprobe von Schüler:innen, die die Filterführung des Papierfragebogens missachtet haben. Jugendliche ohne Migrationshintergrund, die in der Regel auch keine weitere Sprache in der Familie erlernt haben, sollten die Fragen zur Erst- und Zweitsprache, einschließlich

der Selbsteinschätzung, eigentlich überspringen. Bei den vorliegenden Fällen handelt es sich vermutlich um Jugendliche, die den Fragebogen nur mit geringer Sorgfalt beantwortet haben, weswegen die skizzierten Hinweise nur bedingt aussagekräftig sind. Gleichzeitig unterstützt diese Besonderheit der Teilstichprobe die grundlegende theoretische Argumentation, da verzerrende Tendenzen besonders stark ins Gewicht fallen sollten, wenn der Antwortprozess – wie in dem speziellen Fall – nur oberflächlich erfolgt.

Es gibt allerdings auch Hinweise auf systematische Unterschiede in den Antworttendenzen bestimmter Gruppen. Ein besonders auffallendes und über nahezu alle durchgeführten Analysen relativ stabiles Muster zeigt Unterschiede nach dem Generationenstatus auf. Demnach schätzen Jugendliche mit eigener Migrationserfahrung ihre Kompetenzen im Deutschen tendenziell akkurater ein als Jugendliche, die in Deutschland geboren wurden und deren Eltern oder Großeltern zugezogen sind. Dies könnte damit zusammenhängen, dass selbst zugewanderte Personen Deutsch oft erst später im Leben als Fremd- oder Zweitsprache lernen und für sie „gute“ L2-Kompetenzen möglicherweise nicht so stark mit der Aufrechterhaltung eines positiven Selbstbilds verknüpft sind, wie für Personen der zweiten und dritten Zuwanderungsgeneration, die von klein auf mit der deutschen Sprache aufwachsen. Dementsprechend ist anzunehmen, dass bei Migrant:innen der ersten Generation verzerrende Einflüsse des Selbstbilds, und damit einhergehend auch die Neigung zur Überschätzung der eigenen L2-Kompetenzen insgesamt, schwächer ausgeprägt sind als bei späteren Generationen. Selbst zugewanderte Migrant:innen nehmen zudem oft an strukturierten Sprachförderprogrammen teil, in denen sie regelmäßig eine objektive Bewertung ihres Leistungsstands und Fortschritts erhalten. Diese Bewertungen begünstigen eine genauere Einordnung der eigenen Sprachkompetenz. Weitere Unterschiede in der Verlässlichkeit der Selbsteinschätzungen könnten mit dem Alter der jeweils betrachteten Untersuchungspopulation zusammenhängen. Aus der bisherigen Forschung gibt es Hinweise darauf, dass die Tendenz der Überschätzung der eigenen Leistung bei Kindern mit zunehmendem Alter abnimmt (vgl. Xia et al. 2024). Demzufolge könnten die Selbsteinschätzungen von jüngeren Untersuchungspopulationen, wie beispielsweise Kindern im Vorschul- oder Grundschulalter, noch fehleranfälliger sein als bei den untersuchten Jugendlichen. Bei Erwachsenen hingegen sollten die Verzerrungen tendenziell geringer ausfallen. Die unterschiedlichen Befundmuster nach Generationenstatus und Alter bedeuten jedoch nicht, dass die Selbsteinschätzungen bei Analysen zu selbst zugewanderten und/oder erwachsenen Migrant:innen bedenkenlos verwendet werden können. Auch hier sind systematische Abweichungen zu erwarten, jedoch vermutlich in geringerem Ausmaß als bei den untersuchten Jugendlichen, die vorwiegend der zweiten und dritten Zuwanderungsgeneration angehören. In der vorliegenden Dissertation konnte diesen Annahmen aufgrund der Zusammensetzung der Stichprobe nicht näher nachgegangen werden. Es wäre jedoch aufschlussreich, insbesondere vor dem Hintergrund der wachsenden Bedeutung der Forschung zu Integrationsprozessen von Geflüchteten, diese Zusammenhänge zukünftig systematisch zu untersuchen.

Insgesamt lassen die unterschiedlichen Hinweise vermuten, dass die grundlegenden Tendenzen für Antwortverzerrungen bei der Selbsteinschätzung von Sprachkompetenzen universell auftreten. Das Ausmaß der potentiellen Verzerrungen, die mit der

Verwendung der selbsteingeschätzten Sprachkompetenzen in Analysen einhergehen, kann allerdings in Abhängigkeit der Zusammensetzung der jeweiligen Untersuchungspopulation variieren. Mit gewissen Abweichungen zwischen subjektiven und objektiven Sprachkompetenzen ist demnach immer zu rechnen. In vielen Fällen werden die interessierenden Zusammenhänge lediglich unter- oder überschätzt, jedoch in der Tendenz korrekt wiedergegeben. Allerdings besteht auch das Risiko, dass manche Beziehungen komplett verschleiert oder falsch dargestellt werden. Letztlich bleibt nicht völlig auszuschließen, dass falsche Schlüsse aus den Analysen gezogen werden könnten. Daher ist es ratsam, im Umgang mit selbsteingeschätzten Sprachkompetenzen besonders vorsichtig zu sein.

Die gesamte Ausrichtung und Argumentation der vorliegenden Dissertation baut auf der Prämisse auf, dass die Selbsteinschätzungen als Indikator für die tatsächlichen Sprachkompetenzen der Befragten verstanden werden. In der Integrationsforschung werden die selbsteingeschätzten Sprachkompetenzen in der Regel genau zu diesem Zweck eingesetzt. Auch aus inhaltlicher Sicht stehen meist die objektiven Sprachkompetenzen im Zentrum des Interesses, zum Beispiel, wenn es um die Bedeutung der L2-Kompetenzen für die Erklärung von ethnischen Bildungsungleichheiten geht. Die subjektiven Sprachkompetenzen können allerdings nicht nur als Substitut für eine objektive Messung dienen, sondern auch für sich genommen eine inhaltlich relevante Variable darstellen (vgl. Esser 1985, S. 215f.). Es deutet Vieles darauf hin, dass Individuen oft davon überzeugt sind, ihre Leistungen wären besser als sie tatsächlich sind; und genau diese Tendenz bilden Selbsteinschätzungen relativ akkurat ab. Die Messung von Selbsteinschätzungen kann folglich Aufschluss darüber geben, was Individuen wirklich denken und was letztendlich auch ihr Handeln beeinflussen sollte. Diese subjektiven Auffassungen bilden die Grundlage für individuelle Handlungsentscheidungen, wie sie nach den etablierten Ausführungen der Wert-Erwartungs-Theorie, deren Anwendung in der soziologischen Bildungs- und Integrationsforschung – aber auch darüber hinaus – sehr verbreitet ist, modelliert werden (z.B. Essers 1999, 2006; Dollmann 2010). Handlungsweisend sind demnach nicht ausschließlich die objektiv vorherrschenden Randbedingungen, sondern gerade auch die individuelle Wahrnehmung dieser Bedingungen in Form von *subjektiven* Erwartungen und Bewertungen seitens der Akteure (vgl. Esser 1999, S. 253). Bei der theoretischen Modellierung des Spracherwerbs ist die Entscheidung für Investitionen in zusätzliche Kompetenzen unter anderem von den bereits erreichten Sprachkompetenzen, dem Status quo, abhängig (vgl. Esser 2006). In diesem Zusammenhang ist beispielsweise zu erwarten, dass die Entscheidung einen Sprachkurs zu besuchen oder in alltäglichen Situationen Deutsch anstelle der Herkunftssprache zu verwenden vielmehr von der subjektiven Wahrnehmung der eigenen Sprachkompetenzen abhängt als von den tatsächlichen Kompetenzen – zumal die Akteure oftmals gar kein akkurates Bild ihrer tatsächlichen Kompetenzen haben, wie die Ergebnisse der Dissertation nahelegen. Vor diesem Hintergrund haben subjektive Sprachkompetenzen zweifellos ihre Daseinsberechtigung und es ist durchaus zielführend, diese in zukünftigen Erhebungen weiterhin – auch zusätzlich zu einer objektiven Sprachmessung – zu erfassen. Die theoretische Bedeutung der subjektiven Messung und das damit verbundene Analysepotential hat in der bisherigen Forschung

wenig Beachtung gefunden. Es könnte sich lohnen, an diese Überlegungen anzuknüpfen und diese aus handlungstheoretischer Perspektive tiefergehend zu erforschen.

Abschließende Bemerkung

Im Rahmen der Dissertation wurden die beiden Themenkomplexe – *Außerfamiliärer Sprachgebrauch und Zweitspracherwerb* und *Validität subjektiver Einschätzungen von Sprachkompetenzen* – weitgehend losgelöst voneinander untersucht. Anknüpfend an die theoretischen Überlegungen zur Bedeutung der subjektiven Wahrnehmung der Randbedingungen für den Prozess des Spracherwerbs, lassen sich die beiden Themenkomplexe aber durchaus auch inhaltlich in Beziehung setzen. Personen, die ihre aktuellen Sprachkompetenzen als hoch oder ausreichend einschätzen, haben eine geringere Motivation aktiv in zusätzliche Maßnahmen zur Kompetenzsteigerung zu investieren, da der Status quo bereits einen hohen Nutzen verspricht und demgegenüber der Grenzertrag weiterer Investitionen gering ausfällt. Gleichzeitig neigen gerade Personen mit vergleichsweise geringen tatsächlichen Sprachkompetenzen eher dazu, ihren Status quo zu überschätzen. Diese beiden Tendenzen bedingen sich gegenseitig, was dazu führen kann, dass Personen mit geringeren Sprachkompetenzen weniger Fortschritte machen, weil sie die Notwendigkeit dazu nicht erkennen, woran sich wiederum aufgrund des eingeschränkten Lernzuwachses längerfristig nur wenig ändert. Dies kann sich zum Beispiel darin äußern, dass Personen bewusst auf die Teilnahme an einem Sprachkurs verzichten, obwohl sie davon in besonderem Maße – sowohl hinsichtlich des Kompetenzerwerbs als auch der Selbsteinschätzung – profitieren würden. Demgegenüber könnten Personen, die ihre Sprachkompetenzen überschätzen, aber auch weniger Hemmungen haben, die jeweilige Sprache im Alltag häufiger zu verwenden, obwohl ihre tatsächlichen Kompetenzen möglicherweise begrenzt sind. Der häufige Sprachgebrauch sollte dabei den Kompetenzerwerb in der betreffenden Sprache begünstigen. Darüber hinaus kann die Verwendung der Sprache, insbesondere wenn die Interaktionspartner die Sprache kompetent beherrschen, nützliche Hinweise für eine akkuratere Einschätzung der eigenen Kompetenzen liefern – zum Beispiel, weil bestimmte sprachliche Defizite erst in der Anwendung auffallen.

Für die Beantwortung der einzelnen Fragestellungen waren die Zusammenhänge zwischen den Themenkomplexen jedoch nicht entscheidend. Aus analytischer Sicht war es daher zielführend, beide Aspekte getrennt zu untersuchen, um die theoretische Argumentation und das empirische Vorgehen gezielt auf die spezifische Teilfragestellung ausrichten zu können. Mit diesem theoriegeleiteten Ansatz hat die Dissertation dazu beigetragen, die Schließung bestehender Forschungslücken voranzutreiben und sowohl inhaltliche als auch methodische Denkanstöße für zukünftige Forschungen zu geben.

Literaturverzeichnis

- Abadie, A., Athey, S., Imbens, G. W., und Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, 138(1):1–35.
- Ackerman, P. L., Beier, M. E., und Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, 33(4):587–605.
- Alavi, S. M. und Akbarian, I. (2008). Validating a self-assessment questionnaire on vocabulary knowledge. *TELL*, 2(2):125–154.
- Alba, R., Logan, J., Lutz, A., und Stults, B. (2002). Only English by the third generation? Loss and preservation of the mother tongue among the grandchildren of contemporary immigrants. *Demography*, 39(3):467–484.
- Alba, R. und Nee, V. (1997). Rethinking assimilation theory for a new era of immigration. *International Migration Review*, 31(4):826–874.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Allison, P. D. (2001). *Missing data*. Thousand Oaks: Sage Publications.
- Aronson, E. (1968). Dissonance theory: Progress and problems. In Abelson, R. P., Aronson, E., McGuire, W. J., Newcomb, T. M., Rosenberg, M. J., und Tannenbaum, P. H., Herausgeber, *Theories of cognitive consistency: A sourcebook*, Seiten 5–27. Chicago: Rand McNally.
- Asfar, D., Born, M. P., Ostrom, J. K., und Van Vugt, M. (2019). Psychological individual differences as predictors of refugees’ local language proficiency. *European Journal of Social Psychology*, 49(7):1385–1400.
- Bachman, L. F. und Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16(4):449.
- Bankston III, C. L. und Caldas, S. (1996). Majority African American schools and social injustice: The influence of de facto segregation on academic achievement. *Social Forces*, 75(2):535–555.
- Bankston III, C. L. und Zhou, M. (1995). Effects of minority-language literacy on the academic achievement of vietnamese youths in New Orleans. *Sociology of Education*, 68(1):1–17.

- Baumert, J. und Schümer, G. (2001). Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb. In Baumert, J., Artelt, C., und Klieme, E., Herausgeber, *PISA 2000: Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*, Seiten 323–407. Opladen: Leske + Budrich.
- Baur, C. und Häussermann, H. (2009). Ethnische Segregation in deutschen Schulen. *Leviathan*, 37(3):353–366.
- Becker, B. (2011). Cognitive and language skills of Turkish children in Germany: A comparison of the second and third generation and mixed generational groups. *International Migration Review*, 45(2):426–459.
- Becker, B., Klein, O., und Biedinger, N. (2013). The development of cognitive, language, and cultural skills from age 3 to 6: A comparison between children of Turkish origin and children of native-born German parents and the role of immigrant parents' acculturation to the receiving society. *American Educational Research Journal*, 50(3):616–649.
- Belli, R. F., Traugott, M. W., und Beckmann, M. N. (2001). What leads to voting overreports? Contrasts of overreporters to validated voters and admitted nonvoters in the American National Election Studies. *Journal of Official Statistics*, 17(4):479–498.
- Biedinger, N. (2007). Entwicklung und Lebensumfeld von Vorschulkindern: Zur Heterogenität von Familien mit türkischem Migrationshintergrund. *Zeitschrift für Türkeistudien*, 20(1):7–24.
- Biedinger, N. (2011). Wie erlernen Kinder die ersten Wortschatz- und Rechenfähigkeiten? Erklärung mit Hilfe eines erweiterten sozialpsychologischen Lernmodells. *Zeitschrift für Soziologie der Erziehung und Sozialisation*, 22(2):153–168.
- Biedinger, N., Becker, B., und Klein, O. (2015). Turkish-language ability of children of immigrants in Germany: Which contexts of exposure influence preschool children's acquisition of their heritage language? *Ethnic and Racial Studies*, 38(9):1520–1538.
- Blossfeld, H.-P. und Roßbach, H.-G. (2019). *Education as a lifelong process: The German National Educational Panel Study (NEPS)*. Wiesbaden: Springer VS.
- Bong, M. und Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review*, 15(1):1–40.
- Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in Western society*. New York: Wiley.
- Brantmeier, C. (2006). Advanced L2 learners and reading placement: Self-assessment, CBT, and subsequent performance. *System*, 34(1):15–35.
- Brantmeier, C. und Vanderplank, R. (2008). Descriptive and criterion-referenced self-assessment with L2 readers. *System*, 36(3):456–477.

- Brantmeier, C., Vanderplank, R., und Strube, M. (2012). What about me?: Individual self-assessment by skill and level of language instruction. *System*, 40(1):144–160.
- Cannell, C. F., Miller, P. V., und Oksenberg, L. (1981). Research on interviewing techniques. *Sociological Methodology*, 12:389–437.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64(8):1–9.
- Carter, T. J. und Dunning, D. (2008). Faulty self-assessment: Why evaluating one’s own competence is an intrinsically difficult task. *Social and Personality Psychology Compass*, 2(1):346–360.
- Chaiken, S. und Trope, Y. (1999). *Dual-process theories in social psychology*. New York: The Guilford Press.
- Charette, M. und Meng, R. (1994). Explaining language proficiency: Objective versus self-assessed measures of literacy. *Economics Letters*, 44(3):313–321.
- Chiswick, B. R. und Miller, P. W. (1995). The endogeneity between language and earnings: International analyses. *Journal of Labor Economic*, 13(2):246–288.
- Chiswick, B. R. und Miller, P. W. (2001). A model of destination-language acquisition: Application to male immigrants in Canada. *Demography*, 38(3):391–409.
- Critcher, C. R. und Dunning, D. (2009). How chronic self-views influence (and mislead) self-assessments of task performance: Self-views shape bottom-up experiences with the task. *Journal of Personality and Social Psychology*, 97(6):931–945.
- DeMaio, T. (1984). Social desirability and survey measurement: A review. In Turner, C. F. und Martin, E., Herausgeber, *Surveying subjective phenomena*, Seiten 257–282. New York: Russell Sage Foundation.
- Diehl, C., Hunkler, C., und Kristen, C. (2016a). Ethnische Ungleichheiten im Bildungsverlauf: Eine Einführung. In Diehl, C., Hunkler, C., und Kristen, C., Herausgeber, *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*, Seiten 3–31. Wiesbaden: Springer VS.
- Diehl, C., Hunkler, C., und Kristen, C. (2016b). *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*. Wiesbaden: Springer VS.
- Diehl, C. und Schnell, R. (2006). “Reactive ethnicity” or “assimilation”? Statements, arguments, and first empirical evidence for labor migrants in Germany. *International Migration Review*, 40(4):786–816.
- Dollmann, J. (2010). *Türkischstämmige Kinder am ersten Bildungsübergang: Primäre und sekundäre Herkunftseffekte*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Dollmann, J. (2017). Ethnische Bildungsungleichheiten. In Becker, R., Herausgeber, *Lehrbuch der Bildungssoziologie*, Seiten 487–510. Wiesbaden: Springer VS.

- Dollmann, J. und Kristen, C. (2010). Herkunftssprache als Ressource für den Schulerfolg? Das Beispiel türkischer Grundschul Kinder. In Allemann-Ghionda, C., Stanat, P., Göbel, K., und Röhner, C., Herausgeber, *Migration, Identität, Sprache und Bildungserfolg*, Seiten 123–146. Weinheim: Beltz.
- Driessen, G. (2002). School composition and achievement in primary education: A large-scale multilevel approach. *Studies in Educational Evaluation*, 28(4):347–368.
- Dubowy, M., Ebert, S., von Maurice, J., und Weinert, S. (2008). Sprachlich-kognitive Kompetenzen beim Eintritt in den Kindergarten: Ein Vergleich von Kindern mit und ohne Migrationshintergrund. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 40(3):124–134.
- Dunn, L. M. und Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (PPVT) – 4th edition*. Bloomington: Pearson Assessments.
- Dunning, D. (2005). *Self-insight: Roadblocks and detours on the path to knowing thyself*. New York: Psychology Press.
- Dunning, D., Johnson, K., Ehrlinger, J., und Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, 12(3):83–87.
- Dustmann, C. und Van Soest, A. (2001). Language fluency and earnings: Estimation with misclassified language indicators. *The Review of Economics and Statistics*, 83(4):663–674.
- Dustmann, C. und Van Soest, A. (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review*, 55(3):473–492.
- Edele, A., Schotte, K., und Stanat, P. (2016). Assessment of immigrant students' listening comprehension in their first languages (L1) Russian and Turkish in grade 9: Test construction and validation. In Blossfeld, H.-P., von Maurice, J., Bayer, M., und Skopek, J., Herausgeber, *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*, Seiten 441–463. Wiesbaden: Springer VS.
- Edele, A., Seuring, J., Schotte, K., Kristen, C., und Stanat, P. (2023). Is the first language a resource, an obstacle, or irrelevant for language minority students' education? In Weinert, S., Blossfeld, G. J., und Blossfeld, H.-P., Herausgeber, *Education, competence development and career trajectories: Analysing data of the National Educational Panel Study (NEPS)*, Seiten 349–367. Cham: Springer.
- Edele, A. und Stanat, P. (2016). The role of first-language listening comprehension in second-language reading comprehension. *Journal of Educational Psychology*, 108(2):163–180.
- Ehrlinger, J. (2008). Skill level, self-views and self-theories as sources of error in self-assessment. *Social and Personality Psychology Compass*, 2(1):382–398.

- Ehrlinger, J. und Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, 84(1):5–17.
- Ehrlinger, J., Johnson, K., Banner, M., Dunning, D., und Kruger, J. (2008). Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent. *Organizational Behavior and Human Decision Processes*, 105(1):98–121.
- Eksner, J. und Stanat, P. (2011). Effects of ethnically segregated learning settings in the United States and Germany. In Gallagher, K. S., Brenwer, D., Goodyear, R., und Bensimon, E., Herausgeber, *Urban education: A model for leadership and policy*, Seiten 826–878. New York: Routledge.
- Espenshade, T. J. und Fu, H. (1997). An analysis of english-language proficiency among U.S. immigrants. *American Sociological Review*, 62(2):288–305.
- Esser, H. (1985). Zur Validität subjektiver Sprachkompetenzmessung. In Sievering, U., Herausgeber, *Arbeitsmigrantenforschung in der Bundesrepublik Deutschland: Methodenprobleme der Datenerhebung*, Seiten 192–226. Frankfurt am Main: Haag + Herchen.
- Esser, H. (1990). „Habits“, „Frames“ und „Rational Choice“: Die Reichweite von Theorien der rationalen Wahl (am Beispiel der Erklärung des Befragtenverhaltens). *Zeitschrift für Soziologie*, 19(4):231–247.
- Esser, H. (1991). Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „rational choice“. In Wittenberg, R., Herausgeber, *Person–Situation–Institution–Kultur: Günter Büschges zum 65. Geburtstag*, Seiten 59–78. Berlin: Duncker und Humblot.
- Esser, H. (1999). *Soziologie: Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt am Main: Campus.
- Esser, H. (2001). *Soziologie: Spezielle Grundlagen. Band 6: Sinn und Kultur*. Frankfurt am Main: Campus.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten*. Frankfurt am Main: Campus.
- Falchikov, N. und Boud, D. (1989). Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 59(4):395–430.
- Farh, J.-L. und Dobbins, G. H. (1989). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology*, 42(4):835–850.

- Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The mode model as an integrative framework. *Advances in Experimental Social Psychology*, 23:75–109.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford: Stanford University Press.
- Finnie, R. und Meng, R. (2005). Literacy and labour market outcomes: Self-assessment versus test score measures. *Applied Economics*, 37(17):1935–1951.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34(10):906–911.
- Freund, P. A. und Kasten, N. (2012). How smart do you think you are? A meta-analysis on the validity of self-estimates of cognitive ability. *Psychological Bulletin*, 138(2):296–321.
- Friedrichs, J. (2008). Ethnische Segregation. In Kalter, F., Herausgeber, *Migration und Integration*, Seiten 380–411. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Gehrer, K., Zimmermann, S., Artelt, C., und Weinert, S. (2012). *The assessment of reading competence (including sample items for grade 5 and 9). Scientific use file 2012, version 1.0.0*. Bamberg: University of Bamberg, National Educational Panel Study.
- Gnambs, T. (2020). *NEPS technical report for reading: Scaling results of starting cohort 4 for grade 9 in special schools (NEPS Survey Paper No. 63)*. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.
- Gresch, C. und Kristen, C. (2011). Staatsbürgerschaft oder Migrationshintergrund? *Zeitschrift für Soziologie*, 40(3):208–227.
- Gärtig, A.-K., Albrecht, P., und Rothe, A. (2010). *Wie Menschen in Deutschland über Sprache denken: Ergebnisse einer bundesweiten Repräsentativerhebung zu aktuellen Spracheinstellungen (IDS Arbeitspapiere und Materialien zur deutschen Sprache, Bd. 40)*. Mannheim: Institut für Deutsche Sprache.
- Haag, N., Böhme, K., Rjosk, C., und Stanat, P. (2016). Zuwanderungsbezogene Disparitäten. In Stanat, P., Böhme, K., Schipolowski, S., und Haag, N., Herausgeber, *IQB-Bildungstrend 2015: Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im zweiten Ländervergleich*, Seiten 431–479. Münster: Waxmann.
- Hans, S. (2010). *Assimilation oder Segregation? Anpassungsprozesse von Einwanderern in Deutschland*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Harder, N., Figueroa, L., Gillum, R. M., Hangartner, D., Laitin, D. D., und Hainmueller, J. (2018). Multidimensional measure of immigrant integration. *Proceedings of the National Academy of Sciences of the United States of America*, 115(45):11483–11488.

- Harmon-Jones, E. und Harmon-Jones, C. (2007). Cognitive dissonance theory after 50 years of development. *Zeitschrift für Sozialpsychologie*, 38(1):7–16.
- Heath, A. F. und Brinbaum, Y. (2007). Guest editorial: Explaining ethnic inequalities in educational attainment. *Ethnicities*, 7(3):291–304.
- Heath, A. F., Rothon, C., und Kilpi, E. (2008). The second generation in western Europe: Education, unemployment, and occupational attainment. *Annual Review of Sociology*, 34(1):211–235.
- Heilenman, K. (1990). Self-assessment of second language ability: The role of response effects. *Language Testing*, 7(2):174–201.
- Henschel, S., Heppt, B., Rjosk, C., und Weirich, S. (2022). Zuwanderungsbezogene Disparitäten. In Stanat, P., Schipolowski, S., Schneider, R., Sachse, K., Weirich, S., und Henschel, S., Herausgeber, *IQB-Bildungstrend 2021: Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*, Seiten 181–219. Münster: Waxmann.
- Henschel, S., Heppt, B., und Weirich, S. (2023). Zuwanderungsbezogene Disparitäten. In Stanat, P., Schipolowski, S., Schneider, R., Weirich, S., Henschel, S., und Sachse, K., Herausgeber, *IQB-Bildungstrend 2022: Sprachliche Kompetenzen am Ende der 9. Jahrgangsstufe im dritten Ländervergleich*, Seiten 299–344. Münster: Waxmann.
- Henschel, S., Heppt, B., Weirich, S., Edele, A., Schipolowski, S., und Stanat, P. (2019). Zuwanderungsbezogene Disparitäten. In Stanat, P., Schipolowski, S., Mahler, N., Weirich, S., und Henschel, S., Herausgeber, *IQB-Bildungstrend 2018: Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich*, Seiten 295–336. Münster: Waxmann.
- Hillmert, S. und Jacob, M. (2005). Institutionelle Strukturierung und inter-individuelle Variation. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 57(3):414–442.
- Holbrook, A. L., Krosnick, J. A., Moore, D., und Tourangeau, R. (2007). Response order effects in dichotomous categorical questions presented orally: The impact of question and respondent attributes. *Public Opinion Quarterly*, 71(3):325–348.
- Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30(2):161–172.
- Hopf, D. (2005). Zweisprachigkeit und Schulleistung bei Migrantenkinder. *Zeitschrift für Pädagogik*, 51(2):236–251.
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications*. New York: Routledge, Taylor & Francis.

- Janssen-van Dieten, A.-M. (1989). The development of a test of Dutch as a second language: The validity of self-assessment by inexperienced subjects. *Language Testing*, 6(1):30–46.
- Kaminska, O. und Foulsham, T. (2016). Eye-tracking social desirability bias. *Bulletin of Sociological Methodology*, 130(1):73–89.
- Kosyakova, Y., Kristen, C., und Spörlein, C. (2022). The dynamics of recent refugees' language acquisition: How do their pathways compare to those of other new immigrants? *Journal of Ethnic and Migration Studies*, 48(5):989–1012.
- Kosyakova, Y. und Laible, M.-C. (2024). Importance of personality traits for destination-language acquisition: Evidence for refugees in Germany. *International Migration Review*, 58(1):347–385.
- Kristen, C. (2008a). Primary school choice and ethnic school segregation in German elementary schools. *European Sociological Review*, 24(4):495–510.
- Kristen, C. (2008b). Schulische Leistungen von Kindern aus türkischen Familien am Ende der Grundschulzeit: Befunde aus der IGLU-Studie. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 48:230–251.
- Kristen, C. (2019). Language assimilation and the education of immigrant students. In Becker, R., Herausgeber, *Research handbook on the sociology of education*, Seiten 519–534. Cheltenham: Edward Elgar Publishing.
- Kristen, C. und Dollmann, J. (2010). Sekundäre Effekte der ethnischen Herkunft: Kinder aus türkischen Familien am ersten Bildungsübergang. In Becker, B. und Reimer, D., Herausgeber, *Vom Kindergarten bis zur Hochschule: Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., und Will, G. (2011). The education of migrants and their children across the life course. In Blossfeld, H.-P., Roßbach, H.-G., und von Maurice, J., Herausgeber, *Education as a lifelong process: The German National Educational Panel Study (NEPS)*, Seiten 121–137. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Kristen, C. und Olczyk, M. (2013). Ethnische Einbettung und Bildungserfolg. In Becker, R. und Schulze, A., Herausgeber, *Bildungskontexte: Strukturelle Voraussetzungen und Ursachen ungleicher Bildungschancen*, Seiten 353–403. Wiesbaden: Springer VS.
- Kristen, C., Olczyk, M., und Will, G. (2016). Identifying immigrants and their descendants in the National Educational Panel Study. In Blossfeld, H.-P., von Maurice, J., Bayer, M., und Skopek, J., Herausgeber, *Methodological issues of longitudinal surveys: The example of the National Educational Panel Study*, Seiten 195–211. Wiesbaden: Springer VS.

- Kristen, C. und Seuring, J. (2021). Destination-language acquisition of recently arrived immigrants: Do refugees differ from other immigrants? *Journal for Educational Research Online*, 13(1):128–156.
- Kristen, C., Seuring, J., und Stanat, P. (2019). Muster und Bedingungen des Erwerbs und Erhalts herkunftssprachlicher Kompetenzen. *Zeitschrift für Erziehungswissenschaft*, 22(S1):143–167.
- Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure: Ein allgemeines Modell des Handelns. *Zeitschrift für Soziologie*, 34(5):344–363.
- Kroneberg, C. (2011). *Die Erklärung sozialen Handelns: Grundlagen und Anwendung einer integrativen Theorie*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5(3):213–236.
- Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50(1):537–567.
- Krosnick, J. A., Holbrook, A. L., Berent, M. K., Carson, R. T., Hanemann, W. M., Kopp, R. J., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Moody, W. R., Green, M. C., und Conaway, M. (2002). The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? *Public Opinion Quarterly*, 66(3):371–403.
- Krosnick, J. A., Narayan, S., und Smith, W. R. (1996). Satisficing in surveys: Initial evidence. *New Directions for Evaluation*, 70:29–44.
- Kruger, J. und Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one’s own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6):1121–1134.
- Kwang, T. und Swann, W. B. (2010). Do people embrace praise even when they feel unworthy? A review of critical tests of self-enhancement versus self-verification. *Personality and Social Psychology Review*, 14(3):263–280.
- LeBlanc, R. und Painchaud, G. (1985). Self-assessment as a second language placement instrument. *TESOL Quarterly*, 19(4):673–687.
- Lenhard, A., Lenhard, W., Segerer, R., und Suggate, S. (2015). *Peabody Picture Vocabulary Test (PPVT) – 4. Ausgabe (Deutsche Version)*. Frankfurt am Main: Pearson Assessments.
- Li, M. und Zhang, X. (2021). A meta-analysis of self-assessment and language performance in language testing and assessment. *Language Testing*, 38(2):189–218.

- Lin, L.-M., Moore, D., und Zabrucky, K. M. (2001). An assessment of students' calibration of comprehension and calibration of performance using multiple measures. *Reading Psychology, 22*(2):111–128.
- Mabe, P. A. und West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3):280–296.
- Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal, 23*(1):129–149.
- Marsh, H. W. (1987). The big-fish-little-pond effect on academic self-concept. *Journal of Educational Psychology, 79*(3):280–295.
- Marsh, H. W. und Parker, J. W. (1984). Determinants of student self-concept: Is it better to be a relatively large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology, 47*(1):213–231.
- Marsh, H. W., Trautwein, U., Lüdtke, O., und Köller, O. (2008). Social comparison and big-fish-little-pond effects on self-concept and other self-belief constructs: Role of generalized and specific others. *Journal of Educational Psychology, 100*(3):510–524.
- Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics, 32*(3):385–397.
- Mouw, T. und Xie, Y. (1999). Bilingualism and the academic achievement of first- and second-generation Asian Americans: Accommodation with or without assimilation? *American Sociological Review, 64*(2):232–252.
- Müller, A. G. und Stanat, P. (2006). Schulischer Erfolg von Schülerinnen und Schülern mit Migrationshintergrund: Analysen zur Situation von Zuwanderern aus der ehemaligen Sowjetunion und aus der Türkei. In Baumert, J., Stanat, P., und Watermann, R., Herausgeber, *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*, Seiten 221–255. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Nauck, B. (2007). Immigrant families in Germany: Family change between situational adaptation, acculturation, segregation and remigration. *Zeitschrift für Familienforschung, 19*(1):34–54.
- Olczyk, M. (2018). *Ethnische Einbettung und schulischer Erfolg: Zur Bedeutung ethnisch segregierter Lebenswelten für den Bildungserwerb von Kindern mit Zuwanderungshintergrund*. Wiesbaden: Springer VS.
- Olczyk, M., Seuring, J., Will, G., und Zinn, S. (2016). Migranten und ihre Nachkommen im deutschen Bildungssystem: Ein aktueller Überblick. In Diehl, C., Hunkler, C., und Kristen, C., Herausgeber, *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten*, Seiten 33–70. Wiesbaden: Springer VS.

- Paetsch, J., Felbrich, A., und Stanat, P. (2015). Der Zusammenhang von sprachlichen und mathematischen Kompetenzen bei Kindern mit Deutsch als Zweitsprache. *Zeitschrift für Pädagogische Psychologie*, 29(1):19–29.
- Paulhus, D. (2003). Self-presentation measurement. In Fernandez-Ballesteros, R., Herausgeber, *Encyclopedia of psychological assessment*, Seiten 858–861. Thousand Oaks: Sage Publications.
- Peetsma, T., Van Der Veen, I., Koopman, P., und Van Schooten, E. (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement*, 17(3):275–302.
- Peirce, B. N., Swain, M., und Hart, D. (1993). Self-assessment, French immersion, and locus of control. *Applied Linguistics*, 14(1):25–42.
- Peyer, E. und Kaiser, I. (2010). Subjektive und objektive Messgrößen zur Erforschung des fremdsprachlichen Leseverstehens: Selbsteinschätzung und Testergebnisse im Vergleich. *Zeitschrift für Fremdsprachenforschung*, 21(1):79–105.
- Portes, A. und Hao, L. (2002). The price of uniformity: Language, family and personality adjustment in the immigrant second generation. *Ethnic and Racial Studies*, 25(6):889–912.
- Portes, A. und Zhou, M. (1993). The new second generation: Segmented assimilation and its variants. *Annals of the American Academy of Political and Social Sciences*, 530(1):74–96.
- Ramm, G., Walter, O., Heidemeier, H., und Prenzel, M. (2005). Soziokulturelle Herkunft und Migration im Ländervergleich. In Prenzel, M., Baumert, J., Blum, W., Lehmann, R. H., Leutner, D., Neubrand, M., Pekrun, R., Rost, J., und Schiefele, U., Herausgeber, *PISA 2003: Der zweite Vergleich der Länder in Deutschland – Was wissen und können Jugendliche?*, Seiten 269–297. Münster: Waxmann.
- Rauch, D. P., Jurecka, A., und Hesse, H.-G. (2010). Für den Drittspracherwerb zählt auch die Lesekompetenz in der Herkunftssprache: Untersuchung der Türkisch-, Deutsch- und Englisch-Lesekompetenz bei Deutsch-Türkisch bilingualen Schülern. In Allemann-Ghionda, C., Stanat, P., und Göbel, K., Herausgeber, *Migration, Identität, Sprache und Bildungserfolg*, Seiten 78–100. Weinheim: Beltz.
- Raudenbush, S. und Bryk, A. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks: Sage Publications.
- Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experiential factors. *Language Testing*, 15(1):1–20.
- Roßmann, J. (2017). *Satisficing in Befragungen: Theorie, Messung und Erklärung*. Wiesbaden: Springer VS.

- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York: Wiley.
- Salentin, K. und Wilkening, F. (2003). Ausländer, Eingebürgerte und das Problem einer realistischen Zuwanderer-Integrationsbilanz. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 55(2):278–298.
- Schaeffer, N. (2004). Conversation with a purpose – or conversation? Interaction in the standardized interview. In Biemer, P. P., Groves, R., Lyberg, L., Mathiowetz, N., und Sudman, S., Herausgeber, *Measurement errors in surveys*, Seiten 367–391. Hoboken: Wiley.
- Scheele, A. F., Leseman, P. P. M., und Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, 31(1):117–140.
- Schild, A. (2024). *Selbsteinschätzungen sprachlicher Kompetenzen im Deutschen jugendlicher Flüchtlinge*. Bamberg: Otto-Friedrich-Universität.
- Schunck, R. (2016). Cluster size and aggregated level 2 variables in multilevel models. *Methods, Data, Analyses*, 10(1):97–108.
- Schwarz, N. (2007). Cognitive aspects of survey methodology. *Applied Cognitive Psychology*, 21(2):277–287.
- Schwarz, N. und Bienias, J. (1990). What mediates the impact of response alternatives on frequency reports of mundane behaviors? *Applied Cognitive Psychology*, 4(1):61–72.
- Schwarz, N., Hippler, H.-J., und Noelle-Neumann, E. (1992). A cognitive model of response-order effects in survey measurement. In Schwarz, N. und Sudman, S., Herausgeber, *Context effects in social and psychological research*, Seiten 187–201. New York: Springer.
- Schwarz, N., Knäuper, B., Hippler, H.-J., Noelle-Neumann, E., und Clark, L. (1991). Rating scales: Numeric values may change the meaning of scale labels. *Public Opinion Quarterly*, 55(4):570–582.
- Schwippert, K., Hornberg, S., Freiberg, M., und Stubbe, T. (2007). Lesekompetenzen von Kindern mit Migrationshintergrund im internationalen Vergleich. In Bos, W., Hornberg, S., Arnold, K.-H., Faust, G., Fried, L., Lankes, E.-M., Schwippert, K., und Valtin, R., Herausgeber, *IGLU 2006: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Seiten 249–269. Münster: Waxmann.
- Schwippert, K., Wendt, H., und Tarelli, I. (2012). Lesekompetenz von Schülerinnen und Schülern mit Migrationshintergrund. In Bos, W., Tarelli, I., Bremerich-Vos, A., und Schwippert, K., Herausgeber, *IGLU 2011: Lesekompetenzen von Grundschulkindern in Deutschland im internationalen Vergleich*, Seiten 191–207. Münster: Waxmann.

- Segeritz, M., Walter, O., und Stanat, P. (2010). Muster des schulischen Erfolgs von jugendlichen Migranten in Deutschland: Evidenz für segmentierte Assimilation? *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 62(1):113–138.
- Shajek, A., Lüdtke, O., und Stanat, P. (2006). Akademische Selbstkonzepte bei Jugendlichen mit Migrationshintergrund. *Unterrichtswissenschaft*, 34(2):125–145.
- Shavelson, R. J., Hubner, J. J., und Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, 46(3):407–441.
- Simon, E. und Neuwöhner, U. (2011). Medien und Migranten 2011: Zielsetzung, Konzeption und Basisdaten einer repräsentativen Untersuchung der ARD/ZDF-Medienkommission. *Media Perspektiven*, 10:458–470.
- Simon, H. A. (1957). *Models of man: Social and rational*. New York: Wiley.
- Skopek, J., Pink, S., und Bela, D. (2013). *Starting cohort 4: Grade 9 (SC4). Scientific use file, version 1.1.0. Data manual*. Bamberg: University of Bamberg, National Educational Panel Study.
- Spolsky, B. (1989). *Conditions for second language learning: Introduction to a general theory*. Oxford: Oxford University Press.
- Stanat, P. (2006). Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der Zusammensetzung der Schülerschaft. In Baumert, J., Stanat, P., und Wattermann, R., Herausgeber, *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit*, Seiten 189–219. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Stanat, P. (2008). Heranwachsende mit Migrationshintergrund im deutschen Bildungswesen. In Cortina, K. S., Baumert, J., Leschinsky, A., Mayer, K. U., und Trommer, L., Herausgeber, *Das Bildungswesen in der Bundesrepublik Deutschland: Strukturen und Entwicklungen im Überblick*, Seiten 691–705. Hamburg: Rowohlt.
- Stanat, P., Rauch, D. P., und Segeritz, M. (2010a). Schülerinnen und Schüler mit Migrationshintergrund. In Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., und Stanat, P., Herausgeber, *PISA 2009: Bilanz nach einem Jahrzehnt*, Seiten 200–230. Münster: Waxmann.
- Stanat, P., Schwippert, K., und Gröhlich, C. (2010b). Der Einfluss des Migrantenanteils in Schulklassen auf den Kompetenzerwerb. Längsschnittliche Überprüfung eines umstrittenen Effekts. In Allemann-Ghionda, C., Stanat, P., Göbel, K., und Röhner, C., Herausgeber, *Migration, Identität, Sprache und Bildungserfolg*, Seiten 147–164. Weinheim: Beltz.
- StataCorp (2015). *Stata statistical software: Release 14*. College Station: StataCorp LP.

- StataCorp (2021). Stata statistical software: Release 17. College Station: StataCorp LLC.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, 21:261–302.
- Steinhauer, H. und Zinn, S. (2016). *NEPS technical report for weighting: Weighting the sample of starting cohort 4 of the National Educational Panel Study (wave 1 to 6)*. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.
- Stocké, V. (2004). Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit: Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie*, 33(4):303–320.
- Strack, F. und Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8(3):220–247.
- Strack, F. und Martin, L. (1987). Thinking, judging, and communicating: A process account of context effects in attitude surveys. In Hippler, H.-J., Schwarz, N., und Sudman, S., Herausgeber, *Social information processing and survey methodology*, Seiten 123–148. New York: Springer.
- Strickhouser, J. E. und Zell, E. (2015). Self-evaluative effects of dimensional and social comparison. *Journal of Experimental Social Psychology*, 59:60–66.
- Strobel, B. (2016). Does family language matter? The role of foreign language use and family social capital in the educational achievement of immigrant students in Germany. *Ethnic and Racial Studies*, 39(14):2641–2663.
- Strobel, B. und Kristen, C. (2015). Erhalt der Herkunftssprache? – Muster des Sprachgebrauchs in Migrantenfamilien. *Zeitschrift für Erziehungswissenschaft*, 18(1):125–142.
- Sudman, S. und Bradburn, N. (1974). *Response effects in surveys: A review and synthesis*. Chicago: Aldine Publishing Co.
- Sudman, S., Bradburn, N., und Schwarz, N. (1996). *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco: Jossey-Bass.
- Swann, W. B. (1983). Self-verification: Bringing social reality into harmony with the self. *Social Psychological Perspectives on the Self*, 2:33–66.
- Swann, W. B. und Read, S. J. (1981). Self-verification processes: How we sustain our self-conceptions. *Journal of Experimental Social Psychology*, 17(4):351–372.
- Taylor, S. (1981). The interface of cognitive and social psychology. In Harvey, J. H., Herausgeber, *Cognition, social behavior, and the environment*, Seiten 189–211. Hillsdale: Erlbaum.

- Thijs, J., Verkuyten, M., und Helmond, P. (2010). A further examination of the big-fish-little-pond effect: Perceived position in class, class size, and gender comparisons. *Sociology of Education*, 83(4):333–345.
- Tourangeau, R. (1984). Cognitive sciences and survey methods. In Jabine, T., Straf, M. L., Tanur, J. M., und Tourangeau, R., Herausgeber, *Cognitive aspects of survey methodology: Building a bridge between disciplines*, Seiten 73–100. Washington D.C.: National Academy Press.
- Tourangeau, R. und Rasinski, K. A. (1988). Cognitive processes underlying context effects in attitude measurement. *Psychological Bulletin*, 103(3):299–314.
- Tourangeau, R., Rips, L. J., und Rasinski, K. A. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press.
- Tourangeau, R. und Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, 60(2):275–304.
- Tourangeau, R. und Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin*, 133(5):859–883.
- Trautwein, U., Lüdtke, O., Marsh, H. W., und Nagy, G. (2009). Within-school social comparison: How students perceive the standing of their class predicts academic self-concept. *Journal of Educational Psychology*, 101(4):853–866.
- Truxillo, D. M., Seitz, R., und Bauer, T. N. (2008). The role of cognitive ability in self-efficacy and self-assessed test performance. *Journal of Applied Social Psychology*, 38(4):903–918.
- Van de Werfhorst, H. G. und Van Tubergen, F. (2007). Ethnicity, schooling, and merit in the Netherlands. *Ethnicities*, 7(3):416–444.
- Van Ewijk, R. und Sleegers, P. (2010). Peer ethnicity and achievement: A meta-analysis into the compositional effect. *School Effectiveness and School Improvement*, 21(3):237–265.
- Van Tubergen, F. und Mentjox, T. (2014). Minority language proficiency of adolescent immigrant children in England, Germany, the Netherlands, and Sweden. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 66(1):241–262.
- Vannette, D. und Krosnick, J. A. (2014). A comparison of mindlessness and survey satisficing. In Ie, A., Ngnoumen, C. T., und Langer, E. J., Herausgeber, *The Wiley Blackwell handbook of mindfulness (Volume I)*, Seiten 312–327. Chichester: Wiley.
- Veerman, G.-J. M. und Dronkers, J. (2016). Ethnic composition and school performance in the secondary education of Turkish migrant students in seven countries and 19 European educational systems. *International Migration Review*, 50(3):537–567.

- Von Hippel, P. T. (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods & Research*, 49(3):699–718.
- Walter, O. (2009a). Ethno-linguale Kompositionseffekte in neunten Klassen: Befunde aus der Klassenstichprobe von PISA 2006. In Prenzel, M. und Baumert, J., Herausgeber, *Vertiefende Analysen zu PISA 2006*, Seiten 169–184. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Walter, O. (2009b). Herkunftsassoziierte Disparitäten im Lesen, der Mathematik und den Naturwissenschaften: Ein Vergleich zwischen PISA 2000, PISA 2003 und PISA 2006. In Prenzel, M. und Baumert, J., Herausgeber, *Vertiefende Analysen zu PISA 2006*, Seiten 149–169. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wells, L. E. und Sweeney, P. D. (1986). A test of three models of bias in self-assessment. *Social Psychology Quarterly*, 49(1):1–10.
- White, I. R., Royston, P., und Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Statistics in Medicine*, 30(4):377–399.
- Widhiarso, W. (2014). Relationship between cognitive ability and accurate self-reporting. *Journal of Educational Sciences and Psychology*, 14(2):85–95.
- Wilson, K. M. (1999). *Validity of global self-ratings of ESL speaking proficiency based on an FSI/ILR-referenced scale (ETS Research Report Series 99-13)*. Princeton: Statistics and Research Division.
- Worbs, S. (2010). *Mediennutzung von Migranten in Deutschland (Integrationsreport 8)*. Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Xia, M., Poorthuis, A. M. G., und Thomaes, S. (2024). Children’s overestimation of performance across age, task, and historical time: A meta-analysis. *Child Development*, 95(3):1001–1022.
- Zander, L., Kreutzmann, M., und Hannover, B. (2017). Peerbeziehungen im Klassenzimmer. *Zeitschrift für Erziehungswissenschaft*, 20(3):353–386.
- Zell, E. und Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science*, 9(2):111–125.

Anhang

Tabelle A.1: Lineare Regression getesteter und selbsteingeschätzter L2-Kompetenzen auf das globale Selbstwertgefühl

	Model 1: L2-Test ^a	Model 2: Globale Selbstein- schätzung ^a	Model 3: Einschätzung Testleistung ^a	
Herkunftsgruppe ^b				
Ehem. SU, 1. Gen.	0.17 (0.05)	*** (0.06)	0.01 (0.06)	0.07 (0.06)
Ehem. SU, 2./3. Gen.	0.16 (0.05)	*** (0.07)	-0.13 (0.07)	-0.06 (0.07)
Türkei, 1. Gen.	0.04 (0.08)	-0.29 (0.11)	** (0.11)	-0.00 (0.11)
Polen, 1. Gen.	0.24 (0.07)	*** (0.08)	0.16 (0.08)	* (0.08)
Polen, 2./3. Gen.	0.11 (0.09)	-0.18 (0.12)	-0.07 (0.11)	-0.07 (0.11)
Other, 1. Gen.	0.14 (0.04)	*** (0.05)	-0.04 (0.05)	0.04 (0.05)
Other, 2./3. Gen.	0.06 (0.04)	-0.32 (0.05)	*** (0.05)	-0.10 (0.05)
Männlich	-0.20 (0.02)	*** (0.03)	-0.21 (0.03)	*** (0.03)
Allgemeine kognitive Fähigkeiten	0.10 (0.01)	*** (0.01)	0.00 (0.01)	0.02 (0.01)
Bildung Eltern ^c				
Mittlerer Abschluss	-0.03 (0.04)	0.01 (0.05)	0.01 (0.05)	0.03 (0.05)
Höherer Abschluss	-0.07 (0.04)	0.07 (0.06)	0.07 (0.06)	0.06 (0.05)
HISEI ^a				
Anzahl Bücher	0.08 (0.01)	*** (0.01)	0.05 (0.01)	** (0.02)
Identifikation mit DE	0.06 (0.02)	** (0.03)	0.11 (0.03)	** (0.03)
Eigenethnische Nachbarschaft	-0.05 (0.01)	*** (0.01)	0.01 (0.01)	-0.00 (0.01)
L2-Gebrauch in Familie	0.03 (0.02)	-0.05 (0.03)	-0.05 (0.03)	-0.05 (0.03)
L2-Gebrauch mit Gleichaltrigen	0.15 (0.03)	*** (0.04)	0.22 (0.04)	*** (0.04)
L2 Medienkonsum	-0.04 (0.03)	0.23 (0.05)	*** (0.05)	0.02 (0.05)
Globales Selbstwertgefühl ^a	0.03 (0.02)	0.18 (0.02)	*** (0.02)	0.07 (0.02)
R ²	0.42	0.18		0.11
N	4,078	4,078		4,078

Anmerkung: Robuste Standardfehler in Klammern; zusätzlich für Schulform kontrolliert.
 *** p<0.001, ** p<0.01, * p<0.05. ^a Standardisierte Variablen. ^b Referenzkategorie:
 Türkei, 2./3. Generation. ^c Referenzkategorie: Kein/niedriger Abschluss.
 Die Analysen beruhen auf den Daten und Berechnungen aus Tabelle 6.4 in Kapitel
 6 und erweitern diese um die Variable zum globalen Selbstwertgefühl.

Tabelle A.2: Lineare Regression getesteter und selbsteingeschätzter L2-Kompetenzen auf die Big Five Persönlichkeitsmerkmale

	Model 1: L2-Test ^a	Model 2: Globale Selbstein- schätzung ^a	Model 3: Einschätzung Testleistung ^a	
Herkunftsgruppe ^b				
Ehem. SU, 1. Gen.	0.16 (0.05)	** (0.06)	-0.01 (0.06)	0.05 (0.06)
Ehem. SU, 2./3. Gen.	0.14 (0.05)	** (0.07)	-0.14 (0.07)	* (0.07)
Türkei, 1. Gen.	0.04 (0.08)		-0.32 (0.11)	** (0.11)
Polen, 1. Gen.	0.23 (0.07)	*** (0.07)	0.14 (0.07)	0.09 (0.08)
Polen, 2./3. Gen.	0.07 (0.09)		-0.23 (0.13)	-0.10 (0.11)
Other, 1. Gen.	0.12 (0.04)	** (0.05)	-0.06 (0.05)	0.03 (0.05)
Other, 2./3. Gen.	0.05 (0.05)		-0.34 (0.07)	*** (0.07)
Männlich	-0.19 (0.03)	*** (0.04)	-0.15 (0.04)	*** (0.03)
Allgemeine kognitive Fähigkeiten	0.10 (0.01)	*** (0.01)	0.00 (0.01)	0.02 (0.01)
Bildung Eltern ^c				
Mittlerer Abschluss	-0.03 (0.04)		0.02 (0.05)	0.03 (0.05)
Höherer Abschluss	-0.06 (0.04)		0.08 (0.06)	0.06 (0.05)
HISEI ^a				
	-0.00 (0.02)		0.01 (0.02)	0.04 (0.03)
Anzahl Bücher	0.08 (0.01)	*** (0.01)	0.04 (0.01)	** (0.02)
Identifikation mit DE	0.06 (0.02)	** (0.03)	0.10 (0.03)	* (0.03)
Eigenethnische Nachbarschaft	-0.05 (0.01)	*** (0.01)	0.01 (0.01)	-0.00 (0.01)
L2-Gebrauch in Familie	0.03 (0.02)		-0.04 (0.03)	-0.06 (0.03)
L2-Gebrauch mit Gleichaltrigen	0.15 (0.03)	*** (0.04)	0.21 (0.04)	*** (0.04)
L2 Medienkonsum	-0.04 (0.03)		0.25 (0.05)	*** (0.05)
Big Five:				
Offenheit für Erfahrungen ^a	0.05 (0.01)	*** (0.02)	0.05 (0.02)	** (0.02)
Gewissenhaftigkeit ^a	-0.04 (0.01)	** (0.02)	0.04 (0.02)	* (0.02)
Extraversion ^a	-0.00 (0.01)		0.05 (0.02)	* (0.02)
Verträglichkeit ^a	0.01 (0.01)		0.01 (0.02)	-0.03 (0.02)
Neurotizismus ^a	-0.00 (0.01)		-0.05 (0.02)	* (0.02)
R ²	0.42		0.16	0.12
N	4,078		4,078	4,078

Anmerkung: Robuste Standardfehler in Klammern; zusätzlich für Schulform kontrolliert.

*** p<0.001, ** p<0.01, * p<0.05. ^a Standardisierte Variablen. ^b Referenzkategorie:

Türkei, 2./3. Generation. ^c Referenzkategorie: Kein/niedriger Abschluss.

Die Analysen beruhen auf den Daten und Berechnungen aus Tabelle 6.4 in Kapitel 6 und erweitern diese um die fünf Variablen der Big Five.