Empirical Studies on Migration

From Language Learning to Integration Processes

A thesis submitted in fulfillment of the requirements
for the degree of Doctor Rerum Politicarum
from the Faculty of Economic Sciences
University of Göttingen

by
Ann-Marie Sommerfeld
born in Konstanz

Göttingen, 2023

First Supervisor: Prof. Dr. Robert Schwager

Second Supervisor: Prof. Dr. Silke Uebelmesser

Third Supervisor: Prof. Dr. Krisztina Kis-Katos

To Claudia. For *everything*.

Abstract

Research over the past decades has documented growing migration flows worldwide; and specifically increased immigration into developed countries of the global north. Against this background, this thesis investigates the role of social context, language learning and, more broadly, education in migration decisions and for assimilation in the host country, with a focus on Germany as a developed receiving country. Migration is conceptualized as a process; from the decision whether and when to learn a foreign language, over the considerations and plans to leave one's home country, to finally the integration process in the host country.

In the four main chapters of this thesis, first the reasons for foreign language learning are classified into human capital investment and consumption motive, and individual and country level determinants of those are analyzed (Chapter 2). Then, the macro level determinants of foreign language learning are investigated with a special focus on the certainty with which the return on investment can be realized through migration (Chapter 3). Individual level determinants of migration aspirations and intentions are analyzed (Chapter 4). Lastly, the effect of children's schooling on their parents' labor market outcomes and other integration measures is analyzed (Chapter 5).

Taken together, this thesis provides evidence that foreign language learning can stem from an investment motive or from a consumption motive, which relies on the idea that education holds consumption value itself, hence generates direct utility (Chapter 2). Crucially, Chapter 3 provides evidence that in the context of migration the timing of language learning depends on certainty of access to the destination country's labor market. While learning the German language abroad is strongly associated with migration from countries whose citizens have guaranteed access to the German labor market, language learning in Germany is positively associated with migration from countries whose citizens face uncertain access. Restricted access to the destination country is one potential barrier due to which migration desires do not realize, female gender is another (Chapter 4). In contrast, strong social ties can be drivers towards migration. This thesis provides further evidence on the idea of women being "tied movers" who are willing to follow their partner abroad, even against their own desire to stay. For men, potential drivers towards migration are education and professional opportunities. Lastly, Chapter 5 provides first evidence that primary schooling of the oldest child in the household positively affects their parents' labor market outcomes as well as other integration measures, like German language skills. An analysis of underlying mechanisms suggests that these results are driven by gains in disposable time and exposure to the German language and culture.

The findings of this thesis highlight the importance of addressing access barriers and opportunities of language learning, and considering family and gender context in migration and integration. Policy recommendations derived from the results of this thesis include subsidies for foreign language learning and increased supply of language learning opportunities to target those seeking human capital investment, as well as policies to support the social and economic integration of migrants in their host country, specifically women.

Keywords: *language learning, new dataset, human capital investment, consumption, language skills, migration, labor market access, international migration, temporary migration, permanent migration, migration aspirations, migration intentions, assimilation, integration, education, schooling, early age, family*

Zusammenfassung

In den letzten Jahrzehnten hat die Forschung weltweit wachsende Migrationsbewegungen und insbesondere Zuwanderung in entwickelte Länder des globalen Nordens dokumentiert. Vor diesem Hintergrund untersucht diese Dissertation welche Rolle das soziale Umfeld, das Sprachenlernen und Bildung im weiteren Sinne bei Migrationsentscheidungen und Assimilation im Zielland spielen. Hierbei liegt der Schwerpunkt auf Deutschland als entwickeltes Einwanderungsland. Migration wird als Prozess verstanden: von der Entscheidung ob und wann eine Fremdsprache erlernt wird, über Überlegungen und Pläne das Heimatland zu verlassen bis hin zum Integrationsprozess im Aufnahmeland.

In den vier zentralen Kapiteln dieser Dissertation werden zuerst die Gründe für das Sprachenlernen in Investitions- und Konsummotive eingeteilt und die individuellen und länderspezifischen Determinanten dieser Motive analysiert (Kapitel 2). Im Anschluss werden die Determinanten des Sprachenlernens auf der Makroebene untersucht, mit besonderem Augenmerk auf die Sicherheit, mit der die Investitionsrendite durch Migration realisiert werden kann (Kapitel 3). Es werden die individuellen Determinanten von Migrationswunsch und Migrationsabsicht analysiert (Kapitel 4). Letztlich werden die Effekte des Schulbesuchs des ältesten Kindes im Haushalt auf die Arbeitsmarktergebnisse und andere Integrationsmaße ihrer Eltern analysiert (Kapitel 5).

Die vorliegende Arbeit liefert Belege, dass Sprachenlernen aus einem Investitionsmotiv oder, im Gegensatz zu den Annahmen der Humankapitaltheorie, aus einem Konsummotiv resultieren kann (Kapitel 2). Ersteres kann sowohl auf dem lokalen als auch auf dem ausländischen Arbeitsmarkt, d. h. im Kontext der Migration, eine positive Investitionsrendite erbringen. Kapitel 3 liefert Belege dafür, dass der Zeitpunkt des Sprachenlernens erheblich von der Gewissheit des Zugangs zum Arbeitsmarkt im Zielland abhängt. Während das Erlernen der Deutschen Sprache im Ausland stark mit der Migration aus Ländern verbunden ist, deren Staatsbürger einen gesicherten Zugang zum deutschen Arbeitsmarkt haben, ist das Erlernen der Sprache in Deutschland positiv mit der Migration aus Ländern verbunden, deren Staatsbürger kein sicherer Zugang gewährleistet ist. Der eingeschränkte Zugang zum Zielland ist eine mögliche Barriere aufgrund derer sich Migrationswünsche häufig nicht realisieren. Begünstigt werden kann Migration andererseits durch starke soziale Beziehungen ins Zielland. Die vorliegende Arbeit liefert weitere Belege für das Konzept von Frauen als "tied movers", die bereit sind ihrem Partner ins Ausland zu folgen, selbst dann wenn sie lieber in der Heimat verbleiben würden. Für Männer hingegen sind Bildung und berufliche Chancen potenzielle Treiber von Migration. Schlussendlich liefert Kapitel 5 den ersten Beleg dafür, dass sich der Schulbesuch des ältesten Kindes im Haushalt positiv auf die Arbeitsmarktergebnisse und weitere Integrationsmaße, wie Sprachkenntnisse, dessen Eltern auswirkt. Eine Analyse der Wirkungsmechanismen deutet darauf hin, dass diese Ergebnisse einerseits auf den Zugewinn an verfügbarer Zeit nach der Einschulung des Kindes und andererseits den Kontakt mit der deutschen Sprache und Kultur zurückzuführen sind.

Die Ergebnisse der vorliegenden Arbeit unterstreichen die Bedeutung von Zugangsbarrieren und Möglichkeiten des Sprachenlernens sowie des familiären Kontext und Unterschiede zwischen Geschlechtern in Migrationsentscheidungen und Integration. Mögliche politische Handlungsempfehlungen, die aus den Ergebnissen dieser Arbeit abgeleitet werden können, sind beispielsweise Subventionen für das Erlernen von Fremdsprachen und ein größeres Angebot an Sprachlernmöglichkeiten, sowie Maßnahmen zur Unterstützung der sozialen und wirtschaftlichen Integration von Migranten in ihrem Gastland, insbesondere Frauen.

Acknowledgments

First and foremost, I would like to thank my supervisor Robert Schwager for offering me this amazing opportunity, for giving me his valuable advice and constant support all while leaving me a free hand over my own research. I never really understood the notion of "roots and wings" but I think I do now.

Furthermore, I would like to express my sincere gratitude to

Silke Uebelmesser, who potentially is the only person in possession of a time-turner. Otherwise I cannot explain all the time, effort and passion she puts in everything she does; and she put in me.

Krisztina Kis-Katos, for being a supportive and insightful advisor.

My co-authors Matthias Huber, Till Nikolka, Panu Poutvaara, and Severin Weingarten for excellent cooperation, enriching discussions and their (relentless) feedback.

My colleagues and friends, Ansgar Quint, Florian Rottner, Raphael Brade, Kamila Danilowicz-Goesele, Julian Derewjankin, and Monika Jackmann for the wonderful atmosphere and working environment, and enlightening conversations over lunch and (horrible) cafeteria coffee. Special thanks to Fabian Koenings and Jana Schuetz for so wonderfully including me into their work group, their encouragement and feedback.

Markus Pawellek, Johannes Kremer, Vivienne Schuenemeyer, Luise Krause, Cassandra Bielak and Nicolle Sommerfeld for being the most supportive, encouraging, and patient friends I could have hoped for during these years. And to Martin Fieles, for staying a reliable friend and support after our master's.

My mother, who is forever strong and brave, for bringing us up on nothing. And for politely telling a primary school prinicipal her two cents about ability tracking, without which this thesis would never have been possible.

My father, who would be immensely proud.

Finally, my endless gratitude goes to my sister Claudia. For everything.

Contents

List of Tables iv

List of Figures viii

List of Abbreviations x

1 General Introduction 1
- 1.1 Language Learning, Migration Decisions and Integration Outcomes . . 2
- 1.2 Thesis Outline 6

2 Language Learning: Human Capital Investment or Consumption? 10
- 2.1 Introduction 10
- 2.2 Survey Set-Up and Data 12
  - 2.2.1 Survey Design 13
  - 2.2.2 Descriptives and (Limits to) Respresentativity 14
- 2.3 Motives of Language Learning 17
  - 2.3.1 Conceptual Framework and Operationalization 17
  - 2.3.2 Expected Associations 19
  - 2.3.3 Graphical Illustrations 21
- 2.4 Estimation Strategy 26
- 2.5 Estimation Results 27
  - 2.5.1 Determinants of the Human Capital Motive 27
  - 2.5.2 Determinants of the Professional Use of German 31
- 2.6 Conclusion 36
- 2.A Appendix: Tables 38
- 2.B Appendix: Figures 50

3 Macro-Level Analysis of Language Learning and Migration 55
- 3.1 Introduction 55
- 3.2 Data and Hypotheses 59
  - 3.2.1 Dependent Variables: The Goethe Institute Dataset 59
  - 3.2.2 Explanatory Variables 63
  - 3.2.3 Hypotheses 64
- 3.3 Empirical Setup 68
- 3.4 Results 70
  - 3.4.1 Main Estimation Results 71
  - 3.4.2 Capitals and Large Cities 73
- 3.5 Country Heterogeneities and Identification Issues 75

3.5.1 OLS Estimations with Country Characteristics 75
3.5.2 Within-Between Estimations 76
3.5.3 Causality Issues 79
3.6 Conclusions 83
3.A Appendix A 85
3.B Appendix B 91
3.B.1 Language Course Pricing 91
3.B.2 Sample Shrinkage due to Missings 91
3.B.3 Graphical Illustration ('Abroad' Specification): Averages per Institutes 92
3.B.4 List of Countries and Institutes 93

4 Migration Aspirations and Intentions 96
4.1 Introduction 96
4.2 Measuring Potential Migration 98
4.2.1 Migration Aspirations 99
4.2.2 Migration Intentions 99
4.2.3 Combining Migration Aspirations and Intentions 99
4.2.4 Determinants of Aspirations and Intentions 103
4.3 Data and Descriptive Statistics 104
4.3.1 Survey Design 104
4.3.2 Descriptive Statistics 108
4.4 Estimation Strategy 113
4.4.1 Estimating Aspirations and Intentions to Migrate 113
4.4.2 Conditionality and Differences 114
4.5 Results 115
4.5.1 Aspirations and Intentions to Migrate 115
4.5.2 Intentions to Migrate Temporarily and Permanently 118
4.5.3 Conditionality and Differences 118
4.6 Conclusion 121
4.A Appendix: Descriptive Statistics and Variable Descriptions 122
4.B Appendix: Further Estimations and Robustness Checks 133

5 The Effect of Schooling on Parental Integration 145
5.1 Introduction 145
5.2 Institutional Background, Estimation Strategy and Data 148
5.2.1 Institutional Background 148
5.2.2 Data 149
5.2.3 Estimation Strategy 154
5.2.4 Descriptive Analysis 156
5.3 Results 159
5.3.1 Channel Analysis 161
5.3.2 Threats to Identification and Sensitivity Analysis 173

5.4 Conclusion 176
5.A Appendix 177

6 General Conclusions **205**
6.1 Policy Implications 205
6.2 Limitations and Future Research 207

Bibliography **209**

List of Tables

2.1 Country Characteristics and Response Rates 14
2.2 Descriptive Statistics: Means of Individual Characteristics by Country Groups 16
2.3 Human Capital Investment: Basic Specifications 28
2.4 Human Capital Investment: Subsample by Age, Gender and Education . 30
2.5 Professional Use: Basic Specifications 32
2.6 Professional Use: Subsamples by Age, Gender and Education 34
2.7 Professional Use: Subsamples by Consumption/Investment 35
2.A Variable Description: Individual Characteristics 38
2.B Variable Description: Country Characteristics 39
2.C Sample Shrinkage 39
2.D Descriptive Statistics: Means of Individual Characteristics by Country Groups (Sample Before Droppings) 40
2.E Descriptive Statistics: Means of Individual Characteristics by Countries 41
2.F Questions on (Main) Reasons for Learning German 42
2.G Tetrachoric Correlation Coefficients 43
2.H Tetrachoric Correlation Coefficients for Categories 44
2.I Human Capital Investment: Basic Specifications with Country FE Interacted with Gender and Age 45
2.J Human Capital Investment: Subsamples by Country Characteristics . . 46
2.K Professional Use: Subsamples by Country Characteristics 47
2.L Human Capital Investment: Basic Specifications without *Education* as Motive and with *Education* as Consumption Motive 48
2.M Professional Use: Basic Specifications without *Education* as Motive and with *Education* as Consumption Motive 49

3.1 Descriptive Statistics ‘Abroad’ 64
3.2 Descriptive Statistics ‘Germany’ 65
3.3 Stepwise Estimation Results for Exam Participation Abroad 72
3.4 Stepwise Estimation Results for Course Participation in Germany 73
3.5 Estimation Results: Largest and Other Cities 74
3.6 Estimation Results: Within-between Random effects Specification . . . 78
3.7 IV Estimation Results for Exam Participation Abroad 81
3.8 IV Estimation Results for Course Participation in Germany 82
3.A Variables Description 86
3.B Descriptive Statistics ‘Abroad’ OLS Specification 87
3.C Descriptive Statistics ‘Germany’ OLS Specification 88

3.D Estimation Results for Exam and Course Participation Abroad and in Germany, Free Move Countries 89
3.E Estimation Results for Exam and Course Participation Abroad and in Germany, Restricted Move Countries 90
3.F Prices of Language Courses at Goethe Institutes and Other Providers in 2015 91
3.G Sample Shrinkage due to Missings: ‘Abroad‘ Specification 91
3.H Sample Shrinkage due to Missings: ‘Germany‘ Specification 92

4.1 Participants per Country — GI Survey 106
4.2 Participants per University — Student Survey 107
4.3 Joint Distribution of Aspirations and Intentions 110
4.4 Main Motivations for Potential Migration by Gender 113
4.5 Aspirations and Intentions 116
4.6 Intentions by Aspirations 119
4.A Variables Description 126
4.B Variables Description (Continued) 127
4.C Joint Distribution of Aspirations and Intentions — Foreign-Born Respondents Only 127
4.D Descriptive Statistics 128
4.E Answered the Question on Migration Aspirations: Selection on Observable Characteristics — GI and Student Sample 129
4.F Joint Distribution of Aspirations and Intentions by Gender 130
4.G Preferred Destination Countries for Potential Migration by Origin Country 131
4.H Main Motivations for Potential Migration by Aspirations and Intentions 132
4.I Aspirations and Intentions — High-Income Countries 133
4.J Aspirations and Intentions — Middle-Income Countries 134
4.K Aspirations and Intentions — Women 135
4.L Aspirations and Intentions — Men 136
4.M Aspirations and Intentions — GI Sample, 6 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above) 137
4.N Aspirations and Intentions — GI Sample, 14 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above) 138
4.O Intentions by Aspirations — GI Sample, 6 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above) 139
4.P Intentions by Aspirations — GI Sample, 14 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above) 140
4.Q Intentions by Aspirations — Women 141

4.R Intentions by Aspirations — Men 142
4.S Intentions by Aspirations — High-Income Countries 143
4.T Intentions by Aspirations — Middle-Income Countries 144

5.1 Differences in Parental Characteristics and Integration Outcomes between Parents whose Oldest Child was Eligible and Not Eligible for Enrollment the Year They Turned Six 152
5.2 Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income 160
5.3 Estimates of Months of Schooling and Enrollment Timing on Parental Employment 161
5.4 Estimates of Months of Schooling and Enrollment Timing on Parental Worries about Personal Finances 162
5.5 Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day 163
5.6 Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours 164
5.7 Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income 166
5.8 Estimates of Months of Schooling and Enrollment Timing on Parental Self-Assessed Health 169
5.9 Estimates of Months of Schooling and Enrollment Timing on Parental Staying Intentions 170
5.10 Estimates of Enrollment Timing on Parental Self-Assessed German Language Abilities 171
5.A Sample Shrinkage due to Missings 177
5.B Variables Description 178
5.C Pre-Treatment Differences in Parental Characteristics and Integration Outcomes between Parents who Enrolled their Children According to Eligibility and Parents who Brought Forward or Postponed Enrollment . 181
5.D Differences in Parental Characteristics and Integration Outcomes between Parents whose Oldest Child was Eligible and Not Eligible for Enrollment the Year They Turned Six (Bandwidth ± 4) 182
5.E Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Formerly Unemployed 183
5.F Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Formerly Unemployed 184
5.G Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Women 185
5.H Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Women 186

5.I Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Women . . . 187
5.J Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Women . . . 188
5.K Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Women . . . 189
5.L Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Parents with only One Child . . . 190
5.M Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Parents with only One Child . . . 191
5.N Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Parents with only One Child . . . 192
5.O Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Parents with only One Child . . . 193
5.P Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Parents with only One Child . . . 194
5.Q Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Germans . . . 195
5.R Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Germans . . . 196
5.S Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Germans . . . 197
5.T Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Germans . . . 198
5.U Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Germans . . . 199
5.V Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Households with only Migrants . . . 200
5.W Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Households with only Migrants . . . 201
5.X Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Households with only Migrants . . . 202
5.Y Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Households with only Migrants . . . 203
5.Z Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Households with only Migrants . . . 204

List of Figures

2.1 Categorization of Reasons to Learn Languages 18
2.2 Main Reasons (left) and Categories (right) for Learning German (n=4891) 22
2.3 Human Capital and Consumption Motives by Countries for Age, Gender and Education Subgroups 24
2.4 Use of German on the Labor Market by Countries for Age, Gender and Education Subgroups 25
2.A Professional Use of German: 1=very unlikely, 5=very likely (n=4891) . 50
2.B Main Reasons for Learning German by Countries (n=4891) 51
2.C Professional Use of German: Subsamples by Countries and Economic Motive 52
2.D Human Capital Investment: Subsamples by Country Groups According to Country Characteristics 53
2.E Professional Use: Subsamples by Country Groups According to Country Characteristics 54

3.1 Germany Specification – Averages per Continent and Year 60
3.2 Abroad Specification – Averages per Countries per Continent and Year . 61
3.A GI in Germany 85
3.B Abroad Specification – Averages per Institutes per Continent and Year . 92

4.1 Combinations of Aspirations and Intentions 101
4.2 Percentage of Respondents with Migration Aspirations — GWP, GI and Student Sample 109
4.A Percentage of Respondents with Migration Aspirations by Age Groups — GWP and GI Sample 122
4.B Percentage of Respondents with Migration Aspirations — GWP, GI Sample and Student Sample, Younger Age Groups 123
4.C Share of Respondents with Migration Intentions — GI and Student Sample 124
4.D Share of Respondents with Migration Intentions — GI and Student Sample, Younger Age Groups 125

5.1 Discontinuity in Months Spent in School (in Year 2 After Cutoff) 150
5.2 Discontinuity in Actual Enrollment in Year 0 (Treatment Assignment) . 153
5.3 Discontinuity in Parental Monthly Income and Parental Employment (in Year 2 After Cutoff) 157
5.4 Discontinuity in Parental Self-Assessed Health and Parental Staying Intentions (in Year 2 After Cutoff) 158

5.5 Means in Parental Working Hours per Week and Parental Hourly Income Over Years 165
5.6 Means in Parental Working Hours per Week and Parental Monthly Net Income Over Years (Germans) 172
5.7 Distribution of Random Treatment Assignment Coefficient on Months Spent in School (N=10000) 175
5.A Means in Parental Integration Outcomes Over Years 179
5.B Actual Enrollment in Year 0 (Treatment Compliance) 180

List of Abbreviations

2SLS	Two Stages Least Squares
ARS	Argentine Peso
AZR	Central Register of Foreign Nationals
BAMF	Federal Office for Migration and Refugees
BIH	Bosnia and Herzegovina
BRL	Brazilian Real
CI	Confidence Interval
CZE	Czech Republic
DE	Germany
DID	Difference In Differences
ECEC	Early Childhood Education and Care
EEA	European Economic Area
ESP	Spain
EU	European Union
EUR	Euro
FE	Fixed Effects
GBR	Great Britain
GDP	Gross Domestic Product
GER	Germany
GNI	Gross National Income
GI	Goethe Institute
GSOEP	German Socioeconomic Panel
GWP	Gallup World Poll
IDN	Indonesia
IND	India
ITA	Italy
ITT	Intention to Treat Effect
IV	Instrumental Variable
IVR	Indulgence Versus Restraint
JPN	Japan
JPY	Yen
KOR	South Korea
LATE	Local Average Treatment Effect
LTO	Long-Term Orientation
MEX	Mexico
MXN	Mexican Peso
NLD	Netherlands
OLS	Ordinary Least Squares
POL	Poland
RDD	Regression Discontinuity Design
RE	Random Effects
ROU	Romania
TRY	Turkish Lira
UKR	Ukraine
UN	United Nations
USA	United States of America

1 | General Introduction

I think massive migration is inevitable. As sea levels rise, as climate change happens, as fertile fields become arid, as wars are fought, people are going to move. They always have.
— MOHSIN HAMID

Migration movements, especially into developed countries of the global north, has always been a prominent topic in media and politics. With ongoing climate change and the ecological and economic issues it entails, migration movements will only increase in the future. In 2021 alone, 2.3 million immigrants[1] entered the European Union (EU) from non-EU countries, and 23.8 million people of the 446.7 million people living in the EU were non-EU citizens (5.3 % of the population) (Eurostat 2021). The USA reported roughly 1.5 million new immigrants[2] in 2021, and a record breaking total of 46.2 million immigrants resided in the country in 2021 (14.2 % of the population) (Bureau 2021).

Against this background, research on migration flows, integration measures – and with it language learning – will grow in importance to policy makers. This thesis sheds some light on the processes and motives of language learning, migration decision and integration with the focus on a receiving developed country, namely Germany. Germany is not only an important receiving country – by the end of 2020, Germany had a migrant population of over 10 million people[3], which is a population share of 13.7 % (Destatis 2020). One million alone are Syrian refugees who entered Germany in mid2010s (BAMF 2016). As Germany faces a massive demographic change with its rapidly aging population and a tremendous shortage of skilled workers, migration constitutes a major chance to address these problems.

There is much to be learned about the complex migration process and this thesis provides insights into the role of language learning, education and social context in migration. In several chapters it contributes to three strands of literature. Chapter 2 and Chapter 3 analyze motives for and timing of foreign language acquisition and highlight its importance for success in both origin and host country labor markets. Chapter 4 enhances the understanding of migration motives and intentions of potential migrants before actual migration occurs. Lastly, Chapter 5 focuses on the assimilation of families in the host country. It provides first evidence that formal schooling of children positively affects their parents integration outcomes, especially on the labor market.

[1] Individuals who were not born within the EU borders.
[2] Individuals who were not born within the US borders.
[3] Individuals who were not born in Germany but regularly reside in Germany.

1.1 Language Learning, Migration Decisions and Integration Outcomes

One prerequisite for communication across country borders – and for international mobility – are foreign language skills. There are two contexts in which foreign language skills have productive value: foreign language skills outside the context of migration and in the context of migration. The former relates to language skills which are foreign relative to the primarily spoken language of the country of residence (e.g. English language skills in Germany). Several studies have found high returns to such foreign language skills in the labor market of some European countries for both immigrants (Isphording 2013; Toomet 2011) and natives (Ginsburgh and Prieto-Rodriguez 2011), depending on the relative scarcity of specific language skills. While in the US there are no or only very small returns to foreign language skills (Fry and Lowel 2003; Saiz and Zoido 2005), Stöhr (2015) has found large returns to expert level English for natives and even more so for immigrants in Germany. Regarding the latter, the literature has long recognized the importance of language skills for migration choice and subsequent integration (e.g. Belot and Ederveen 2012; Grogger and Hanson 2011; Mayda 2010; Ortega and Peri 2013). The host country's main language is often foreign relative to the native language of immigrants, and learning it holds tremendous benefits for the immigrant. Language skills among immigrants improve their labor market outcomes, i.e. by having a positive effect on earnings (Chiswick and Miller 1995; Dustmann and Van Soest 2001, 2002) and employment probabilities (Budría et al. 2019; Dustmann and Fabbri 2003). But the effects go beyond the factor labor. Positive effects of language skills have been found with regards to social integration (Aldashev et al. 2009) as well as for education and health (Aoki and Santiago 2018). Further, better language skills increase the probability of intermarriage and reduce the likelihood of living in an ethnic enclave (Bleakley and Chin 2010).

Due to all these benefits of host country language skills of immigrants and foreign language skills in a domestic labor market, learning a foreign language can be viewed as an investment. According to the human capital theory (Becker 1964; Schultz 1960), acquiring human capital is a costly investment which is expected to lead to monetary returns, e.g. by increasing wages or employment probabilities. Individuals choose the human capital that maximizes their expected net present value of income. In the literature, this framework has been enlarged to include migration and therefore expected returns which can realize in the domestic and the foreign labor market (Sjaastad 1962). Foreign language skills can be seen as a specific type of human capital – and hence language learning as investment in human capital. Yet, empirical evidence suggests that the human capital model alone cannot fully explain individual choices. Expected labor market returns do not seem to be the only determinant of the choices, as individuals often choose more education or other types than would be optimal according to the human capital theory (Canton and de Jong 2005; Oosterbeek and Van Ophem 2000; Oosterbeek and Webbink 1995). Besides irrational choices, a possible explanation for this is that education holds consumption value itself, which generates direct utility (Kodde and Ritzen 1984; Lazear

1977; Schaafsma 1976).

Chapter 2 of this thesis contributes to the relatively scarce literature on language learning motives. Most of the literature measures foreign language skills on the aggregate level, not language learning itself, and does so after migration has taken place. Yet, by observing language learning instead of language skills, distinctions between investment and consumption motives of language learning, and timing of language learning in the migration context (which will be relevant in Chapter 3) are possible. Further, it allows to differentiate between language learning in school (which is most likely driven by school curricula and parents' preferences rather than individual choices) and adult age language learning (which is very likely based on individual choices). Using data on language learning, Chapter 2 shows that, at the micro level, adult individuals can have very different reasons for language learning – i.e. investment as well as consumption motive.

At the macro level, common language skills have been found to positively affect trade (Isphording 2013; Lohmann 2011; Melitz and Toubal 2014) as well as migration flows (Aparicio Fenoll and Kuehn 2016). Here, the focus is on migration rather than trade. Given the positive effect on migration success in terms of economic and social integration, migrants have strong incentives to learn the language spoken in their host country. Yet, the decision to learn a foreign language is influenced by the ease of learning and the certainty of investment payoff. Regarding the first, the farther a language is linguistically from one's native language, the more costs and effort learning it entails. Individuals who are linguistically far away from the language spoken in a certain country might thus refrain from learning the language and migrating there. Belot and Ederveen (2012) and Adserà and Pytliková (2015) portrait a negative relationship between the linguistic distance of origin and host country language and migration flows.[4] Regarding the second – the certainty of investment payoff – migrants from different origin countries face different degrees of access to their destination country and destination country's labor market. E.g. while within the EU and Schengen Area citizens have relatively free access to the labor market of other member countries, citizens from outside the EU face considerable access restrictions. Concerning the decision to learn a foreign language, this could result in two different mechanisms: On the one hand, citizens who enjoy freedom of movement (i.e. free access to their destination country and labor market) may find it easier to simply move to their destination and look for employment, without having acquired language skills beforehand. Hence, they may delay their language learning until after their arrival. Citizens from countries with restricted movement (i.e. restricted access to their destination country and labor market) may try to build up language skills before their arrival to facilitate the process of finding employment from abroad and to bridge potential access restrictions (*bridge-the-gap mechanism*). On the other hand, citizens who enjoy freedom of movement can safely invest into language skills before

[4] When it comes to trade flows, there is a positive relationship between trade flows and the trading partners sharing a common official or spoken language (Egger and Lassmann 2012; Head et al. 2010) or speaking languages which are linguistically close (Isphording 2013; Melitz and Toubal 2014).

migration, since they are certain they can reap the benefits of their investment. Citizens with restricted access face the uncertainty of not being able to work in their destination even if they acquire the necessary language skills beforehand. They may therefore delay the investment until this uncertainty has resolved and prefer to learn the language after their migration (*certainty-of-investment mechanism*).

Chapter 3 contributes to the literature by comparing both mechanisms and presenting evidence in favor of the *certainty-of-investment mechanism*. It shows that German language learning abroad is strongly associated with migration from countries whose citizens have secure access to the German labor market (i.e. countries within the EU and Schengen Area). Contrary, language learning in Germany is strongly associated with migration from countries whose citizens face uncertain access (i.e. countries outside the EU and Schengen Area) to the German labor market.

Like language learning, migration, too, is a costly investment. The Roy-Borjas model formulates – in an extension of the human capital theory – that individuals who are able to migrate intend to migrate if their expected utility from relocating abroad, net of migration costs, exceeds their expected utility from staying (Borjas 1987). Some results presented in this thesis speak in favor of the Roy-Borjas model when it comes to migration decisions. E.g. in Chapter 2 the probability of the investment motive for language learning is larger for the younger age group since the net gain of migration decreases with age as age lowers the time to recoup the investment made by migration and increases the migration costs faced by the individual (Sjaastad 1962). Further, investment in language learning and migration is a risky endeavor since the benefits are not certain. Hence, individuals with a higher willingness to take risks are more likely to learn German from an investment motive (Chapter 2).

Yet, just as not all language learning could be explained by human capital theory, the same holds for migration decisions. Migration decisions are subject to several constraints, such as visa requirements, liquidity constraints, and social ties at home, which can all prevent migration desires from being realized. Many individuals, especially from developing countries, would like to migrate under ideal circumstances, yet will never do so because they face migration barriers. E.g. aggregated data from the 2018 wave of the Gallup World Poll[5] shows that 25 % of surveyed Mexicans desire to emigrate, yet in the following year of 2019 less than 1 % of the Mexican population actually left the country. On the other hand, migration can also be evoked by migration drivers, even when one would ideally not like to migrate, e.g. by a partner wanting to emigrate or career reasons. Migration is not only an individual decision but happens in social context (Stark and Bloom 1985), hence family ties are bound to influence the migration decision. Strong family ties at home can increase migration costs and be a factor which impedes aspirations to migrate from being realized, especially for women. Ruyssen and Salomone (2018) have shown that women, specifically in countries where gender discrimination is comparably high, are more likely to aspire to migrate, but they also face high costs and

[5]The Gallup World Poll (GWP) is a representative repeated cross-sectional survey conducted in countries worldwide.

obstacles in realizing these aspirations.

To identify such barriers and drivers towards migration, potential migrants need to be observed before actual migration happens. While a broad strand of literature evaluates immigrants in destination countries (e.g. Abramitzky et al. 2012; Beine et al. 2011; Borjas 1987; McKenzie and Rapoport 2010), others survey individuals at their origin (e.g. Bertoli et al. 2022; Bertoli and Ruyssen 2018; Manchin and Orazbayev 2018; Papapanagos and Sanfey 2001; Ruyssen and Salomone 2018; Uebelmesser 2006). The former approach only observes a selective subsample of those who realized their migration plans, i.e. who actually migrated. The latter allows to identify all potential migrants prior to this selection, yet comes with its own difficulties. The question is how to measure potential migration in countries of origin, and the literature is divided on this. Some studies use migration intentions, which express considerations and plans to migrate subject to constraints (e.g. De Jong 2000; Friebel et al. 2013; Uebelmesser 2006; Van Dalen and Henkens 2012). More recent studies use migration aspirations, which describe a desire to migrate under the absence of any barriers (e.g. Bertoli et al. 2022; Bertoli and Ruyssen 2018; Docquier et al. 2020; Ruyssen and Salomone 2018).

Chapter 4 of this thesis contributes to this literature by combining both measures – aspirations and intentions. This sheds light on a migration pattern which has received little attention so far: migration intentions among those without aspirations to migrate. Further, it allows to identify potential drivers and barriers towards the realization of migration desires. In addition, the approach in Chapter 4 differentiates between intentions to migrate permanently versus temporarily, which provides evidence that temporary and return migration constitute a considerable share of migration (cf. Delogu et al. 2018). The results of Chapter 4 show that family ties abroad can be an encouraging factor towards migration as they imply easier access to information about labor market opportunities and financial support. Also, a partner who wants to emigrate can evoke migration considerations and plans among individuals, e.g. if joint emigration is necessary to maintain the relationship. This is most prominent among women. Hence, this thesis provides further evidence that women are often "tied movers" who follow their partner abroad even if they would ideally like to stay in their home country.

As such "tied movers" women often face worse labor market returns in their host country and hold potential for spillovers of their acquired language knowledge into professional use at the labor market (cf. Chapter 2). Hence arises the question: How can the integration of migrants in general – and specifically women – into the host country's labor market be fostered? As discussed above, a broad literature focuses on the positive effects of host country language skills on labor market outcomes (Aldashev et al. 2009; Bleakley and Chin 2004; Chiswick and Miller 1995; Dustmann 1994; Dustmann and Fabbri 2003; Dustmann and Van Soest 2001). However, integration into host countries is a complex process that spans economic outcomes as well as social and cultural assimilation (Constant and Zimmermann 2008; Facchini et al. 2015). Further, for migrant families, integration happens within a family context where parents influence the integration process of their children and vice versa. While the effect of school attendance and attendance of early childhood education and care facilities (ECEC, e.g. kindergarten) on migrant children themselves has been extensively studied (Bleakley and Chin 2008; Cornelissen et al. 2018; Felfe and Lalive 2018), relatively few studies have investigated how attendance of such facilities affect migrant parents. Using data from Norway, Drange and Telle (2015) found no effect of migrant children's ECEC attendance on their parents' labor market and education outcomes. Exploiting regional differences in the availability of ECEC facilities in German states as exogenous sources of variation, Gambaro et al. (2021) estimate the effect of ECEC attendance by refugee children on their parents' integration outcomes. They find positive effect on labor market outcomes and language ability, and an index of overall integration.

Chapter 5 contributes to this literature by providing the first evidence of a positive effect of children's primary schooling on parents' integration into the host country. It shows that schooling of the oldest child increases parents' labor market participation and returns – especially for mothers. Further, it increases parental health, staying intentions and German language skills. Chapter 5 also contributes to the literature on mechanisms through which assimilation to the host countries' language and culture happens. An analysis of underlying mechanisms suggests that besides the gained disposable time upon the school enrollment of the oldest child in the household, parents also benefit from the exposure to the German language and culture. In their "three E's" Chiswick and Miller (2005) and Isphording and Otten (2014) differentiate between three main channels of language learning, which can also be applied more broadly to assimilation measures: economic incentives, exposure, and efficiency. Here, the focus is on the exposure channel. Through the schooling of their oldest child, parents have direct personal contact to teachers, administrators as well as other children and parents. This offers the opportunity to build social networks and practice the local language (Facchini et al. 2015; Martinovic et al. 2009, 2015). Further, indirect contact to the culture and language via community meetings and their children learning the language fosters cultural assimilation and develops language skills (Avitabile et al. 2013; Dustmann 1996).

1.2 Thesis Outline

This section provides a brief summary of the topics and research questions addressed in this thesis, as well as the main findings of each chapter.

As discussed, foreign language skills are not only an important prerequisite for successful migration, but also hold potential benefits at home. Thus, in Chapter 2, the focus is on the reasons for language learning and whether it is viewed as human capital investment or consumption. To observe language learning, rich data from the Goethe Institutes (GI) are used. The GI is a German cultural association which aims to promote German culture and language around the globe. Currently, it is present in 98 countries with a total of 158 institutes in different cities. Through offering language services it is an important supplier of German language courses and exams worldwide.[6] Data on reasons for language learning and sociodemographic characteristics of language course participants (as well as their migration aspirations and plans, which will be used in Chapter 4) have been collected via surveys conducted at Goethe Institutes in 14 countries.[7] Using this data, individuals are categorized to have either an investment or consumption motive, based on their main reason for learning the German language. Roughly 55 % of the participants' main reason for language learning belong to the investment motive, whereas roughly 40 % belong to the consumption motive. Binary probit estimations are used to analyze individual and country level determinants of language learning motives and expected use of German language skills in the labor market. Results reveal that language learners are heterogeneous within and between countries. On the individual level, younger age and a job that is linked to more internationally applicable skills and a higher need for communication are positively related to the human capital investment motive. Female gender, children and a native German partner make the consumption motive more likely. On the country level, larger linguistic and geographic distances increase the likelihood of language learning as human capital investment. Higher income level, on the contrary, allows more for language learning as consumption. While policy measures are needed to target those seeking human capital investment (e.g. subsidies for foreign language learning or, in general, more language learning opportunities), spillovers from the consumption motive to a professional use of language skills might be of interest, too. The latter emerge mostly among younger women with a native German partner in a "tied movers" context. Even though the consumption motive is the main reason for them to learn a foreign language, a professional use of the language is not unlikely. As a consequence, policy measures aiming at this group not only support their social, but also their economic integration in Germany. Other factors that contribute to a higher probability of a professional use of German include being in education without a university degree, working in a job with internationally applicable skills and high communication needs, and having a risk-prone personality.

After establishing that there are human capital investment as well as consumption motives for language learning, the focus is on the former in Chapter 3. Macro-level drivers of adult language learning are analyzed, with a particular focus on migration. For this, German administrative data on migrant flows and stocks and student immigration is matched with language course and exam participation in Goethe Institutes in Germany and abroad. Data on language course and exam participation in Germany and abroad are collected from the GI yearbooks of 1992 to 2006.[8] Within this time period, exam participation abroad is observed for 137 Goethe Institutes in 76 countries, as well as language course participation in Germany for 157 different nationalities. Fixed effects (FE) regressions indicate that language learning abroad is strongly associated with migra-

[6]For more information on the GI see Uebelmesser et al. (2018a).

[7]For more information on the survey, survey design and discussion of its limitations see Chapter 2.2 and Chapter 4.3.

[8]For more information on the data and discussion of its limitations, see Chapter 3.2.1.

tion from countries whose citizens have secure access to the German labor market (i.e. countries within the European Union or Schengen Area). Language learning in Germany, on the other hand, is positively associated with migration from countries whose citizens face uncertain access (i.e. countries outside the European Union and Schengen Area). Additionally, the positive association of migrant stocks with language learning found for countries with uncertain access can also be related to uncertainty considerations as a large migrant stock indicates that in the past, immigrants of a given country were successful in entering Germany. Overall, these findings provide evidence that the mechanism of certainty of investment plays an important role in encouraging preparatory language learning for immigrants from countries with guaranteed access to Germany. Meanwhile, the lack of a positive association between language learning in countries with uncertain access to Germany and migration from these countries supports the need for policy intervention. To shed some light on country heterogeneity, the location fixed effects are substituted with a vector of time-invariant country characteristics, which include linguistic, geographic and cultural distance measures. Linguistic and cultural distance to Germany is negatively correlated with learning of the German language in general. The correlation between geographic distance and language learning is more varied, and depends on whether access to Germany is restricted or free. Lastly, some tentative arguments in favor of a causal interpretation are provided and the analysis is complemented by an instrumental variable (IV) exercise. The main results related to the role of uncertainty of access hold.

After seeing the correlation between migrant flows and language learning, in Chapter 4 the focus is on individual level determinants of aspirations and intentions to emigrate. Data from two multinational surveys are used: one among language course participants in Goethe Institutes in 14 countries (which is the same data set as used in Chapter 2) and one among university students in 6 countries. Differentiation between intentions to migrate temporarily and permanently reveals that between 25 % (Goethe Institute sample) and 34 % (student sample) respondents intend to emigrate only temporarily. This has important implications for migration research as data on temporary migration is generally rare, and integration measures might differ in their scope and effectiveness when the targeted population intends to only stay in their host country temporarily. Further, the data allows to observe within-individual and between-individual differences of migration aspiration and intention. Multinomial probit regressions show that individual level determinants for aspirations and intentions vary between genders. Women are more influenced by family ties, while men are more influenced by education and career prospects. An analysis of main motivations to emigrate reveals that family considerations primarily motivate women who intend to migrate without aspirations to do so, while men often intend to migrate only temporarily for work or studies. Given that highly educated individuals are in general more likely to emigrate, women who might otherwise pursue an ambitious career could be disproportionately affected by the prospect of being a “tied mover”. As early investments in careers are made before the realizations of own and partner’s job opportunities abroad are revealed, this would affect the incentives well beyond the group that actually emigrates. Those findings have several implications. First, a better understanding of the length of the intended stay is valuable for both sending and receiving countries. Sending countries could identify which share of planned emigration is about permanent migration and which is about brain circulation. Receiving countries could use this information to guide integration policies since those depend on the expected length of stay. Second, within-individual differences between migration aspiration and intention can reveal barriers and drivers towards migration. While the desire to leave one's country permanently might originate in pull factors abroad, push factors in the country of origin may result in intentions to emigrate even if one would ideally like to stay.

After individuals realize their migration aspirations and follow through with plans to migrate to another country, the final aspect of the migration process is successful integration into the host country. In Chapter 5 it is analyzed in how far children's formal schooling holds the potential to enhance the integration of migrant parents. The analysis links administrative records on primary school enrollment cutoff dates with micro data from the German Socioeconomic Panel (GSOEP). It exploits the age-at-enrollment policies in the German federal states as exogenous source of variation in children's school enrollment timing. A regression discontinuity design around the school enrollment cutoff and an instrumental variable approach are utilized to show that schooling of the oldest child in the household improves the integration outcomes of their parents – both on the extensive margin (i.e. school enrollment) and on the intensive margin (i.e. months of schooling). The results show that schooling of the oldest child positively affects parents' integration outcomes along several dimensions such as labor market outcomes, staying intentions and German language skills. An analysis of underlying mechanisms suggests that the results are due to both gains in disposable time upon the child's school enrollment, and exposure to the German language and culture. Overall, these findings provide the first evidence that children's school attendance promotes parental integration, which could help policymakers identify effective policies for successful integration.

Chapter 6 concludes with a discussion of policy implications and limitations of the analyses.

2 | Language Learning: Human Capital Investment or Consumption?[1]

with Matthias Huber[2]
and Silke Uebelmesser[3]

Abstract. *This paper focuses on foreign language learning as human capital investment or consumption. We apply the human capital investment framework to foreign language learning and enlarge it by the consumption motive. Based on a novel dataset of close to 5,000 language course participants in 14 countries worldwide, we estimate individual and country level determinants of the different motives for language learning and of the expected use of language skills in the labor market. We highlight possible spillovers from the consumption motive to a professional use, which emerge mostly in a "tied mover" context. This provides guidance for targeted language policies.*

Keywords: *language learning, new dataset, human capital investment, consumption*

2.1 Introduction

Foreign language skills have a productive value in two different contexts. First, one can think of language skills of natives and immigrants, which are foreign relative to the main language of the country of residence. Whereas studies have found no or only very small returns to foreign language skills in the US (Saiz and Zoido 2005), high returns to those skills show up in the labor market of some European countries for immigrants (Isphording 2013; Toomet 2011) as well as for natives (Ginsburgh and Prieto-Rodriguez 2011). As the latter authors point out, these returns often depend on the relative scarcity of specific language skills. In Germany, there are large returns to expert level English for natives and even more so for immigrants (Stöhr 2015).

Second, foreign language skills can also be viewed in the context of migration. The host country's main language is often foreign relative to the main language of the immigrants. Researchers and policy makers alike emphasize the importance of immigrants' skills of the host country's language for integration into the labor market

[1] The Version of Record of this article is published in *Empirica* 49.4, 897–948 (2022), and is available online at https://doi.org/10.1007/s10663-022-09548-7.

[2] Federal Office for Migration and Refugees and Friedrich Schiller University Jena

[3] Friedrich Schiller University Jena and CESifo

of the host country. More specifically, language skills improve labor market outcomes of migrants by increasing earnings (see e.g Chiswick and Miller 1995; Dustmann and Van Soest 2001) and employment probabilities (Budría et al. 2019; Dustmann and Fabbri 2003) and by improving occupational choices (Aldashev et al. 2009).

All this evidence can be put in the light of the human capital theory (Becker 1964; Schultz 1960). According to this theory, individuals choose the human capital that maximizes their expected net present value of income. Acquiring human capital is a costly investment which is expected to lead to monetary returns via increased wages or increased employment probabilities by fostering the individual productivity. This framework has been enlarged to include migration and therefore expected returns which can realize in the domestic and the foreign labor market (Sjaastad 1962); it can be further extended easily to comprise foreign language skills as a specific type of human capital.

When looking empirically at individual choices, however, the human capital model is not able to fully explain the data. Individuals often choose more education or other types than would be optimal according to the human capital theory (Canton and de Jong 2005; Oosterbeek and Van Ophem 2000; Oosterbeek and Webbink 1995). If we ignore irrational choices, expected labor market returns do not seem to be the only determinant of the choices. One explanation for the observed pattern is that education or, broadly speaking, learning has a consumption value and generates direct utility (Kodde and Ritzen 1984; Lazear 1977; Schaafsma 1976). This consumption value can be defined as "the private, intended, non-pecuniary return to education" (see Alstadsaeter 2011). Individuals may then choose a quantity or type of education which leads to lower monetary returns than other possible choices (Alstadsaeter 2011; Arcidiacono 2004).

In this paper, we focus on different motives of foreign language learning. Language learning leads to a particular form of skills that can be acquired in many different contexts, e.g. at school, university, but also in language courses. While choices about language acquisition at school are often determined by the school's curriculum or parents' preferences, participating in a language course offered by a university or a private provider as an adult is more directly related to the individual's human capital or consumption motive. For a better understanding of the different motives, in a first step, we study the determinants of the human capital motive of language learning. In a second step, we look at the determinants of a professional use of German in the labor market. While we expect a positive relation between the human capital motive and a professional use, we are particularly interested in possible spillovers from a given consumption motive to a professional use on the one hand and possible barriers that might hinder a professional use despite a given investment motive on the other hand.

To the best of our knowledge, we are the first to apply the human capital framework enlarged by the consumption motive to foreign language learning. For this, we use unique survey data of almost 5,000 language course participants collected in 14 countries at institutes of the Goethe Institute.[4] The sample is very likely not representative for the

[4]The GI is a German cultural institute, which offers language courses worldwide and is an important part of the foreign cultural policy of the German government. In addition, the GI is engaged in cultural exchange

populations in the respective countries, as the participants are relatively young and highly skilled. For policy makers in the home country, and equally in Germany, this group might, however, be particularly interesting as those individuals are often more mobile and more open to international experiences, both professional and private. Furthermore, the dataset is very suitable for analysing our questions of interest about the motives of language learning and possible spillovers. We are thus able to contribute to a better understanding of foreign language acquisition in a cross-country perspective and to provide guidance for policy makers for targeted language policies.

We use binary probit estimations to study individual and country level determinants of the human capital investment motive and of the use of German language skills in a professional environment. In order to identify heterogeneities, we also have a closer look at subgroups based on age, gender and education and consider differences across countries. While a younger age and a job that is linked to more internationally applicable skills and a higher need for communication is positively related to the human capital motive, female gender, children and a native German partner make the consumption motive more likely. At the country level, we find that larger linguistic and geographic distances increase the likelihood of language learning as human capital investment; a higher income level, on the contrary, allows more for language learning as consumption. We also show that a given human capital investment motive does not necessarily match with a high probability of professional use. We find spillovers from a given consumption motive to a professional use mostly in a "tied mover" context, i.e. for women and in the presence of children and a German native partner.

The remainder of the paper is structured as follows. In Section 2.2, we describe the survey set-up and discuss selection issues and (limits to) the dataset's representativity. Section 2.3 introduces the conceptual framework and provides graphical illustrations. Section 2.4 explains the empirical strategy. In Section 2.5, we present estimation results for the determinants of the human capital motive of language learning and the probability of a professional use and discuss spillover effects. Section 2.6 concludes.

2.2 Survey Set-Up and Data

We address our questions of interest based on survey data which we collected from language course participants at the Goethe Institute. The survey was conducted between June and December 2018 and included questions on socio demographic characteristics, education, professional background, language skills, previous migration experience and future migration plans as well as questions on the reasons of learning the German language. In the following, we explain the design of the survey, the data collection process and possible limitations.

---

and provides information about German culture and society (Auswärtiges Amt and Goethe-Institut 2004). While the main funding is provided by the Federal Foreign Office, language courses are financed by fees (Goethe Institute 2014). In 2021, the GI was present in 98 countries with a total of 158 institutes (Goethe Institute 2021). For more details on language learning at the GI, see Uebelmesser et al. (2018a).

2.2.1 Survey Design

For our analysis, we selected 19 institutes in 14 countries. The choice of the countries was motivated by the wish to capture cross-country variations in several dimensions. Table 2.1 gives an overview of the selected countries and the main characteristics on which we based this selection to assure a heterogeneous sample of countries: geographic distance to the German speaking region, linguistic distance to the German language, average income level as categorized by the World Development Indicator in 2018 (World Bank 2021) and the absence (or presence) of migration barriers vis-à-vis the German speaking region. The presence of a large institute measured by the number of course participants was of further importance for the selection of countries. In Indonesia and South Korea, we had the opportunity to conduct the survey in more than one institute.

To reduce the issue of (non-)selection of participants into the survey, we undertook several measures to achieve a high response rate. First, the survey was translated into the main language of each country. In India, the questionnaire was in English. Additionally, we provided English and German questionnaires upon request in every country. Second, we opted for a pen-and-paper survey as this allowed for a more direct involvement with the participants. Third, each participant could take part in a lottery to win a free language course at the given institute (limited to one language course per country). Fourth, we encouraged participation in further ways, which differed between European and non-European countries. In European countries, a team member of the research project was present at the institutes for at least one unit of each course offered during a given week and handed out the questionnaire to all present participants. Most of the participants filled in the questionnaire during the course break or after the course unit, others took it home and returned it later to the team member. In non-European countries, team members were not present in person to conduct the survey. Instead, the printed questionnaires were sent by mail to the institutes and were then distributed by the course instructors. To reduce the time and effort of the instructors and other GI officials and to minimize the probability of errors in the distribution process, we prepared envelopes for each course containing the questionnaires. In Mexico, the questionnaires were distributed during the process of course inscription for the upcoming course term.

All those measures combined resulted in high participation numbers and high response rates. Table 2.1 gives an overview of the numbers and rates by country. In European countries, the response rates ranged from 67 % to 99 %. In these countries (except the Netherlands), the response rate is based on the number of questionnaires distributed to all participants who were present during the language lesson. In non-European countries (and the Netherlands), on the contrary, the response rate is based on the number of registered course participants. In those countries, the response rates ranged from 59 % to 72 %. It is not so straightforward to compare the response rates for the European and the non-European countries for the following reason: As not all registered course participants are present at every lesson, the number of registered participants is, by definition, equal or larger than the number of those who were present at the lesson when the survey took place. For the same number of collected questionnaires, therefore this leads to lower

Table 2.1: *Country Characteristics and Response Rates*

Countries	Ling. close	Geogr. close	Income (GNI/capita)	Absence of migr. barriers	Notes	Partici-pants	Response rate
Netherlands	Yes	Yes	High	Yes	Linguistically close country	139	0.67*
Great Britain	Yes	Yes	High	Yes/No (Brexit)	Linguistically close country	480	0.88
Spain	No	Yes	High	Yes	(Recently) high-unempl. c.	611	0.83
Italy	No	Yes	High	Yes	(Recently) high-unempl. c.	371	0.86
Czech Rep.	No	Yes	High	Yes	New EU country (since 2004)	481	0.83
Poland	No	Yes	High	Yes	New EU country (since 2004)	236	0.69
Romania	No	Yes	Upper-middle	Yes	New EU country (since 2007)	327	0.87
Bosnia	No	Yes	Upper-middle	No	Close, non-EU country	270	0.99
Ukraine	No	Yes	Lower-middle	No	Close, non-EU country	782	0.93
Japan	No	No	High	No	Developed, non-Europ. country	293	0.59*
South Korea	No	No	High	No	Developed, non-Europ. country	470	0.65*
Mexico	No	No	Upper-middle	No	Emerging markets	491	0.60
Indonesia	No	No	Lower-middle	No	Developing country	883	0.55*
India	Yes/No (English)	No	Lower-middle	No	Developing country	830	0.72*

Note: * Response rates based on registered course participants, not actual attendance. Income levels are categorized as by the World Development Indicator in 2018 (World Bank 2021). Languages are linguistically close if they are Germanic languages. Countries are geographically close if the country is in Europe. Absence of migration barriers are defined by freedom of movement for workers within the European Union and the European Economic Area (EEA). For variable descriptions, see Table 2.B.

response rates in non-European institutes, which can be interpreted as lower bounds, when compared to response rates in European institutes.[5] A further exception is Mexico where the response rate (60 %) is based on the number of distributed questionnaires during the process of course inscription.

2.2.2 *Descriptives and (Limits to) Respresentativity*

In total, 6,664 language course participants submitted valid questionnaires. Of those, we excluded 1,773 observations because of missing information in the variables utilized in our analysis. Our final sample therefore contains 4,891 individuals. Table 2.A explains the individual level variables and Table 2.C details the dropping of observations in a step by step way. While missings in individual characteristics only lead to a drop of roughly 220 observations, more missings are related to the questions on the international applicability of skills and the importance of communication skills in professional life, respectively. The most important drop is due to missing information about the main reasons of language learning, which is our main variable for constructing the human capital and the consumption motives. In order to see whether dropping observations due to missings introduces a bias, we present descriptives separately for our final sample in

[5] In those eight European institutes where a team member was present, we know the actual attendance numbers as well as the registration numbers. It turns out that on average 75% of registered participants were present.

Table 2.2 and the sample before the droppings in Table 2.D. T-tests for the means do not point towards significant differences.[6,7]

Focusing on the means of the total sample in Table 2.2,[8] we see that 78 % of the participants are younger than 35 years. The majority of course participants is female (57 %) and has no partner (61 %), with some variations across country groups. While the share of partners with German as native language is very small in non-EU countries (3 %) (except Japan with 11 %), it is rather large in the EU (on average 9 %, but in particular due to Great Britain with 27 % and the Netherlands with 17 %).

The young average age of the participants, in particular, in the non-EU countries might be responsible for the overall low share of those with children (only 13 %). This might also explain the high share of those who indicate that they are in education (45 % overall, over 50 % in non-EU countries, but only 29 % in EU countries). Most of the other participants are active in the labor market (42 %). Of those, 69 % state a high level of international applicability of their skills; and more than three quarter of them see a high importance of communication skills in their professional life. The majority of course participants has a university degree (61 %). In combination with those still in education, part of whom will likely receive a university degree in the future, the large majority of course participants is highly skilled.

Participants in the EU are on average slightly less risk prone, but also slightly less patient. Almost half of them have high English language proficiency while the numbers are much lower for non-EU participants. Overall, it shows that individuals from non-EU countries (within or outside Europe) are more similar to one another than to individuals from EU countries with the exception of partner status and children where differences are more pronounced.

When looking at the descriptives, it becomes obvious that the sample is not representative for the populations at large. The participants in the survey are relatively young and highly skilled. For policy makers in the home country, and equally in Germany, this group might, however, be particularly interesting as those individuals are often more mobile and more open to international experiences both related to their professional and their private activities. Furthermore, the dataset is very suitable for the purpose of our study given its focus on language learning. In the analyzes, we will nevertheless consider subgroups based on age and on education (next to gender) to understand any potential differences and to assess their relevance for the results.

For a generalisation of our insights about motives of adult age language learning, a further issue concerns the question whether language course participants at the GI differ from those at other language institutes. Self-selection of participants, in general, could take place based on the following three characteristics (see also Uebelmesser et al.

[6]The only exception is the variable “English speaker” with a share of 0.348 in the final sample and a share of 0.327 in the sample before the droppings (significant at the 5% level).

[7]An analysis of the missings shows that the significant determinants of not answering the question on the main reason do not follow a pattern. We are therefore confident that this does not imply a selection issue. The results are available from the authors upon request.

[8]See Table 2.E for descriptives by countries.

Table 2.2: *Descriptive Statistics: Means of Individual Characteristics by Country Groups*

Variable	EU	non-EU		Total
		European	non-European	
	n=2040	*n=754*	*n=2097*	*n=4891*
Age: under 35 years	0.67	0.81	0.87	0.78
Gender: male	0.40	0.38	0.40	0.40
Gender: female	0.58	0.59	0.56	0.57
Gender: n/a	0.03	0.03	0.04	0.03
Children	0.19	0.17	0.07	0.13
No partner	0.44	0.58	0.78	0.61
Partner (native German)	0.09	0.03	0.03	0.05
Partner (other native)	0.46	0.39	0.19	0.34
Occ.: low applicability	0.16	0.11	0.11	0.13
Occ.: high appl./low comm. skills	0.08	0.05	0.04	0.06
Occ.: high appl./high comm. skills	0.31	0.19	0.16	0.23
Occ.: in education	0.29	0.51	0.58	0.45
Occ.: other occ./no answer	0.15	0.13	0.11	0.13
University degree	0.74	0.56	0.51	0.61
Risk attitude	6.09	6.77	6.49	6.36
Patience	5.80	6.59	6.25	6.11
English speaker	0.48	0.35	0.22	0.35

Note: For variable descriptions, see Table 2.A and for the grouping of the countries see Table 2.B.

2018a): willingness or ability to pay, location, and age. Selection on willingness to pay could occur if the prices of courses at the GI differed significantly from prices of other equally suitable learning options. Prices could be higher if one considers the GI as a premium provider of language courses because of its semi-official status and its long tradition and good reputation. Prices could also be lower because of funding by the German government. Both arguments are not fully convincing, however: When looking at current prices of courses offered by the GI and by competitors, the prices do not indicate that the GI is usually the most expensive provider in the market. At the same time, language courses are priced to be self-financing, that is they are not financed by government funding. As to location, institutes are usually located in capitals and other major cities, which might lead to an underrepresentation of language learners from rural areas. However, the bias is likely attenuated by the fact that institutes also offer intensive courses taught en bloc. Participants who do not live in the vicinity of the respective institute, may stay there for the duration of the course only. Still, we cannot rule out that other providers of language courses are more present in rural areas.

Admittedly, language services offered by the GI are only one way for adults to acquire skills in the German language. Naturally, there are a large number of alternatives, including universities, private language schools, and internet platforms. This might give rise to additional concerns regarding the self-selection of language learners into courses offered by the GI, particularly based on age. Again, the bias might be less severe: On the one hand the GI has complemented its course offer by online and blended learning courses, which combine traditional and online learning, since 2010 (Goethe Institute

2011). On the other hand, as our data show, courses offered by the GI are attractive to younger language learners; an overrepresentation of older participants cannot be observed in our sample. In sum, we conclude that there is no strong evidence that participants of language courses offered by the GI systematically differ from participants of language courses offered by other providers.

Despite all the caveats mentioned above, this sample suits well our purposes given our research interest in the motives of foreign language learning. It provides many individual level information as well as information about language learning motivations for a sample of close to 5,000 individuals in 14 countries. At the same time, we are aware of the limitations as to its generalisation. In particular, we acknowledge that we cannot say anything about selection into language courses or more generally, the decision of adults to learn a foreign language versus the decision not to do this and, despite the arguments brought forward above, we cannot fully rule out selection bias relative to other providers.

2.3 Motives of Language Learning

Our dataset informs us about the reasons of language learning. In the following, we will explain how we derive the human capital motive and the consumption motive from these reasons, discuss the expected associations and provide some graphical illustrations.

2.3.1 Conceptual Framework and Operationalization

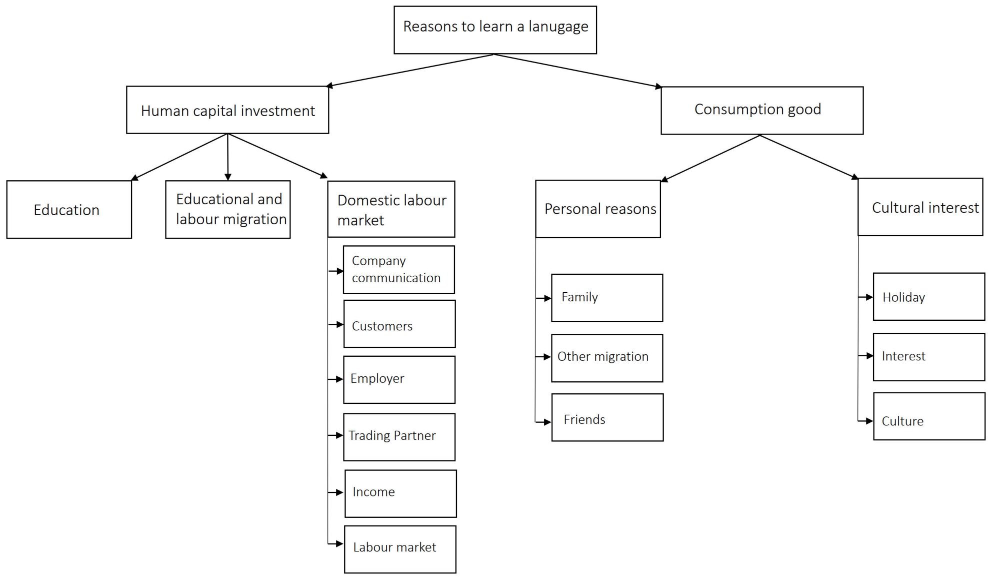
In the survey among course participants, all respondents have in common that they decided to learn the German language. We now examine the motives behind this decision. All participants answered the following multiple response question: *Why are you learning German?* Afterwards, participants were asked to choose the main reason among the stated reasons. Table 2.F provides the reply options.

We categorize the main reasons according to Figure 2.1 and use this categorization as the basis for our analysis. In a first step, we aggregate the 14 reasons presented in the questionnaire into the five categories *education, educational and labor migration, domestic labor market, personal reasons* and *cultural interest*. In a second step, we further aggregate these categories into the two motives *human capital investment* and *consumption*.

On the one hand, language learning can be an investment in human capital, i.e. language skills can be used in a productive way such that there are (expected) monetary returns to these skills. In our context, we use a broad definition of monetary returns and consider all categories which contain reasons related to the *domestic labor market* or the foreign labor market via *labor migration*.

On the other hand, language learning can be seen as a consumption good with non-monetary returns that leads to a direct increase in utility, either immediately or later. We define reasons as being related to the consumption good which belong to the categories *personal reasons* and *cultural interest*.

Figure 2.1: *Categorization of Reasons to Learn Languages*



Note: See Table 2.F in the Appendix for the exact wording of the question in the questionnaire and for country-specific details.

Additionally, learning a language can happen in the context of domestic *education* or foreign education via *educational migration*. This happens either directly by adding language skills to the human capital stock, or indirectly if language skills positively affect the accumulation of other human capital, e.g. by opening up better education possibilities in destination countries if language skills are a requirement for education. We see the human capital motive as the most relevant one when it comes to the link with education. This view is also supported by the observation that (higher) education is costly for the individuals in terms of opportunity costs and, even more important in many countries, tuition fees. Still, we acknowledge that education can also have a consumption aspect (we will come back to this point at the end of this section).

To base our categorization and the further aggregation to motives not only on the general considerations outlined above, we take the observed correlations between reasons and categories as further guidance to compensate for the lack of literature.

As can be seen in Table 2.G, all reasons belonging to the *personal reasons* and *cultural interest* categories are positively and significantly correlated. They are then further aggregated to the consumption motive. The reasons which are part of the *domestic labor market* category are also all significantly and positively correlated. To study the correlation of the the single reason categories *education* and *education and labor migration* category, we look at Table 2.H. We find support for relating both the *domestic labor market* category and the *education and labor migration* category to the human capital investment motive. As to the *education* category, the picture is somewhat ambiguous. There is a clear positive correlation with the *education and labor migration* category, which induces us to aggregate it in the human capital motive as well. At the same time, we observe a negative, albeit smaller, correlation with the *domestic labor market* category. Therefore, when presenting our main regression results in Chapter 2.5, we complement them with specifications where we move the *education* category to the consumption motive and other specifications were we exclude all participants whose main reason is *education*.

2.3.2 Expected Associations

In the following, we discuss what we expect for the association between individual-specific and country-specific explanatory variables on the one hand and the two motives on the other hand. Due to a lack of related research, our general approach and our choice of variables is guided by studies focusing on other forms of human capital investment, not language learning, based on the human capital theory (Becker 1964; Schultz 1960; Sjaastad 1962) or studies on the determinants of language proficiency of immigrants (Chiswick and Miller 2015).[9]

Looking first at individual characteristics, age is an important factor with an expected negative effect on human capital investment according to the human capital theory. To put it differently, the older the individual the less time there is to recoup the investment. Analogously, the older the participant in a language course, the less time for the monetary returns to realize. In addition, the costs of learning a language grow with age as the required effort increases. All of this makes the human capital motive less likely. This line of argument finds some support in the literature on language acquisition of immigrants in their host country (see Chiswick and Miller 2015 who stress "efficiency", which relates to age, as an important determinant of language learning).[10]

A higher level of education, on the contrary, can be expected to decrease the cost of acquiring language skills by increasing the efficiency of learning (see again Chiswick and Miller 2015 and Footnote 10). Furthermore, foreign language skills might increase the productivity of other skills in the labor market in a complementary way and therefore positively affect the overall benefits. This makes the human capital investment more likely. Related to this, international applicability of education makes it also more likely that opportunities specifically on the foreign labor market emerge where language skills lead to benefits. This might be all the more important for occupations, where communication skills play an important role. When we focus on this complementary view, we therefore expect an overall positive correlation with the human capital investment motive for those in the labor market whose skills are internationally applicable and whose professional communication needs are high. There is, however, one caveat: It could be that, despite a high level of education coupled with internationally applicable skills in communication

[9]It is important to note that our analysis differs from these studies in one important way: in our case, the alternative to learning for investment purposes is learning for consumption purposes, and not "no learning" at all.

[10]Chiswick and Miller (2015) focus on the three "Es". They comprise exposure, which refers to the environment in which the migrants live and communicate, economic incentives, which cover a mix of internal and external factors such as planned duration of stay and expected earnings gains, and efficiency, which next to age at migration include the level of education and similar characteristics that enhance individuals' abilities to learn.

intensive jobs, German language skills are not needed for a productive use because work relationships rely on a high proficiency in English. In this case, we would expect that German language learning and English language proficiency are substitutes making the human capital motive less likely for those with very good English language skills.

For risk proneness and patience, we expect a positive correlation with the human capital motive following the investment literature starting with Becker (1964). Monetary returns to language learning are uncertain and realise, if at all, in the future. So more risk-prone and more patient individuals should be more likely to learn a language with an investment motive.

We also predict that female participants have a lower probability of the investment motive than male participants. The situation of women on the domestic and foreign labor market is often worse in terms of labor market participation and wages. Furthermore, in the migration context, women are more likely the "tied movers" (Geist and McManus 2011; Mincer 1978) who join the male labor migrant with an a priori lower own probability of labor market participation. This makes it more likely that the consumption motive dominates.

Closely related to the gender aspect, children might make it more difficult to realize benefits of language learning on the labor market. In particular, we expect that this is important for women who carry most often a larger burden of care work. If we consider the partnership status, we do not expect a significant association between a partner with a non-German native language and the investment motive. A partner with German as native language might be negatively correlated with the investment motive, as opportunities for consumption seem to be more likely, e.g. migration to the home country of the partner or communication with the partner as well as families and friends.

When looking at country level determinants, a larger linguistic distance might mean larger costs of language learning and therefore a smaller likelihood of observing an investment motive. Following a similar line of reasoning, a larger geographic distance – possibly linked with migration restrictions – might also make an investment motive less likely. Returns realize less easily as labor market contacts to German speaking countries are less frequent. At the same time, German-language skills might be less frequent in countries characterised by a larger linguistic or geographic distance allowing reaping higher returns. Predictions, therefore, are not clear. Cultural distance is, in many cases, related to linguistic and geographic distance. There are however exceptions, e.g. Australia, which justify a separate consideration with similar predictions as for linguistic and geographic distance, however. We expect individuals from higher income countries to have the means to see learning a foreign language as a consumption good and not as a way to reap monetary benefits. This should make the consumption motive more likely.

Indicating a reason for language learning, which we aggregate to the investment motive, is not exactly the same as using the German language at work with a high probability. We expect a positive relation between the two. But it is also possible that returns on the labor market will not realize in the near future – if at all. At the same time, it is possible that acquiring German language skills for consumption purposes opens up opportunities in the labor market.

For illustration, we take a look at the two migration related reasons *educational and labor migration* and *other migration*, where the probability of a professional use of German might diverge – at least in the short run. Let us start with *other migration*, which belongs to the consumption motive (see Figure 2.1), and is often related to individuals who migrate as "tied movers". At first, they acquire foreign language skills because a good command of the host country's language increases utility by facilitating communication and integration in the new destination. Only later, the foreign language skills might be used in a professional environment, which we interpret as spillover from the consumption motive. On the other hand, individuals might prepare for *educational migration* with the purpose of using the language skills only to acquire other skills abroad before returning to their home country. That means that they do not have in mind a professional use despite their clearly given investment motive.[11]

When studying the determinants of a professional use of German, we will pay special attention to those individuals with a consumption motive. Following our reasoning from above, this group likely includes women, and individuals with a German native partner and with children.

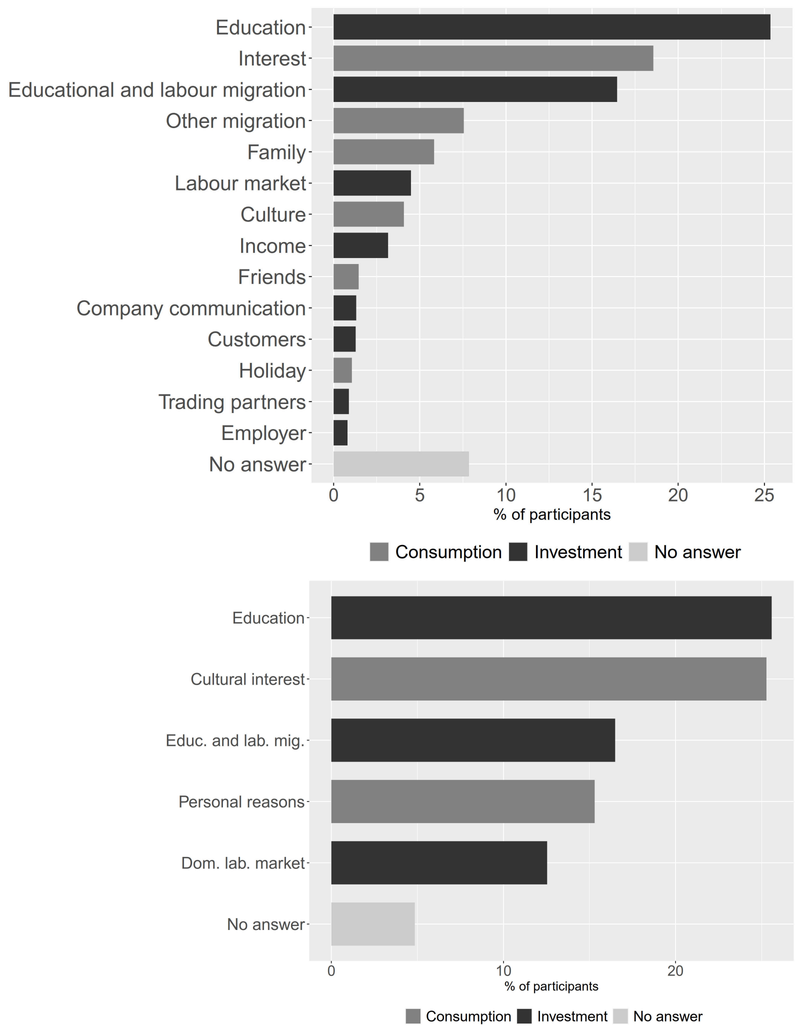
2.3.3 Graphical Illustrations

Figure 2.2 provides an overview of the relative importance of the different main reasons and categories.[12] Looking at categories, we see that one quarter of participants indicated either *education* or *cultural interest* as their motivation behind their decision to learn German, followed by the categories related to *educational and labor migration* (16.5 %), *personal reasons* (15.3 %) and the *domestic labor market* (12.5 %). The grey and black colors in Figure 2.2, in addition, allow assessing the relative importance of the human capital motive and the consumption motive. Roughly 40% of the participants indicated a main reason belonging to the consumption motive, while for 55% the main reason is part

[11] In order to shed more light on those with a migration related reason to learn German, we have made use of additional information in our dataset about migration intention and main reasons for migration in case of intention. Of those who stated *other migration* as main reason for language learning, 70% gave reasons related to the partner or cultural interest as their main migration reason. Considering those with *educational or labor migration* as main reason for language learning, 74% stated education or reasons related to the labor market as their main migration reason. It is interesting in our context that of those 74%, a bit more than one third (27.5%) indicated education as their main migration reason and a bit less than one half (46.5%) mentioned the labor market as their main migration reason.

[12] Given our main interest in motives (see the estimations in Chapter 2.5.1), we impute motives and categories if the respondents gave reasons that belong only to one motive or category, respectively. (As an example, suppose that someone indicates "Family" and "Culture" as reasons. We do not know the main reason, nor can we tell the category of the main reason as it could be *person reasons* or *cultural interest*. As both reasons are, however, part of the consumption motive, we can assign that motive to that participant.) This allows us to increase the sample size as for the final sample we only drop observations if we neither have information about the respondent's main reason nor are able to impute the motive. It is however possible that there are missings if we look at main reasons or categories as is the case in Figure 2.2.

Figure 2.2: *Main Reasons (left) and Categories (right) for Learning German (n=4891)*



of the human capital investment motive.[13]

Given our focus on cross-country differences with a special interest in possible differences between EU and non-EU countries and on heterogeneities based on age, gender and education, we also present graphical illustrations for subgroups.

In Figure 2.B in the Appendix, we present an overview of the main reasons by countries and show that there is a large heterogeneity. This is in particular obvious when looking at *education* and *family*. The share of participants that indicated *education* ranges from 4.2 % to 57.4 % and the share that indicated *family* from 0.3 % to 26.3 %.

When we take a look at the main reasons with the highest share in each country, we can see that there are four single reasons that make it to the top of at least one country. In India, Indonesia, Korea, Mexico and Ukraine, the largest share of participants indicated *education* as their main reason to study German. These countries have in common, that

[13]In this representation, numbers do not add up to 100% due to the “No answer” category for the main reasons. See also the preceding footnote.

they are not members of the EU and their income is relatively low with Korea as an exception. The six countries where most of the participants indicated *interest* in languages as their main reason have the opposite in common: the Czech Republic, Great Britain, Italy, Japan, Poland and Spain are the countries with the highest income in the sample, and all are members of the EU, except for Japan. In Romania and Bosnia and Herzegovina the most important main reason is *educational and labor migration* and in the Netherlands it is *family*.

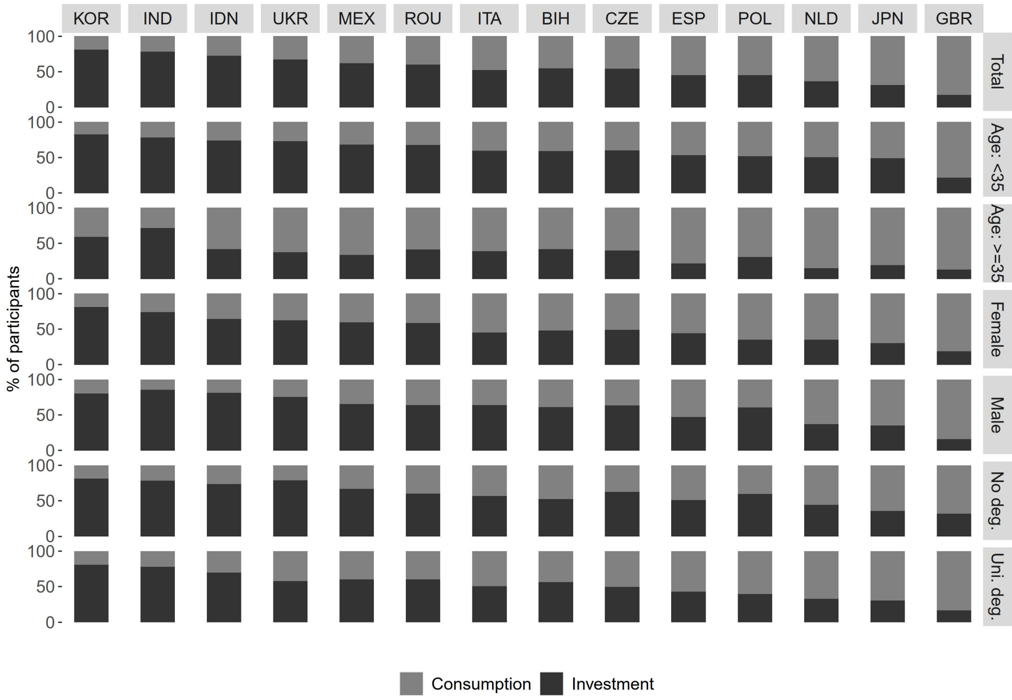
Aggregating the reasons to categories according to Figure 2.1, we can again see the heterogeneity across countries similar to what we observed for the main reasons. This means that the main reason with the largest share often translates into the category with the largest share. That is the case for the Czech Republic, Spain, Poland, Japan and Italy, where the large share of *interest* translates into the category *cultural interest*. The same holds for Korea, Indonesia, Ukraine and Mexico with *education*, for Romania with *educational and labor migration*, and for the Netherlands with *family* which translates into the category *personal interest*. There are only three countries, for which this pattern does not hold: Bosnia and Herzegovina, India and Great Britain.

Finally, Figure 2.3, upper part, gives the distribution of the investment motive and the consumption motive by countries. The variation across countries is large and the share of human capital investment as main motive ranges from 17.1 % to 80.1 % (and vice versa for the consumption motive). We see three groups of countries. First, the investment motive is much more important than the consumption motive in Korea, India, Indonesia, Ukraine, Mexico and Romania. Second, the shares for investment and consumption motives are much more equal with a slight tendency towards investment in Italy, Bosnia and Herzegovina and the Czech Republic. Third, in Spain, Poland, the Netherlands, Japan and Great Britain the consumption motive is more important than the investment motive.

As to differences across age groups, we see that for all countries with the exception of India, the consumption motive becomes more important for individuals older than 35 years of age compared to the full sample. In most countries, it is even more important than the investment motive. The three countries with the highest shares of the consumption motive are Great Britain, Japan and the Netherlands. It is important to note, however, that the share of older people is relatively low in some countries (see Table 2.E in the Appendix). When looking at the younger age group, the pattern is relatively close to the full sample for the three groups of countries reflecting the large share of younger individuals in the sample. The investment motive dominates in most countries except in Great Britain.

Considering gender, there are no strong differences between male and female participants in most countries. If at all, the investment motive seems to be slightly more important for men. There are also no pronounced differences between participants with and without a university degree. If at all, the human capital motive is more probable for those without a university degree (yet) than for those with a university degree. This could be partially driven by participants who are still in education and see language skills as complementary to their acquisition of other human capital.

Figure 2.3: *Human Capital and Consumption Motives by Countries for Age, Gender and Education Subgroups*



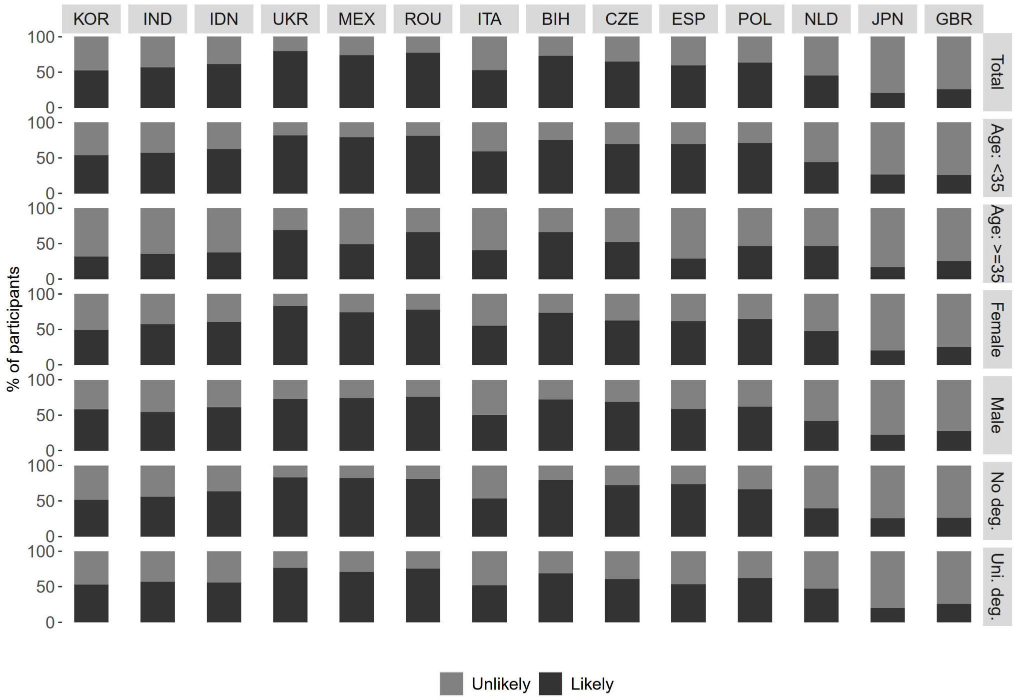
Note: The countries are arranged in descending order according to the share of participants indicating a main reason categorized as human capital investment.

Overall, we conclude that differences in the gender and education composition cannot explain much of the differences in the relative importance of the human capital motive and the consumption motive across countries. Age, however, seems to play an important role for the two motives behind the decision to learn a foreign language. The different composition of the participants in the different countries as far as their age is concerned translates – at least partially – into the observed cross-country differences of the importance of the two motives. In the empirical analysis in Chapter 2.5, we will complement the cross-country perspective by an investigation of the within-country variation.

We are also interested in the question whether a human capital motive indeed leads to a professional use of German. According to our data, the probability of using German in the labor market is on average quite high with 3.67 on a scale from 1 (very unlikely) to 5 (very likely) (see Figure 2.A in the Appendix). Around 60 % indicate that they will likely or very likely use German in the labor market. These shares can be expected if one assumes that those with investment motives also indicate a high probability of professional use, and those with consumption motives give a low or medium probability. We take this as the benchmark for our comparison, when investigating the heterogeneity of responses across countries and for our subsamples in the following.

The distribution again varies across countries as presented in Figure 2.4, upper part.

Figure 2.4: *Use of German on the Labor Market by Countries for Age, Gender and Education Subgroups*



Note: We aggregate the five point scale to the binary variable "Professional use of German" where values 1 to 3 correspond to "unlikely" and values 4 and 5 to "likely". As in Figure 2.3 the countries are ordered by their share of the human capital investment motive with the country with the highest share, i.e. Korea, at the very left and the country with the lowest share, i.e. Great Britain, at the very right.

In the group of countries with a very high share of the human capital investment motive, we also expect a very high share of participants that indicate a high probability of using German in the labor market. In Korea, India and Indonesia, however, this share is much smaller than we expect and also smaller than in Ukraine, Mexico and Romania, where participants indicate the highest probability of a professional use compared to all other countries. A similarly mixed picture emerges for the next group of countries. In Italy, the share is smaller than expected, while it is larger in Bosnia and Herzegovina (75 %) and the Czech Republic (65 %). The last group of countries, when we follow the grouping used before, includes two countries with Spain and Poland, where we expect the share of those with a high probability to be smaller than 50 %, while it is actually around 60 %. The same holds on a somewhat lower level for the Netherlands and Great Britain, while the probability of professional use is smaller in Japan than expected.

There seem to be factors at play that hinder those with a human capital investment motive to think that they will be able to use German in a work related context, and which vice versa make those with a consumption motive expect a professional use of their foreign language skills. Before we examine in detail possible determinants, we again look at the distributions by age, gender and education.

Figure 2.4 shows that the pattern for the younger age group closely follows the pattern for the full sample. This does not hold for older participants where the likelihood of a professional use is much lower, in general. This emphasizes the role of age for the likelihood of using German on the labor market and in the educational context. The latter is supported by the pattern for those who do not have (yet) a university degree, while those with a degree indicate a somewhat lower probability of a professional use. On the contrary, there seems to be no – or in some countries (Italy, Korea, Netherlands) only a small – relationship between gender and the likelihood of using German in the labor market.

The graphical analyzes above provides some evidence that there is no perfect correlation between the human capital investment motive and a high likelihood of using German in the labor market. Nevertheless, as Figure 2.C shows, there is a positive correlation between those two: In all countries, the share of those with a human capital investment motive is larger among the participants with a high likelihood for professional use.

In Chapter 2.5, we will first investigate the determinants of the human capital investment motive and, second, try to better understand the imperfect relationship between the human capital motive and the professional use of German. To put it differently, we want to see what makes participants with investment motives to abstain from indicating a high likelihood of using German in the labor market and what creates spillovers from the consumption motive to the labor market.

2.4 Estimation Strategy

We explore individual-specific and country-specific determinants of the human capital investment motive when learning German on the one hand and of a professional use of that language on the other hand. In particular, we are interested if the descriptive evidence found above for age, gender and education continues to hold after controlling for other factors.

We estimate the probability of both of our outcome variables via maximum likelihood method in a binary probit model:

$$Pr(G_i|\, X_i, C_i) = \alpha + \beta\,'\, X_i + \gamma\,'\, C_i + \varepsilon_i \tag{2.1}$$

where G_i takes a value of 1 if respondent i states to have a human capital investment motive and 0 otherwise when considering the determinants of the motive or, alternatively, takes a value of 1 if respondent i states a high probability for professional use of German and 0 otherwise when studying the use of the German language in the labor market. X_i represents a set of individual-specific explanatory variables of respondent i as presented in Table 2.A in the Appendix following our theoretical considerations in Section 2.3.2. C_i captures either country level factors as listed in Table 2.B or country fixed effects to control for country-specific heterogeneities. ε_i is an idiosyncratic error term. Standard errors are heteroscedasticity robust White standard errors.

Country level factors are control variables that are based on the country characteristics as described in Table 2.1 (see also Section 2.3.2). They include linguistic differences by distinguishing between Germanic or non-Germanic official languages. They also consider economic differences by distinguishing lower-middle and upper-middle income countries as well as high income countries. Further, we categorize the countries into three groups based on their geographic distance to Germany, but also on the absence or presence of migration barriers: one group consists of EU countries, which are close to Germany and for which migration restrictions are non-existent, the second group comprises non-EU countries. This group is further subdivided into European countries, where the geographic distance to Germany is still rather small, but migration to Germany, Austria and also Switzerland is much more restricted, and non-European countries, where migration restrictions are equally relevant and, in addition, the geographic distance to German speaking countries is much larger. As geographical distance does not always correspond to cultural distance, we further add two variables capturing this based on Hofstede and Minkov (2013): one variable about the distance in "long term orientation" (LTO) to Germany and one variable about the distance in "indulgence versus restraint" (IVR).

2.5 Estimation Results

2.5.1 Determinants of the Human Capital Motive

We present our main results in Table 2.3. Column 1 includes individual-specific characteristics only. When adding country-specific controls in Column 2 via country-fixed effects and in Column 3 via country-specific characteristics, the goodness of fit measured with the McFadden Pseudo R2 and the percentage of correctly predicted observations increase.[14]

Country Level Determinants

Before focusing on the individual characteristics, we take a closer look at the country-specific characteristics in Column 3. The probability of the investment motive decreases when the language spoken in the country is a Germanic language (i.e. English in Great Britain and India, and Dutch in the Netherlands) in comparison to a non-Germanic language. The benefits of learning German do not seem to be very large for those with another Germanic language as mother tongue. Given the linguistic closeness of these languages, speakers of Dutch (and German) have relatively low costs of learning English, which is the most spoken foreign language of the world ("lingua franca"). At the same time, Dutch and English allow its speakers a relatively easy access to German. Both might reduce the need for formal learning of German at adult age.[15]

[14]In Table 2.I, we include fixed effects interacted with age and gender. As the model fit is unchanged, we opt for the country-fixed effects without interactions as our main specification. This allows studying the importance of gender and age more explicitly.

[15]In Chapter 2.5.1, we discuss how individual proficiency of the English language relates to the motive of German language learning in order to see whether both languages are substitutes.

Table 2.3: *Human Capital Investment: Basic Specifications*

	Dependent Variable: Human Capital Investment		
	(1)	(2)	(3)
Age: under 35 years	0.225^{***}	0.157^{***}	0.189^{***}
	(0.023)	(0.023)	(0.023)
Gender: female	-0.068^{***}	-0.082^{***}	-0.072^{***}
	(0.013)	(0.013)	(0.013)
Children	-0.033	-0.071^{***}	-0.053^{**}
	(0.025)	(0.025)	(0.025)
Partner (native German)	-0.441^{***}	-0.360^{***}	-0.402^{***}
	(0.027)	(0.034)	(0.030)
Partner (other native)	-0.043^{***}	-0.022	-0.032^{**}
	(0.016)	(0.016)	(0.016)
Occ.: high appl./low comm. skills	0.101^{***}	0.106^{***}	0.110^{***}
	(0.029)	(0.027)	(0.028)
Occ.: high appl./high comm. skills	0.114^{***}	0.119^{***}	0.125^{***}
	(0.020)	(0.020)	(0.020)
Occ.: in education	0.212^{***}	0.186^{***}	0.213^{***}
	(0.025)	(0.026)	(0.025)
University degree	0.023	0.045^{***}	0.051^{***}
	(0.017)	(0.017)	(0.017)
Risk attitude	0.011^{***}	0.006^{**}	0.007^{**}
	(0.003)	(0.003)	(0.003)
Patience	0.010^{***}	0.005^{*}	0.007^{**}
	(0.003)	(0.003)	(0.003)
English speaker	-0.042^{***}	-0.013	-0.025^{*}
	(0.014)	(0.014)	(0.015)
Germanic language			-0.094^{***}
			(0.020)
Non-EU (European)			-0.070^{*}
			(0.038)
Non-EU (Non-European)			0.058^{***}
			(0.020)
Upper-middle income			-0.039
			(0.029)
High income			-0.115^{***}
			(0.024)
Cultural distance: LTO			-0.158^{***}
			(0.046)
Cultural distance: IVR			0.057
			(0.037)
Country FE	No	Yes	No
McFadden Pseudo R2	0.13	0.17	0.15
Percent. correctly predicted	69.5	71.3	71.0
Observations	4,891	4,891	4,891

Note: Average marginal effects. Reference category for the occupation categories is "occupation with low internationally applicable skills and low or high communication needs" (for other reference categories, see Table 2.A). Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

For language learners from non-EU countries outside Europe, the probability of the human capital investment motive is significantly larger than for those from European countries, both in and outside the EU. This can be related to several reasons. First, language skills are often a prerequisite for legal migration to German-speaking countries from non-EU countries particularly outside Europe. This makes it more likely for language course participants from these countries to acquire language skills for investment purposes, as we saw in Figure 2.3. Second, geographic proximity, which is given for all European countries, can be related to a larger migrant stock in German-speaking countries due to previous migration. This makes the consumption motive of language learning more likely for participants from these countries. Both reasons might explain the observed differences between European, EU- and non-EU countries on the one hand and non-European countries on the other hand. Considering a more direct measure of cultural proximity shows that a larger distance with respect to long term orientation (LTO) is related to a smaller probability of the human capital motive and more importance attributed to the consumption motive.

Finally, the country-wide income level plays an important role: participants from higher income countries are on average less likely to learn German for investment purposes. The higher the average income level the more likely participants have the means to see learning a foreign language as a consumption good and not as a way to reap monetary benefits. Obviously, there is a large overlap between EU countries and high-income countries (see also Table 2.1) which is reflected in the results here.

Individual Level Determinants

On the individual level, there are only few differences between Columns 2 and 3. As there is a higher goodness of fit in Column 2 with country fixed effects, we use that specification for the discussion about the results in the following and for the estimations by age, gender and education subgroups in Table 2.4.

In line with the human capital theory, the probability of the investment motive is larger for the younger age group. Being under 35 years of age leads to a 15.7 % point increase; this relationship is also robust within gender subsamples, while it only holds for those with a university degree when considering education subsamples.

We find less language learning for investment purposes for women. Having a native German partner also reduces the probability of the investment motive in comparison to singles. The absolute size of the average marginal effect is 36.0 % point in the full sample and thus almost three times as large as the age effect of the younger age group. A native German partner increases the opportunities where the consumption motive of German language skills seems to be more likely, e.g. communication with the partner as well as family members and friends. This relationship is robust within all subsamples. Also the presence of children makes the investment motive less likely. This is however mostly limited to women and those without a university degree.

The probability of the investment motive is larger by 18.6 % point for course participants in education compared to those who are in the labor market (with low internationally

Table 2.4: *Human Capital Investment: Subsample by Age, Gender and Education*

	Dependent Variable: Human Capital Investment					
	(1) female	(2) male	(3) age: < 35	(4) age: ≥ 35	(5) no uni. degree	(6) university degree
Age: under 35 years	0.147^{***}	0.172^{***}			-0.057	0.171^{***}
	(0.031)	(0.037)			(0.063)	(0.024)
Gender: female			-0.090^{***}	-0.051^{*}	-0.085^{***}	-0.082^{***}
			(0.015)	(0.027)	(0.021)	(0.016)
Children	-0.083^{**}	-0.052	-0.025	-0.055^{*}	-0.319^{***}	-0.036
	(0.034)	(0.039)	(0.044)	(0.029)	(0.088)	(0.026)
Partner (native German)	-0.354^{***}	-0.358^{***}	-0.405^{***}	-0.239^{***}	-0.298^{***}	-0.351^{***}
	(0.038)	(0.069)	(0.045)	(0.033)	(0.107)	(0.031)
Partner (other native)	-0.028	-0.022	-0.015	-0.024	0.018	-0.037^{**}
	(0.022)	(0.025)	(0.019)	(0.032)	(0.031)	(0.019)
Occ.: high appl./	0.073	0.133^{***}	0.128^{***}	0.070	0.167	0.101^{***}
low comm. skills	(0.045)	(0.035)	(0.031)	(0.051)	(0.082)	(0.030)
Occ.: high appl./	0.108^{***}	0.131^{***}	0.120^{***}	0.123^{***}	0.122^{*}	0.120^{***}
high comm. skills	(0.028)	(0.030)	(0.023)	(0.036)	(0.063)	(0.022)
Occ.: in education	0.190^{***}	0.173^{***}	0.232^{***}	0.320	0.308^{***}	0.194^{***}
	(0.034)	(0.040)	(0.029)	(0.272)	(0.078)	(0.028)
University degree	0.043^{*}	0.048^{*}	0.071^{***}	-0.048		
	(0.023)	(0.027)	(0.019)	(0.048)		
Risk attitude	0.006	0.006	0.002	0.020^{***}	0.007	0.006
	(0.004)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)
Patience	0.002	0.009^{**}	0.004	0.007	-0.0004	0.007^{**}
	(0.004)	(0.004)	(0.003)	(0.006)	(0.005)	(0.003)
English speaker	-0.019	-0.0004	-0.005	-0.033	0.013	-0.024
	(0.020)	(0.021)	(0.016)	(0.029)	(0.023)	(0.018)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
McFadden Pseudo R2	0.15	0.20	0.11	0.15	0.08	0.20
Percent. correctly predicted	68.8	74.9	71.2	72.9	71.7	71.8
Observations	2,810	1,934	3,807	1,084	1,885	3,006

Note: Average marginal effects. Reference category for the occupation categories is "occupation with low internationally applicable skills and low or high communication needs" (for other reference categories, see Table 2.A). Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

applicable skills). Not surprisingly, this relationship does not hold for participants in the older age group. We also observe a large probability of the investment motive for those in the labor market who have highly internationally applicable skills. When considering the subsamples, this pattern holds for those who also have high communication needs and is more mixed when communication is less important. Similarly, having a university degree increases the probability of the investment motive. While English language proficiency has the expected negative relation with the investment motive as long as we do not control for country fixed effects, this relation becomes insignificant in the presence of these fixed effects. Apparently, variation within countries is not so large.

Human capital investment is a risky endeavour as the benefits are not certain, which shows up in the positive and significant coefficient. Looking at the subsamples, we find a positive relation for risk attitude only for the older age group. Patience is an important characteristic as well with regard to human capital investment, where benefits realize much later – if they realize at all. This shows up in the positive correlation between patience and the probability of having an investment motive, which is however more pronounced when country fixed effects are not included. This seems to be mainly due to male participants and those with a university degree.

Figure 2.D graphically displays the results by country groups (see Table 2.J for the estimates). As can be seen, differences are not large.

In order to see how sensitive our results are to our allocation of the *education* category to the investment motive (see the discussion in Chapter 2.3.1), we run estimations where we exclude all those with *education* as their main reason and where we categorize *education* as part of the consumption motive (see Table 2.L). The results for the individual characteristics are qualitatively the same as in Table 2.3. Not too surprisingly given that *education* is particularly important for some country groups, there are a few changes when it comes to the country characteristics. Given that we mostly focus on specifications with country fixed effects, this does, however, not affect our regression analyzes. Overall, we see that the specifications with a different treatment of the *education* category are inferior to our main specification in terms of model fit (McFadden Pseudo R2 and the percentage of correctly predicted observations).

2.5.2 Determinants of the Professional Use of German

Apart from the reasons behind their decision to learn German, participants also indicated the probability of using their foreign language skills in the labor market. If opportunities arise to use them in a professional environment, they have a productive value independent from the main reason behind the learning decision. Based on our expected associations and our graphical illustrations (see Sections 2.3.2 and 2.3.3), we want to inquire what makes the professional use of German language skills more likely. For this, we estimate its determinants and try to identify possible spillovers from the consumption motive.

Country and Individual Level Determinants

We present our main results in Table 2.5. Column 1 includes the same individual-specific characteristics and country fixed effects as in Table 2.3, Column 2 additionally adds the investment motive dummy and Column 3 includes dummies for the categories with *education* as reference category instead of the investment motive dummy. Column 4 re-estimates Column 2 with country characteristics instead of country fixed effects.[16]

When adding the investment motive in Table 2.5, Column 2, and the categories in Column 3, the goodnes of fit measured with the McFadden Pseudo R2 and the percentage

[16]Note that the sample is slightly smaller in Column 3 due to the imputation of the categories and not of the motives as otherwise done. This leads to some missings as described in Section 2.3.1.

Table 2.5: *Professional Use: Basic Specifications*

	Dependent variable: Professional use of German			
	(1)	(2)	(3)	(4)
Investment		0.211^{***}		0.222^{***}
		(0.015)		(0.015)
Domestic labor market			0.052^{**}	
			(0.024)	
Educational and labor migration			0.033	
			(0.022)	
Personal reasons			-0.147^{***}	
			(0.024)	
Cultural interest			-0.198^{***}	
			(0.021)	
Age: under 35 years	0.113^{***}	0.078^{***}	0.069^{***}	0.094^{***}
	(0.023)	(0.022)	(0.022)	(0.021)
Gender: female	0.015	0.032^{**}	0.033^{**}	0.039^{***}
	(0.014)	(0.013)	(0.014)	(0.013)
Children	-0.049^{*}	-0.034	-0.039	-0.018
	(0.026)	(0.025)	(0.026)	(0.024)
Partner (native German)	0.043	0.100^{***}	0.093^{***}	0.074^{**}
	(0.031)	(0.028)	(0.031)	(0.028)
Partner (other native)	0.032^{*}	0.038^{**}	0.034^{**}	0.034^{**}
	(0.017)	(0.016)	(0.017)	(0.016)
Occ.: high appl./low comm. skills	0.039	0.014	0.007	0.011
	(0.033)	(0.033)	(0.034)	(0.033)
Occ.: high appl./high comm. skills	0.113^{***}	0.087^{***}	0.078^{***}	0.086^{***}
	(0.021)	(0.021)	(0.022)	(0.021)
Occ.: in education	0.169^{***}	0.129^{***}	0.136^{***}	0.135^{***}
	(0.025)	(0.025)	(0.026)	(0.025)
University degree	0.007	-0.003	0.002	-0.002
	(0.018)	(0.017)	(0.018)	(0.017)
Risk attitude	0.022^{***}	0.020^{***}	0.020^{***}	0.019^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
Patience	0.004	0.003	0.003	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
English speaker	0.011	0.017	0.022	0.017
	(0.015)	(0.015)	(0.015)	(0.015)
Germanic lang.				-0.165^{***}
				(0.021)
Non-EU (European)				0.018
				(0.038)
Non-EU (Non-European)				-0.158^{***}
				(0.020)
Upper-middle income				-0.001
				(0.029)
High income				-0.136^{***}
				(0.024)
Cultural distance: LTO				0.087^{*}
				(0.045)
Cultural distance: IVR				0.103^{***}
				(0.038)
Country FE	Yes	Yes	Yes	No
McFadden Pseudo R2	0.12	0.15	0.15	0.14
Percent. correctly predicted	68.1	70.1	70.2	70.0
Observations	4,891	4,891	4,654	4,891

Note: Average marginal effects. Reference category for the occupation categories is "occupation with low internationally applicable skills and low or high communication needs" (for other reference categories, see Table 2.A). Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

of correctly predicted observations improves. Furthermore, some variables turn significant, e.g. being female or having a native German partner. As we expected, an investment motive increases the probability of a professional use of German significantly by 21.1 % points. Categories that belong to the consumption motive are negatively associated compared to the reference category *education*. Within the investment categories, individuals that indicate a reason that refers to the *domestic labor market* are more likely to have a high probability of professional use of German in comparison to those who indicate reasons which are part of the *education* category and also of the *educational and labor migration* category.

The country characteristics in Column 4 are similarly correlated to a professional use of German as in Table 2.3, where we considered the determinants of the human capital investment motive, as far as linguistic and economic factors are concerned. On the contrary, a larger geographic distance as we have for non-European countries is now negatively associated with a larger probability of a professional use and a larger cultural distance is now positively associated. The latter holds, in particular, for the IVR measure, which was insignificant in the analyzes above.

The results for individual characteristics are partially in line with previous results on the human capital investment motive (see also Table 2.6 for subgroups by age, gender and education), but there are three important differences. First, the relationship between being female and the probability of a professional use of German now turns positive and significant, but only among the younger individuals and those with a university degree. Second, while having a partner with German as native language is associated negatively with the investment motive, we find a positive and significant relationship for the professional use of German, but more pronounced among younger respondents and those without a university degree. It also holds for both gender subsamples. Children, however, are not significantly related. Third, a university degree and an occupation with highly applicable skills, but no high communication needs are not significantly associated with the probability of a professional use, while both coefficients were significant and positive in the investment motive specification.

As to the different country groups, we see many similarities, but also some differences (see Figure 2.E and Table 2.K). First, Germanic countries stand out as individual characteristics there mostly do not play a role for the professional use of German language skills. Second, EU countries and high-income countries are very similar and the same holds for culturally close, upper-middle income and non-Germanic countries. There are also comparable patterns for lower-middle income countries and culturally more distant countries.

In order to assess the sensitivity of our results relative to our allocation of the *education* category to the investment motive (see the discussion in Chapter 2.3.1), we also run estimations where we exclude all those with *education* as their main reason and where we categorize *education* as part of the consumption motive (see Table 2.M). Results for both the individual characteristics and the country characteristics are qualitatively very similar to those in Table 2.5. Overall, we see that the specification with *education* as part

Table 2.6: *Professional Use: Subsamples by Age, Gender and Education*

	Dependent variable: Professional use of German					
	(1) female	(2) male	(3) age: <35	(4) age: ≥ 35	(5) no uni. deg.	(6) uni. deg.
Investment	0.186^{***}	0.255^{***}	0.162^{***}	0.381^{***}	0.154^{***}	0.248^{***}
	(0.019)	(0.025)	(0.017)	(0.032)	(0.024)	(0.019)
Age: under 35 years	0.091^{***}	0.059^{*}			0.187^{***}	0.058^{**}
	(0.029)	(0.036)			(0.080)	(0.023)
Gender: female			0.034^{**}	0.002	0.057	0.021^{***}
			(0.015)	(0.028)	(0.022)	(0.017)
Children	-0.025	-0.058	-0.007	-0.003	0.123^{*}	-0.049^{*}
	(0.033)	(0.040)	(0.048)	(0.029)	(0.058)	(0.026)
Partner (native German)	0.087^{**}	0.133^{**}	0.107^{***}	0.071	0.020^{***}	0.111
	(0.035)	(0.048)	(0.034)	(0.050)	(0.100)	(0.030)
Partner (other native)	0.015	0.074^{***}	0.074^{***}	-0.062^{*}	0.057	0.031^{*}
	(0.021)	(0.027)	(0.019)	(0.032)	(0.032)	(0.019)
Occ.: high appl./	0.051	-0.026	0.021	-0.016	0.137	0.006
low comm. skills	(0.047)	(0.049)	(0.041)	(0.051)	(0.119)	(0.034)
Occ.: high appl./	0.098^{***}	0.066^{*}	0.091^{***}	0.092^{***}	0.179^{***}	0.086^{**}
high comm. skills	(0.027)	(0.036)	(0.027)	(0.036)	(0.054)	(0.023)
Occ.: in education	0.135^{***}	0.110^{***}	0.165^{***}	-0.065	0.179^{***}	0.139^{**}
	(0.031)	(0.042)	(0.029)	(0.134)	(0.075)	(0.028)
University degree	-0.019	-0.004	0.010	-0.003		
	(0.022)	(0.028)	(0.019)	(0.042)		
Risk attitude	0.019^{***}	0.022^{***}	0.023^{***}	0.009	0.030^{***}	0.015^{***}
	(0.004)	(0.005)	(0.004)	(0.006)	(0.005)	(0.004)
Patience	0.002	0.005	0.003	0.008	0.0003	0.005
	(0.004)	(0.005)	(0.003)	(0.006)	(0.005)	(0.003)
English speaker	0.038^{**}	-0.008	0.024	0.001	0.018	0.020
	(0.019)	(0.023)	(0.017)	(0.030)	(0.024)	(0.018)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
McFadden Pseudo R2	0.16	0.14	0.11	0.22	0.11	0.17
Percent. correctly predicted	70.9	69.8	70.0	74.9	71.0	70.2
Observations	2,810	1,934	3,807	1,084	3,006	1,885

Note: Average marginal effects. Reference category for the occupation categories is "occupation with low internationally applicable skills and low or high communication needs" (for other reference categories, see Table 2.A). Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

of the consumption motive is inferior to the specification where it is part of the investment motive in terms of model fit, while the specification where we drop those with *education* as main reason is comparable to it.

Table 2.7: *Professional Use: Subsamples by Consumption/Investment*

	Dependent variable: Professional use of German			
	(1) consumption good	(2) hum. capital investment	(3) consumption good	(4) hum. capital investment
Age: under 35 years	0.137^{***}	-0.038	0.092^{***}	-0.038
	(0.029)	(0.033)	(0.029)	(0.033)
Gender: female	0.046^{**}	0.015	0.055^{***}	0.016
	(0.022)	(0.017)	(0.021)	(0.017)
Children	-0.017	-0.045	-0.002	-0.045
	(0.033)	(0.044)	(0.032)	(0.043)
Partner (native German)	0.121^{***}	0.069	0.138^{***}	0.070
	(0.036)	(0.082)	(0.035)	(0.082)
Partner (other native)	0.039	0.037^{*}	0.031	0.037^{*}
	(0.026)	(0.021)	(0.026)	(0.021)
Occ.: high appl./low comm. skills	0.068	-0.031	0.057	-0.031
	(0.046)	(0.048)	(0.045)	(0.048)
Occ.: high appl./high comm. skills	0.073^{**}	0.094^{***}	0.056^{*}	0.095^{***}
	(0.031)	(0.030)	(0.030)	(0.030)
Occ.: in education	0.152^{***}	0.132^{***}	0.126^{***}	0.133^{***}
	(0.039)	(0.034)	(0.039)	(0.034)
University degree	-0.028	0.017	-0.040	0.017
	(0.029)	(0.021)	(0.028)	(0.021)
Risk attitude	0.017^{***}	0.021^{***}	0.014^{***}	0.021^{***}
	(0.005)	(0.004)	(0.005)	(0.004)
Patience	0.002	0.003	0.001	0.003
	(0.004)	(0.004)	(0.004)	(0.004)
English speaker	0.010	0.026	0.007	0.026
	(0.022)	(0.019)	(0.022)	(0.019)
Other investm. reason			0.226^{***}	-0.005
			(0.023)	(0.017)
Country FE	Yes	Yes	Yes	Yes
McFadden Pseudo R2	0.08	0.14	0.08	0.18
Percent. correctly predicted	71.8	69.2	72.1	71.6
Observations	2,067	2,824	2,067	2,824

Note: Average marginal effects. Reference category for the occupation categories is "occupation with low internationally applicable skills and low or high communication needs" (for other reference categories, see Table 2.A). Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Spillovers

Three individual characteristics – female gender, partner with German as native language and a younger age – are important determinants for spillovers from a consumption motive to a professional use of German as can be seen in Table 2.7. In addition, it shows in

Column (3) that being still in education (but without a university degree) or in occupation with internationally applicable skills and high communication needs and also a risk proneness lead to a higher probability of a professional use of German, when the main motive of language learning is consumption. We further include in this specification if the respondent indicated a least one reason that we categorize as an investment reason, which is positively related to the likelihood of a professional use of German.

Overall, it seems that while language learning has a larger consumption value for younger women with a native German partner, who might be considered “tied movers”, a professional use of German language skills is not unlikely. This holds especially if the investment motive plays a role as well (even though not the main one).

2.6 Conclusion

While the productive value of language skills has been shown in previous literature, our contribution is to highlight that it is not enough to focus on the human capital aspect of language learning. For the full picture, we enlarge the human capital framework by adding the consumption motive of foreign language learning. Based on a unique dataset collected from close to 5,000 language course participants in 14 countries worldwide, we analyze language learning in a cross-country perspective.

Our results show that the group of language learners is heterogeneous within and between countries. From the perspective of German speaking countries, two points of interest emerge: First, the human capital motive is particularly relevant for course participants in the context of education and the labor market, both abroad or at home. Policy measures targeting this group, such as subsidies for foreign language learning or, in general, more language learning opportunities, should therefore be one focus.[17]

Second, there are possible spillovers from the consumption motive to a professional use of German in the labor market, which might be of interest for policy makers as well. This group mostly comprises younger women with a native German partner, who might be considered “tied movers”. Even though the consumption motive is the main reason for them to learn a foreign language, a professional use of the language is not unlikely. As a consequence, policy measures aiming at this group not only support their social, but also their economic integration in Germany.

The immigration related regulations of recent years in Germany can be viewed in the context of our findings: The “A1 requirement” for family reunification, which became effective in Germany in 2007, introduced the requirement that spouses from non-EU countries must have basic knowledge of German at the A1 level before being granted permission to live in Germany with their partners. By establishing a minimum level of language proficiency of migrants, this regulation lies the basis for the spillover effects from language learning for consumption reasons to an application of the acquired skills in the labor market. In contrast to this, the new Skilled Immigration Act effective since 2020

[17] For macro level analyzes of language learning opportunities and migration to Germany based on aggregate data from the GI, see Huber and Uebelmesser (2023).

facilitates access of skilled workers from third countries to the German labor market and, by doing so, reduces uncertainty related to the returns of investing in the language of the destination country (see Uebelmesser et al. 2022 for an analysis about the consequences for language learning incentives based on macro data from the GI). Overall, the two policies address the two different motives: the consumption motive in the former case and the investment motive in the latter case. Due to the spillover effects identified above, they foster language learning in the migration context and lead to better integrated individuals.

While language learning related to migration is of importance, we need to keep in mind that there are also reasons for language learning in the absence of any migration intention both for investment or consumption purposes. Only when considering all contexts, do we get the full picture.

2.A Appendix: Tables

Table 2.A: *Variable Description: Individual Characteristics*

Variable name	Type	Description
Age: under 35 years	Binary	Indicates whether participant is under 35 years, based on age group according to the ranges: *under 18, 18 to 24, 25 to 34, 35 to 49, 50 to 64, 65 and older*. Reference category is 35 years and older.
Gender: female	Categorical	Indicates whether respondent's gender is female. Reference category is male. Includes a third category of respondents indicating *"No answer/prefer not to say"* or if response is missing that is not reported in the result tables.
Partner: native German	Categorical	Indicates whether respondent has a partner with German as native language. Reference category is single.
Partner: other native	Categorical	Indicates whether respondent has a partner with other native languages than German. Reference category is single.
Children	Binary	Indicates whether respondent has any children. Reference category is no children.
University degree	Binary	Indicates whether respondent has a university degree based on highest educational qualification: *no university degree (no degree, school diploma which cannot lead to higher education, school diploma which can lead to higher education), university degree below PhD, PhD*. Reference category is no university degree.
Occ.: in labor market	Categorical	Indicates whether respondent's main occupation is in labor market based on the category on main occupation: *employee/civil servant with non-highly skilled job, employee/civil servant with highly skilled job, self-employed graduate (lawyer, doctor, ...)/freelance, other self-employed*.
Occ.: in education	Categorical	Indicates whether respondent's main occupation is in education based on the category on main occupation: *pupil, student, student apprentice/(unpaid) trainee or apprentice*.
Occ.: other/no answer	Categorical	Indicates whether respondent's main occupation is not answered or other based on the category on main occupation: *unemployed, housewife/househusband, retiree* or *other*. Not reported in the result tables.
International applicability of skills	Binary	Measures respondents' self-evaluated international applicability of skills (*"If you work abroad, you may be able to use only some of your acquired skills there. How much of your education or professional skills do you think you can use abroad? "*) on a 5 point scale from 1 for *"none"* to 5 for *"all"*. Respondents indicating a value under 4 are categorized as *"low applicability"* and respondents indicating a value of 4 and above as *"high applicability"*.
Importance of communication skills	Binary	Measures respondents' self-evaluated importance of communication skills in professional life (*"How important are communication skills in your professional life?"*) on a 6 point scale from 1 for *"not at all important"* to 6 for *"very important"*. Respondents indicating a value under 5 are categorized as *"low communication skills"* and respondents indicating a value of 5 and above as *"high communication skills"*.
Risk attitude	Numerical (0-10)	Measures respondents' self-reported willingness to take risks (*"Would you describe yourself as someone who tries to avoid risks (risk-averse) or as someone who is willing to take risks (risk-prone)?"*) on a 11 point scale from 0 for *"risk-averse"* to 10 for *"risk-prone"*.
Patience	Numerical (0-10)	Measures respondents' self-reported patience (*"Would you describe yourself as an impatient or a patient person in general?"*) on a 11 point scale from 0 for *"very impatient"* to 10 for *"very patient"*.
English speaker	Binary	Indicates whether respondent is a native speaker of the English language or has command of the English language comparable to a native speaker, indicated by the highest language level (5) on a 5 point scale of 1 for *"basic knowledge"* to 5 for *"very high language level"*. Reference category is no English speaker.

Table 2.B: *Variable Description: Country Characteristics*

Variable name	Type	Description
Germanic language	Binary	Indicates whether the language spoken in the country in which the survey took place is a Germanic language (i.e. English in Great Britain and India, and Dutch in the Netherlands). Reference category is non-Germanic.
European / EU	Categorical	Indicates whether the country in which the survey took place is a European country that is a member of the EU (Netherlands, Great Britain, Spain, Italy, Czech Republic, Poland and Romania), a European country that is not a member of the EU (Ukraine and Bosnia and Herzegovina), or a non-European country (Japan, Korea, Mexico, India and Indonesia). Reference category is EU-country.
Income	Categorical	Indicates income levels of the country in which the survey took place: High income (Czech Republic, Great Britain, Italy, Japan, Netherlands, Poland, South Korea and Spain), upper-middle income (Bosnia and Herzegovina, Mexico and Romania), and lower-middle income (India, Indonesia and Ukraine). *Source: World Development Indicator in 2018 (World Bank 2021).*
Cultural Distance to Germany: LTO	Rate (0–1)	Distance in long term orientation index to Germany by country, normalized between 0 (lowest distance) to 1 (largest distance), as of 2013 or latest year available. *Source: Hofstede and Minkov (2013).*
Cultural Distance to Germany: IVR	Rate (0–1)	Distance in indulgence restraint index to Germany by country, normalized between 0 (lowest distance) to 1 (largest distance), as of 2013 or latest year available. *Source: Hofstede and Minkov (2013).*

Table 2.C: *Sample Shrinkage*

Step	Action	EU	non-EU		Total
			European	non-European	
1	Completed questionnaires	2,645	1,052	2,967	6,664
2	Remove missings: Age	2,631	1,047	2,958	6,636
3	Remove missings: Children and partner	2,616	1,030	2,928	6,574
4	Remove missings: University degree	2,593	1,016	2,902	6,511
5	Remove missings: Risk attitude	2,582	1,014	2,876	6,472
6	Remove missings: Patience	2,579	1,010	2,869	6,458
7	Remove missings: English speaker	2,579	1,009	2,856	6,444
8	Remove missings: Applic. and comm. skills	2,375	971	2,821	6,167
9	Remove missings: Reasons for lang. learn.	2,306	892	2,765	5,963
10	Remove missings: Main reason for lang. learn.	2,040	754	2,097	4,891

Note: No missings for gender and occupations (see Table 2.A for further explanations).

Table 2.D: *Descriptive Statistics:*
Means of Individual Characteristics by Country Groups (Sample Before Droppings)

Variable	EU	non-EU		Total
		European	non-European	
	n=2645	*n=1052*	*n=2967*	***n=6664***
Age: under 35 years *(28)*	0.63	0.79	0.88	0.77
Gender *(0)*				
Gender: male	0.39	0.39	0.39	0.39
Gender: female	0.58	0.58	0.57	0.58
Gender: n/a	0.03	0.03	0.04	0.04
Children *(2)*	0.20	0.19	0.06	0.14
Partner *(70)*				
No partner	0.44	0.57	0.79	0.62
Partner (native German)	0.09	0.03	0.03	0.05
Partner (other native)	0.48	0.40	0.18	0.33
Occupation *(290)**				
Occ.: low appl.	0.17	0.11	0.10	0.13
Occ.: high appl./low comm. skills	0.08	0.05	0.04	0.05
Occ.: high appl./high comm. skills	0.31	0.20	0.16	0.22
Occ.: in education	0.28	0.48	0.59	0.45
Occ.: other occ./no answer	0.17	0.16	0.12	0.14
University degree *(76)*	0.75	0.56	0.52	0.62
Risk attitude *(61)*	6.10	6.81	6.52	6.40
Patience *(52)*	5.83	6.60	6.30	6.16
English speaker *(14)*	0.47	0.34	0.20	0.33

Note: Number of missing observations per variable in parentheses. Missing observations excluded from means.
* Missings due to variables "International applicability of skills" and "Importance of communication skills".

Table 2.E: *Descriptive Statistics: Means of Individual Characteristics by Countries*

Variable	BIH n=(143)	CZE n=(362)	ESP n=(521)	GBR n=(380)	IDN n=(705)	IND n=(491)	ITA n=(272)	JPN n=(225)	KOR n=(309)	MEX n=(367)	NLD n=(86)	POL n=(154)	ROU n=(265)	UKR n=(611)	Total n=(4891)
Age: under 35 years	0.75	0.70	0.75	0.50	0.96	0.97	0.65	0.40	0.93	0.83	0.60	0.68	0.72	0.83	0.78
Gender: male	0.48	0.35	0.41	0.48	0.45	0.37	0.39	0.32	0.33	0.45	0.44	0.34	0.33	0.36	0.40
Gender: female	0.51	0.63	0.57	0.48	0.48	0.61	0.58	0.67	0.66	0.51	0.53	0.61	0.65	0.61	0.57
Gender: n/a	0.01	0.02	0.02	0.04	0.07	0.02	0.03	0.02	0.01	0.03	0.02	0.05	0.02	0.03	0.03
Children	0.20	0.24	0.12	0.22	0.04	0.03	0.16	0.25	0.04	0.09	0.21	0.18	0.22	0.17	0.13
No partner	0.52	0.36	0.62	0.32	0.81	0.94	0.41	0.37	0.91	0.63	0.38	0.42	0.46	0.60	0.61
Partner (native German)	0.02	0.05	0.05	0.27	0.03	0	0.05	0.11	0	0.04	0.17	0.03	0.02	0.03	0.05
Partner (other native)	0.45	0.58	0.33	0.41	0.16	0.05	0.54	0.52	0.09	0.34	0.44	0.55	0.52	0.37	0.34
Occ.: low appl.	0.06	0.21	0.08	0.20	0.08	0.04	0.15	0.44	0.08	0.10	0.19	0.25	0.17	0.12	0.13
Occ.: high appl./low comm. skills	0.13	0.07	0.08	0.09	0.04	0.00	0.08	0.07	0.03	0.08	0.09	0.08	0.07	0.04	0.06
Occ.: high appl./high comm. skills	0.22	0.31	0.25	0.45	0.13	0.13	0.22	0.12	0.07	0.40	0.30	0.29	0.35	0.18	0.23
Occ.: in education	0.41	0.31	0.42	0.06	0.68	0.77	0.34	0.14	0.67	0.32	0.23	0.29	0.31	0.54	0.45
Occ.: other occ./no answer	0.19	0.10	0.17	0.19	0.08	0.06	0.21	0.24	0.15	0.11	0.19	0.09	0.10	0.12	0.13
University degree	0.59	0.66	0.70	0.95	0.28	0.57	0.71	0.86	0.44	0.71	0.71	0.73	0.70	0.56	0.61
Risk attitude	7.01	5.64	6.36	5.52	6.96	6.59	6.89	5.28	5.62	6.91	6.34	5.82	6.21	6.72	6.36
Patience	6.75	5.55	5.87	5.34	6.07	6.57	6.09	6.24	6.18	6.22	5.74	5.75	6.38	6.55	6.11
English speaker	0.41	0.49	0.57	0.26	0.32	0.02	0.43	0.16	0.07	0.47	0.51	0.58	0.57	0.33	0.35

Table 2.F: *Questions on (Main) Reasons for Learning German*

(i) Why are you learning German? (Multiple choices)	**Labels (in the paper)**
☐ Study/education/training/PhD[b]	Education
☐ Internal company communication	Company communication
☐ German speaking trading partners	Trading partners
☐ German speaking customers	Customers
☐ Higher income in the [country]	Income
☐ Requirement/support of the employer	Employer
☐ Other considerations regarding career/labor market in [country]	Labor market
☐ Partner or family	Family
☐ Social environment/friends	Friends
☐ (Possible) move to a German speaking c. for professional reasons[a]	Educational and labor migr.
☐ (Possible) move to a German speaking c. for other reasons[a]	Other migration
☐ Requirement for visa[a]	
☐ Holiday	Holiday
☐ Culture (film, literature,...)	Culture
☐ Interest in languages	Interest
☐ Other: [free-text]	

(ii) Look at your answers and circle the main reason why you are learning German.

Note:

[a] Note that in Japan, Bosnia and Herzegovina, Great Britain and Poland we have not distinguished between "(Possible) move to a German speaking country for professional reasons" and "(Possible) move to a German speaking country for other reasons". In these cases, we have imputed the reasons by making use of the main reason for a potential move to a German speaking country, which the respondents answered in the survey as well. The same method was applied for the category "Requirement for visa" in all surveys.

[b] In Indonesia, we split the category "Study/education/training/PhD" into "Study/education/PhD" and "(Vocational) Training", but re-merged it for our analysis.

Other reasons were categorized according to the free-text field if possible.

Table 2.G: *Tetrachoric Correlation Coefficients*

	Interest	Culture	Holiday	Friends	Family	Other migrat.	Educ., labor migration	Educat.	Labor market	Income	Employer	Customer	Trade partner
Interest													
Culture	0.52***												
Holiday	0.30***	0.52***											
Friends	0.27***	0.33***	0.41***										
Family	-0.14***	-0.09***	0.06***	0.24***									
Other migration	0.10***	0.09***	0.17***	0.23***	0.37***								
Educ., lab. mig.	0.04***	0.07***	-0.04**	0.07***	-0.24***	0.03**							
Education	-0.13***	-0.06***	-0.17***	-0.08***	-0.38***	-0.09***	0.27***						
Labor market	0.10***	0.06***	0.09***	0.05***	-0.16***	0.00	0.10***	-0.04***					
Income	0.03**	0.03*	0.04***	0.08***	-0.22***	-0.01	0.21***	0.11***	0.31***				
Employer	0.00	0.05***	0.05***	0.10***	-0.08***	-0.08***	-0.02	-0.12***	0.17***	0.25***			
Customer	-0.02	0.00	0.04***	0.04***	-0.10***	-0.09***	0.05***	-0.19***	0.15***	0.36***	0.45***		
Trade partner	-0.12***	-0.04**	0.08***	0.04***	-0.13***	-0.04***	0.03**	-0.13***	0.11***	0.30***	0.42***	0.70***	
Company comm.	-0.07***	-0.03*	-0.04***	0.06***	-0.09***	-0.07***	0.07***	-0.10***	0.13***	0.28***	0.51***	0.61***	0.57***

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.H: *Tetrachoric Correlation Coefficients for Categories*

	Cultural interest	Education	Domestic labor market	Education, labor migration
Cultural interest				
Education	-0.17***			
Domestic labor market	0.04***	-0.06***		
Education, labor migr.	0.04***	0.27***	0.15***	
Personal reasons	0.16***	-0.19***	-0.11***	-0.01

Note: $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.I: *Human Capital Investment: Basic Specifications with Country FE Interacted with Gender and Age*

	Dependent variable: Human capital investment		
	(1)	(2)	(3)
Age: under 35 years	0.157^{***}	0.025	0.019
	(0.023)	(0.085)	(0.085)
Gender: female	-0.082^{***}	-0.083^{***}	-0.091
	(0.013)	(0.013)	(0.068)
Children	-0.071^{***}	-0.074^{***}	-0.074^{***}
	(0.025)	(0.026)	(0.026)
Partner (native German)	-0.360^{***}	-0.356^{***}	-0.352^{***}
	(0.034)	(0.034)	(0.034)
Partner (other native)	-0.022	-0.020	-0.020
	(0.016)	(0.016)	(0.016)
Occ.: high appl./low comm. skills	0.106^{***}	0.105^{***}	0.103^{***}
	(0.027)	(0.028)	(0.028)
Occ.: high appl./high comm. skills	0.119^{***}	0.120^{***}	0.117^{***}
	(0.020)	(0.020)	(0.020)
Occ.: in education	0.186^{***}	0.189^{***}	0.187^{***}
	(0.026)	(0.026)	(0.026)
University degree	0.045^{***}	0.047^{***}	0.048^{***}
	(0.017)	(0.017)	(0.017)
Risk attitude	0.006^{**}	0.007^{**}	0.007^{**}
	(0.003)	(0.003)	(0.003)
Patience	0.005^{*}	0.005^{*}	0.005^{*}
	(0.003)	(0.003)	(0.003)
English speaker	-0.013	-0.011	-0.009
	(0.014)	(0.014)	(0.014)
Country FE	Yes	Yes	Yes
Country FE $\times$ Age	No	Yes	Yes
Country FE $\times$ Gender	No	No	Yes
McFadden Pseudo R2	0.17	0.17	0.17
Percent. correctly predicted	71.3	71.3	71.3
Observations	4,891	4,891	4,891

Note: Average marginal effects. Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.J: *Human Capital Investment: Subsamples by Country Characteristics*

	Dependent variable: Human capital investment									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Language		Income			Geography			Culture	
						EU	non-EU			
	Germanic	non-Germanic	Lower-middle	Upper-middle	High		European	non-European	Close	Far
Age: under 35 years	0.067	0.163^{***}	0.204^{***}	0.197^{***}	0.195^{***}	0.131^{***}	0.145^{*}	0.115^{***}	0.213^{***}	0.139^{***}
	(0.047)	(0.026)	(0.055)	(0.058)	(0.027)	(0.029)	(0.066)	(0.043)	(0.033)	(0.032)
Gender: female	-0.056^{**}	-0.086^{***}	-0.118^{*}	-0.067^{***}	-0.103^{***}	-0.059^{***}	-0.068^{***}	-0.103^{***}	-0.085^{***}	-0.098^{***}
	(0.025)	(0.015)	(0.021)	(0.035)	(0.019)	(0.021)	(0.033)	(0.019)	(0.018)	(0.019)
Children	−0.079	-0.076^{***}	0.003	−0.043	-0.020^{***}	-0.094^{***}	−0.091	−0.007	-0.069^{**}	-0.079^{*}
	(0.056)	(0.029)	(0.047)	(0.064)	(0.032)	(0.035)	(0.061)	(0.045)	(0.034)	(0.038)
Partner (native German)	-0.348^{***}	-0.332^{***}	-0.501^{***}	-0.497^{***}	-0.511^{***}	-0.276^{***}	-0.286^{***}	-0.486^{***}	-0.463^{***}	-0.369^{***}
	(0.061)	(0.046)	(0.080)	(0.076)	(0.035)	(0.036)	(0.099)	(0.069)	(0.051)	(0.043)
Partner (other native)	−0.002	−0.020	-0.035^{**}	−0.087	−0.054	0.008	0.005	−0.061	−0.031	−0.019
	(0.038)	(0.019)	(0.030)	(0.040)	(0.022)	(0.023)	(0.038)	(0.029)	(0.022)	(0.024)
Occ.: high appl./low comm. skills	0.045	0.111^{***}	0.173	0.051^{***}	0.128^{**}	0.090	0.059^{***}	0.190^{***}	0.121^{***}	0.135
	(0.071)	(0.032)	(0.038)	(0.075)	(0.039)	(0.043)	(0.055)	(0.041)	(0.035)	(0.044)
Occ.: high appl./high comm. skills	0.142^{***}	0.122^{***}	0.132^{*}	0.097^{***}	0.126^{***}	0.118^{***}	0.086^{**}	0.121^{***}	0.154^{***}	0.122^{***}
	(0.039)	(0.024)	(0.030)	(0.056)	(0.028)	(0.031)	(0.050)	(0.027)	(0.027)	(0.031)
Occ.: in education	0.249^{***}	0.187^{***}	0.239^{**}	0.147^{***}	0.206^{***}	0.182^{***}	0.111^{***}	0.309^{***}	0.238^{***}	0.204^{***}
	(0.065)	(0.028)	(0.045)	(0.066)	(0.036)	(0.040)	(0.062)	(0.038)	(0.034)	(0.040)
University degree	0.024	0.036^{*}	0.041^{**}	0.121	0.066	0.032	0.012	0.014^{***}	0.094^{*}	0.045^{*}
	(0.033)	(0.019)	(0.027)	(0.049)	(0.025)	(0.029)	(0.039)	(0.025)	(0.023)	(0.026)
Risk attitude	0.006	0.006	0.011	-0.0002^{**}	0.007	0.004	0.002	0.0004^{**}	0.011	0.006
	(0.006)	(0.004)	(0.005)	(0.008)	(0.004)	(0.005)	(0.008)	(0.005)	(0.004)	(0.005)
Patience	0.007	0.006^{*}	0.006	0.009	0.007	0.002	0.0003	0.006^{**}	0.009	0.004
	(0.006)	(0.003)	(0.005)	(0.007)	(0.004)	(0.004)	(0.007)	(0.004)	(0.004)	(0.004)
English speaker	0.049	−0.019	−0.011	−0.003	−0.009	−0.016	−0.012	0.007	−0.017	−0.026
	(0.038)	(0.017)	(0.025)	(0.034)	(0.021)	(0.021)	(0.035)	(0.024)	(0.019)	(0.021)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	957	3,720	775	1,807	2,309	2,040	754	2,097	2,627	2,264

Note: Average marginal effects. Heteroscedasticity robust White standard errors in parentheses. For the groupings, see Table 2.B; culturally close countries to Germany are Bosnia and Herzegovina, Czech Republic, Indonesia, Italy, Japan, South Korea and Ukraine and culturally more distant countries are Great Britain, India, Mexico, Netherlands, Poland, Romania and Spain based on the sum of distances of the two culture distance measures (LTO and IVR). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.K: *Professional Use: Subsamples by Country Characteristics*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Language		Income			Geography			Culture	
						EU	non-EU			
	Germanic	non-Germanic	Lower-middle	Upper-middle	High		European	non-European	Close	Far
Investment	0.215***	0.225***	0.152***	0.199***	0.257***	0.234***	0.172***	0.196***	0.198***	0.227***
	(0.042)	(0.017)	(0.026)	(0.035)	(0.022)	(0.023)	(0.035)	(0.026)	(0.021)	(0.023)
Age: under 35 years	−0.071	0.100***	0.101**	0.117*	0.050*	0.049*	0.041	0.148***	0.062*	0.073**
	(0.049)	(0.025)	(0.054)	(0.052)	(0.028)	(0.028)	(0.053)	(0.047)	(0.032)	(0.031)
Gender: female	−0.0002	0.040**	0.061	0.022***	0.013	0.025	0.110***	0.010	0.043**	0.017
	(0.032)	(0.016)	(0.023)	(0.032)	(0.020)	(0.021)	(0.031)	(0.022)	(0.019)	(0.020)
Children	−0.044	−0.029	−0.003	0.024	−0.062*	−0.065*	0.013	−0.012	−0.024	−0.059
	(0.062)	(0.028)	(0.053)	(0.051)	(0.034)	(0.034)	(0.048)	(0.055)	(0.035)	(0.038)
Partner (native German)	0.068	0.134***	0.237	0.023***	0.092**	0.082**	0.143*	0.189***	0.200***	0.050
	(0.058)	(0.036)	(0.047)	(0.106)	(0.036)	(0.037)	(0.054)	(0.059)	(0.036)	(0.042)
Partner (other native)	0.024	0.028	0.066	0.033**	0.027	0.034	0.002	0.064**	0.046**	0.027
	(0.049)	(0.018)	(0.032)	(0.036)	(0.023)	(0.023)	(0.035)	(0.031)	(0.023)	(0.025)
Occ.: high appl./low comm. skills	0.126	−0.013	0.100	−0.050	−0.008	−0.061	0.031	0.111*	0.052	−0.042
	(0.086)	(0.038)	(0.061)	(0.073)	(0.045)	(0.046)	(0.062)	(0.063)	(0.044)	(0.050)
Occ.: high appl./high comm. skills	−0.005	0.085***	0.170	0.048***	0.045	0.050	0.154***	0.068	0.138***	0.026
	(0.056)	(0.024)	(0.034)	(0.048)	(0.031)	(0.031)	(0.034)	(0.042)	(0.028)	(0.033)
Occ.: in education	0.069	0.126***	0.263*	0.102***	0.070**	0.066*	0.154***	0.200***	0.146***	0.121***
	(0.064)	(0.027)	(0.042)	(0.057)	(0.036)	(0.038)	(0.052)	(0.043)	(0.033)	(0.039)
University degree	0.049	0.011	0.022	−0.018	−0.005	−0.021	−0.020	0.039	−0.013	0.016
	(0.039)	(0.019)	(0.029)	(0.042)	(0.026)	(0.028)	(0.032)	(0.030)	(0.024)	(0.026)
Risk attitude	−0.003	0.020***	0.022***	0.020***	0.020***	0.018***	0.030***	0.020***	0.027***	0.012***
	(0.008)	(0.004)	(0.006)	(0.007)	(0.005)	(0.005)	(0.007)	(0.005)	(0.005)	(0.005)
Patience	−0.004	0.005	−0.005	0.003	0.008*	0.005	0.004	0.0004	0.006	−0.0003
	(0.007)	(0.003)	(0.005)	(0.006)	(0.004)	(0.004)	(0.006)	(0.005)	(0.004)	(0.004)
English speaker	0.074	0.015	−0.003	0.022	0.031	0.022	−0.009	0.032	0.012	0.019
	(0.049)	(0.017)	(0.027)	(0.031)	(0.022)	(0.021)	(0.031)	(0.027)	(0.020)	(0.022)
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	892	3,525	735	1,715	2,204	1,984	726	1,944	2,501	2,153

Note: Average marginal effects. Heteroscedasticity robust White standard errors in parentheses. For the groupings, see Table 2.B; culturally close countries to Germany are Bosnia and Herzegovina, Czech Republic, Indonesia, Italy, Japan, South Korea and Ukraine and culturally more distant countries are Great Britain, India, Mexico, Netherlands, Poland, Romania and Spain based on the sum of distances of the two culture distance measures (LTO and IVR). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.L: *Human Capital Investment:*
Basic Specifications without Education as Motive and with Education as Consumption Motive

	Dependent variable: Human capital investment			
	without *education* as motive		*education* as consumption motive	
	(1)	(2)	(3)	(4)
Age: under 35 years	0.125^{***}	0.156^{***}	0.084^{***}	0.109^{***}
	(0.023)	(0.023)	(0.020)	(0.020)
Gender: female	-0.095^{***}	-0.079^{***}	-0.059^{***}	-0.039^{***}
	(0.016)	(0.016)	(0.013)	(0.013)
Children	-0.066^{**}	-0.046^{*}	-0.046^{*}	-0.027
	(0.027)	(0.028)	(0.024)	(0.025)
Partner (native German)	-0.309^{***}	-0.357^{***}	-0.234^{***}	-0.278^{***}
	(0.029)	(0.024)	(0.022)	(0.017)
Partner (other native)	-0.013	-0.019	0.003	-0.0001
	(0.019)	(0.019)	(0.016)	(0.017)
Occ.: high appl./low comm. skills	0.114^{***}	0.110^{***}	0.092^{***}	0.081^{**}
	(0.034)	(0.035)	(0.033)	(0.035)
Occ.: high appl./high comm. skills	0.130^{***}	0.131^{***}	0.104^{***}	0.096^{***}
	(0.025)	(0.025)	(0.024)	(0.025)
Occ.: in education	0.104^{***}	0.147^{***}	-0.008	0.025
	(0.029)	(0.029)	(0.025)	(0.025)
University degree	0.058^{***}	0.083^{***}	0.055^{***}	0.084^{***}
	(0.021)	(0.021)	(0.017)	(0.018)
Risk attitude	0.009^{**}	0.009^{**}	0.008^{***}	0.008^{***}
	(0.004)	(0.004)	(0.003)	(0.003)
Patience	0.007^{**}	0.010^{***}	0.005^{*}	0.009^{***}
	(0.003)	(0.003)	(0.003)	(0.003)
English speaker	-0.025	-0.042^{**}	-0.024^{*}	-0.041^{***}
	(0.017)	(0.018)	(0.014)	(0.014)
Germanic lang.		-0.023		0.077^{***}
		(0.023)		(0.021)
European (Non-EU)		-0.051		0.005
		(0.043)		(0.038)
Non-European		-0.003		-0.062^{***}
		(0.024)		(0.021)
Upper-middle income		0.013		0.096^{***}
		(0.034)		(0.030)
High income		-0.095^{***}		-0.005
		(0.030)		(0.023)
Cultural distance: LTO		-0.121^{**}		-0.035
		(0.053)		(0.048)
Cultural distance: IVR		0.031		-0.012
		(0.044)		(0.037)
Country FE	Yes	No	Yes	No
McFadden Pseudo R2	0.14	0.10	0.09	0.05
Percent. correctly predicted	67.2	63.5	69.7	68.0
Observations	3,651	3,651	4,891	4,891

Note: Average marginal effects. Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 2.M: *Professional Use:*
Basic Specifications without Education as Motive and with Education as Consumption Motive

	Dependent variable: Professional use of German			
	without *education* as motive		*education* as consumption motive	
	(1)	(2)	(3)	(4)
Investment	0.226^{***}	0.239^{***}	0.157^{***}	0.168^{***}
	(0.017)	(0.016)	(0.015)	(0.014)
Age: under 35 years	0.081^{***}	0.096^{***}	0.099^{***}	0.116^{***}
	(0.023)	(0.022)	(0.022)	(0.022)
Gender: female	0.036^{**}	0.041^{***}	0.024^{*}	0.030^{**}
	(0.016)	(0.016)	(0.014)	(0.014)
Children	-0.027	-0.013	-0.042^{*}	-0.025
	(0.026)	(0.025)	(0.025)	(0.024)
Partner (native German)	0.098^{***}	0.081^{***}	0.074^{**}	0.049
	(0.030)	(0.030)	(0.029)	(0.029)
Partner (other native)	0.033^{*}	0.031	0.032^{*}	0.026
	(0.019)	(0.019)	(0.017)	(0.017)
Occ.: high appl./low comm. skills	0.021	0.020	0.024	0.023
	(0.035)	(0.035)	(0.033)	(0.032)
Occ.: high appl./high comm. skills	0.080^{***}	0.080^{***}	0.098^{***}	0.099^{***}
	(0.023)	(0.023)	(0.021)	(0.021)
Occ.: in education	0.110^{***}	0.115^{***}	0.170^{***}	0.177^{***}
	(0.028)	(0.028)	(0.025)	(0.024)
University degree	-0.023	-0.027	-0.002	-0.005
	(0.020)	(0.020)	(0.018)	(0.018)
Risk attitude	0.019^{***}	0.018^{***}	0.021^{***}	0.020^{***}
	(0.004)	(0.004)	(0.003)	(0.003)
Patience	0.003	0.004	0.003	0.004
	(0.003)	(0.003)	(0.003)	(0.003)
English speaker	0.013	0.015	0.016	0.017
	(0.017)	(0.017)	(0.015)	(0.015)
Germanic lang.		-0.175^{***}		-0.199^{***}
		(0.023)		(0.021)
Non-EU (European)		-0.035		0.0002
		(0.043)		(0.038)
Non-EU (Non-European)		-0.172^{***}		-0.135^{***}
		(0.022)		(0.020)
Upper-middle income		-0.030		-0.023
		(0.033)		(0.029)
High income		-0.193^{***}		-0.159^{***}
		(0.029)		(0.024)
Cultural distance: LTO		0.039		0.051
		(0.049)		(0.045)
Cultural distance: IVR		0.096^{**}		0.121^{***}
		(0.042)		(0.038)
Country FE	Yes	No	Yes	No
McFadden Pseudo R2	0.16	0.15	0.13	0.13
Percent. correctly predicted	70.8	70.4	69.7	69.4
Observations	3,651	3,651	4,891	4,891

Note: Average marginal effects. Heteroscedasticity robust White standard errors in parentheses. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

2.B Appendix: Figures

Figure 2.A: *Professional Use of German: 1=very unlikely, 5=very likely (n=4891)*

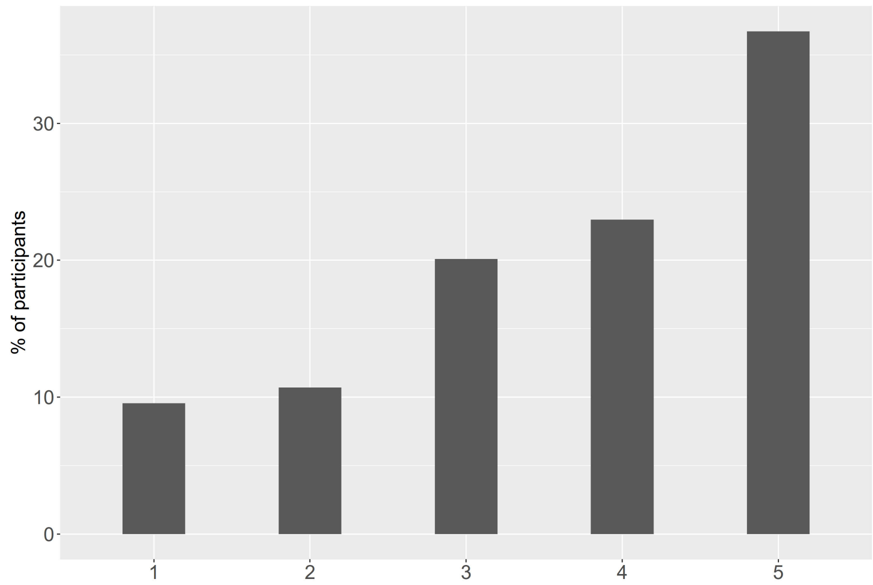


Figure 2.B: *Main Reasons for Learning German by Countries (n=4891)*

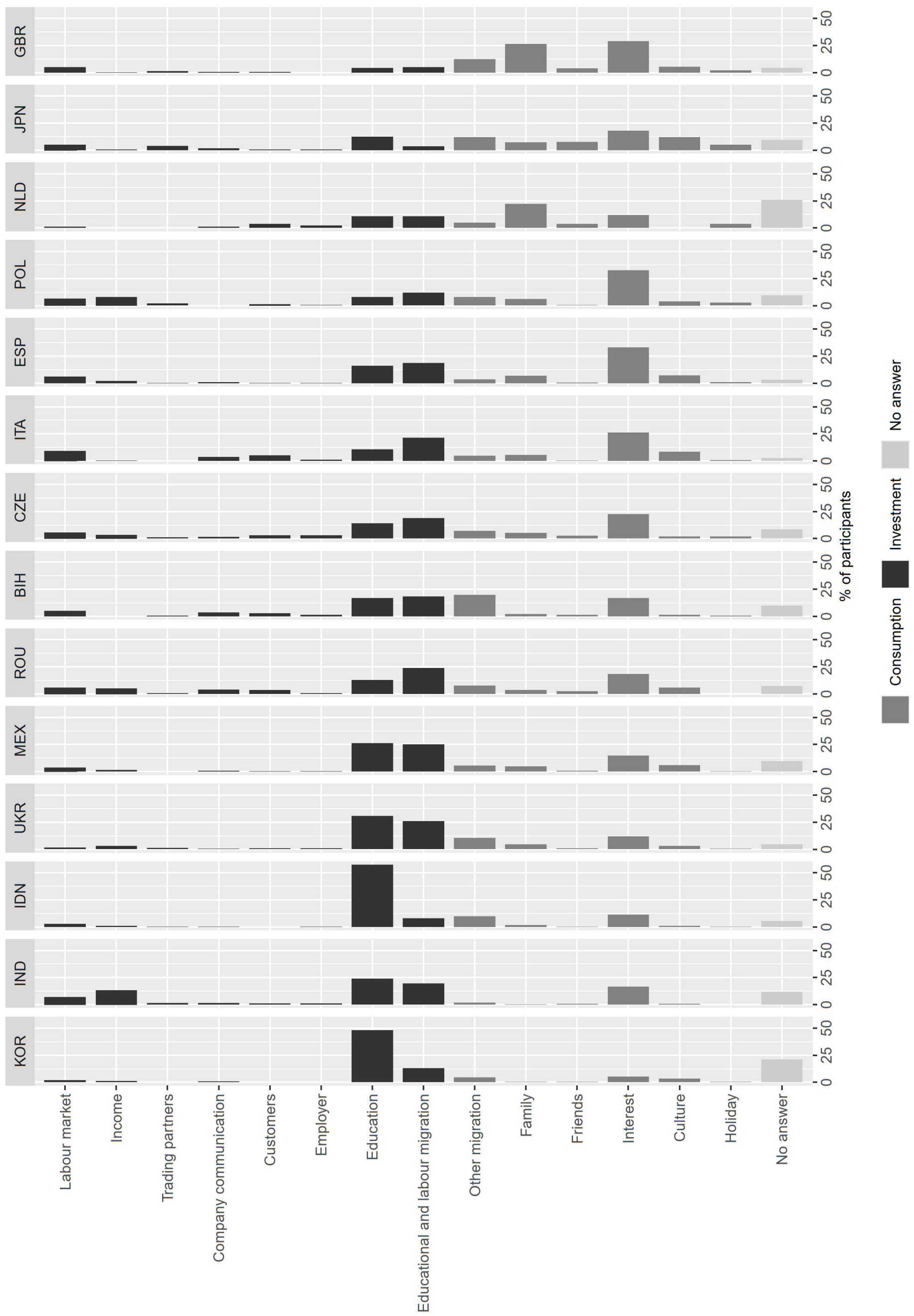


Figure 2.C: *Professional Use of German: Subsamples by Countries and Economic Motive*

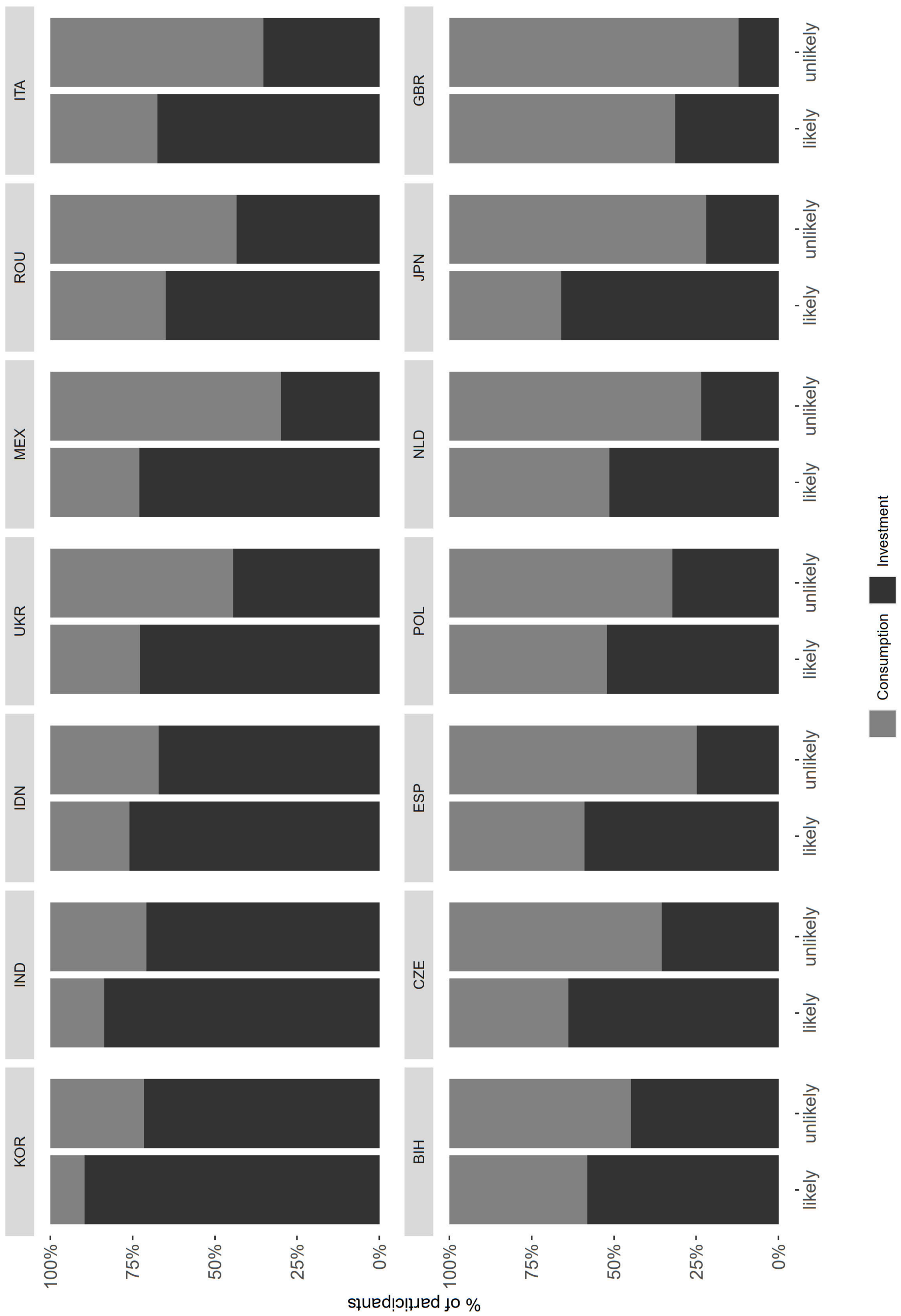
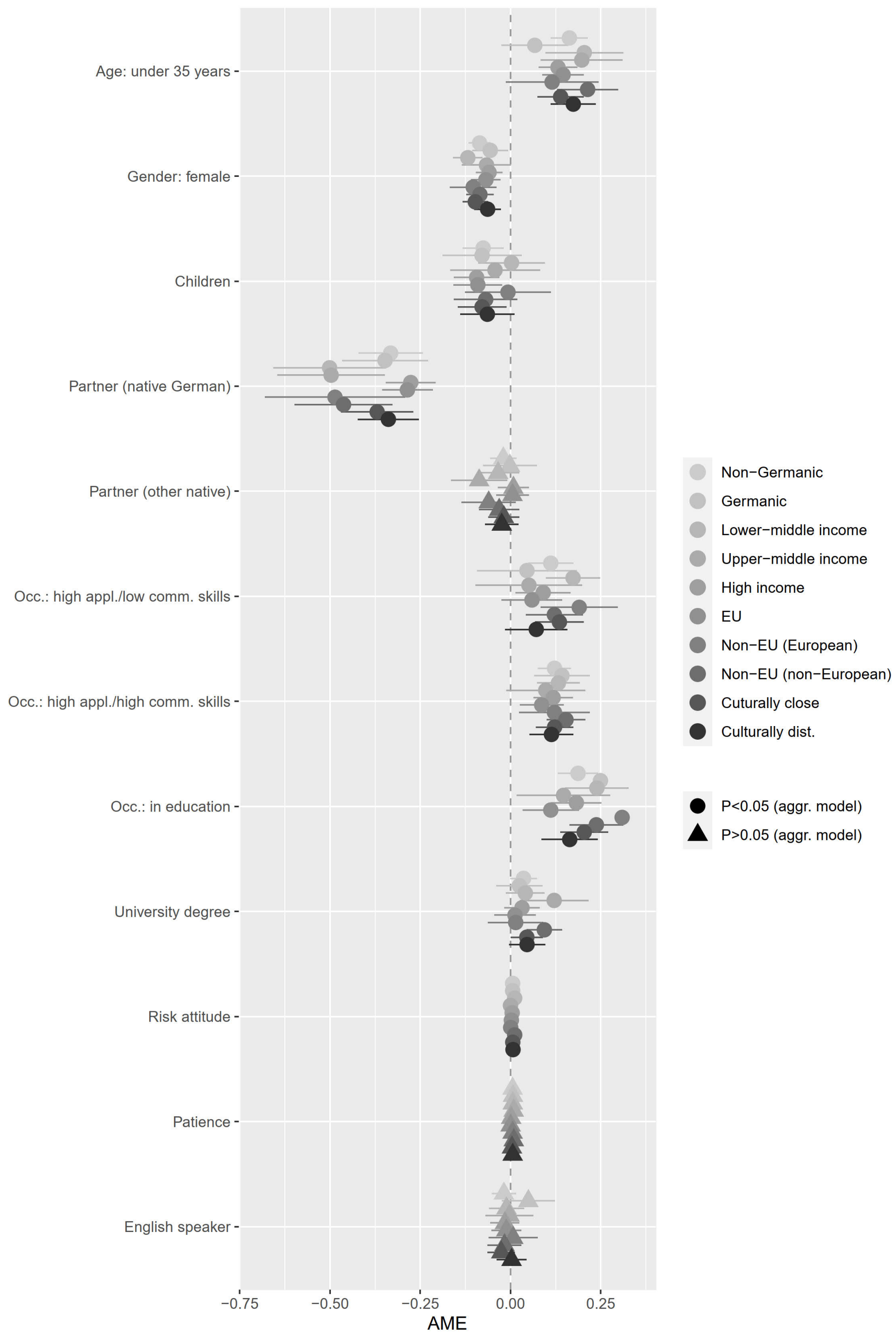
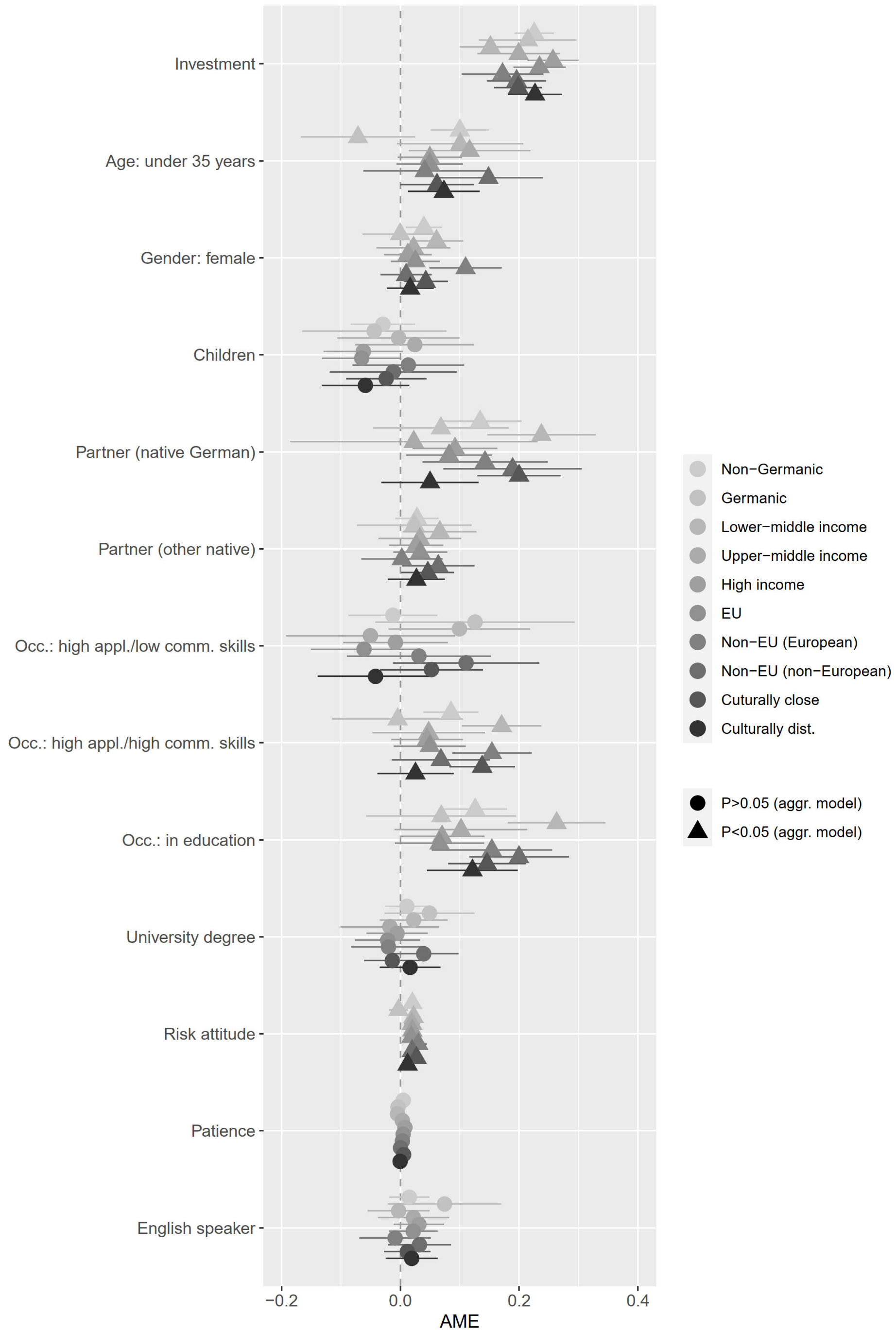


Figure 2.D: *Human Capital Investment: Subsamples by Country Groups According to Country Characteristics*



Note: Average marginal effects and 95 % CI. Shapes according to the p-values of the aggregated model, see Table 2.3 Column 2. For detailed estimation results, see Table 2.J in the Appendix.

Figure 2.E: *Professional Use: Subsamples by Country Groups According to Country Characteristics*



Note: Average marginal effects and 95 % CI. Shapes according to the p-values of the aggregated model, see Table 2.5 Column 3. For detailed estimation results see Table 2.K in the Appendix.

3 A Macro-Level Analysis of Language Learning and Migration[1]

with SILKE UEBELMESSER[2]
and SEVERIN WEINGARTEN[3]

Abstract. *This article investigates the macro-level drivers of adult age language learning with a focus on migration based on a new dataset on German language learning in 77 countries (including Germany) for 1992-2006. Fixed effects regressions show that language learning abroad is strongly associated with immigration from countries of the European Union and the Schengen Area whose citizens enjoy free access to Germany, while language learning in Germany is strongly associated with immigration from countries with restricted access. The different degrees of uncertainty about access to Germany seem to be of importance for preparatory language learning. To shed light on country heterogeneities, we substitute the location fixed effects with a vector of country characteristics, which include several distance measures among others, and we estimate a random effects model. Last, we provide some tentative arguments in favor of a causal interpretation. The main results related to the role of uncertainty are mostly unaffected. The Skilled Immigration Act from 2020 removes this uncertainty with potential positive effects on preparatory language learning and economic and social integration.*

Keywords: *language skills, language learning, migration, labor market access, new dataset*

3.1 Introduction

Language skills are a crucial prerequisite for communication across country borders and for the international mobility of people and goods. There are obvious benefits for migrants and linguistic minorities to learn the primary language in their country of residence. Related to the labor market, language skills have a large positive effect on wages (Dustmann and Van Soest 2001, 2002) and employment probabilities (Dustmann and Fabbri 2003) with possible gender differences (Yao and van Ours 2015).[4] The

[1]The Version of Record of this article is published in *German Economic Review* 23.2, 181–232 (2021), and is available online at https://doi.org/10.1515/ger-2020-0067.

[2]Friedrich Schiller University Jena and CESifo

[3]Friedrich Schiller University Jena

[4]For an overview of the labor market effects, see Chiswick and Miller (2007b).

effects are, however, not limited to the factor labor. They have been found for social integration (Aldashev et al. 2009) more broadly as well as for education and health (Aoki and Santiago 2018). Better language skills increase the probability of intermarriage and reduce the likelihood of living in an ethnic enclave (Bleakley and Chin 2010). In addition, common language skills positively affect trade as shown by Lohmann (2011), Isphording and Otten (2013) and Melitz and Toubal (2014) among others.

Overall, the literature has long recognized the importance of language skills for migration choice and subsequent integration by controlling for common languages of origin and destination countries (e.g. Belot and Hatton 2012; Grogger and Hanson 2011; Mayda 2010; Ortega and Peri 2013). More recently, Adserà and Pytliková (2015) and Belot and Ederveen (2012) show that linguistic distance, which can be interpreted as the difficulty associated with learning another language, has an important effect on international migration flows. However, studies which focus on the concept of linguistic distance neglect the possibility that migrants acquire the language of the host country and overcome the negative effects of linguistic distance. An exception is Aparicio Fenoll and Kuehn (2016) who find evidence that language education at school can affect migration decisions. Their results show that the study of language learning can add value to a literature which has previously focused on linguistic properties. While linguistic properties are beyond the reach of policy makers, language learning is not. It can be part of school curricula; similarly, employers or the governments of destination countries can encourage or require it.

The aim of this study is to investigate the determinants of language learning of adults in the context of migration, using data for German language learning for the period 1992-2006. Based on earlier research on migration choice and migrants' language skills, we hypothesize that language learning is positively associated with migration flows and migration stocks. Furthermore, we put a focus on the interplay of language learning and the ease of access to the German labor market. We hypothesize that uncertainty about the access affects language learning. If potential migrants are not certain that they will ultimately migrate, they also cannot know for sure that that their language learning investment will pay off. As uncertainty is larger for migrants from countries with restricted access than for migrants from countries with free access, we separately look at language learning for these two groups. We are, in particular, interested in understanding whether more uncertainty leads to more language learning in the home country, that is before migration, or in Germany after arrival. This is a topic of large policy relevance, especially against the background of the new Skilled Immigration Act effective since March 2020, which aims at facilitating migration of skilled workers from third countries to Germany.

To the best of our knowledge, our study is the first to explore the determinants of adult age language learning at the country level in the context of migration.[5] There is, however, a large number of studies in the migration field that explore the determinants

[5]The study by Ginsburgh et al. (2017) considers language learning also at the country level, however, with a focus on the causal impact of bilateral trade.

of individual migrants' language skills. Chiswick and Miller (2015: Section 4) offer an extensive review of the literature and divide the determinants of language skills into three categories, which they dub "the three E's": exposure summarizes the environment in which migrants live and communicate, efficiency captures age at migration, level of education and similar characteristics that enhance individuals' abilities to learn, and, lastly, economic incentives cover a mix of internal and external factors such as planned duration of stay and expected earnings gains. Since most of "the three E's" vary on the level of the individual migrant, studies on the determinants of migrants' language skills often use censuses or surveys to obtain micro-level data.

While these studies provide important insights into the relationship between individual characteristics and language skills, they have one important limitation: typically, they explain self-reported language skills of the respondents at the time of the collection of the data. By doing so, they ignore the timing of language learning: Foreign language acquisition at early ages occurs primarily at school or as a consequence of parents' preferences. For adults on the contrary, the decision to learn a language is more likely made in light of a specific migration decision. This difference between early age and adult age learning has implications for the causal interpretation of the relationship between language skills and migration. Early age learning is unlikely affected by migration decisions that are made later in life; rather it is quite plausible that early age learning builds up language skills. Migrants may very well sort into the destination countries where these skills are most useful. On the contrary, migration decisions very likely affect the incentives for adult age learning. Studies that are based on measures of language skills at some point in time after migration cannot distinguish between early age and adult age learning. As a consequence, they cannot disentangle pre-migration language skills that caused migrants to sort into a particular destination country and language learning that occurred as a consequence of the migration decision. For convenience, we will refer to these channels as the sorting channel and the incentive channel. From the point of view of the policy maker, an understanding of the incentive channel is highly relevant, because it allows the targeting of language learning opportunities at groups of immigrants who are more likely to lack necessary language skills due to a lack of incentives.

Our study wants to isolate the incentive channel by using participation in language courses and exams rather than individual language skills as the dependent variable. The data were collected from the yearbooks of the Goethe Institute, a German association that maintains cultural institutes and provides German language courses in many countries of the world (see Uebelmesser et al. 2018a for details).[6] From this dataset, we use the number of language exam participants in 137 institutes located in 76 countries for the period from 1992 to 2006 and, separately, the number of language course participants of 157 nationalities in Germany for the same period. We argue that especially the exam participation data from the GI are a reasonably good proxy for language learning in the

[6]In this paper, we stick to the following convention: when referring to the association of the Goethe Institute we use the abbreviation "GI". When talking about a specific branch of the GI in Germany or abroad, we refer to it as "institute".

wider populations of the countries where the institutes are located. The exams are widely recognized, do not require participation in a language course at a Goethe institute, and the institutes themselves are accessible to learners of all demographics.

Results from fixed effects panel data estimations indicate that language learning abroad is strongly associated with total immigration and immigration of students from countries whose citizens enjoy free access to Germany. On the contrary, immigration from countries with restricted access is associated with language learning in Germany. The different degrees of uncertainty about access to Germany seem to be of importance. The positive association of migrant stocks with language learning, which we find for countries with restricted access, can also be related to uncertainty considerations as a large stock indicates that in the past, immigrants of a given country were successful in entering Germany. In order to also shed some light on the country heterogeneities, we extend the analysis in two ways: We substitute the location fixed effects with a vector of time-invariant country characteristics, which include linguistic, geographic and cultural distance measures among others. Furthermore, we estimate a random effects model that separates within and between-country effects. While a few between-country differences show up, the main results related to the migration variables are mostly unaffected.

The estimation approach allows identifying correlations between language learning and immigration. Based on these results, a causal interpretation is not possible as immigration can affect language learning, but language learning can also affect immigration. We address this point in more detail and provide some arguments in favor of a causal interpretation, including an instrumental variable estimation. In light of several important caveats, which we discuss, we are careful, however, to abstain from seeing this as anything more than a first exercise.

While the focus of this paper is on immigrants' acquisition of the host country's language, the paper can be more broadly related to the literature on returns to language skills. This includes, first, studies of the returns to foreign language skills (relative to the main language of the country of residence). Whereas there are no or only very small returns to foreign language skills in the US (Fry and Lowel 2003; Saiz and Zoido 2005), high returns to foreign language skills have been found for immigrants in some European countries (Isphording 2013; Toomet 2011) as well as for natives (Ginsburgh and Prieto-Rodriguez 2011).[7]

Second, studies have pointed out differences of the returns to acquired versus native language skills. Melitz and Toubal (2014) analyze the impact of common native language versus common spoken language versus common official language for bilateral trade. The focus lies on identifying the separate roles of language as a means to ease communication and as a proxy for a shared ethnicity and resulting trust. Despite the important cultural component of a common native language (for a discussion of culture and language against

[7]More recently, studies have pointed out the benefits of learning a particular regional language, such as Catalan in Catalonia (Rendon 2007), and the regional and national differences in the benefits of learning a national language in a linguistically diverse country, such as English in India (see Azam et al. 2013; Shastry 2012).

the background of a common ancestry, see Spolaore and Wacziarg 2016), they suggest that common spoken languages and the larger ease of communication play a substantial role for bilateral trade (see also Egger and Toubal 2016). Ginsburgh et al. (2017) consider the role of bilateral trade for language learning. Our paper can be seen as transferring this to a setting with a focus on the role of migration.

The paper is organized as follows. Section 3.2 gives an overview of the datasets used in our analysis, in particular of the language learning data, and outlines our hypotheses. Section 3.3 describes our empirical set-up. In Section 3.4 we discuss our main results. Section 3.5 considers country heterogeneities and identification issues. Section 3.6 concludes with a summary and policy recommendations.

3.2 Data and Hypotheses

Our data on language learning activities on the one hand, and migration on the other hand are drawn from several sources. We begin by introducing our dataset on language learning at the Goethe institutes and continue with other variables. We finally derive hypotheses as the basis for the subsequent analyses.

3.2.1 Dependent Variables: The Goethe Institute Dataset

The Goethe Institute is a German association that promotes the study of the German language and culture abroad. Since its foundation in 1951, it has been maintaining institutes worldwide. At many of these institutes, locals can study the German language and obtain language certificates which are widely recognized. The GI is mainly funded by the German government and through course fees (Goethe Institute 2019b). The GI reports key statistics about language learning in its yearbooks. The datasets from which our dependent variables are taken were created by digitizing these statistics.[8]

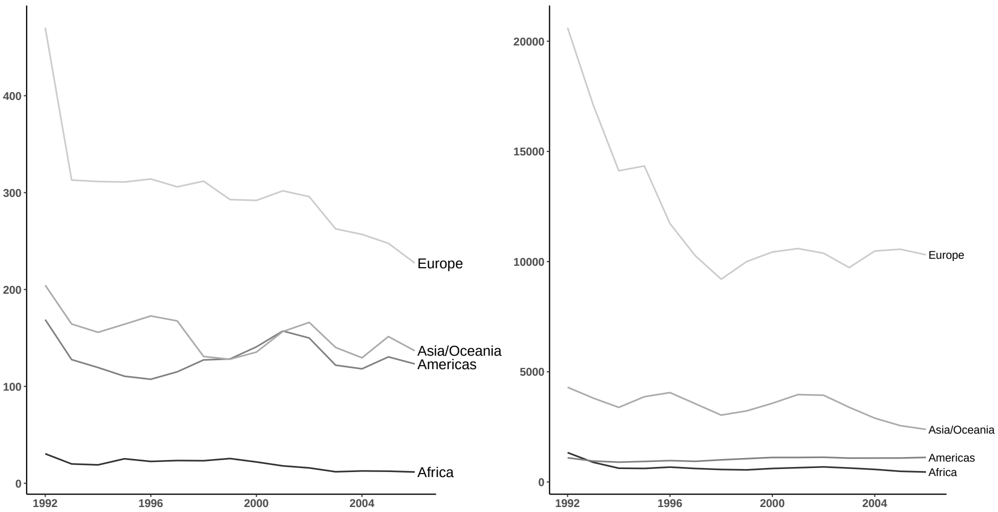
The first dataset covers language learning abroad. It reports yearly observations of course and exam participation at Goethe institutes around the world. We use the exam participation variable from this dataset, because it has two advantages over course participation: First, the exams are standardized, cover fixed amounts of learning content (e.g. the A1/1 level in the Common European Framework of Reference for Languages, CEFR) and are therefore not affected by decisions of individual Goethe institutes. Second, exam participation is open to learners who did not participate in a language course at one of the Goethe institutes. As a consequence, it may act as a better proxy for the extent of language learning activities in the respective country.

The second dataset reports the yearly number of language course participants summed up for all Goethe institutes in Germany and disaggregated by the nationality of the participants. Figure 3.A shows the distribution of the German institutes and informs about their presence during the time period from 1992 to 2006. This is the period of our

[8]The data were manually transferred to spreadsheets and checked for plausibility. Country and city names were harmonized to match them with country level and city level datasets from other sources. See Uebelmesser et al. (2018a) for details and Uebelmesser et al. (2018b,c) for the data.

Figure 3.1: *Germany Specification – Averages per Continent and Year*

(a) *Average Registrations* **(b)** *Average Migration to Germany*



analysis as data from both datasets and migration data for a large number of countries are available. At the same time, this period pre-dates regulation about the “A1 requirement” for family reunification, which was introduced in 2007.

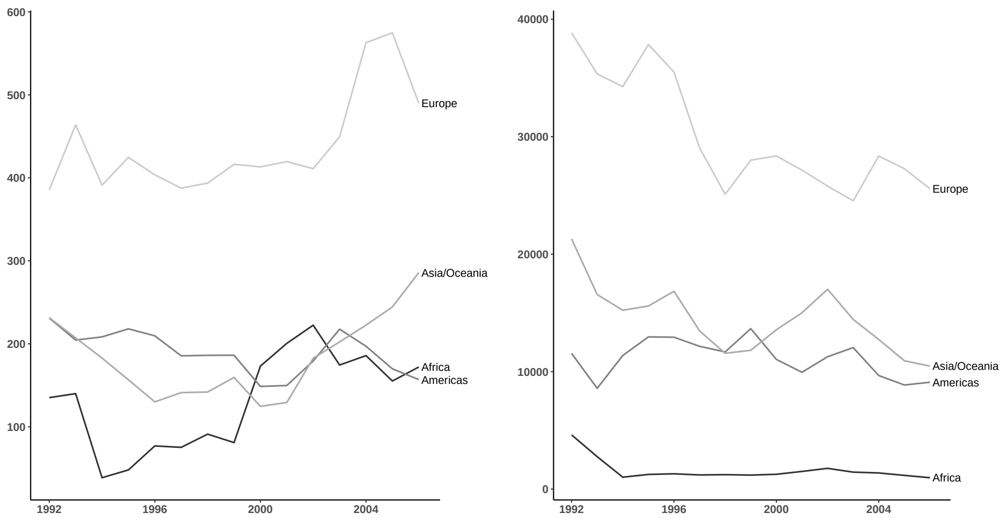
While the aim of the analysis in this paper is to address language learning and its relation with migration in general, our data are limited to learners that come into contact with a Goethe institute. Naturally, there are a number of other language learning opportunities, including universities, private language schools, and internet platforms. We want to address two possible issues. First, there could be concerns about the general importance of the GI as a provider of language courses and exams and about a link with migration to Germany. Figures 3.1 and 3.2 provide an overview of the numbers of exams and course participants in our dataset and relate them to the number of migrants to Germany. As our sample is not balanced, the numbers are averages per countries per continent and year. For illustration, let us look at the numbers for Europe for 2006: With the ‘Germany’ specification, we find that in 2006, there were on average almost 230 course participants in each of the 36 European countries in our sample (amounting to a total of close to 8,200 course registrations). In the same year, about 10,000 individuals migrated on average to Germany from the respective countries. Analogously with the ‘Abroad’ specification, there were on average almost 490 exams taken in each of the 20 European countries with one or more institutes in our sample (amounting to a total of almost 9,800 exams). In the same year, about 25,600 individuals migrated on average to Germany from the respective countries.[9]

Even though the data on exams and courses do not allow any conclusion about the market shares of the GI in the different countries, the data show that the numbers of exams and courses taken is non-negligible. Furthermore, in the context of our study, our

[9]As we use data on course participation at the institute level in our ‘Abroad’ specification, figure 3.B in appendix 3.B.3 is based on averages per institutes per continent and year.

Figure 3.2: *Abroad Specification – Averages per Countries per Continent and Year*

(a) *Average Exams* **(b)** *Average Migration to Germany*



language data, which very likely only cover part of German language learning, make it less likely to find any association. To say it differently, any association which we might find can be seen as an 'a fortiori' association. Also note that for language learning abroad, we partially alleviate this problem by using exam numbers. They cover a larger group of learners who may have learned the language elsewhere but have their skills certified at a Goethe institute (for language learning in Germany, this data is not available).

Second, the multitude of alternative learning opportunities gives rise to further concerns regarding the self-selection of language learners into courses offered by the Goethe institutes. Three characteristics on which self-selection may be based are willingness or ability to pay, location, and age.

Selection on willingness to pay could occur if the prices of courses at the Goethe institutes differed significantly from the costs of other equally suitable learning options. On the one hand, one might suspect the Goethe institutes to be somewhat of a premium provider of language courses, because they are part of a semi-official German organization with a long tradition and a good reputation. Such a status would allow them to charge higher prices. On the other hand, one might suspect courses to be particularly cheap, because the majority of the Goethe institutes' funds comes from the German government. There are several reasons why both concerns might not be warranted. First, in conversations with us, employees of the GI have stated that language courses are priced to be self-financing and that government funding is used for non language course related activities only. Second, prices of language courses do not point towards any systematic deviation relative to competitors' prices. Historical price data on language courses are not available to the best of our knowledge. For a rough idea, Table 3.F in appendix 3.B.1 contains current price data on comparable language courses offered by the Goethe institutes and by other institutes in six cities in different countries. While the data are far from complete or representative, they do not indicate that the Goethe institutes are usually the most expensive provider in the market.[10] Additionally, according to employees of the GI, the price policies of the individual institutes take the prices of local competitors into account and are at least indirectly related to the countries' income levels. It should also be noted that we have 137 institutes in 76 countries in our sample; courses offered by institutes in Germany are attended by 157 nationalities. At the country level, there is no indication that the demand for courses comes from higher-income countries only; this does, however, not rule out that within countries, participants might be self-selected. Still, a priori there is no reasons to expect that results are upward biased.

Goethe institutes are usually located in capitals and other major cities. The lack of institutes in rural areas is likely to lead to an under-representation of language learners from these areas among participants at the Goethe institutes. However, the bias need not be as large as one would initially expect: Goethe institutes offer both extensive and intensive language courses. Extensive courses are based on weekly lessons and last for several months, but intensive courses are taught en bloc. Participants of intensive courses do not necessarily have to live in the vicinity of the respective institute. They may also stay there for the duration of the course only. This holds even more so for exam participation where a day-trip to the institute might not be so uncommon.

The language courses taught by the Goethe institutes are a traditional "offline" form of language learning (at least during the years of our analysis). At the other end of the spectrum, there are pure online courses like those offered by "myngle" or "babble". The latter kind of courses may be more attractive to a younger generation of language students, which is more familiar with using the internet in general. While this difference may lead to an over-representation of older participants in language courses at the GI, the advent of online language learning platforms in the late 2000s falls in the very last years of the period of observation (1992-2006) of our analysis. Consequently, we are confident that the age bias only has a small, if any, effect on our results.

Last, in order to isolate the incentive channel, i.e. how migration intention affects language learning incentives, the possible relation between children age and adult age language learning needs to be discussed. Participating in a course or exam at a Goethe institute does not rule out that some language skills have already been acquired before and possibly for reasons not related to migration. Rather it provides a complementary perspective to the sorting channel with a focus on deliberate decisions by adults who want to continue learning, brush up their skills and possibly work towards a certificate. Furthermore, there is evidence that a substantial part of course and exam participants start learning the German language as an adult. The number of courses offered at the basic A level exceeds by a factor of at least 4 the number of courses offered at the advanced C level in most institutes.[11]

More generally, it should be noted that systematic differences across countries, for

[10] Data comes from the websites of the course providers. The websites of non Goethe institute providers were found by searching Google for "language learning" and the name of the respective city in the native language of the respective country.

[11] See the websites of the institutes for information about the current course programs.

example, related to the organization of courses or the composition of the group of language learners are captured by our fixed effect approach. In addition, using exam numbers in our 'Abroad' specification alleviates some of the concerns mentioned above.

3.2.2 Explanatory Variables

In the 'Abroad' specification, the explanatory variables are characteristics of the countries where the respective institutes are located or characteristics of the relationships of these countries with Germany. In the 'Germany' specification, we use the characteristics of the countries of origin of the language course participants and these countries' relationships with Germany instead.

Migration. Data on migrant flows and stocks come from the German Federal Statistical Office (Destatis). Yearly immigration flows for our period of observation from 1992 to 2006 are available from the German "Wanderungsstatistik". It documents the number of citizens of each country that relocate their primary residence to Germany in a given year. Since residence registration is mandatory in Germany, we expect this measure to appropriately reflect legal immigration. Data on migrant stocks are taken from the Central Register of Foreign Nationals ("Ausländerzentralregister", AZR).

Student migration. Data on student migration come from Destatis as well. A direct measure of the number of immigrants, who migrate for the purpose of studying, is not available for Germany. Instead, we rely on aggregate university statistics ("Hochschulstatistik") and use the number of foreign students who are enrolled in their first semester at a Germany university as a proxy for student immigration flows. Focusing on first year students implies that this number very likely does not include Erasmus or other exchange students as those students normally do not go abroad during their very first year.[12]

Other control variables. We proxy the intensity of trade relationship by total trade revenues, imports plus exports, including tourism, with each country of interest in a given year. Trade revenues are provided by Destatis. Data on country and city populations come from the UN World Population Prospects and World Urbanization Prospects datasets, respectively. As a measure of countries' gross domestic product (GDP) we use the expenditure side real GDP ('rgdpe') from the Penn World Table (v8.0).

In some specifications, we add further country-specific characteristics (instead of country or institute fixed effects). They comprise geographic, linguistic and cultural distance measures as well as information about the educational level or the literacy rate and the sectoral composition. We also use data on the prevalence of the German language in the origin country and of the immigrants' native language in the world. In addition, we control for political rights and civil liberties and the presence of a visa required for entry

[12]There are good reasons for excluding this group of students who plan to stay only for a short term, mostly for one or two semesters, as their motives for language learning might be rather different (depending, among others, on the language of the study program). These short term students are also in most cases not included in the 'Wanderungsstatistik", which is used to capture yearly immigration flows (see above) as registration is only mandatory for stays which are longer than three months. On the contrary, foreign students who stay for a longer period, possibly planning to acquire their degree in Germany, are included.

Table 3.1: *Descriptive Statistics 'Abroad'*

	Free Move		Restr. Move	
Variable	Mean	Std Dev.	Mean	Std Dev.
Exam Participants (abroad)	244.61	505.88	139.34	222.08
Immigration ($\times 10^3$)	16.35	20.06	8.86	15.70
Migrant Stock ($\times 10^3$)	179.49	201.15	111.92	377.87
Student Immig. ($\times 10^3$)	1.32	0.93	0.55	0.85
Trade ($\times 10^6$)	56.90	34.96	11.57	22.17
GDP per capita ($\times 10^3$)	25.74	5.65	11.11	11.43
Population ($\times 10^6$)	37.02	23.78	180.71	323.84
Population City ($\times 10^6$)	2.33	2.37	5.06	5.53
Num. obs.	360		1119	

Note: All values are rounded to two decimal places. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. See Table 3.A in the Appendix for descriptions of the variables.

into Germany. A detailed description of these and all other variables and the respective sources can be found in Table 3.A.

Merging the datasets for our dependent and explanatory variables for the time period 1992-2006 leaves us with 137 institutes in 76 countries for our 'Abroad' specification resulting in 1,479 observations and 157 nationalities for our 'Germany' specification resulting in 2,261 observations.[13] As the ease of access to Germany and the German labor market is an important factor for immigration (see also Section 3.2.3 below), Tables 3.1 and 3.2 report descriptive statistics for the 'Abroad' and the 'Germany' datasets separated for countries with free movement, i.e. countries which belong to the EU or to the Schengen area, and countries with restricted movement vis-à-vis Germany.

3.2.3 Hypotheses

Previous studies that explore the relationship between language and migration fall into one of two groups as outlined in Section 3.1: The first group connects linguistic characteristics or aggregate measures of language skills to migration flows between locations, usually between countries (see, for example, Adserà and Pytliková 2015; Belot and Ederveen 2012). The seconds group connects individual language skills to migration decisions or to a measure of migration success (examples are Dustmann and Van Soest 2001, 2002 for the effect of language skills on wages). None of these studies look at language learning decisions explicitly, but a clear set of predictions regarding the motivation of those decisions emerges from their results. While both groups of studies form the basis for

[13] See Tables 3.G and 3.H for a step-by-step explanation of the sample construction and appendix 3.B.4 for a list of the countries and the time periods for which data are used.

Table 3.2: *Descriptive Statistics 'Germany'*

	Free Move		Restr. Move	
Variable	Mean	Std Dev.	Mean	Std Dev.
Course Participants (in DE)	411.91	501.90	107.41	300.70
Immigration ($\times 10^3$)	10.52	17.34	3.14	9.80
Migrant Stock ($\times 10^3$)	119.46	153.05	31.38	169.76
Student Immig. ($\times 10^3$)	0.84	0.78	0.17	0.46
Trade ($\times 10^6$)	36.78	33.09	2.95	9.88
GDP per capita ($\times 10^3$)	27.37	9.69	7.77	9.82
Population ($\times 10^6$)	19.69	21.32	39.32	141.17
Num. obs.	236		2025	

Note: All values are rounded to two decimal places. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. There are occurrences of zero-values in the 'Free Move' subsample for Language Students (in DE) (0.42 % Zeros), and in the 'Restricted Move' subsample for Language Students (in DE) (14.37 % Zeros), Immigration ($\times 10^3$) (0.20 % Zeros), and First-year Students ($\times 10^3$) (11.80 % Zeros). See Table 3.A in the Appendix for descriptions of the variables.

our theoretical considerations, our empirical approach is closer to that of the first group because we look at aggregate data. As a consequence our hypotheses aim to establish links between aggregate measures on the basis of individual motivations.

Migration Flows and Stocks

Immigration. We begin with our most basic hypothesis: In line with the literature, we regard language learning as an investment (cf. Chiswick and Miller 2007b).[14] Given the large number of potential benefits of language proficiency and the robust results regarding the effect of proficiency on migration success in terms of economic and social integration, migrants should have an incentive to learn the language spoken in their host country.

Previous survey based studies find that a considerable number of migrants speak their host country's language (see Section 3.1). These language skills may have been acquired as a result of the decision to migrate (incentive channel), or they may pre-date the migration decision if they have originally been required for another reason (sorting channel). Typically, survey studies cannot distinguish between the two channels, because the surveys are taken in the host country and do not ask when and how the language was learned.

In our study, we observe language learning directly. A positive association between immigration and course and exam participation would isolate the incentive channel, where individuals at adult age actively acquire skills of the language of their (prospective) host

[14]Language learning can also have a consumption aspect, for example, when it is related to a general interest in languages or in foreign culture (see Huber et al. 2022).

country. While individuals may, of course, learn a language as a child and continue to learn as an adult, we argue in section 3.2.1 that our data are a good proxy for adult age language learning as the number of courses offered at the basic A level exceeds in most institutes the number of courses offered at the advanced C level by a factor of at least 4.

It remains an open question whether individuals decide to acquire host country language skills before or after migration. We will return to this issue in Section 3.2.3.

Student Migration. Students from other countries who immigrate to study at German universities are a special subgroup of immigrants, but a similar incentive rationale applies to them. As most study programs are offered in German, German language skills are a precondition for study success (and a minimum level of skills is normally required). In addition, students may even be more motivated to learn the language than other groups of immigrants, because German law makes it easy for them to stay if they can find employment within 18 months after completing their studies.

Again, a positive association between student immigration and course and exam participation would isolate the incentive to learn German in the context of the decision to study in Germany. As for immigrants, the timing of language learning remains an open question, i.e. whether students decide to acquire host country language skills before or after migration. We will return to this issue in Section 3.2.3.

Migrant Stocks. There are several reasons why not only current migration flows, but also the presence of migrants from a particular sending country in Germany should increase language learning by those who live in that country or who migrated from that country to Germany.

For language learning in Germany, migrants who arrived in previous years may still be taking language courses; either because they did not have the opportunity or motivation to do so earlier or because they are continuing their education.

For language learning abroad, a large migrant community in Germany may lead to increased interest in the German language by those who remained at home, but belong to the social circle of those who migrated to Germany. These individuals may simply be curious about the German culture, but the migration experience of others may also lead them to prepare their own (potential) migration to Germany in the long-term.

We expect the presence of a large number of migrants from a particular sending country in Germany to be positively associated with language learning.

Migration and Minority Language Concentrations. Minority language concentration[15] is often considered to improve the ability of migrants to find work in their host country and build social ties to others who speak their native language. As a consequence, speaking the host country's language may be less important for migrants who live in minority language concentrations. Several studies find that minority language concentrations are associated with lower levels of language proficiency (Chiswick and Miller 2007a; Espenshade and Fu 1997; Isphording and Otten 2013; Lazear 1999). This finding may be the result of a reduced incentive to learn the host country's language or a result of

[15] Other authors use the terms "ethnic enclaves" (e.g. Danzer and Yaman 2013) or "migrant networks" (e.g. Bertoli and Fernández-Huertas Moraga 2015).

the sorting of migrants with lower language skills into minority language concentrations.

We use the migrants stock, i.e. the number of citizens of a sending country who live in Germany, as a proxy for the size of the respective minority language concentration. If we find that the presence of a larger number of migrants from a particular sending country in Germany or, similarly, a larger number of students, weakens the association between migration and language learning, this would isolate the incentive channel and provide evidence of a decreased incentive to learn for migrants from countries with large minority language concentrations in Germany. If, on the contrary, the association is strengthened this might point towards early cohort migrants in Germany informing potential migrants about the benefits of German language skills supporting the incentive channel.

Freedom of Movement

Previous studies have found evidence that host country's language skills yield considerable wage premia and provide other integration related benefits to migrants. Language learning can thus be seen as an investment. However, migrants from different sending countries face different degrees of access to the German labor market. In particular, EU citizens can work freely anywhere in the EU; citizens from non-EU countries, which are part of the Schengen area, also enjoy freedom of movement. Citizens from other countries, on the contrary, face considerable restrictions.[16] Overall, citizens of countries which are members of the EU or the Schengen area benefit from the right to enter Germany visa-free to look for work and easier recognition of their degrees. This goes often hand in hand with higher cultural similarities, and not the least, a smaller travel time to Germany compared to most migrants from other countries.

One can think of two opposing mechanisms when it comes to how this difference in access affects language learning incentives of migrants. On the one hand, citizens who enjoy freedom of movement may find it easier to simply come to Germany and look for employment without having acquired language skills beforehand. As a consequence they may delay their language learning until after their arrival. Instead, citizens from countries where the movement is restricted may try to build up language skills before their arrival to facilitate the process of finding employment from abroad and to bridge the restrictions described above. We will refer to this mechanism as bridge-the-gap.

On the other hand, citizens of countries belonging to the EU or the Schengen area can safely invest into language skills before migrating to Germany, because they can be sure to access Germany once they plan to do so. This allows them to find employment more easily and reap the benefits of their investment.[17] On the contrary, citizens from

[16] While there is a large overlap between countries belonging to the EU and those belonging to the Schengen area, there are some deviations. Looking at 2021, for example, Ireland is an EU country, but not part of the Schengen area; Iceland, Switzerland and Norway, on the contrary, belong to the Schengen area, but are not member countries of the EU. In our analysis, we are careful to distinguish for each year between countries for which the principle of free movement applied and countries for which it did not apply (for example for Bulgaria and Romania, the transitional period is taken into account.)

[17] Countries who join the EU create this perspective for their citizens even if full labor market access is only granted after a transition period.

other countries face the uncertainty of not being able to work in Germany even if they acquire the necessary language skills. They may therefore delay the investment until this uncertainty has resolved and prefer to learn German after their arrival. We will refer to this mechanism as certainty-of-investment.

Since our data covers both language learning abroad and in Germany, it will allow us to identify which mechanism is more important. If the bridge-the-gap mechanism is more important, we should find a stronger association between migration and language learning in the home country, i.e. before migration, for citizens who face restrictions and a stronger association between migration and language learning in Germany, i.e. after migration, for citizens who enjoy free movement. If the certainty-of-investment mechanism is more important, we should find the opposite. This latter mechanism is also relevant against the background of the new Skilled Immigration Act effective since March 2020, which facilitates migration of skilled workers from third countries to Germany and thus reduces uncertainty related to the migration decision.

3.3 Empirical Setup

We are interested in the relationship between language learning on the one hand and several migration related variables on the other hand, where language learning is proxied by two different variables: exam participation at institutes abroad and course participation at institutes in Germany.

In the estimations where the dependent variable is language learning abroad (‘Abroad’ specification), we observe one exam participation number for each institute in each year. This gives rise to a two-level geographical structure, where most explanatory variables are available at the country level, but where each country may contain several institutes. This approach also avoids the problem of abrupt changes in country level aggregates of our dependent variable when institutes open and close. In the estimations where the dependent variable is language learning in Germany (‘Germany’ specification), the geographical structure is more straightforward. We observe one number of course participants in Germany for each nationality in each year and run our estimations at the country level.

We use ordinary least squares (OLS) regressions with institute/country fixed effects and year fixed effects for our main estimations. For exam participation abroad we estimate

$$
\begin{aligned}
P_{ijt}^{j} = {} & \alpha + \beta_1 \text{Immig}_{jt} + \beta_2 \text{StudImmig}_{jt} + \beta_3 \text{MigStock}_{jt} + \beta_4 x_{jt} \\
& + \gamma_1 \, \text{Immig}_{jt} \times \text{MigStock}_{jt} + \gamma_2 \, \text{StudImmig}_{jt} \times \text{MigStock}_{jt} \\
& + \eta_t \, D_t + \eta_i \, D_i + u_{ijt}
\end{aligned}
\tag{3.1}
$$

and for course participation in Germany, the estimation Equation is

$$
\begin{aligned}
P_{jt}^{DE} = {} & \delta + \mu_1 \text{Immig}_{jt} + \mu_2 \text{StudImmig}_{jt} + \mu_3 \text{MigStock}_{jt} + \mu_4 x_{jt} \\
& + \lambda_1 \, \text{Immig}_{jt} \times \text{MigStock}_{jt} + \lambda_2 \, \text{StudImmig}_{jt} \times \text{MigStock}_{jt} \\
& + \nu_t \, D_t + \nu_j \, D_j + u_{jt}
\end{aligned}
\tag{3.2}
$$

where the indices reflect the dimensions across which variables vary. The country where the institute is located is indicated with superscript j in the 'Abroad' specification and DE in the 'Germany' specification. Index i denotes the city of the institute in the 'Abroad' specification, while the city is not known for institutes in the 'Germany' specification. Index j refers to the country of residence or the home country of course and exam participants as well as migrants, respectively, and t is the time index. Whenever variables capture a bilateral relation, like trade or migration, or are related to a distance measure, like geographic distance, this refers to country j and Germany.

With P representing exam or course participation, the 'Abroad' specification (3.1) is about the correlation of exam participation in institute i in country j in year t by individuals from the same country j with the right-hand variables; analogously, the 'Germany' specification (3.2) examines the correlation between course participation in DE by individuals from country j in year t with the right-hand variables.

$Immig_{jt}$ and $StudImmig_{jt}$ denote, respectively, general and student migration flows from country j to Germany in year t and $MigStock_{jt}$ denotes the stock of migrants from country j in Germany in year t. γ_1 and γ_2 capture, respectively, the interaction between general migration flows and stocks and between student migration flows and stocks. This holds analogously for λ_1 and λ_2. Further control variables related to country j in year t (trade revenues with Germany, country GDP and country population size) are captured by the vector x_{jt}. In Equation (3.1), we further control for the population of the city where the institute is located.[18]

D_i, D_j, and D_t are institute, country, and year dummies, respectively. In Equation (3.1) institute level dummies capture both country fixed and institute fixed effects. α is an intercept and u_{ijt} and u_{jt} are error terms. In both estimations, the errors are assumed to be clustered at the country level.[19]

With the chosen fixed effects approach, we want to capture possible country-specific, time-invariant effects. This allows reducing the possible problem of omitted variable bias as we might not be able to directly control for all relevant country characteristics. The drawback is that the coefficients are only identified by variation within institutes/countries and that the role of country characteristics is concealed. To address these issues and to assess the robustness of the results, in Section 3.5.1, we substitute the country and institute dummies with a vector c_j of time-invariant country characteristics, which include linguistic, geographic and cultural distance measures among others. Furthermore, we extend the analysis to a random effects model that separates within and between-country effects in Section 3.5.2.

The estimation approach outlined so far allows identifying correlations between language learning and immigration. A causal interpretation is however not possible as immigration can affect language learning, but language learning can also affect immigra-

[18]We also include an interaction between immigration flows and country GDP. This is left out in Equation (3.1) and Equation (3.2) to make the presentation clearer.

[19]Cluster-robust standard errors are calculated according to Cameron et al. (2011) using the R package multiwayvcov (version 1.2.2).

tion. We have called the former the incentive channel and the latter the sorting channel (see the Introduction where we relate this to adult age language learning and children age language learning, respectively).[20] In Section 3.5.3, we address this point in more detail and provide some arguments in favor of a causal interpretation, including an instrumental variable estimation. In light of several important caveats, which we discuss, we are careful, however, to abstain from seeing this as anything more than a first exercise.

All non-dummy, non-rate and non-scale variables in our regressions are in logs so that coefficients can be interpreted as elasticities (see Table 3.A for a detailed description of the variables). Our main variables related to language learning and migration enter in levels, which are then logged. The main advantage of this log-level approach is that it allows variation from countries of different sizes and with completely different magnitudes of language learning and migration to drive the results of our model. An estimation in non-logged levels would suffer from considerable heteroskedasticity and the results would necessarily be driven by a small number of countries that send a large number of migrants to Germany. While immigration from these countries may be economically more relevant because of its magnitude, our main interest is in identifying the mechanisms that drive language learning more generally and, thus, in using variation from as many countries as possible to identify our coefficients. Additionally, an estimation in non-logged levels would require that we specify to which extent institutes in cities of different sizes are exposed to changes in our country level explanatory variables. For example, an absolute change in immigration from France should, in absolute terms, have a larger effect on exam participation in Paris than on exam participation in Nancy. Paris is a larger city and the institute there has a larger catchment area. A log-log estimation does away with this concern, because it assumes that both institutes experience the same relative rather than absolute change in exam participation as a result of a relative increase in migration.[21]

3.4 Results

In this section, we present the results of our fixed effects estimations. All regressions cover the time period 1992-2006 and are run separately for countries without restrictions of movement vis-à-vis Germany (‘Free Move’), i.e. countries which are members of the European Union or belong to the Schengen area, and countries with restricted movement (‘Restricted Move’).[22]

[20] A similar line of reasoning applies for the relation between language learning and trade (see Ginsburgh et al. 2017).

[21] A small proportion of the values of our dependent and independent variables are zero (see Tables 3.1 and 3.2). For a log-log estimation we had to eliminate these values. To do so, we replaced all zero values with 0.5 before applying the log.

[22] We have also run the regressions separately for EU countries and non-EU countries. The results do hardly change; they are available from the authors upon request.

3.4.1 Main Estimation Results

Table 3.3 shows the results of our 'Abroad' estimations. For institutes in the 'Free Move' countries, we find a positive coefficient for immigration flows. It reduces in size as we add further migration related variables, but remains significant in all specification. We also find a similarly sized significant coefficient for student immigration. In our preferred specification 'Free Move 3', a 1% change in immigration corresponds to a 0.3% change in exam participation, while a 1% change in student immigration corresponds to a 0.4% change. The correlation is stronger for countries with a larger migrant stock in Germany and with a higher GDP per capita. In contrast to the results for the 'Free Move' countries, neither general nor student immigration flows are significantly associated with exam participation in the 'Restricted Move' countries. Instead, we find a large and highly significant coefficient for migrant stocks. After adding controls in 'Restricted Move 3', a 1% change in migrant stocks corresponds to a 0.7% change in exam participation. In summary, our abroad estimations indicate a strong link between language learning and migration flows for 'Free Move' countries, which is replaced by a strong link between migrant stocks and language learning for 'Restricted Move' countries.

Table 3.4 shows the results of our 'Germany' estimations. We find no significant results for 'Free Move' countries. Furthermore the R^2 of even the most basic specification 'Free Move 1' is 97%, indicating that country and year-specific fixed effects already explain most of the variation in language course participation by citizens of 'Free Move' countries in Germany.[23] For citizens from 'Restricted Move' countries, we find a positive coefficient for immigration flows. It remains stable in size and significance as we add our additional migration related variables. In our preferred specification 'Restricted Move 3', a 1% increase in immigration flows corresponds to a 0.3% increase in language course participation. The coefficient for student immigration is also positive, significant and of similar magnitude as the coefficient for general immigration.

When comparing our preferred specifications from Tables 3.3 and 3.4, we observe a clear mirror pattern in the results for migration flows. For 'Free Move' countries, both general and student migration are significantly associated with language learning outside of Germany, while for 'Restricted Move' countries[24] there is a significant correlation between both general and student migration and language learning in Germany. This pattern is in line with our expectation (see Section 3.2.3) that individuals from countries with unrestricted and with restricted access to Germany face different trade-offs when it comes to the decision whether to learn German before or after migrating to Germany. The results suggest that there is a strong certainty-of-investment mechanism at play, where membership in the EU or in the Schengen area encourages pre-migration language learning by guaranteeing access to the German labor market. In contrast, we find no evidence of a bridge-the-gap mechanism that encourages citizens of countries outside

[23] We return to this issue when we discuss the results for an OLS estimation with country characteristics (instead of country dummies) and for a within-between model in Sections 3.5.1 and 3.5.2.

[24] To be precise, this holds for citizens of 'Restricted Move' countries, because the Goethe institutes identify language learners in Germany by their nationality.

Table 3.3: *Stepwise Estimation Results for Exam Participation Abroad*

	Free Move 1	Free Move 2	Free Move 3	Restr. Move 1	Restr. Move 2	Restr. Move 3
Immigration	0.54^{***}	0.37^{***}	0.32^{***}	0.07	-0.31^{*}	-0.27
	(0.19)	(0.12)	(0.12)	(0.17)	(0.16)	(0.17)
Imm. $\times$ Mig. Stock		0.28^{***}	0.18^{*}		-0.14	-0.12
		(0.08)	(0.09)		(0.08)	(0.09)
Imm. $\times$ GDP per capita		0.22	0.25^{**}		-0.03	-0.06
		(0.16)	(0.10)		(0.12)	(0.12)
Student Imm.		0.55^{***}	0.44^{***}		0.19	0.13
		(0.14)	(0.14)		(0.12)	(0.11)
Student Imm. $\times$ Mig. Stock		-0.40^{**}	-0.19		0.05	0.03
		(0.16)	(0.18)		(0.05)	(0.05)
Migrant Stocks		-0.51	-0.10		0.81^{***}	0.66^{***}
		(0.43)	(0.37)		(0.22)	(0.24)
GDP per capita		-1.62^{**}	-1.48^{**}		0.43^{*}	0.29
		(0.66)	(0.74)		(0.24)	(0.26)
Trade			-0.17			0.26
			(0.35)			(0.19)
Population			3.54			-1.96
			(3.13)			(1.78)
Population City			3.47			1.68
			(3.07)			(1.23)
Institute-fixed effects	✓	✓	✓	✓	✓	✓
Year-fixed effects	✓	✓	✓	✓	✓	✓
Adj. R^2	0.92	0.93	0.93	0.69	0.71	0.72
Num. obs.	360	360	360	1119	1119	1119
Num. institutes	38	38	38	111	111	112
Num. countries	17	17	17	67	67	67
Num. years	15	15	15	15	15	15

Note: Standard errors are clustered on the country level. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the ‘Free Move’ and ‘Restricted Move’ samples, respectively, on the basis of the year of accession. Institute-fixed effects capture also country-fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

the EU or the Schengen area to invest into language skills early to facilitate job seeking before their arrival in Germany.

The positive association of migrant stocks with language learning abroad, which we find for ‘Restricted Move’ countries (see Table 3.3), can also be related to uncertainty considerations: A large stock indicates that – at least in the past – many immigrants of a given country were successful in entering Germany. At the same time, these settled immigrants can help potential new immigrants to reduce uncertainty. Both can be expected to be positively related to preparatory language learning.

Table 3.4: *Stepwise Estimation Results for Course Participation in Germany*

	Free Move 1	Free Move 2	Free Move 3	Restr. Move 1	Restr. Move 2	Restr. Move 3
Immigration	0.20	0.14	0.16	0.22^{***}	0.24^{***}	0.26^{***}
	(0.15)	(0.12)	(0.14)	(0.07)	(0.09)	(0.09)
Imm. × Mig. Stock		0.04	0.02		0.03	0.03
		(0.05)	(0.04)		(0.02)	(0.02)
Imm. × GDP per capita		0.37	0.46^{**}		0.04	0.04
		(0.25)	(0.22)		(0.04)	(0.04)
Student Imm.		0.43^{**}	0.32^{*}		0.18^{***}	0.17^{***}
		(0.21)	(0.19)		(0.05)	(0.04)
Student Imm. × Mig. Stock		-0.03	-0.05		0.00	-0.00
		(0.06)	(0.06)		(0.02)	(0.02)
Migrant Stocks		0.04	-0.08		-0.06	-0.07
		(0.54)	(0.45)		(0.12)	(0.11)
GDP per capita		0.37	-0.77		0.44^{**}	0.39^{**}
		(0.47)	(0.79)		(0.18)	(0.17)
Trade			0.61			0.09
			(0.43)			(0.06)
Population			1.01			-0.50
			(3.24)			(0.61)
Country-fixed effects	✓	✓	✓	✓	✓	✓
Year-fixed effects	✓	✓	✓	✓	✓	✓
Adj. R^2	0.97	0.97	0.97	0.89	0.89	0.89
Num. obs.	236	236	236	2025	2025	2025
Num. countries	25	25	25	145	145	145
Num. years	15	15	15	15	15	15

Note: Standard errors are clustered on the country level. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

3.4.2 Capitals and Large Cities

A potential concern with our 'Abroad' specification is that institutes in cities with different characteristics attract language learners from different sub-groups of the population of the respective country and that our estimations throw all of these sub-groups together. In particular, institutes in large cities may attract a large number of language learners that work for the government or for large businesses. While we cannot control for the individual characteristics of different language learners in our macro level study, we can use city level data to shed light on these differences. Table 3.5 presents the results of four estimations for language learning abroad, where we split the sample based on city population. 'Largest' includes the largest city in each country; 'Other' contains the remaining cities.

The sample split does indeed reveal a difference between cities that are and cities

Table 3.5: *Estimation Results: Largest and Other Cities*

	Free Move		Restr. Move	
	Largest	Other	Largest	Other
Immigration	0.22	0.03	-0.03	-0.46
	(0.15)	(0.22)	(0.18)	(0.30)
Imm. × Mig. Stock	0.33^{***}	0.57^{***}	-0.14	0.04
	(0.11)	(0.21)	(0.09)	(0.10)
Imm. × GDP per capita	0.75^{**}	-0.76^{***}	0.02	-0.16
	(0.29)	(0.22)	(0.12)	(0.21)
Student Imm.	0.19	0.75^{***}	0.24^{*}	-0.08
	(0.21)	(0.22)	(0.13)	(0.14)
Student Imm. × Mig. Stock	-0.04	0.17	0.10^{*}	-0.07
	(0.18)	(0.46)	(0.05)	(0.10)
Migrant Stocks	0.26	0.94	0.73^{***}	-0.53
	(1.02)	(0.84)	(0.25)	(0.38)
GDP per capita	-1.40	-4.49^{*}	0.35	0.93
	(1.14)	(2.62)	(0.22)	(0.74)
Trade	-0.33	0.36	0.30	0.36
	(0.54)	(0.79)	(0.22)	(0.34)
Population	1.15	21.39^{***}	-1.27	0.84
	(5.41)	(4.14)	(2.07)	(2.27)
Population City	4.37^{*}	-0.50	1.83	-0.29
	(2.60)	(2.50)	(1.43)	(0.91)
Institute-fixed effects	✓	✓	✓	✓
Year-fixed effects	✓	✓	✓	✓
Adj. R^2	0.95	0.93	0.75	0.68
Num. obs.	169	191	737	382
Num. institutes	17	21	67	44
Num. countries	17	10	67	21
Num. years	15	15	15	15

Note: Standard errors are clustered on the country level. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. Institute-fixed effects capture also country-fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

that are not the largest in their country. For the relationship between migration flows and language learning in 'Free Move' countries, the largest cities seem to be responsible for the results regarding general flows when looking at those countries with a larger migrant stock and those with higher GDP per capita. While for the smaller cities, the positive correlation between migration flows and language learning is also there for countries with a larger migrant stock, we observe a negative correlation for countries with higher GDP per capita. Language learning may be particularly important for migrants connected to large businesses or the government which are more common in the largest cities of each country, especially in high-GDP countries. The smaller cities are, however, responsible for the results regarding student migration. This may be the case because smaller cities have higher relative student populations.

This last point may also apply to language learning outside of the EU or the Schengen area, where the positive association between migrant stocks and language learning is driven entirely by the largest cities. Growing migrant stocks in Germany may create additional demand for language skills among those individuals in big business or the government who establish ties with Germany through the diaspora.

3.5 Country Heterogeneities and Identification Issues

While our fixed effects approach eliminates potential biases from time-invariant institute-specific and country-specific factors that are correlated with language learning and our explanatory variables, it also prevents us from exploring between-country differences. However, these differences may be interesting in their own right. Within-country changes in our variables are often smaller than the between-country differences and a pure within-country analysis could therefore miss important parts of the "bigger picture". Naturally, between-country comparisons have to be interpreted very carefully, because they are not robust to the omission of time-invariant variables that are correlated with both the left-hand side and the right-hand side of the estimation equation. To address these points, we proceed in two ways: First, we run OLS regressions where we substitute the institute and country fixed effects with time-invariant country characteristics. Second, we estimate a within-between model.

3.5.1 OLS Estimations with Country Characteristics

For the OLS specification with time-invariant country characteristics, we consider the 'Free Move' countries and the 'Restricted Move' countries separately as different characteristics might be relevant for the two groups of countries. Also data availability is a larger issue for the countries with restricted access to Germany; this is also the reasons why some of the variables are included as time-invariant variables even though they possibly vary across time (Table 3.A informs about the chosen years).

For the 'Free Move' countries, we include different geographic, linguistic and cultural distance measures as well as information about the educational level and the sectoral composition. Following Ginsburgh et al. (2017), we also add information about the importance of the German language in the origin country and of the immigrants' native language in the world. For the 'Restricted Move' countries, we also include distance measures (due to data availability we cannot control for cultural distance, however). The importance of the German and the native language as well as the sectoral composition are also captured. The educational level is extended by the literacy rate. In addition, we control for a political rights index and a civil liberties index and the presence of a visa requirement for Germany. Tables 3.B and 3.C provide the descriptive statistics.

We would expect that linguistic and cultural distance is negatively correlated with language learning in general, while geographic distance might differently affect where to learn the language and this might also depend on whether access to Germany is free or restricted. More German speakers in the home country, which indicates some closer economic and social links with Germany, should lead to larger incentives to learn German; more speakers of the native language in the world should, on the contrary, decrease the benefits of learning a foreign language. A higher educational level in the origin country might be positively correlated with learning German. If a visa is required, this makes migration and the benefits from language learning more uncertain and should thus be negatively related to language learning, especially abroad.

Tables 3.D and 3.E present the results. As the samples become smaller because of the data issues mentioned above, the fixed effect estimations for the smaller sample are shown next to the respective OLS specifications for easier comparison. We find that the results for the immigration variables do not differ in an important way for the fixed effect and the OLS specifications. It shows that the coefficients of the population and population-city variables are now highly significant in the OLS estimation; this was apparently absorbed by the institute and country fixed effects. Looking at the country characteristics, a few patterns can be observed. The linguistic and cultural distance measures mostly show the expected negative link, while the picture for geographic distance is more mixed. The share of German speakers in the home country is not significantly correlated with learning German, while the result for the world speakers of the immigrants' natives language is mixed. Visa requirements decrease the incentives to learn German abroad in the 'Restricted Move' countries in line with the certainty-of-investment mechanism discussed above. Overall, the OLS specification allows gaining further insights into the role of heterogeneities across countries. At the same time, the main results related to the migration variables are mostly unaffected.

3.5.2 Within-Between Estimations

To further explore between-country differences, we complement our fixed effects models with a random effects (RE) specification. We use Bell and Jones' (2014) re-formulation of an RE estimator proposed by Mundlak (1978). Instead of removing all between country variance from the data, the RE estimator uses it to identify the between coefficients (subscript B). As with the FE estimator, the within coefficients (subscript W) are only identified by within-country variation. For exam participation abroad, we estimate

$$\begin{aligned} P^{j}_{ijt} = {} & \alpha + \beta_W \, (x_{jt} - \bar{x}_j) + \beta_B \, \bar{x}_j \\ & + \gamma_{WW} (\mathrm{Z}_{jt} - \overline{\mathrm{Z}}_j) \times (\mathrm{MigStock}_{jt} - \overline{\mathrm{MigStock}}_j) \\ & + \gamma_{WB} \, \overline{\mathrm{Z}}_j \qquad\quad \times (\mathrm{MigStock}_{jt} - \overline{\mathrm{MigStock}}_j) \\ & + \gamma_{BW} \, (\mathrm{Z}_{jt} - \overline{\mathrm{Z}}_j) \times \overline{\mathrm{MigStock}}_j \\ & + \gamma_{BB} \;\, \overline{\mathrm{Z}}_j \qquad\quad \times \overline{\mathrm{MigStock}}_j \\ & + \delta_W \, (\mathrm{Pop}_{ijt} - \overline{\mathrm{Pop}}_i) + \delta_B \, \overline{\mathrm{Pop}}_{ij} \\ & + u_i + u_j + u_t + u_{ijt} \end{aligned} \tag{3.3}$$

and for course participation in Germany

$$
\begin{aligned}
P_{jt}^{DE} &= \delta + \mu_W\,(x_{jt} - \bar{x}_j) + \mu_B\,\bar{x}_j \\
&+ \lambda_{WW}(\mathrm{Z}_{jt} - \overline{\mathrm{Z}}_j) \times (\mathrm{MigStock}_{jt} - \overline{\mathrm{MigStock}}_j) \\
&+ \lambda_{WB}\,\overline{\mathrm{Z}}_j \qquad\qquad\;\; \times (\mathrm{MigStock}_{jt} - \overline{\mathrm{MigStock}}_j) \\
&+ \lambda_{BW}\,(\mathrm{Z}_{jt} - \overline{\mathrm{Z}}_j) \times \overline{\mathrm{MigStock}}_j \\
&+ \lambda_{BB}\;\overline{\mathrm{Z}}_j \qquad\qquad\;\; \times \overline{\mathrm{MigStock}}_j \\
&+ u_j + u_t + u_{jt}
\end{aligned}
\tag{3.4}
$$

where Z stands for $Immig$ and $StudImmig$ and bars indicate averages.[25] This within-between specification also disentangles the interaction between immigration (and analogously for student immigration) and migrant stocks into three components: One component that is driven by the within variation of both variables (WW) and two additional components, each of which is driven by the within variation of one but by the between variation of the other variable (WB and BW). For example, the within-between interaction "Immigration (B) $\times$ Migrant Stocks (W)" can be interpreted as the change in the effect of the country average of immigration on language learning that is brought about by growing (or shrinking) migrant stocks. A fourth interaction term is identified by the between variation of both variables (BB), but this term has no equivalent in the FE estimation, where it would have been captured by the institute fixed or country fixed effects. u_i, u_j, u_t, u_{jt} and u_{ijt} are error terms (random effects). These specifications are extensions of Bell and Jones' (2014) Equation (12). We add the interaction terms and the additional error term u_i to account for our nested geographical structure of institutes within countries.

We would expect most of the coefficients that are identified by within variation to be the same in our RE and FE estimations, because the regression models are very similar. The two differences are the disentangled interaction terms discussed above and the different error structure. In addition to the FE models' idiosyncratic error term, u_{ijt} or u_{jt}, the RE models allow for random errors on the levels of institutes, countries, and years. When looking at Table 3.6, we see our expectation of identical within coefficients largely confirmed. In particular, we find highly significant within-coefficients for general immigration and student immigration in the 'Abroad, Free Move' specification and in the 'Germany, Restricted Move' specification (cf. also Tables 3.3 and 3.4). This also holds for the within-coefficient of migrant stock in the 'Abroad, Restricted Move' specification.

The within-between model adds several insights. First and in addition to the significant within-coefficient of immigration and student immigration discussed above, the between-coefficients of both, immigration and student immigration, are large and significant in both the 'Abroad' and the 'Germany' specification for 'Restricted Move' countries. This implies that countries that, averaged across the entire period of observation, sent more students to Germany also showed more language learning activity, both abroad and in

[25] The interaction between migration and GDP per capita is, again, left out in Equations (3.3) and (3.4) to make the presentation clearer. Where applicable, the same notation is used as in Equations (3.1) and (3.2).

Table 3.6: *Estimation Results: Within-between Random effects Specification*

	Abroad		Germany	
	Free Move	Restr. Move	Free Move	Restr. Move
Immigration (W)	0.32^{***}	-0.27^{***}	0.21^{*}	0.31^{***}
	(0.10)	(0.09)	(0.11)	(0.05)
Immigration (B)	-0.12	0.47^{*}	-0.07	0.28^{**}
	(0.88)	(0.23)	(0.66)	(0.12)
Imm. (W) × GDP per capita (W)	0.04	1.62^{***}	-0.43	0.08
	(0.60)	(0.44)	(0.66)	(0.14)
Imm. (W) × GDP per capita (B)	0.20	-0.09	0.96^{***}	0.13^{***}
	(0.34)	(0.07)	(0.35)	(0.03)
Imm. (B) × GDP per capita (W)	0.26	-0.01	0.29	-0.01
	(0.25)	(0.11)	(0.24)	(0.03)
Imm. (B) × GDP per capita (B)	0.26	-0.12	-0.73	0.01
	(0.82)	(0.08)	(0.48)	(0.02)
Imm. (W) × Mig. Stock (W)	-1.22	-0.02	-0.15	0.14
	(0.93)	(0.28)	(0.85)	(0.13)
Imm. (W) × Mig. Stock (B)	0.20^{***}	-0.06	0.07	0.07^{***}
	(0.08)	(0.05)	(0.09)	(0.02)
Imm. (B) × Mig. Stock (W)	0.02	-0.61^{***}	0.58	-0.05
	(0.39)	(0.13)	(0.51)	(0.06)
Imm. (B) × Mig. Stock (B)	0.07	0.15	0.03	-0.05^{*}
	(0.42)	(0.14)	(0.16)	(0.02)
Student Imm. (W)	0.49^{***}	0.07	0.23	0.17^{***}
	(0.13)	(0.05)	(0.15)	(0.03)
Student Imm. (B)	0.57	0.54^{***}	-0.41	0.38^{***}
	(0.78)	(0.16)	(0.39)	(0.06)
Student Imm. (W) × Mig. Stock (W)	2.40	-0.89^{***}	1.09	-0.03
	(1.63)	(0.16)	(1.82)	(0.13)
Student Imm. (W) × Mig. Stock (B)	-0.27^{**}	0.01	-0.08	0.02
	(0.13)	(0.03)	(0.08)	(0.01)
Student Imm. (B) × Mig. Stock (W)	0.06	0.45^{***}	-0.69	0.00
	(0.38)	(0.12)	(0.55)	(0.06)
Student Imm. (B) × Mig. Stock (B)	-0.34	-0.16	-0.12	0.04
	(0.55)	(0.15)	(0.16)	(0.03)
Migrant Stocks (W)	-0.24	0.52^{***}	-0.19	-0.05
	(0.42)	(0.12)	(0.49)	(0.08)
Migrant Stocks (B)	0.22	-0.50^{***}	-0.49	-0.20^{*}
	(0.75)	(0.18)	(0.42)	(0.10)
GDP per capita (W)	-1.53^{***}	0.51^{***}	-1.31^{**}	0.31^{***}
	(0.55)	(0.15)	(0.53)	(0.09)
GDP per capita (B)	-2.79	-0.31	1.31	0.39^{***}
	(2.07)	(0.21)	(1.27)	(0.09)
Trade (W)	-0.17	0.13	0.68^{***}	0.08^{**}
	(0.29)	(0.09)	(0.25)	(0.04)
Trade (B)	1.14^{*}	-0.02	-1.21^{***}	0.12^{**}
	(0.67)	(0.16)	(0.41)	(0.06)
Population (W)	2.99	-2.86^{***}	-1.60	-1.36^{***}
	(2.36)	(0.68)	(2.37)	(0.33)
Population (B)	-1.59^{*}	-0.27^{**}	2.78^{***}	0.33^{***}
	(0.93)	(0.13)	(0.44)	(0.06)
Population City (W)	2.58^{**}	1.47^{***}		
	(1.02)	(0.43)		
Population City (B)	0.66^{***}	0.36^{***}		
	(0.18)	(0.10)		
Num. obs.	360	1119	236	2025
Num. institutes	38	111	–	–
Num. countries	17	67	25	145
Num. years	15	15	15	15
Var: institute-level	0.41	0.54	–	–
Var: country-level	0.45	0.20	0.69	0.30
Var: year-level	0.02	0.00	0.01	0.03
Var: Residual	0.08	0.32	0.09	0.49

Note: Standard errors were calculated from a bootstrap ($n = 10,000$); significance levels are based on basic confidence intervals calculated from the same bootstrap sample. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Germany. There are at least three potential explanations for this result. First, in the long term, (student) migration may lead to increased interest in the German language and culture abroad. Second, strong cultural links may not be the consequence but the cause of both language learning activities and migration. While time-invariant cultural links are controlled out of our within coefficients, they do enter into the between-coefficients. Third, (student) migrants may learn German several years before or after staying in Germany. As discussed above, this effect would not be picked up by our unlagged within-coefficient, but may show up in the between-coefficient.

Second, we see a significant within-between-coefficient for the interaction of immigration and student immigration, respectively, with the migrant stock in the 'Abroad, Free Move' specification (and partially in the 'Germany, Restricted Move' one). In countries with, on average, a larger migrant stock, a change in immigration is positively associated with language learning, while the opposite holds for a change in student immigration. The incentive effects thus diverge. Minority language concentration decreases students' incentives to learn the language in the home country before migration. This could indicate that they rely more on help from fellow citizens at the beginning of their stay in Germany. On the contrary, immigrants with a likely larger labor market interest increase their language learning effort in the presence of a larger stock of migrants. One reason could be that the fellow citizens make them aware of the importance of language skills.

Third, the random effects model allows us to explicitly compare the variances of its various error terms. The comparison shows that most of the variance that is not attributed to our explanatory variables can be attributed to time-invariant factors on the city level in the 'Abroad' estimations and on the country level in both estimations. The variances of the respective error terms are larger than the variances of the other entity level error terms and also larger than the residual variance in all but the 'Germany, Restricted Move' sample. Fixed effects on the year level do not seem to play a role as the variance of the respective error term is very small in all estimations.

3.5.3 Causality Issues

Above, we present evidence of a positive association between language learning and several migration related variables. For a causal interpretation, the issue of reverse causality needs to be addressed. In this section, we provide arguments why we think that this association might be driven by a causal effect of those variables on language learning and complement this by an instrumental variable exercise.

With respect to language learning in Germany, a reverse causal effect of course participation on our migration related variables is unlikely, because language learners are already located in Germany by definition. Theoretically, the availability of post-migration language courses might affect the decision whether and where to migrate. However, since courses that teach migrants the language of the destination country are likely to be available in any potential destination, the effect of availability on migration can be neglected in our application.

With respect to language learning outside of Germany, three channels of reverse causality could be particularly relevant: the opening and closing of institutes, the participant recruitment and capacity planning of the institutes, and changes in the individual motivation of participants. First, the opening and closing of institutes may be motivated by the presence of migration flows. This channel of causality would affect the results of our estimations if these estimations "compared" participation in cities in which language courses or exams were offered with zero participation in cities where this was not the case. However, our dataset only includes observations for city-year combinations where language courses were offered and is, thus, not susceptible to the endogenous opening and closing of institutes.[26] Second, the presence of large migration flows from a particular country may motivate institutes in that country to advertise more heavily to "capture" a larger share of the outgoing migrants. While we cannot measure the intensity of advertising of individual institutes, there is no indication that the institutes follow such a strategy. In our conversations with officials at the Goethe institutes, they repeatedly stated that they attempt to adjust to local demand rather than to actively encourage outgoing migrants to participate in courses. Third, the language learning experience at the Goethe institutes may motivate individuals, who initially take the course for non-migration related reasons, to move to Germany. While this may be a good description of the experience of a small number of language learners, the migration choice literature seems to agree that the key determinants of migration decisions are others, income differentials and migration policies in particular (Bertoli and Fernández-Huertas Moraga 2015; Grogger and Hanson 2011; Ortega and Peri 2013).

In the next step, we want to address this in a somewhat more systematic way. We are fully aware that reverse causality issues and endogeneity issues more generally affect several of our main migration variables. Also trade, which we include as a control variable, likely suffers from this problem as Ginsburgh et al. (2017) have demonstrated. Trade can affect language learning, but better language skills can also foster trade. To overcome this problem, they implement an IV strategy based on bilateral trade shares weighted by trading partners' respective ratios of native speakers of the target language. As we are not able to do a full-fledged IV analysis taking all possible endogeneities into account, we only consider the relation of immigration and language learning in more detail.[27] This exercise can therefore only give a first indication; at the same time, it complements the arguments brought forward above.

Our instruments are based on the sectoral composition, geographic distance and tertiary education shares. The first stage of the 2 stages least squares (2SLS) estimation

[26] See Huber and Uebelmesser (2023) for an analysis of the effect of language learning opportunities via the opening of institutes on migration to Germany and an extensive discussion of arguments supporting the exogeneity of the decisions to open or close institutes relative to migration flows to Germany.

[27] Concerning trade, the endogeneity issue might be less relevant in our specific context. Language learning helps to overcome linguistic barriers, but it is unlikely to have a sizeable short run effect on trade volumes. Such an effect would imply that a large proportion of individuals begin language courses for non-trade related reasons, but then decide to use their newly acquired skills to establish trade relationships with Germany. While this may happen in individual cases it is unlikely to happen often, because trade is usually conducted through firms and has considerable set-up costs.

Table 3.7: *IV Estimation Results for Exam Participation Abroad*

	Free Move			Rest. Move		
	1 stage	2 stage	FE	1 stage	2 stage	FE
Immigration		0.42^{*}	0.32^{**}		-0.59	-0.13
		(0.23)	(0.14)		(0.38)	(0.17)
Student Imm.	0.69^{***}	0.29^{*}	0.33^{***}	0.23^{***}	0.16	0.07
	(0.18)	(0.15)	(0.11)	(0.05)	(0.12)	(0.11)
Migrant Stocks	1.35^{*}	0.12	0.25	0.55^{***}	0.83^{**}	0.58^{***}
	(0.81)	(0.49)	(0.54)	(0.15)	(0.37)	(0.22)
Tertiary Enrolment × Pop.	0.86^{**}			0.04		
	(0.34)			(0.03)		
Geographic Dist. × Pop.	-6.63^{***}			0.69		
	(1.90)			(0.57)		
Agriculture Sector × Pop.	3.94			0.23		
	(3.85)			(0.37)		
GDP per capita	0.74	-1.03^{*}	-0.86	0.19	0.17	0.06
	(1.09)	(0.59)	(0.66)	(0.13)	(0.23)	(0.22)
Trade	-0.41	-0.03	-0.14	-0.01	0.30	0.31
	(0.27)	(0.34)	(0.35)	(0.13)	(0.23)	(0.23)
Population	-38.33^{**}	4.12	4.21	2.11	-1.41	-2.74
	(16.41)	(3.70)	(3.78)	(1.34)	(1.99)	(1.90)
Population City	0.24	4.06	4.25^{*}	-0.34	1.99	2.34^{*}
	(0.87)	(2.70)	(2.54)	(0.37)	(1.29)	(1.31)
Institute-fixed effects	✓	✓	✓	✓	✓	✓
Year-fixed effects	✓	✓	✓	✓	✓	✓
Adj. R^2	0.98	0.93	0.93	0.98	0.70	0.71
Num. obs.	360	360	360	1035	1035	1035
Num. institutes	38	38	38	104	104	104
Num. countries	17	17	17	62	62	62
Num. years	15	15	15	15	15	15

Note: Standard errors are clustered on the country level. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. Institute fixed effects capture also country fixed effects. All values, which are not in percentage, are in logs. Wald test comparing the model including instruments and excluding instruments shows that the instruments are jointly significant with an F-value of 58.229 (p < 2.2e-16) for the 'Free Move' countries, and an F-value of 45.866 (p < 2.2e-16) for the 'Restricted Move' countries. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

works well if we take the F-statistics as guide for the joint significance of the instruments (see the notes in Tables 3.7 and 3.8). Results of the second stage confirm our main FE results: In the 'Abroad' specification in Table 3.7, the IV and the FE results are very close. In particular, we find that immigration affects language learning abroad in the 'Free Move' countries and does not do so in the 'Restricted Move' countries. We also find similar IV and FE results in the 'Germany' specification (Table 3.8). For the 'Restricted Move' countries, the results might require some further thoughts given the relatively large second stage immigration coefficient.

As already noted, we see this as an exercise and not a full-fledged analysis. Two caveats are particularly important to keep in mind: First, the exogeneity of the instruments

Table 3.8: *IV Estimation Results for Course Participation in Germany*

	Free Move			Rest. Move		
	1 stage	2 stage	FE	1 stage	2 stage	FE
Immigration		0.12	0.08		1.13^{*}	0.22^{**}
		(0.32)	(0.17)		(0.67)	(0.10)
Student Imm.	0.39^{**}	0.40	0.42^{*}	0.06^{**}	0.10	0.17^{***}
	(0.18)	(0.27)	(0.24)	(0.03)	(0.08)	(0.04)
Migrant Stocks	1.01	0.32	0.39	0.58^{***}	−0.62	−0.06
	(0.61)	(0.57)	(0.37)	(0.13)	(0.46)	(0.11)
Tertiary Enrolment × Pop.	0.12			0.04^{*}		
	(0.13)			(0.02)		
Geographic Dist. × Pop.	−1.94			0.64		
	(1.47)			(0.48)		
Agriculture Sector × Pop.	10.21^{***}			0.42^{*}		
	(2.16)			(0.23)		
GDP per capita	−0.33	−0.33	−0.39	−0.12	0.31^{*}	0.24
	(0.77)	(0.95)	(0.77)	(0.10)	(0.16)	(0.17)
Trade	−0.21	0.34	0.34	-0.09^{**}	0.17	0.08
	(0.30)	(0.42)	(0.42)	(0.05)	(0.11)	(0.08)
Population	-92.76^{***}	0.78	0.96	-8.35^{*}	−0.72	−0.39
	(27.88)	(3.49)	(3.37)	(4.58)	(0.74)	(0.63)
Country-fixed effects	✓	✓	✓	✓	✓	✓
Year-fixed effects	✓	✓	✓	✓	✓	✓
Adj. R^2	0.99	0.97	0.97	0.97	0.85	0.88
Num. obs.	236	236	236	1780	1780	1780
Num. countries	25	25	25	129	129	129
Num. years	15	15	15	15	15	15

Note: Standard errors are clustered on the country level. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. All values, which are not in percentage, are in logs. Wald test comparing the model including instruments and excluding instruments shows that the instruments are jointly significant with an F-value of 47.515 (p < 2.2e-16) for the 'Free Move' countries, and an F-value of 36.239 (p < 2.2e-16) for the 'Restricted Move' countries. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

can be discussed. There are arguments in its favor: A larger share of the agricultural sector makes the economy more dependent on the weather and more generally on climatic conditions. Adverse developments might push individuals out of their home country without directly affecting their language learning incentives. In a similar vain, given a migration intention, geographic distance might affect the choice of the destination country and, only after that choice is made, the language learning incentives. Tertiary enrolment shares could be expected to affect immigration plans in particular for 'Restricted Move' countries as a higher education level might enable migrants to benefit from preferential visa regulations. We acknowledge that the exogeneity argument is weaker here as a higher education level could also directly affect language learning incentives for some individuals. Second, other variables might also be endogenous. We already mentioned trade. Student immigration could also be concerned in a similar way as immigration, even though one could argue that the endogeneity issue might be somewhat less relevant. Learning German in a Goethe institute by young people can of course be triggered by the wish to complement the language education at school and reflect a true interest in the language. Still it is more likely that this is done in preparation of a future stay in Germany possibly for studying purposes.

Overall, we see the different arguments brought forward in this section as providing some indications about a causal relation; in light of the caveats, however, we are careful to abstain from seeing this as anything more than that.

3.6 Conclusions

In this paper, we use a new dataset collected from the yearbooks of the German Goethe Institute on the extent of German language learning around the world. To the best of our knowledge this is the first large-scale dataset on adult age language learning. We use the number of language exam participants in 137 institutes located in 76 countries and the number of language course participants from 157 nationalities in Germany for the period from 1992 to 2006. Both measures vary considerably between and within institutes and countries and this variance is not explained by differences in population alone.

With this data, we investigate the determinants of course and exam participation abroad and in Germany. Our results complement those of studies based on individual level datasets, which investigate the determinants of language skills of migrants, but not their actual learning decisions. Language skills and migration can be linked through a sorting and an incentive channel and skill-based studies cannot disentangle the two. We observe language learning rather than language skills, which may or may not have been acquired in the context of a migration decision. This allows us to focus on the incentive channel.

Using fixed effects regressions, we find that language learning at Goethe institutes abroad is strongly associated with immigration from countries whose citizens have certain access to the German labor market, but that this is not the case for those who do not. Instead, language learning in Germany is positively associated with immigration of countries whose citizens face uncertainty regarding their access. This suggests that the certainty-of-investment mechanism plays an important role which makes preparatory language learning much more likely for immigrants from the EU and Schengen Area, whose citizens are granted certain access to Germany, than from other countries. Language learning at institutes in countries with limited access to Germany is instead strongly associated with migrant stocks in Germany.

The lack of a positive association between language learning in countries with restricted access to Germany and migration from these countries to Germany supports the notion that policy interventions may be necessary. In 2007 Germany introduced the requirement that spouses from non-EU countries must have basic knowledge of German at the A1 level before being granted a visa to live in Germany with their partners. This regulation established a minimum level of language proficiency of migrants. The focus of the Skilled Immigration Act from March 2020 is different: It aims at facilitating migration of skilled workers from third countries to Germany. In the context of our study, this can be seen as removing part of the uncertainty related to access to the German labor market and thus to the returns of investment in language skills. The evidence derived in this study points towards a positive effect on language learning before migrating to Germany and thus better prospects for economic and social integration.

3.A Appendix A

Figure 3.A: *GI in Germany*

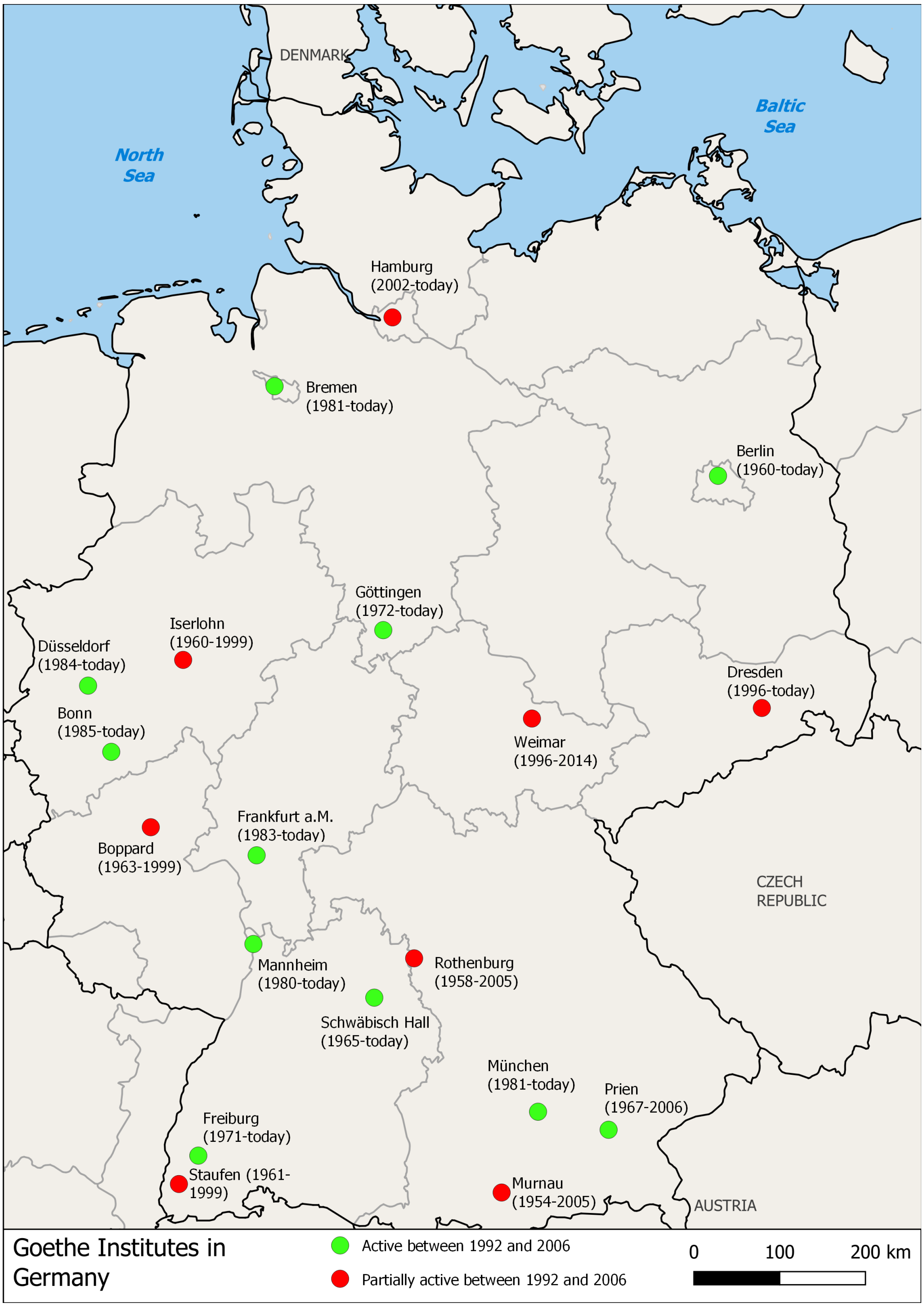


Table 3.A: *Variables Description*

Variable	Type	Description
Exam Participants (abroad)	Numerical (Log)	Yearly number of exam participation at the Goethe institutes by country. *Source: Uebelmesser et al. (2018b).*
Course Participants (in DE)	Numerical (Log)	Yearly number of language course participants at the Goethe institutes in Germany, disaggregated by the nationality of the participants. *Source: Uebelmesser et al. (2018c).*
Immigration	Numerical (Log)	Yearly immigration flows (number of citizens of each country that relocate their primary residence to Germany in a given year). *Source: German "Wanderungsstatistik".*
Student Immigration	Numerical (Log)	Number of foreign students who are enrolled in their first semester at a Germany university. *Source: German "Hochschulstatistik".*
Migration Stocks	Numerical (Log)	Yearly migration stock in Germany. *Source: Central Register of Foreign Nationals ("Ausländerzentralregister", AZR).*
GDP per capita	Numerical (Log)	Yearly expenditure-side real GDP. *Source: Penn World Table (v8.0).*
Trade	Numerical (Log)	Yearly total trade revenues (imports plus exports, including tourism) between each country of interest and Germany. *Source: Destatis.*
Population	Numerical (Log)	Population per country and year. *Source: UN World Population Prospects.*
Population City	Numerical (Log)	Population per (institute) city and year. *Source: World Urbanization Prospects.*
Visa Required for Germany	Binary	Indicates whether a visa is required for entry into Germany, as of 2000. *Source: Immigration Policy Index.*
Linguistic Distance to Germany	Rate (0-1)	Distance of native language to German by country, normalized between 0 (lowest distance) to 1 (largest distance). *Source: Melitz and Toubal (2014).*
Geographic Distance to Germany	Rate (0-1)	Distance to Germany by country, normalized between 0 (lowest distance) to 1 (largest distance). *Source: CEPII's GeoDist database.*
Cultural Distance to Germany: LTO	Rate (0-1)	Distance in long-term orientation index to Germany by country, normalized between 0 (lowest distance) to 1 (largest distance), as of 2013 or latest year available. *Source: Hofstede and Minkov (2013).*
Cultural Distance to Germany: IVR	Rate (0-1)	Distance in indulgence-restraint index to Germany by country, normalized between 0 (lowest distance) to 1 (largest distance), as of 2013 or latest year available. *Source: Hofstede and Minkov (2013).*
Speakers of German (%)	Rate (0-1)	Percentage of population which speaks German by country. *Source: Ginsburgh et al. (2017).*
World Speakers of Native Lang.	Numerical (Log)	Total speakers worldwide of native language by country. In case of multiple native languages maximum is chosen. *Source: Visual Capitalist, Languages.*
Tertiary Enrolment	% Gross	Ratio of total enrollment in tertiary education, regardless of age, to the population of the age group that officially corresponds to the level of education shown, as of 2000. *Source: Worldbank.*
Literacy Rate	Rate (0-1)	Percentage of literate population aged 15 and above by country, as of 2019 or latest year available. *Source: Worldbank, WorldAtlas.*
Political Rights & Civil Liberties Index	Scale (1-7)	Political rights & civil liberties index by country, as of 2000. *Source: Freedom House Index.*
Agriculture Sector (%)	Rate (0-1)	Agriculture, forestry, and fishing, value added per worker (constant 2010 US$) as a share of all sectors by country, as of 2000 (or closest year available). *Source: Worldbank.*
Industry Sector (%)	Rate (0-1)	Industry (including construction), value added per worker (constant 2010 US$) as a share of all sectors by country, as of 2000 (or closest year available). *Source: Worldbank.*
Service Sector (%)	Rate (0-1)	Services, value added per worker (constant 2010 US$) as a share of all sectors by country, as of 2000 (or closest year available). (Residual from Agriculture and Industry). *Source: Worldbank.*

Table 3.B: *Descriptive Statistics 'Abroad' OLS Specification*

	Free Move		Restr. Move	
Variable	Mean	Std Dev.	Mean	Std Dev.
Exam Participants (abroad)	247.35	516.53	134.80	221.66
Immigration ($\times 10^3$)	16.98	20.26	9.02	15.77
Migrant Stock ($\times 10^3$)	186.30	202.76	119.70	394.59
Student Imm. ($\times 10^3$)	1.36	0.93	0.58	0.88
Trade ($\times 10^6$)	56.99	35.64	11.68	22.79
GDP per capita ($\times 10^3$)	25.60	5.70	11.15	11.34
Population ($\times 10^6$)	38.19	23.61	192.78	336.23
Population City ($\times 10^6$)	2.36	2.41	4.76	4.34
Visa Required for Germany	-	-	0.52	0.50
Linguistic Distance	0.79	0.13	0.89	0.10
Geographic Distance ($\times 10^2 km$)	8.87	4.93	66.49	39.42
Cultural Distance: LTO	52.74	13.46	-	-
Cultural Distance: IVR	49.32	16.20	-	-
Speakers of German (%)	0.10	0.13	0.03	0.06
World Speakers of Native Lang. ($\times 10^6$)	394.57	400.25	500.42	472.13
Tertiary Enrolment (% Gross)	53.97	8.57	32.25	24.19
Literacy Rate	-	-	89.92	11.60
Political Rights Index	-	-	2.94	1.88
Civil Liberties Index	-	-	3.28	1.55
Agriculture Sector (%)	0.18	0.04	0.17	0.14
Industry Sector (%)	0.38	0.05	0.45	0.12
Service Sector (%)	0.44	0.06	0.38	0.10
Num. obs.	345		1021	

Note: All values are rounded to two decimal places. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. There are occurrences of zero-values in the 'Free Move' subsample for Speakers of German (%) (9.86 % Zeros); and in the 'Restricted Move' subsample for Speakers of German (%) (62.98 % Zeros), and Visa Required for Germany (47.89 % of Zeros). See Table 3.A in the Appendix for descriptions of the variables.

Table 3.C: *Descriptive Statistics ‘Germany’ OLS Specification*

	Free Move		Restr. Move	
Variable	Mean	Std Dev.	Mean	Std Dev.
Course Participants (in DE)	452.84	517.20	109.47	258.87
Immigration ($\times 10^3$)	11.64	18.07	3.48	10.20
Migrant Stock ($\times 10^3$)	132.35	157.53	37.17	189.03
Student Imm. ($\times 10^3$)	0.90	0.81	0.20	0.51
Trade ($\times 10^6$)	37.19	33.80	3.32	10.59
GDP per capita ($\times 10^3$)	25.87	6.88	7.95	9.42
Population ($\times 10^6$)	21.37	21.96	45.66	156.57
Visa Required for Germany	-	-	0.69	0.46
Linguistic Distance	0.79	0.14	0.90	0.09
Geographic Distance ($\times 10^2 km$)	10.75	5.64	62.18	35.14
Cultural Distance: LTO	49.02	15.25	-	-
Cultural Distance: IVR	52.96	16.94	-	-
Speakers of German (%)	0.19	0.27	0.03	0.10
World Speakers of Native Lang. ($\times 10^6$)	284.59	396.02	380.59	411.70
Tertiary Enrolment (% Gross)	54.10	10.68	20.54	19.85
Literacy Rate	-	-	83.38	19.27
Political Rights Index	-	-	3.68	2.12
Civil Liberties Index	-	-	3.78	1.68
Agriculture Sector (%)	0.17	0.06	0.14	0.11
Industry Sector (%)	0.40	0.06	0.49	0.17
Service Sector (%)	0.43	0.08	0.37	0.13
Num. obs.	210		1621	

Note: All values are rounded to two decimal places. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the ‘Free Move’ and ‘Restricted Move’ samples, respectively, on the basis of the year of accession. There are occurrences of zero-values in the ‘Free Move’ subsample for Language Students (in DE) (0.48 % Zeros), and Speakers of German (%) (12.86 % Zeros); and in the ‘Restricted Move’ subsample for Language Students (in DE) (9.93 % Zeros), Immigration ($\times 10^3$) (0.06 % Zeros), Student Imm. ($\times 10^3$) (8.08 % Zeros), Visa Required for Germany (30.72 % of Zeros), and Speakers of German (%) (84.82 % Zeros). See Table 3.A in the Appendix for descriptions of the variables.

Table 3.D: *Estimation Results for Exam and Course Participation Abroad and in Germany, Free Move Countries*

	Abroad		In Germany	
	FE	OLS	FE	OLS
Immigration	0.33^{***}	0.66^{***}	0.19	0.22
	(0.12)	(0.17)	(0.15)	(0.23)
Imm. × Mig. Stock	0.19^{**}	0.17^{**}	-0.01	0.13^{*}
	(0.09)	(0.07)	(0.05)	(0.08)
Imm. × GDP per capita	0.27^{**}	0.07	0.50^{**}	0.74^{***}
	(0.12)	(0.24)	(0.23)	(0.21)
Student Imm.	0.43^{***}	0.19	0.26	1.33^{***}
	(0.14)	(0.15)	(0.19)	(0.22)
Student Imm. × Mig. Stock	-0.22	-0.08	-0.07	-0.10
	(0.20)	(0.15)	(0.08)	(0.12)
Migrant Stocks	-0.10	0.29	-0.14	-0.97^{***}
	(0.36)	(0.21)	(0.40)	(0.13)
GDP per capita	-1.73^{**}	-1.62^{***}	-1.11	1.25^{***}
	(0.73)	(0.36)	(0.87)	(0.42)
Trade	-0.05	0.63^{**}	0.91^{*}	-0.69^{***}
	(0.36)	(0.28)	(0.47)	(0.21)
Population	3.26	-2.20^{***}	1.48	1.78^{***}
	(3.02)	(0.44)	(3.45)	(0.20)
Population City	3.36	0.74^{***}		
	(3.15)	(0.09)		
Linguistic Distance to Germany		-1.68		-8.57^{***}
		(1.71)		(1.65)
Geographic Distance to Germany		5.45		23.71^{***}
		(3.45)		(5.33)
Cultural Distance to Germany: LTO		-2.54		1.46^{*}
		(1.60)		(0.75)
Cultural Distance to Germany: IVR		-5.70^{***}		-1.81^{*}
		(0.73)		(0.97)
Speakers of German (%)		1.25		-0.27
		(1.19)		(0.50)
World Speakers of Native Lang.		0.39^{***}		-0.27^{***}
		(0.14)		(0.04)
Tertiary Enrollment		-0.75		-1.05
		(0.83)		(1.00)
Agriculture Sector (%)		17.65^{***}		-3.62
		(2.09)		(4.31)
Industry Sector (%)		8.02^{*}		-5.95^{*}
		(4.14)		(3.07)
Institute-fixed effects	✓	–	–	–
Country-fixed effects	–	–	✓	–
Year-fixed effects	✓	✓	✓	✓
Adj. R^2	0.93	0.79	0.97	0.93
Num. obs.	345	345	210	210
Num. institutes	37	37	–	–
Num. countries	16	16	22	22
Num. years	15	15	15	15

Note: Standard errors are clustered on the country level. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. Institute fixed effects capture also country fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 3.E: *Estimation Results for Exam and Course Participation Abroad and in Germany, Restricted Move Countries*

	Abroad		In Germany	
	FE	OLS	FE	OLS
Immigration	-0.20	0.22	0.27^{**}	0.18^{**}
	(0.19)	(0.20)	(0.12)	(0.09)
Imm. × Mig. Stock	-0.10	-0.03	0.03	-0.04^{***}
	(0.09)	(0.06)	(0.03)	(0.01)
Imm. × GDP per capita	-0.03	-0.04	0.01	0.02
	(0.14)	(0.07)	(0.06)	(0.02)
Student Imm.	0.09	0.41^{***}	0.15^{***}	0.27^{***}
	(0.11)	(0.13)	(0.05)	(0.05)
Student Imm. × Mig. Stock	0.01	0.01	0.01	0.03^{**}
	(0.05)	(0.06)	(0.02)	(0.02)
Migrant Stocks	0.63^{***}	-0.04	-0.13	-0.08
	(0.24)	(0.13)	(0.13)	(0.08)
GDP per capita	0.03	-0.33^{*}	0.40^{*}	0.32^{***}
	(0.24)	(0.20)	(0.22)	(0.10)
Trade	0.33	-0.00	0.09	0.02
	(0.23)	(0.14)	(0.07)	(0.06)
Population	-2.79	-0.34^{**}	-0.39	0.48^{***}
	(2.09)	(0.16)	(0.63)	(0.09)
Population City	2.30^{*}	0.29^{***}		
	(1.30)	(0.11)		
Visa Required for Germany		-0.52^{*}		-0.30^{*}
		(0.30)		(0.18)
Linguistic Distance to Germany		1.64		0.20
		(2.09)		(0.96)
Geographic Distance to Germany		2.19^{**}		-0.87
		(1.05)		(0.78)
Speakers of German (%)		0.04		-0.19
		(2.22)		(0.94)
World Speakers of Native Lang.		0.04		0.06
		(0.07)		(0.04)
Tertiary Enrollment		-0.16		0.15
		(0.22)		(0.10)
Literacy Rate		-0.00		0.01
		(0.01)		(0.00)
Political Rights Index		0.04		0.03
		(0.10)		(0.09)
Civil Liberties Index		-0.06		-0.08
		(0.14)		(0.11)
Agriculture Sector (%)		1.10		-0.57
		(1.06)		(0.53)
Industry Sector (%)		1.01		-0.41
		(1.34)		(0.52)
Institute-fixed effects	✓	–	–	–
Country-fixed effects	–	–	✓	–
Year-fixed effects	✓	✓	✓	✓
Continent-fixed effects	✓	✓	✓	✓
Adj. R^2	0.70	0.29	0.88	0.82
Num. obs.	1021	1021	1621	1621
Num. institutes	103	103	–	–
Num. countries	61	61	118	118
Num. years	15	15	15	15

Note: Standard errors are clustered on the country level. Observations from countries that joined the EU or the Schengen area during the period of observation are assigned to the 'Free Move' and 'Restricted Move' samples, respectively, on the basis of the year of accession. Institute fixed effects capture also country fixed effects. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

3.B Appendix B

3.B.1 Language Course Pricing

Table 3.F: *Prices of Language Courses at Goethe Institutes and Other Providers in 2015*

City	Provider	Course Type	Price / Hour	Currency
Mexico City	GI	Extensive	146.67	MXN
Mexico City	Tecnologico de Monterrey	Extensive	90.91	MXN
Buenos Aires	GI	Extensive A1.1	80	ARS
Buenos Aires	Sprachzentrum Buenos Aires	Extensive A1.1	65	ARS
Rio de Janeiro	GI	Extensive A1	56.21	BRL
Rio de Janeiro	Baukurs	Extensive A1	49.17	BRL
Lisbon	GI	Extensive	5.67	EUR
Lisbon	ilnova	Extensive	6.17	EUR
Ankara	GI	Extensive A1	10.16	TRY
Ankara	Hitit Education Institutions	Extensive A1	11.88	TRY
Tokyo	GI	Intensive	1541.67	JPY
Tokyo	German Office	Intensive	1971.43	JPY

3.B.2 Sample Shrinkage due to Missings

'Abroad' Specification

Table 3.G: *Sample Shrinkage due to Missings: 'Abroad' Specification*

Step	Action	Obs.	Cntrs.	Inst.	Years
1	Complete Registration info 1992-2006	1689	87	155	15
2	Remove joint reporting cases	1594	86	152	15
3	Remove Greece	1553	85	148	15
4	Remove obs. with < 5 registrations	1527	85	147	15
5	Remove Missings: Migration Flows	1507	79	142	15
6	Remove Missings: Migration Stocks	1507	79	142	15
7	Remove Missings: Foreign First-year Students	1507	79	142	15
8	Remove Missings: Trade	1505	78	141	15
9	Remove Missings: Country Population	1505	78	141	15
10	Remove Missings: City Population	1479	76	137	15
11	Remove Missings: GDP	1479	76	137	15

Note: Joint reporting refers to cases where information about language services in the annual reports is reported jointly for two or more institutes without clarifying which of these institutes actually offered language services (see Uebelmesser et al. 2018a for details). For historical reasons, Greece is an outlier with respect to course and exam participation.

'Germany' Specification

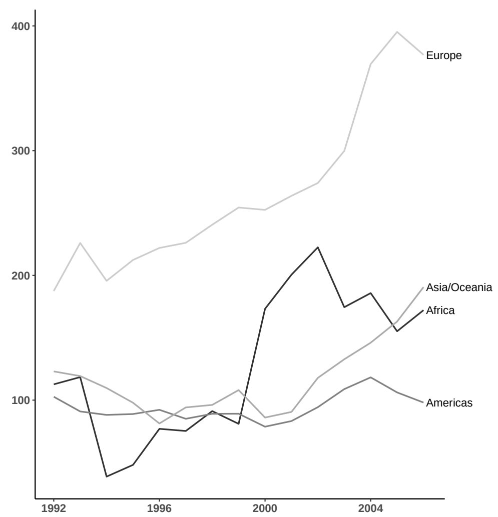
Table 3.H: *Sample Shrinkage due to Missings: 'Germany' Specification*

Step	Action	Obs.	Cntrs.	Years
1	Complete participants info 1992-2006	2864	201	15
2	Remove Missings: Migration Flows	2275	159	15
3	Remove Missings: Migration Stocks	2274	159	15
4	Remove Missings: Foreign First-year Students	2268	157	15
5	Remove Missings: Trade	2261	157	15
6	Remove Missings: Country Population	2261	157	15
7	Remove Missings: GDP	2261	157	15

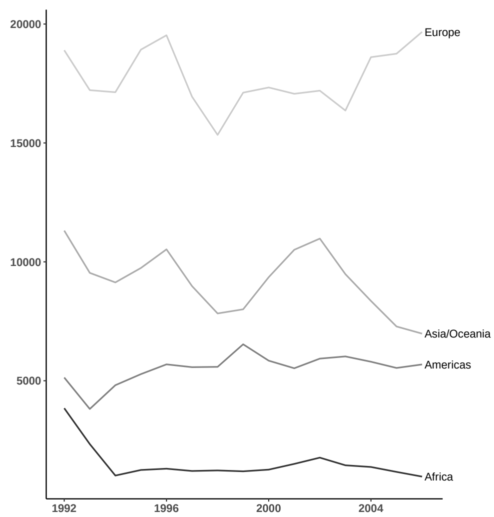
3.B.3 Graphical Illustration ('Abroad' Specification): Averages per Institutes

Figure 3.B: *Abroad Specification – Averages per Institutes per Continent and Year*

(a) *Average Exams*



(b) *Average Migration to Germany*



3.B.4 List of Countries and Institutes

'Abroad' Specification

Africa

Côte d'Ivoire: Abidjan (1995-2000, 2003-2006); Cameroon: Yaoundé (1992-2006); Egypt: Alexandria (1992-1993), Cairo (1992-1993); Ethiopia: Addis Ababa (1992-2001, 2003-2006); Ghana: Accra (1992-2006); Kenya: Nairobi (1992-2006); Morocco: Casablanca (1992-1993), Rabat (1992-1993); Nigeria: Lagos (1992-2006); Senegal: Dakar (1992-2006); South Africa: Johannesburg (1996-2006); Sudan: Khartoum (1992-1996); Tanzania, United Republic of: Dar es Salaam (1993-1997); Togo: Lomé (1995-2006); Tunisia: Tunis (1992-2005); Zimbabwe: Harare (1994)

Americas

Argentina: Buenos Aires (1992-1999, 2001-2006), Córdoba (1992-1998), Mendoza (1992-1995), San Juan (1992-1993); Bolivia (Plurinational State of): La Paz (1992-2006); Brazil: Belo Horizonte (1992-1995), Brasília (1992-1996), Curitiba (1992-2003), Porto Alegre (1992-2006), Rio de Janeiro (1992-2006), Salvador (1992-2006), Sao Paulo (1992-2003); Canada: Toronto (1992-2006), Vancouver (1993-1999); Chile: Valparaiso (1992-1993); Colombia: Bogotá (1992-2006), Medellín (1992-1993); Costa Rica: San José (1992-1999); Mexico: Guadalajara (1992-2006), Mexico City (1992-2006); Peru: Lima (1992-2006); United States of America: Ann Arbor (1994-1999), Atlanta (1992-2006), Boston (1992-2006), Chicago (1992-2006), Cincinnati (1994, 1996-1997), Houston (1992-1999), Los Angeles (1995), San Francisco (1992, 1995-2006), Washington (1999, 2002-2003); Uruguay: Montevideo (1992-2002, 2004-2006); Venezuela (Bolivarian Republic of): Caracas (1992-2003, 2005-2006)

Asia

Bangladesh: Dhaka (1993-2006); China: Beijing (1992-2006), Hong Kong (1993-2006); Georgia: Tiflis (1995-2006); India: Bangalore (1992-1994, 1996-2006), Chennai (1992-2006), Hyderabad (1992-1993), Kolkata (1992-2006), Mumbai (1992-2006), New Delhi (1992-2006), Pune (1992-2006); Indonesia: Bandung (1992-1999), Jakarta (1992-1996, 1998-1999), Surabaya (1992-1995); Iran (Islamic Republic of): Tehran (2006); Israel: Jerusalem (1992, 1995-1997), Tel Aviv (1992-2001, 2003, 2005-2006); Japan: Kyoto (1992-1993, 2002-2006), Osaka (1992-1993, 2002-2006), Tokyo (1992-2006); Jordan: Amman (1993-2001, 2003-2006); Kazakhstan: Almaty (1996-2006); Korea (Republic of): Seoul (1992-2006); Lebanon: Beirut (2002-2006); Malaysia: Kuala Lumpur (1992-1993, 1995-2006); Nepal: Kathmandu (1992-1996); Pakistan: Karachi (1992-1998, 2001-2002, 2005), Lahore (1992, 1994-1996); Philippines: Manila (1992-2006); Singapore: Singapore (1992-1995, 1997-2006); Sri Lanka: Colombo (1992-2006); Syrian Arab Republic: Damascus (1992-2006); Thailand: Bangkok (1992-2006); Turkey: Ankara (1992-2006), Istanbul (1992-2006), Izmir (1992-2006); Uzbekistan: Tashkent (1999-2006); Viet Nam: Hanoi (1998-2006)

Europe

Belarus: Minsk (1996-2006); Belgium: Brussels (1992-2006); Bosnia and Herzegovina:

Sarajevo (2002-2006); Bulgaria: Sofia (1992-2006); Croatia: Zagreb (2006); Czechia: Prague (1993-2006); Denmark: Aarhus (1992, 1994-1995), Copenhagen (1992-1996, 1998-2005); Finland: Helsinki (1992-2006), Tampere (1992-1996), Turku (1992-1995); France: Bordeaux (1992-2005), Lille (1992-2000), Lyon (1992-2006), Marseille (1992-1997), Paris (1992-2006), Toulouse (1992-2006); Hungary: Budapest (1992-2006); Ireland: Dublin (1992-2006); Italy: Genoa (1992-1998), Milan (1992-2004), Naples (1992-2005), Palermo (1992-1996), Rome (1993-2005), Turin (1992-2005); Latvia: Riga (1994-2006); Netherlands: Amsterdam (1992-2005), Rotterdam (1992-2005); Norway: Bergen (1992-1995), Oslo (1992-2006); Poland: Krakow (1999-2006), Warsaw (1992-2006); Portugal: Coimbra (1992-1996), Lisbon (1992-2001), Porto (1992-2001); Romania: Bucharest (1992-2006); Russian Federation: Moscow (1992-2006), St. Petersburg (1996-2006); Slovakia: Bratislava (1993-2006); Spain: Barcelona (1992-2006), Madrid (1992-2006); Sweden: Gothenburg (1993-1995), Stockholm (1993-2003, 2005-2006); Ukraine: Kiev (1994-2006); United Kingdom of Great Britain and Northern Ireland: Glasgow (1992-2006), London (1992-2006), Manchester (1992-2001)

Oceania

Australia: Melbourne (1992-2006), Sydney (1992-2006); New Zealand: Wellington (1992-2006)

'Germany' Specification

Africa

Angola (1992-2006); Burundi (1992-2006); Benin (1992-2006); Burkina Faso (1992-2006); Botswana (1992-2006); Central African Republic (1992-2006); Côte d'Ivoire (1992-2006); Cameroon (1992-2006); Congo (Democratic Republic of the) (1992-2004); Comoros (1992-2006); Cabo Verde (1992-2006); Djibouti (1992-2006); Egypt (1992-2006); Ethiopia (1992-2006); Gabon (1992-2006); Ghana (1992-2006); Guinea (1992-2006); Gambia (1992-2006); Guinea-Bissau (1992-2006); Equatorial Guinea (1992-2006); Kenya (1992-2006); Liberia (1992-2006); Lesotho (1992-2006); Morocco (1992-2006); Madagascar (1992-2006); Mali (1992-2006); Mozambique (1992-2006); Mauritania (1992-2006); Mauritius (1992-2006); Malawi (1992-2006); Namibia (1992-2006); Niger (1992-2006); Nigeria (1992-2006); Rwanda (1992-2006); Sudan (1992-2006); Senegal (1992-2006); Sierra Leone (1992-2006); Eswatini (1992-2006); Chad (1992-2006); Togo (1992-2006); Tunisia (1992-2006); Tanzania, United Republic of (1992-2006); Uganda (1992-2006); South Africa (1992-2006); Zambia (1992-1997, 1999-2006); Zimbabwe (1992-1994, 1996-2006)

Americas

Argentina (1992-2006); Antigua and Barbuda (1992-2006); Bahamas (1992-2006); Belize (1992-2006); Bolivia (Plurinational State of) (1992-2006); Brazil (1992-2006); Barbados (1992-2006); Canada (1992-2006); Chile (1992-2006); Colombia (1992-2006); Costa Rica (1992-2006); Dominica (1992-2006); Dominican Republic (1992-2006); Ecuador (1992-2006); Grenada (1992-2006); Guatemala (1992-2006); Honduras (1992-2006); Jamaica (1992-2006); Saint Kitts and Nevis (1992-2006); Saint Lucia (1992-2006);

Mexico (1992-2006); Panama (1992-2006); Peru (1992-2006); Paraguay (1992-2006); El Salvador (1992-2006); Trinidad and Tobago (1992-2006); Uruguay (1992-2006); United States of America (1992-2006); Saint Vincent and the Grenadines (1992-2006); Venezuela (Bolivarian Republic of) (1992-2006)

Asia

Armenia (1998-2006); Azerbaijan (1998-2006); Bangladesh (1992-2006); Bahrain (1992-2006); Brunei Darussalam (1992-2006); Bhutan (1992-2006); China (1992-2006); Cyprus (1992-2006); Georgia (1998-2006); Indonesia (1992-2006); India (1992-2006); Iran (Islamic Republic of) (1992-2006); Iraq (1992-2006); Israel (1992-2006); Jordan (1992-2006); Japan (1992-2006); Kazakhstan (1998-2006); Kyrgyzstan (1998-2006); Cambodia (1992-2006); Korea (Republic of) (1992-2006); Kuwait (1992-2006); Lao People's Democratic Republic (1992-2006); Lebanon (1992-2006); Sri Lanka (1992-2006); Maldives (1992-2006); Mongolia (1992-2006); Malaysia (1992-2006); Nepal (1992-2006); Oman (1992-2006); Pakistan (1992-2006); Philippines (1992-2006); Qatar (1992-2006); Saudi Arabia (1992-2006); Singapore (1992-2006); Syrian Arab Republic (1992-2006); Thailand (1992-2006); Tajikistan (1998-2006); Turkmenistan (1998-2006); Turkey (1992-2006); Uzbekistan (1998-2006); Viet Nam (1992-2006); Yemen (1992-2006)

Europe

Albania (1992-2006); Austria (1992-2006); Belgium (1992-2006); Bulgaria (1992-2006); Bosnia and Herzegovina (1993-2006); Belarus (1998-2006); Switzerland (1992-2006); Czechia (1993-2006); Denmark (1992-2006); Spain (1992-2006); Estonia (1998-2006); Finland (1992-2006); France (1992-2006); United Kingdom of Great Britain and Northern Ireland (1992-2006); Greece (1992-2006); Croatia (1992-2006); Hungary (1992-2006); Ireland (1992-2006); Iceland (1992-2006); Italy (1992-2006); Lithuania (1998-2006); Luxembourg (1999-2006); Latvia (1998-2006); Moldova (Republic of) (1998-2006); Macedonia (the former Yugoslav Republic of) (1994-2006); Malta (1992-2006); Netherlands (1992-2006); Norway (1992-2006); Poland (1992-2006); Portugal (1992-2006); Romania (1992-2006); Russian Federation (1998-2006); Slovakia (1993-2006); Slovenia (1993-2006); Sweden (1992-2006); Ukraine (1998-2006)

Oceania

Australia (1992-2006); Fiji (1992-2006); New Zealand (1992-2006)

4 | Migration Aspirations and Intentions

with MATTHIAS HUBER[1]
and TILL NIKOLKA[2]
and PANU POUTVAARA[3]
and SILKE UEBELMESSER[4]

Abstract. *We carried out multinational surveys among language course participants and students to analyze aspirations and intentions to emigrate. We identify two groups that have been largely neglected in previous research on migration aspirations and intentions: those who intend to migrate temporarily and those who intend to migrate permanently without aspiring to do so. Analyzing main motivations to emigrate shows that family considerations motivate women who intend to migrate without aspirations to do so, while men often intend to migrate only temporarily for work or studies.*

Keywords: *international migration, temporary migration, permanent migration, migration aspirations, migration intentions*

4.1 Introduction

The reasons why people leave their country of origin have been studied for many decades. A large body of literature evaluates immigrants in destination countries (e.g. Abramitzky et al. 2012; Adserà and Pytliková 2015; Beine et al. 2011; Borjas 1987; McKenzie and Rapoport 2010). Other studies survey individuals at their origin before actual migration takes place, which allows identifying potential future migrants (e.g. Bertoli et al. 2022; Bertoli and Ruyssen 2018; Manchin and Orazbayev 2018; Papapanagos and Sanfey 2001; Ruyssen and Salomone 2018; Uebelmesser 2006).

We sample individuals prior to potential migration in two multinational surveys: one among language course participants in 14 countries and one among university students in six countries. Unlike other studies, we exploit within-individual differences between the desire to migrate permanently (*aspirations*) and actual considerations and plans to migrate either permanently or temporarily (*intentions*) to reveal barriers and drivers

[1]Federal Office for Migration and Refugees and Friedrich Schiller University Jena
[2]German Youth Institute and CESifo
[3]University of Munich, ifo Institute, CESifo, CReAM and IZA
[4]Friedrich Schiller University Jena and CESifo

towards migration.[5] While aspirations, like intentions, emerge from a cost-benefit analysis of potential migration, aspirations can differ from intentions given various constraints which the individuals face. They are not necessarily formed in a consecutive order, i.e. intentions to migrate do not necessarily imply desire to do so and vice versa. In our data there is a considerable share of respondents whose current migration intentions differ from their aspirations (around 40 % of the language course sample, and 50 % of the university student sample).

Our analysis using multinomial probit estimations suggests that individual level determinants for aspirations and intentions vary between genders. For women in our data the main determinant of differences between aspirations and intentions are family ties. Compared to single women or women in a relationship with a native partner, women with a non-native partner are more likely to intend to migrate permanently despite their desire to stay. This is in line with former literature stating that women are predominantly influenced in their migration decisions by family ties (Mincer 1978; Munk et al. 2022). For men, we find education and career prospects to be of larger importance for migration intentions (cf. Geist and McManus 2011; McKinnish 2008). A closer look at self-reported motivations for potential migration underlines those findings. The percentage of respondents who state educational or professional reasons as their main motivation is consistently larger among men, and the percentage of those who state their partner or other family as their main motivation is consistently larger among women.

Individuals' intentions to stay in their host country temporarily or permanently differ depending on their main motivations for potential migration. We find that those planning to migrate primarily for educational or career reasons are most likely to plan to return. Those planning to migrate for family reasons are more likely to intend to migrate permanently. A better understanding of the length of the intended stay is valuable for both origin and destination countries. For origin countries concerned about the emigration of the high-skilled, it allows identifying which share of planned emigration is about brain drain and which is about brain circulation. Destination countries, instead, can use the information to guide integration policies as optimal policies depend on the expected duration of stay.

While there are some studies on return migration (e.g. Dustmann 2003), they use data on individuals after their realized emigration or return to their origin country. Data on temporary migration is generally rare, as are analyses of potential future migrants who intend to stay only temporarily in their destination country. This is so despite evidence that repeat and return migrants make up a considerable share of overall migrants (Dustmann and Görlach 2016). In our data, too, roughly 25 % of the language course sample and 34 % of the university student sample intend to migrate only temporarily.

Our finding that a considerable share of individuals intends to emigrate only temporarily, rather than deciding to return to the home country due to a negative shock, has

[5]While this terminology is widely used, some studies refer to similar concepts, but use different terminology. We will subsume them under those two terms. For an extensive list of terms and questions employed in research on potential migration, see Aslany et al. (2021).

important implications for migration research. Previous research on migration aspirations and intentions has relied on the Gallup World Poll, which does not observe migration intentions among those without aspirations and lacks a clear differentiation of potential future migration into temporary and permanent migration.[6] This means that migration intentions may be systematically underestimated: GWP questions on migration intentions miss both those with migration aspirations who intend to migrate only temporarily, and those who intend to migrate even if they would not ideally like to live permanently in another country.

A major limitation of our data is that the surveys are restricted to subsets of the population that are more likely to emigrate than the population on average. Comparing migration aspirations among our survey respondents to those in the GWP shows that the gap is considerably smaller in European countries than in non-European ones. It is also smaller among young respondents. Nonetheless, even among the younger age group our survey respondents are more likely to aspire to emigrate in all countries. Overall, the respondents constitute a selected, but policy-wise important group of potential migrants.

To sum up, our paper contributes to the existing literature in several ways. First, it sheds light on migration patterns which have received little attention so far, i.e. temporary migration and migration intentions among those without aspirations to migrate permanently. Second, it reveals individual level barriers and drivers towards migration that are not captured by previous research that does not ask about migration intentions from those who do not aspire to migrate permanently. Third, it allows identifying subgroups which are more affected by such barriers and drivers in their migration choice. Given that the highly educated are more likely to emigrate (Grogger and Hanson 2011), women who might otherwise pursue an ambitious career could be disproportionately affected by the prospect of being a "tied mover". As early investments in careers are made before the realizations of own and partner's job opportunities abroad are revealed, this would affect the incentives well beyond the group that actually emigrates.

The remainder of the paper is structured as follows. In Section 4.2, we review the literature on potential migration and introduce our migration choice model. Section 4.3 describes our data sets and provides some descriptive statistics. Section 4.4 shows the estimation strategy. Section 4.5 presents the results and Section 4.6 concludes.

4.2 Measuring Potential Migration

The literature on potential migration relies heavily on survey data and should in the best case scenario convey migration aspirations and intentions reliably. Potential future migrants are identified based on differently framed survey questions.

[6] Data used in this study are obtained from Gallup (2018); see Gallup (2021) for a description of the methodology and codebook.

4.2.1 Migration Aspirations

Several recent studies utilize GWP data to determine potential migration (see Bertoli et al. 2022; Bertoli and Ruyssen 2018; Manchin and Orazbayev 2018; Ruyssen and Salomone 2018), making use of a measure we will subsequently call migration aspirations. The GWP asks respondents *"Ideally, if you had the opportunity, would you like to move permanently to another country, or would you prefer to continue living in [country in which the survey takes place]?"* and gives the response options *"Like to stay in [country in which the survey takes place]"* or *"Like to move to another country"*. This describes aspirations to migrate in a hypothetical, ideal world scenario under the absence of any barriers and makes no statement about concrete considerations or plans.[7]

4.2.2 Migration Intentions

Some studies utilize migration intentions, which express considerations or plans to migrate in a real world scenario, subject to constraints. Though there has been some debate on whether intentions can predict future behavior (Abel et al. 2019; Bassett and Lumsdaine 2001; Bertrand and Mullainathan 2001; Manski 1990), in the economic literature, intentions are frequently used to predict behavior (e.g. Brown 2006; Chandon et al. 2005; Falck et al. 2017; Goda et al. 2012) and migration research is no exemption here (e.g. Dustmann 2008). Under the 'best case' hypothesis, i.e. when respondents state their intentions based on rational expectations, intentions are indeed the best predictor of future behavior (Manski 1990). Burda et al. (1998) assume that intentions are a "monotonic function of the underlying driving variables which motivate migration", and studies frequently find migration intentions to strongly predict subsequent behavior (e.g. Tjaden et al. 2019).

4.2.3 Combining Migration Aspirations and Intentions

In our survey, we ask about respondents' aspirations to migrate permanently in the same way as in the GWP questionnaire (see Section 4.2.1). In direct succession, and independent of their answer to the former, all respondents are asked about their migration intentions: *"Tick the statement that applies to your current situation"*. Out of five options, response options *"I would not move to another country under any circumstances"* and *"In principle, I would move to another country, but I have not thought about it in the last 12 months"* are classified as the respondent having no intentions to migrate, and the remaining response options *"I have been thinking about moving to another country in the last 12 months, but have no specific plans"*, *"I am planning a move to another country"*, and *"I already have a date for my planned move to another country"* are classified as

[7]Measures of aspiration to realize certain desires given the absence of barriers have been used in different economic fields. E.g., Fortin et al. (2015) differentiate between educational aspiration and expectation, where the former describes the desire to attend post-graduate educational institutions after high school, and the latter assesses the likelihood of attending. Graham and Pozuelo (2022), following a similiar understanding of aspirations, observe aspirations regarding migration, education and occupation to show how high aspirations are positively linked to life outcomes such as health and professional development.

the respondent having migration intentions.[8] Those with migration intentions are further divided by whether their intentions relate to temporary or permanent migration.[9]

Combining those questions, we categorize potential migrants in six mutually exclusive combinations of aspirations and intentions (see Panel A of Figure 4.1). As a comparison, we show in Panel B of Figure 4.1 the limited options that are covered in the Gallup World Polls, due to questions regarding migration intentions being asked only from those with migration aspirations, and the questions regarding migration intentions only relating to permanent migration.[10]

According to the Roy-Borjas model (Borjas 1987), individuals who are able to migrate intend to migrate if their expected utility from relocating abroad, net of migration costs, exceeds their expected utility from staying. Then, both migration aspirations and migration intentions emerge as the result of a cost-benefit analysis of potential migration. However, only migration intentions are subject to constraints and drivers according to the real world situation of the individual, while migration aspirations are not. Such constraints, such as visa requirements, liquidity constraints, and social ties at home, may prevent migration intentions from being realized even when one would ideally like to migrate. At the same time, migration intentions can also be evoked by migration drivers even when one would ideally not like to migrate, like a partner wanting to emigrate and career reasons. Consequently, migration aspirations and intentions do not necessarily match for several reasons.[11]

First, the question on migration aspirations aims at a hypothetical scenario (*"Ideally, if you had the opportunity, would you like [....]"*), while migration intentions refer to the current situation which is affected by migration constraints and drivers.

Second, aspirations to migrate are not restricted to any specific timescale, while migration intentions are. The GWP as well as our measure of intentions ask about considerations within the last 12 months or fixed plans which are likely to be executed

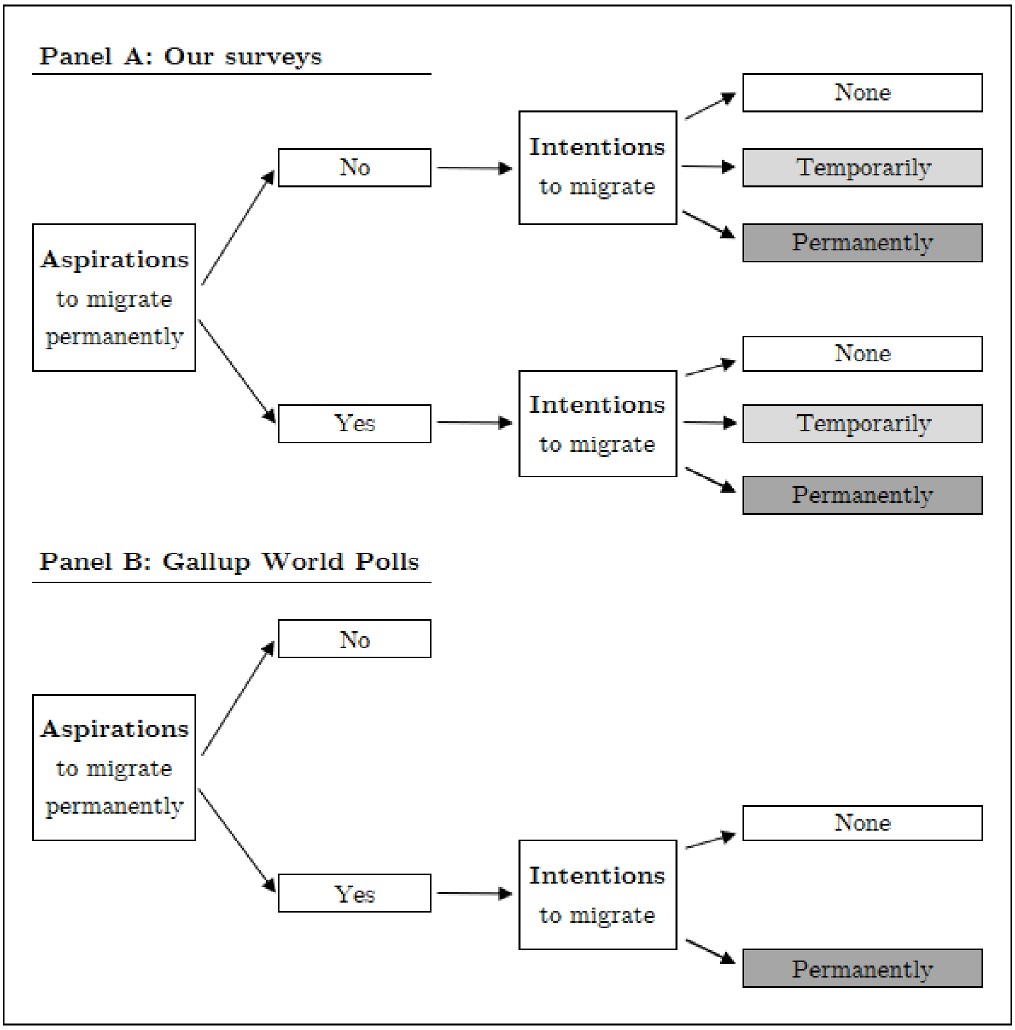
[8]This builds on Ajzen's (1991) *Theory of Planned Behavior*, which differentiates between intentions, which do not include any concrete actions, and behavior, which include concrete actions.

[9]Those respondents with migration intentions who state that they would most likely stay in their preferred destination country for more than 5 years or state that their return to the country in which the survey took place after a temporary stay in their preferred destination country is unlikely are classified as having intentions to migrate permanently; the rest is classified as having intentions to migrate temporarily. Those who state no migration intentions are not asked those questions.

[10]The 2009-2012 waves of the GWP included also two questions referring to 'temporary migration' when it comes to migration aspirations (*"Ideally, if you had the opportunity, would you like to go to another country for temporary work, or not?"* and *"Ideally, if you had the opportunity, would you like to go to another country to study or to participate in a work-study program, or not?"*). If one of those questions was answered affirmatively, respondents were also asked about their intentions to migrate temporarily in the next 24 months (*"Are you planning to go to another country for temporarily work in the next 24 months?"* or *"Are you planning to go study or participate in a work-study program in another country in the next 24 months?"*) and whether they had done any preparations regarding those plans. However, those questions on intentions to migrate temporarily were only asked for selected countries in the 2009 and 2010 waves and again suffer from the conditionality in responses.

[11]Even so, some studies use the question on migration aspirations as a measure for migration intentions, e.g. Bertoli and Ruyssen (2018), while others also include alternative questions on planning and preparation activities, e.g. Ruyssen and Salomone (2018).

Figure 4.1: *Combinations of Aspirations and Intentions*



within the near future, hence whenever we refer to 'intentions', we refer to this measure of current intentions to migrate. Aspiring to migrate permanently yet not having current intentions to do so does not need to be a conflict, but rather is potentially just a question of looking for the optimal timing. Some individuals, for example, who would first like to finish their education or wait for their kids to leave home before migrating permanently, would state to have migration aspirations despite not actually intending to migrate for years to come. Further, the question on migration aspirations is restricted to permanent migration. The GWP as well as our measure of migration aspirations include the word 'permanently', hence, whenever we refer to 'aspirations', we implicitly mean aspirations to migrate permanently. Individuals who aspire to migrate temporarily are inclined to answer negatively. Indeed, utilizing the 2009-2012 waves of GWP data, which in addition to the question on aspirations referring to 'permanent migration' includes questions on migration aspirations referring to 'temporary migration', Delogu et al. (2018) show that those aspiring to migrate temporarily are a considerable share of potential migrants.[12]

Third, and related to the previous reason, choices regarding migration aspirations and intentions are not necessarily made in a consecutive order, and one may not be conditional on the other. Aspirations to migrate are not necessarily followed by intentions to migrate, and intentions to migrate are not necessarily based on aspirations. In fact, many individuals, especially from developing countries, would like to migrate under ideal circumstances, yet will never do so. Aggregated GWP data from the 2018 wave, for example, shows that 25.0 % of surveyed Mexicans would like to migrate under ideal circumstances, yet only 310,000 Mexicans emigrated in 2019, which is a mere 0.2 % of the 127.6 million residents Mexico had in 2019.[13] Further, some individuals might form intentions and only afterwards decide whether they want to stay abroad temporarily or permanently. Alternatively, some form intentions to migrate only if they have the option to migrate temporarily (e.g. students' choice to do a semester or course abroad might be based on the option to stay temporarily, and they would not have formed those intentions if returning after their studies was not an option). As different orders of choice are possible, we do not model those choices as being consecutive, but rather as being simultaneous – resulting in our 6 combinations of aspirations and intentions shown in Panel A of Figure 4.1. This is in stark contrast to the GWP data, which assumes a conditionality in choice. In the GWP, only those respondents who state positive migration aspirations are asked subsequent questions regarding their migration plans within the next 12 months (*"Are you planning to move permanently to another country in the next 12 months, or not?"*) and whether they have already prepared for this move (*"Have you done any preparation for this move (for example applied for residency or visa, purchased the ticket, etc.)?"*). While those questions indeed indicate how far the respondents are in their respective migration decision making process, they exclude all respondents who did not state aspirations to migrate permanently (Panel B of Figure 4.1).

Allowing for differences between aspirations and intentions and extending the time horizon to temporary and permanent intentions, our survey design gives access to two groups of potential migrants which have not received much attention so far. By avoiding the conditionality in responses problem, we observe those who do not report aspirations to migrate permanently but nonetheless consider or plan to migrate permanently. By introducing a time horizon to the question on intentions, we further observe both those who do not report aspirations to migrate permanently but consider or plan to migrate temporarily; and those who do report aspirations to migrate permanently but again only consider or plan to migrate temporarily. Those groups are not only an important target of migration research and policy, with most studies relying on GWP data they have also been largely neglected so far in migration research.

[12]Still, most works utilize the GWP question referring to 'permanent migration' (Bertoli et al. 2022; Manchin and Orazbayev 2018) or combine the questions without specifically differentiating between them (Bertoli and Ruyssen 2018; Ruyssen and Salomone 2018).

[13]Aspirations and behavior fall apart in other contexts as well, e.g. career aspirations. Dhar et al. (2022) show that 72 % of surveyed Indian girls expect to be employed in a white-collar job by the age of 25, yet corresponding shares among young women in India are actually much lower.

4.2.4 Determinants of Aspirations and Intentions

Based on the *Human Capital Theory of Migration*, the net gain of migration decreases with age as it lowers the time to recoup the investment made by migration (Sjaastad 1962). In addition, older individuals face higher migration costs in general as they lose specific human capital as well as their social and professional networks while having larger difficulties in adapting to a new language and a new environment (Belot and Ederveen 2012).

According to this human capital theory, the net gain of migration also depends on education and returns to human capital, i.e. income differences, in different countries (Borjas 1987). Empirically, highly educated individuals have generally been found to be more likely to migrate (see, e.g., Borjas et al. 2019; Docquier et al. 2014; Grogger and Hanson 2011).

Overall, the uncertainty about potential returns and costs makes migration a risky choice. Individuals who are more willing to take risks are expected to be more likely to consider migration (Dustmann et al. 2020; Jaeger et al. 2010). Lastly, the respondent's patience is included as a control, as it might be relevant for weighing costs and benefits of migration.

Yet, migration is not only an individual decision but happens in the context of a social environment (Stark and Bloom 1985), hence family ties are bound to influence the migration decision – though the channels through which this happens can lead to different outcomes. Strong family ties in the source country, such as marriage, long-term relationship or children, could be a restricting factor regarding emigration. Family ties abroad, on the contrary, could work as an encouraging factor as they imply easier access to information about labor market opportunities, as well as financial and emotional support. This all substantially lowers costs and risks related to migration (Manchin and Orazbayev 2018). Also, a partner who wants to emigrate can evoke migration considerations and plans among individuals, e.g. if joint emigration is necessary to maintain the relationship. Hence, generally, migration decisions are coordinated within households or families, yet that coordination seems to be more binding for women (Munk et al. 2022).

While migration can be evaluated ex-post positively as well as negatively, migration-related soft-skills and experience do not depend on such subjective evaluation and can reduce migration costs and constraints. Either way, repeat migration constitutes a considerably large share of overall migration flows (e.g. DaVanzo 1983) and hence, having previously migrated has been shown to be a strong predictor of subsequent migration (Uebelmesser 2006).[14]

Table 4.B in the Appendix presents the definitions of the explanatory variables we

[14]Leaving one's home country is a very different decision compared with leaving a host country, and for the foreign born, emigration plans can constitute repeat migration as well as return migration to their home country. Since we cannot distinguish between the two, we exclude foreign born respondents (i.e. those who have been born in a country different from the country in which the survey took place) from our main analysis. We present the joint distribution of migration aspirations and intentions for foreign born respondents in Table 4.C in the Appendix.

derive from our data. Before discussing how we utilize those determinants within our estimation strategy, we present our data in the next Section.

4.3 Data and Descriptive Statistics

We conducted two multinational surveys: one among language course participants at the Goethe Institute[15] in 14 countries and one among university students in six countries. The survey at the GI captures a self-selected subset of the population, many of whom can be expected to be relatively far in the migration process. The survey at universities allows us to assess the generalizability of our GI results for a population of young and highly skilled individuals.

4.3.1 Survey Design

Survey at the Goethe Institute

We conducted a survey among language course participants at the GI. The survey consisted of a pen and paper questionnaire containing a wide range of questions on socio demographic characteristics, education and labor market status, language skills as well as migration plans and previous migration experience. The survey took place between June and December 2018. In order to minimize potential language barriers, the questionnaires were translated into the main language of each country by professional translators and double-checked by at least one native speaker of each language. In India, the questionnaire was in English. As we aimed at a heterogeneous sample of countries, we identified groups of countries based on combinations of the following characteristics: geographic distance to Germany, linguistic distance to the German language, economic development and absence or presence of migration barriers vis-à-vis Germany. We conducted the survey in at least one large institute (in terms of course participants) for each group (see Table 4.1).

In order to maximize the response rates, we took two measures. First, a member of the project team was present during the survey in all institutes in all European countries. During one week of the course term, the pen and paper questionnaires were personally and directly distributed to all course participants present in the classroom. Participants then either filled in the questionnaire during the course break or after the course, or they took it home and returned it at a second course session within that week. For non-European countries, we sent the questionnaires by mail to the institutes, where the teachers distributed and collected the questionnaires, which were then sent back to Germany. In

[15]The Goethe Institute is a German cultural association which aims to promote German culture and language around the globe. For that purpose, the Goethe Institute is present in 98 countries, with a total of 158 institutes. It offers language services, i.e. language courses and standardized exams, and provides information about the German culture and society with events and libraries (Goethe Institute 2019a). See also Uebelmesser et al. (2018a) for further background information. In this paper, we stick to the following convention: when referring to the association of the Goethe Institute, we use the abbreviation “GI”. When talking about a specific branch of the GI in Germany or abroad, we refer to it as “institute”.

order to minimize errors in distributing the questionnaires, we prepared envelopes for each course containing the questionnaires, which were distributed to the respective teachers of the courses. Second, for each country, we raffled off one free language course at the survey institute in order to incentivize participation. Those measures resulted in response rates ranging from 56.7 to 86.4 % in European countries, and 47.9 to 59.9 % in non-European countries.[16] Table 4.1 gives an overview of the countries the survey was conducted in and their respective characteristics, as well as the target population, the number of respondents who returned the survey, the number of respondents without missings in relevant questions, this number as percentage of the target population and the number of respondents in the final sample, separately for each country.[17] Of the 6,664 language course participants at institutes in 14 countries who participated in the survey, 1,554 individuals were excluded from the analysis due to missings in relevant questions or because they were foreign born. This leaves us with a final sample of 5,110 individuals.[18] Descriptive statistics can be found in Table 4.D in the Appendix.

Survey among University Students

Additionally, we conducted a survey among university students, which was designed similarly to the GI survey and contained the same questions on socio demographic characteristics, education and labor market status, language skills as well as migration plans and previous migration experience. For the survey, three European and three non-European countries were chosen, which were also part of the GI sample. The survey was conducted between April 1, 2019 and April 7, 2020 in all universities that agreed to participate, either at the university level or in selected faculties. As the survey was conducted as an online survey, local university staff sent invitation e-mails including a link to the survey to students. Again, to minimize potential language barriers, the survey was offered in the local language and in English in all countries, apart from India, where

[16]In all European countries except the Netherlands, these numbers refer to actually distributed questionnaires. In the Netherlands and in non-European countries, the response rate is related to registered course participants, as we do not know the number of course participants who were present when the questionnaires were distributed. In the European institutes where a member of the project team was present, not all registered participants attended every lesson of their course, i.e. the number of registered participants is much larger than the number of present participants in many cases. Therefore, the response rate for non-European countries and the Netherlands, which is related to the number of registered course participants and not to the number of present course participants, gives a lower bound. In Mexico, the response rate is related to the distributed questionnaires during the course inscription.

[17]Respondents were excluded if they had not answered the questions on migration aspirations and intentions or the questions on their birth country, their age or highest educational degree, whether they were a student, their prior migration experience, their relationship status or their willingness to take risks and patience. In order to be included in the final sample, on which we do our analysis, respondents needed to answer all of the questions above and not be foreign born.

[18]A concern about our analysis relates to self-selection on the dependent variables, especially the question on migration aspirations. We address this by regressing a dummy for answering the question on migration aspirations on several questions which are featured on the first page of the survey (see Table 4.E in the Appendix). The results reveal that respondents of age 25 and above and women are less likely to answer the question on migration aspiration.

Table 4.1: *Participants per Country — GI Survey*

Countries	Income (GNI/capita)	EU member	Target population	Returned survey	Without missings	Response rate	Final sample
Bosnia	Upper-middle	No	272	270	235	86.4 %	204
Czechia	High	Yes	580	481	439	75.7 %	327
India	Lower-middle	No	1156	830	688	59.5 %*	687
Indonesia	Lower-middle	No	1619	883	775	47.9 %*	761
Italy	High	Yes	432	371	314	72.7 %	281
Japan	High	No	494	293	255	51.6 %*	241
Mexico	Upper-middle	No	800	491	427	53.4 %**	411
Netherlands	High	Yes	208	139	118	56.7 %*	90
Poland	High	Yes	344	236	200	58.1 %	186
Romania	High	Yes	372	327	279	75.0 %	265
South Korea	High	No	719	470	431	59.9 %*	418
Spain	High	Yes	737	611	525	71.2 %	454
Ukraine	Lower-middle	No	841	782	645	76.7 %	614
United Kingdom	High	Yes (leaving)	548	480	399	72.8 %	171
Total			9,122	6,664	5,730	-	5,110

Note: Response rates are the percentage of respondents who answered the questions on migration aspirations and intentions as well as the questions on their birth country, their age and highest educational degree, whether they were a student, prior migration experience, their relationship status and their willingness to take risks and patience among those which attended the course the day the survey took place and hence were handed a questionnaire (target population). The final sample, on which we do our analysis, includes all respondents who provided answers to those questions and are not foreign born. *For India, Indonesia, Japan, the Netherlands and Korea response rates are based on registered course participants, not actual attendance. **For Mexico, the response rate is related to the distributed questionnaires during the course inscription. High-income countries include countries which have a GNI per capita larger than $12,535 in current US-Dollars, as of 2020 (Czechia, Italy, Japan, Netherlands, Poland, Romania, South Korea, Spain, and the United Kingdom); middle-income countries (upper-middle and lower-middle) are countries which have a GNI per capita of $1,036 to $12,535 in current US-Dollars, as of 2020 (Bosnia, India, Indonesia, Mexico, and Ukraine). EU membership as of 2018.

the survey was only available in English. To incentivize participation, individuals could take part in a lottery, which was embedded in the questionnaire. One prize in each survey was a cash payout of EUR 100, and there was also an opportunity to participate in two other lotteries with additional prizes, which depended on choices that respondents made. The largest single prize won among all participants was EUR 250.

Table 4.2 reports the target population, the number of respondents who started the survey, the number of respondents without missings in relevant questions, this number as percentage of the target population and the number of respondents in the final sample, separately for each university. Of the 3,716 students who participated in the survey and answered the questions on migration aspirations and intentions as well as questions on their country of birth, age, education, prior migration experience, relationship status,

Table 4.2: *Participants per University — Student Survey*

Universities	Target population	Started survey	Without missings	Response rate	Final sample
Czechia					
Masaryk University	2,255	495	370	16.4 %	226
University of Ostrava	2,684	373	254	9.5 %	226
University of Economics Prague	3,917	551	397	10.1 %	323
India					
IIT Kanpur	5,261	929	390	7.4 %	388
Ashoka University	1,452	57	39	2.7 %	38
Indonesia					
Institut Pertanian Bogor	n.a.*	26	11	n.a.*	11
Universitas Indonesia	n.a.*	16	4	n.a.*	4
Institut Teknologi Bandung	3,481	323	133	3.8 %	130
Politeknik Manufaktur Bandung	n.a.*	55	29	n.a.*	29
Universitas Padjadjaran	n.a.*	18	8	n.a.*	8
Italy					
Università Cattolica del Sacro Cuore	11,799	360	245	2.1 %	235
Mexico					
El Colegio de Mexico	368	147	112	30.4 %	102
Centro de Investigacion y Docencia Economicas	476	59	48	10.1 %	46
Instituto Tecnologico Autonomo de Mexico	5,032	716	535	10.6 %	511
Universidad Nacional Autonoma de Mexico	2,854	623	419	14.7 %	381
Spain					
Carlos III University of Madrid	8,282	718	499	6.0 %	406
University of Barcelona	6,712	260	176	2.6 %	137
Universitat Autonoma de Barcelona	1,915	96	47	2.4 %	41
Total	56,488	5,822	3,716	-	3,242

Note: Response rates are the percentage of students who answered the questions on migration aspirations and intentions as well as the questions on their birth country, their age and highest educational degree, prior migration experience, their relationship status and their willingness to take risks and patience among those which received an invitation e-mail (target population). The final sample, on which we do our analysis, includes all respondents who provided answers to those questions and are not foreign born. *For Indonesia response rate is calculated for Institut Teknologi Bandung only, since information on the size of the target population is not available for the rest of the universities.

willingness to take risks and patience, 474 individuals were excluded from the analysis because they were foreign born. This leaves us with a final sample of 3,242 students.[19] Descriptive statistics can be found in Table 4.D in the Appendix.

[19]Again we address the concern that our analysis relates to self-selection on the dependent variables by regressing a dummy for answering the question on migration aspirations (see Table 4.E in the Appendix). The results reveal that respondents who are more patient are somewhat more likely to answer the question.

4.3.2 Descriptive Statistics

Both resulting data sets have limitations, and neither is representative of the general population. Our study focuses on specific self-selected groups, as the surveyed individuals are either participants of language courses or university students. Both groups are more likely to be better educated and more likely to aspire to migrate than the overall population, the latter especially applying to language course participants.

While the percentage of women and those with migration experience as well as measures of willingness to take risks and patience in our GI and student samples are fairly comparable (see Table 4.D in the Appendix), other individual characteristics differ.[20] As can be expected, occupational status and age differ between samples: 32.7 % of the GI sample are students compared to 100.0 % in the student sample, and the percentage of respondents in the age group of 18 to 24 years is 38.7 % in the GI sample compared to 76.8 % in the student sample. The percentage of respondents with children is higher in the GI sample than in the student sample, and the distribution over nationalities differs. Both samples are, however, comparable when it comes to the percentages of migration aspirations and intentions among the younger age groups. Thus, in addition to running our main analysis on both samples separately, we will also restrict the GI sample to the younger age groups (under 35 years of age) in our subsequent analysis and compare the results.

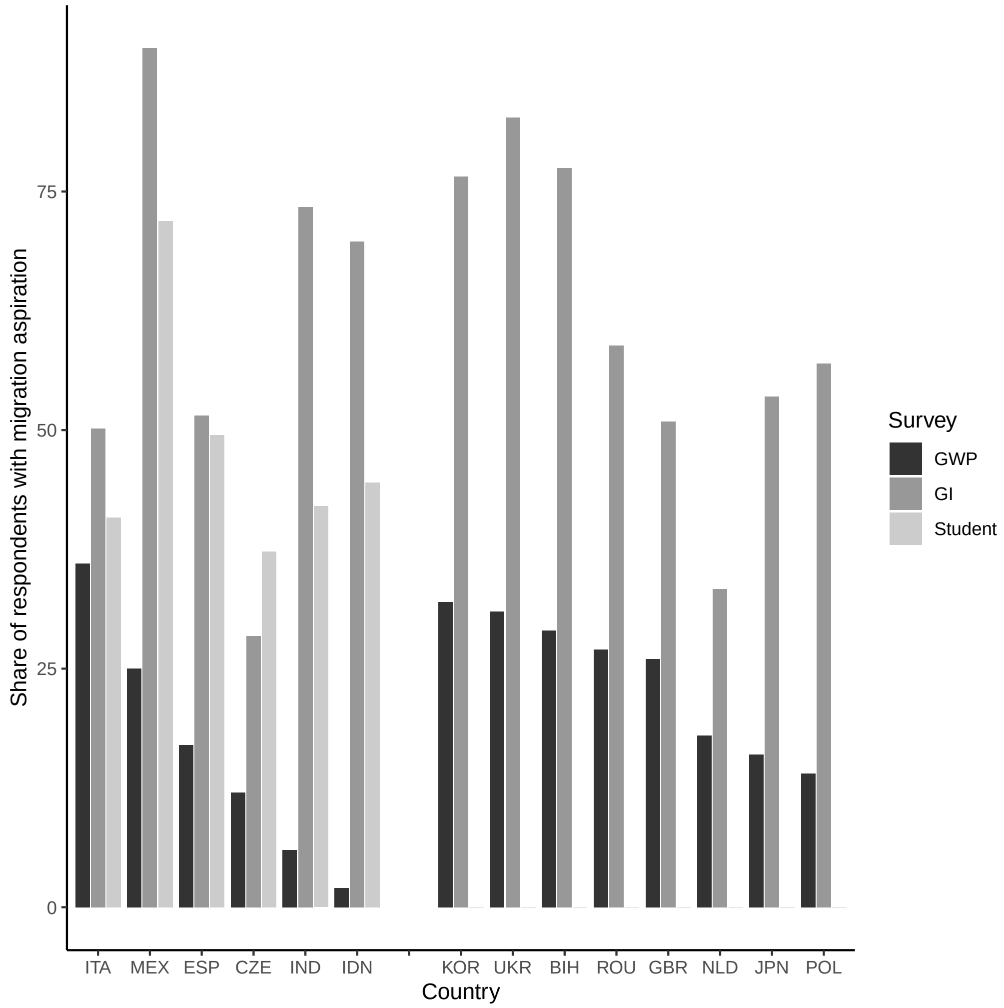
Migration Aspirations

A comparison of migration aspirations for each country based on data from the GWP and data from our GI and student samples illustrates that individuals in both our samples are more likely to aspire to migrate than the overall population. In Figure 4.2, we see that in all countries the percentage of people who would ideally like to migrate permanently is much higher in the GI and student samples than in the GWP data: In the GI sample, we find the highest percentage of respondents who would ideally like to migrate permanently in Mexico with 90 % (compared to 25 % in the GWP data) and the lowest percentage in Czechia with 28 % (compared to 12 % in the GWP data). The largest discrepancy between the two groups can be observed for India (73 % in GI versus 6 % in GWP) and Indonesia (70 % in GI versus 2 % in GWP), while the smallest discrepancy can be observed for Italy (50 % in GI versus 36 % in GWP). In the student sample, the percentage of respondents with migration aspirations is higher than in the GWP in all countries as well. Yet, as expected, it is smaller than in the GI sample, with the exception of Czechia.

The over-representation of individuals with migration aspirations in our GI sample relative to the GWP data can also be seen in Figure 4.A in the Appendix, where we plot the percentage of individuals with migration aspirations in the GWP data and in the GI sample by age groups. For all countries and age groups, the percentage of individuals with migration aspirations in the GI sample exceeds the corresponding percentage in

[20]Note that both the GI and the student samples have high percentages of respondents with migration experience (33.3 % and 34.6 %, respectively). This can be explained by how we defined migration experience – at least one prior stay abroad for at least three consecutive months.

Figure 4.2: *Percentage of Respondents with Migration Aspirations — GWP, GI and Student Sample*



Note: This figure compares the percentages of respondents with migration aspirations between the GWP data, and the GI and student samples. Data from the GWP refers to the 2018 wave, apart from data for Spain, Italy, the Netherlands, and the United Kingdom that refers to the 2017 wave, as data for 2018 was not available.

the GWP data. At the same time, we find for most countries that the percentages are most comparable for younger age groups and comparability decreases for middle aged and older age groups. With respect to their stated migration aspirations, the younger individuals in the GI sample are thus closer to the general population of the same age group in the respective countries than the older individuals.

Due to the different age structure in the student sample and the GI sample, the comparison with the student sample is most appropriate when the GI sample and the GWP data are also restricted to the younger age group. Figure 4.B in the Appendix shows that within the younger age groups, the percentage of respondents with migration aspirations in the student sample is most comparable to the percentage of respondents with migration aspirations in the GI sample. Still, it is smaller for all countries, with the

Table 4.3: *Joint Distribution of Aspirations and Intentions*

Panel A: GI sample	No intentions to migrate		Intentions to migrate temporarily		Intentions to migrate permanently		*Total*	
No aspirations to migrate permanently	1127	(22.1)	466	(9.1)	150	(2.9)	1743	(34.1)
Aspirations to migrate permanently	602	(11.8)	790	(15.5)	1975	(38.6)	3367	(65.9)
Total	1729	(33.9)	1256	(24.6)	2125	(41.5)	5510	(100.0)
Panel B: Student sample								
No aspirations to migrate permanently	1071	(33.0)	415	(12.8)	74	(2.3)	1560	(48.1)
Aspirations to migrate permanently	483	(14.9)	609	(21.3)	509	(15.7)	1682	(51.9)
Total	1554	(47.9)	1105	(34.1)	583	(18.0)	3242	(100.0)

Note: This table shows numbers of observation with percentages in parentheses. Row and column *Total(s)* show row and column totals; percentages of total sample size in the panel are in parentheses.

exception of Czechia. Only in Italy, the percentage of student respondents with migration aspirations is lower than in the GWP data.

Migration Intentions

A direct comparison of migration intentions between our two samples and the GWP data is not possible. Due to the problem of conditionality in responses (see Section 4.2.3) we would need to restrict the observations to only those with migration aspirations. For this reason, we focus on the GI and student samples only. Migration intentions are differently distributed in the GI and student sample. Figure 4.D in the Appendix shows for the younger age groups that the share of those without migration intentions is larger in the student sample than in the GI sample, and the share of those who intend to migrate permanently is consistently larger in the GI sample than in the student sample. Shares of those who intend to migrate temporarily are more comparable between the two samples.

Joint Distribution of Migration Aspirations and Intentions

To observe in how far aspirations and intentions to migrate match in our data, we check their joint distribution in Table 4.3. Indeed, all six outcomes shown in Figure 4.1 occur in both samples.[21]

[21] Since we excluded foreign born respondents (i.e. those who have been born in a country different from the country in which the survey took place) from our two samples, we check the joint distribution of migration aspirations and intentions for them separately (see Table 4.C in the Appendix). Also among foreign born respondents all outcomes occur in both samples. However, among foreign born respondents a higher percentage has migration aspirations and a higher percentage has migration intentions compared to the native born respondents in both the GI and student sample.

Table 4.3 reveals several important patterns. First, 38.6 % of respondents in the GI sample and 15.7 % of respondents in the student sample both aspire and intend to migrate permanently, and 22.1 % in the GI sample and 33.0 % in the student sample neither aspire nor intend to migrate. This makes in the GI sample 60.7 % for whom aspirations and current intentions match perfectly, and leaves 39.3 % of respondents for whom they do not. In the student sample, aspirations and intentions match only for 48.7 % of respondents, suggesting that GWP approach would miss important patterns of migration aspirations and intentions for slightly more than half of all student respondents. Second, similar to Delogu et al. 2018, those intending to migrate temporarily are a substantial percentage of potential migrants that is overlooked in surveys asking only for intentions to migrate permanently. Indeed, we find that 24.6 % of respondents in the GI sample and 34.1 % of respondents in the student sample intend to migrate only temporarily. We will come back to this in Section 4.5.2. Third, we observe clear differences between aspirations and intentions to migrate permanently: 11.8 % of respondents in the GI sample and 14.9 % in the student sample aspire to migrate permanently yet have no current intentions to do so and 2.9 % of respondents in the GI sample and 2.3 % in the student sample intend to migrate permanently despite not desiring to do so.[22] We will come back to this in Section 4.5.3.

Among those respondents with migration intentions, we take a closer look at their preferred destination country. Table 4.G in the Appendix shows the top 5 preferred destination countries by origin country for the GI and student samples. As expected, in the GI sample Germany is by far the most common destination, followed by other German speaking destinations (Austria and Switzerland) as well as the United States and the United Kingdom. In the student sample, there is a remarkable similarity in preferred destinations across all surveyed countries. The United States is the most popular destination, the United Kingdom ranks second and Germany ranks third in all countries. Canada is the fourth most popular destination and France the fifth most popular one.

When we compare country of origin and preferred destination within individuals, we observe certain patterns regarding barriers towards migration. We see that among respondents whose freedom of movement is limited, 13.7 % (GI sample) to 18.5 % (student sample) state no current considerations or plans to migrate despite their aspirations to do so.[23] At odds with Docquier et al. (2015) and Delogu et al. (2018) who assume that all migration aspirations would realize if visa restrictions were abolished, even among respondents who enjoy freedom of movement (i.e. who would migrate within the EU and

[22]For comparison, Bertoli and Ruyssen (2018) analyze a subset of their observed GWP waves of 2007 and 2011. They find that of those who stated migration aspirations only 14.3 % also answered affirmatively to the question *"Are you planning to move to another country in the next 12 months, or not?"*, and only 42.7 % were taking concrete steps towards migration. Ruyssen and Salomone (2018) show that while 16.0 % of their women subsample of the GWP stated migration aspirations, only 4.0 % of those stated that they had done any preparation for this move.

[23]We consider freedom of movement to be limited if a respondent would migrate from a country outside the EU and European Economic Area (including Switzerland) to within it, or vice versa, or between two countries outside the EU and EEA (including Switzerland).

EEA, including Switzerland), 9.6 % in the GI sample and 9.5 % in the student sample state no current considerations or plans to migrate despite their aspirations to do so.[24]

Our framework allows us to identify two groups of potential migrants which have not been observed by studies relying on GWP data. By avoiding the conditionality in responses problem, we observe those who have no desire to migrate permanently, i.e. those who do not state aspirations to migrate permanently under ideal circumstances but who nonetheless consider or plan to migrate permanently (2.9 % of respondents in the GI sample, and 2.3 % in the student sample). By introducing a time horizon to the question on intentions, we observe those who do not aspire to migrate permanently yet intend to migrate temporarily (9.1 % of respondents in the GI sample, and 12.8 % in the student sample) and those who aspire to migrate permanently yet only intend to migrate temporarily (15.5 % of respondents in the GI sample, and 21.3 % in the student sample).

Main Motivations for Intended Migration

Exploring motivations for potential migration can shed some light on the reasons why individuals intend to migrate permanently versus temporarily, or why they intend to migrate despite having no aspirations to do so. In our surveys, respondents who stated intentions to migrate were asked to name their preferred destination country for a potential move abroad and indicate the main reason for such a move (see Table 4.B in the Appendix for a detailed description). Hence, numbers of observations differ from sample totals. In the GI sample we observe main reasons for potential migration for 1248 of 1929 women with intentions to migrate, and for 827 of 1349 men with intentions to migrate. In the student sample we observe main reasons for potential migration for all 940 women with intentions to migrate, and for 739 of 741 men with intentions to migrate.[25]

Table 4.4 shows how the percentages of main motivations differ between those with intentions to migrate temporarily and those with intentions to migrate permanently by gender in both samples. Educational reasons, including studies for a university degree, are the most common main motivation to emigrate. This holds among those with intentions to migrate temporarily and among those with intentions to migrate permanently. Professional reasons, such as the prospect of higher income or a more interesting job or a transfer by the employer, are the second most common motivation among those who intend to migrate temporarily. Combined, educational and professional reasons are the main motivations to intend to migrate for a larger percentage of men than women, and play a more prominent role among those who intend to migrate temporarily. Among those with intentions to migrate permanently despite having no aspirations to do so, the percentage of respondents who state either educational or professional reasons is consistently larger

[24]Numbers of observation differ from sample totals since respondents who stated *“I would not move to another country under any circumstances”* did not answer the question on their preferred destination and not all of those who intend to migrate indicated their preferred destination either. Hence, we observe the freedom of movement criterion for 4355 respondents in the GI sample and 2338 respondents in the student sample.

[25]The difference in response rates is explained as the student survey was carried out online, and the respondent was reminded to answer the question if it was left unanswered. The GI survey was conducted as a paper survey without this possibility.

Table 4.4: *Main Motivations for Potential Migration by Gender*

	Women		Men	
	Intentions to migrate temporarily	Intentions to migrate permanently	Intentions to migrate temporarily	Intentions to migrate permanently
Panel A: GI sample				
	(n = 486)	*(n = 762)*	*(n = 305)*	*(n = 522)*
Educational reasons	43.2	37.9	39.0	34.9
Professional reasons	26.3	22.8	34.4	28.9
Partner or other family	9.5	15.2	4.3	9.4
Other reasons	21.0	24.1	22.3	26.8
Panel B: Student sample				
	(n = 629)	*(n = 311)*	*(n = 469)*	*(n = 270)*
Educational reasons	37.4	28.0	43.7	36.3
Professional reasons	29.6	27.0	27.5	28.1
Partner or other family	5.4	10.3	4.1	4.5
Other reasons	27.6	34.7	24.7	31.1

Note: Each panel of this table shows column percentages. n show column totals.

among men than among women (see Table 4.H in the Appendix, which shows how the percentages of main motivations for intentions to move temporarily or permanently are distributed among those with and without aspirations to migrate permanently). Reasons related to partner or other family – though least commonly stated as the main motivation across all groups – are more strongly linked to intentions to migrate permanently than temporarily and the percentages of respondents who state these reasons larger among women than among men. For those without aspirations to migrate permanently, the percentage of respondents who state partner or other family as the main motivation for intentions to migrate permanently is consistently larger among women (see Table 4.H in the Appendix).

4.4 Estimation Strategy

In Section 4.3 we showed how migration aspirations and intentions are related in our data and which main motivations respondents state for the latter. In this section, we present our estimation strategy to study individual level determinants of aspirations and intentions, and how aspirations predict intentions.

4.4.1 Estimating Aspirations and Intentions to Migrate

In Section 4.5.1, we explore individual level determinants of aspirations and intentions and how they relate. For this, we estimate the probability of having aspirations to migrate in a binary probit model via maximum likelihood method:

$$Pr(\text{aspirations}_i \mid X_i, C_i) = \alpha_1 + \beta_1{}' X_i + \gamma_1{}' C_i + \varepsilon_{1,i} \tag{4.1}$$

where aspirations_i takes a value of 1 if respondent i states having aspirations to migrate and 0 otherwise. X_i represents a set of individual-specific explanatory variables of respondent i: gender, age, university degree, student, migration experience, partner and children, willingness to take risks and a measure of patience. C_i is a country dummy. $\varepsilon_{1,i}$ is an idiosyncratic error term. Table 4.B in the Appendix presents the definitions of all explanatory variables, and Table 4.D shows descriptive statistics.

Also, we estimate the probability of having intentions to migrate. Since we differentiate between time-horizons of migration intentions, we adapt a multinomial probit model and estimate it with the maximum likelihood method:

$$Pr(\text{intentions}_i \mid X_i, C_i) = \alpha_2 + \beta_2{}' X_i + \gamma_2{}' C_i + \varepsilon_{2,i} \tag{4.2}$$

where intentions_i can take any of the following outcomes: no migration intentions, intentions to migrate temporarily or intentions to migrate permanently. To investigate to what extent aspirations to migrate predict intentions to migrate, we also estimate a variation of Equation (4.2) that includes respondent $i's$ aspirations to migrate as an explanatory variable:

$$Pr(\text{intentions}_i \mid X_i, C_i) = \alpha_3 + \zeta * \text{aspirations}_i + \beta_3{}' X_i + \gamma_3{}' C_i + \varepsilon_{3,i} \tag{4.3}$$

We hypothesize that ζ should be positive, as we expect intentions to migrate, both temporarily and permanently, to correlate with migration aspirations. Again, X_i is a set of the above listed individual-specific explanatory variables, C_i are controls, and $\varepsilon_{2,i}$ and $\varepsilon_{3,i}$ are idiosyncratic error terms.

In Section 4.5.2, we take a closer look at Equations (4.2) and (4.3) and exploit the differentiation between intentions to migrate temporarily and permanently in the multinomial probit model. This allows us to focus on the differences in determinants for different time horizons for the intended stay abroad.

4.4.2 Conditionality and Differences

In Section 4.5.3, we explore whether determinants of intentions to migrate, either temporarily or permanently, differ between those with and without aspirations to migrate. We estimate Equation (4.2) conditional on the respondent's aspirations_i in a multinomial probit model with the maximum likelihood method:

$$Pr(\text{intentions}_i \mid \text{aspirations}_i,\ X_i, C_i) = \alpha_4 + \beta_4{}' X_i + \gamma_4{}' C_i + \varepsilon_{4,i} \tag{4.4}$$

with a special focus on those who have no underlying migration aspirations – a group of potential migrants which has been largely neglected so far because of the conditionality problem mentioned above. Results are shown for the GI and student samples separately.

4.5 Results

4.5.1 Aspirations and Intentions to Migrate

To analyze which individual level factors determine the probability of migration aspirations we first estimate Equation (4.1). Column (1) in Table 4.5 shows the results. Second, we estimate Equations (4.2) and (4.3) to explore determinants of the probability of migration intentions. Columns (2) to (5) in Table 4.5 show the results.[26] We run all regressions separately for each sample to check how our findings apply to different populations, and we find largely similar results. As expected, both samples show a clear positive relation between aspirations and intentions. Those with aspirations to migrate permanently are 3.0 times (GI sample) to 3.8 times (student sample) as likely to intend to migrate temporarily, and 15.3 times (student sample) to 23.2 times (GI sample) as likely to intend to migrate permanently compared to those without aspirations. Still, the relation between aspirations and intentions is far from perfect and other determinants remain important in explaining intentions after controlling for aspirations.

Comparing the joint distribution of aspirations and intentions by gender (see Table 4.F in the Appendix) shows that women are just as likely to aspire to migrate as men (GI sample) or slightly more likely (student sample). Hence, being a woman is positively associated with the likelihood of aspirations in the student sample only, and this relationship holds for women from middle-income (non-EU) countries only as we see when we compare results between middle-income and high-income (EU) countries (see Table 4.I and 4.J in the Appendix). However, this does not translate into higher intentions among women. We do not find women to be more likely to intend to migrate in either sample. In the GI sample, women are even significantly less likely to intend to migrate permanently after controlling for their aspirations. This finding is in line with what Ruyssen and Salomone (2018) have shown with GWP data: women, especially in countries where gender discrimination is comparably high, are more likely to aspire to migrate, but they also face costs and obstacles in realizing these aspirations.

Both social ties at home and networks abroad are relevant for migration aspirations and intentions. Manchin and Orazbayev (2018) find that networks abroad account for about 18 % of variation in international migration intentions, and social networks in the country of residence account for only 2-4 %. While we do not focus on social networks in such a broad sense but only on family ties – i.e. partner and children – we too find that family ties abroad play a prominent role in explaining migration intentions. Having a non-native partner makes it 1.9 times (GI sample) to 3.6 times (student sample) as likely to intend to migrate permanently after controlling for underlying aspirations. Subsampling by gender sheds more light on the importance of such family ties (see Tables 4.K and 4.L in the Appendix). The positive linkage between non-native partner and intentions to migrate permanently is large and robust for women in both samples – a finding in line with the theory that especially women are often so-called "tied movers" who follow

[26] As discussed in Section 4.2.3 and shown in Panel A of Figure 4.1 intentions to migrate temporarily and intentions to migrate permanently are understood as mutually exclusive outcomes.

Table 4.5: *Aspirations and Intentions*

Panel A: GI sample	*Binomial probit*	*Multinomial probit*		*Multinomial probit*	
	(1) Aspirations to migrate permanently	(2) Intentions to migrate temporarily	(3) Intentions to migrate permanently	(4) Intentions to migrate temporarily	(5) Intentions to migrate permanently
Migration aspiration				3.019^{***}	23.153^{***}
				(0.086)	(0.107)
Gender: woman	1.064	1.055	0.913	1.009	0.822^{**}
	(0.040)	(0.081)	(0.073)	(0.084)	(0.084)
Age: under 18 years	1.257^{***}	0.812	1.291^{*}	0.749^{*}	1.104
	(0.082)	(0.164)	(0.144)	(0.170)	(0.163)
Age: 25 to 34 years	1.049	1.032	1.408^{***}	1.058	1.481^{***}
	(0.066)	(0.132)	(0.122)	(0.137)	(0.139)
Age: 35 to 49 years	0.845^{**}	0.518^{***}	0.750^{*}	0.551^{***}	0.836
	(0.084)	(0.168)	(0.156)	(0.173)	(0.179)
Age: 50 to 64 years	0.567^{***}	0.146^{***}	0.331^{***}	0.169^{***}	0.430^{***}
	(0.114)	(0.260)	(0.220)	(0.266)	(0.259)
Age: 65 years or above	0.347^{***}	0.065^{***}	0.099^{***}	0.084^{***}	0.191^{***}
	(0.163)	(0.416)	(0.412)	(0.419)	(0.458)
University degree	0.849^{***}	1.312^{**}	0.903	1.412^{***}	1.047
	(0.055)	(0.109)	(0.100)	(0.113)	(0.114)
Student	0.949	1.008	0.818^{*}	1.018	0.830
	(0.061)	(0.119)	(0.109)	(0.124)	(0.124)
Migration experience	1.055	1.247^{**}	1.291^{***}	1.268^{**}	1.316^{***}
	(0.046)	(0.092)	(0.085)	(0.095)	(0.097)
Partner: native	1.027	0.930	1.089	0.919	1.083
	(0.054)	(0.107)	(0.101)	(0.111)	(0.115)
Partner: non-native	1.086	1.050	1.742^{***}	1.089	1.919^{***}
	(0.082)	(0.173)	(0.157)	(0.179)	(0.182)
Children	0.914	0.807	0.950	0.827	1.041
	(0.074)	(0.160)	(0.139)	(0.164)	(0.163)
Willingness to take risks	1.066^{***}	1.093^{***}	1.134^{***}	1.075^{***}	1.099^{***}
	(0.010)	(0.020)	(0.018)	(0.020)	(0.020)
Patience	0.992	0.981	0.979	0.983	0.982
	(0.008)	(0.017)	(0.015)	(0.018)	(0.018)
Number of observations	5110	5110		5110	
Country FE	✓	✓		✓	
Correctly predicted values (%)	72.5	52.1		62.3	
McFadden Pseudo R2	0.13	0.09		0.20	
Panel B: Student sample					
Migration aspiration				3.836^{***}	15.295^{***}
				(0.091)	(0.143)
Gender: woman	1.156^{***}	1.055	1.123	0.982	0.998
	(0.049)	(0.086)	(0.107)	(0.090)	(0.115)
Age: 25 to 34 years	0.979	1.195	0.994	1.215	1.030
	(0.070)	(0.122)	(0.151)	(0.127)	(0.164)
Age: 35 to 49 years	0.869	0.754	0.720	0.775	0.792
	(0.138)	(0.247)	(0.307)	(0.256)	(0.333)
University degree	0.934	1.331^{***}	1.175	1.436^{***}	1.337^{**}
	(0.052)	(0.092)	(0.114)	(0.096)	(0.124)
Migration experience	1.254^{***}	1.726^{***}	1.803^{***}	1.636^{***}	1.632^{***}
	(0.051)	(0.090)	(0.112)	(0.094)	(0.121)
Partner: native	0.846^{***}	0.877	1.005	0.956	1.147
	(0.051)	(0.091)	(0.114)	(0.095)	(0.124)
Partner: non-native	1.235^{*}	1.420	3.500^{***}	1.456	3.609^{***}
	(0.124)	(0.242)	(0.252)	(0.254)	(0.279)
Children	1.071	0.628^{*}	0.457^{**}	0.564^{**}	0.364^{***}
	(0.134)	(0.242)	(0.325)	(0.251)	(0.357)
Willingness to take risks	1.045^{***}	1.126^{***}	1.089^{***}	1.112^{***}	1.071^{**}
	(0.011)	(0.021)	(0.025)	(0.022)	(0.027)
Patience	0.995	1.008	1.022	1.015	1.030
	(0.010)	(0.017)	(0.021)	(0.018)	(0.023)
Number of observations	3242	3242		3242	
Country FE	✓	✓		✓	
Correctly predicted values (%)	65.5	52.1		57.2	
McFadden Pseudo R2	0.08	0.05		0.14	

Note: This table shows risk ratios with standard errors in parentheses. Specification (1) estimates the probability of aspirations to migrate permanently via binomial probit; the reference category is 'no aspirations to migrate permanently'. Specifications (2) and (4) estimate the probability of intentions to migrate temporarily, and specifications (3) and (5) estimate the probability of intentions to migrate permanently via multinomial probit; the reference category is 'no intentions to migrate'. See Table 4.3 for the number of observations for each response category. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

their partner (Geist and McManus 2011; Mincer 1978). As individuals' early career investments are made before their and their partners' job opportunities abroad are revealed, the prospect of becoming a "tied mover" can influence women's incentives to invest in their career, even beyond the group that actually migrates. As a consequence, especially highly educated women could be deterred from pursuing an ambitious career.

Further, we find a positive linkage between migration experience and intentions. This linkage remains robust after controlling for aspirations only among women (see Table 4.K in the Appendix). This gender difference could be rooted in the different roles family ties abroad play in migration decisions for men and women. Another possible explanation are differences in risk preferences between men and women. While an increase in willingness to take risks increases the likelihood of both aspirations and intentions for both genders, on average, in the GI sample, men have a higher willingness to take risks (with a mean of 0.656) than women (with a mean of 0.631). The same holds in the student sample, in which the average willingness to take risks is 0.537 for women and 0.568 for men.[27] Though those differences in means are not large, Welch two-sample t-tests reveal that both differences are strongly significant. 66.7 % of men in the GI sample intend to migrate either temporarily or permanently, and 65.5 % of women do. A simple OLS regression on this outcome shows that ceteris paribus the marginal effect of an increase of 0.025 in willingness to take risks (which is the difference between women's and men's average in the GI sample) can explain roughly 27 % of the gender difference in migration intentions. Based on a similar calculation, in the student sample the marginal effect of an increase of 0.031 in willingness to take risks accounts for roughly 18 % of the gender differences in intentions to migrate.

The wide age distribution of the GI sample allows for a closer look at how age is linked to aspirations and intentions. As expected, the likelihood of both aspirations and intentions to emigrate consistently decreases with age. In both samples, the decrease sets in from the age of 35 onwards. In the student sample age does not show much variation, with 76.8 % of respondents being between 18 and 24 years old.

Overall, the influence of individual characteristics on migration aspirations and intentions differs more strongly between genders than between our two samples. In an attempt to make the GI and student samples even more comparable, we next restrict the GI sample to the younger age groups (under 35 years of age) and to the 6 countries which are surveyed in the student sample. Results (see Table 4.M in the Appendix) are largely comparable between the GI sample restricted by age and countries, and the student sample.[28] The same holds for the GI sample restricted to the younger age groups but including all 14 countries (see Table 4.N in the Appendix).

[27] Since the scales on which willingness to take risks are measured differ between both samples (the scale ranges between 0 and 10 in the GI sample, and between 1 and 10 in the student sample; for details see the variables description in Table 4.B in the Appendix), we changed the scales such that they both range between 0 and 1.

[28] We do not restrict the student sample by age since only 5.1 % of the student sample are 35 years or older.

4.5.2 Intentions to Migrate Temporarily and Permanently

In line with Delogu et al. (2018), which shows with GWP data that individuals who intend to migrate temporarily are a considerable percentage of potential migrants, we show in Table 4.3 that 24.6 % (GI sample) to 34.1 % (student sample) of respondents intend to migrate temporarily. Hence, in Table 4.5 we compare how individual characteristics are linked to intentions to migrate temporarily versus permanently.

University graduates are 1.4 times as likely to intend to migrate temporarily as their less educated counterparts in both the GI and student sample (after controlling for aspirations), while no such positive linkage is found between degree and aspirations. This is in line with Docquier et al. (2014), who argue that college educated individuals do not necessarily show higher rates of aspirations to migrate, even though their actual emigration rates are much larger compared to those of the less educated. Such findings might be driven by the pull of a more international labor market and better professional opportunities abroad. It might as well be due to individuals seeking further education abroad. Indeed, Table 4.4 shows that the percentage of respondents who state educational reasons, such as studies abroad, or professional reasons as the main motivation for potential migration is consistently larger among those who intend to migrate temporarily compared to those who intend to migrate permanently, except for men in the student sample. Here, too, gender differences occur. The positive association between degree and intentions to migrate temporarily is largely driven by men in both samples (see Table 4.L in the Appendix).

Having a non-native partner significantly increases the likelihood of intentions to migrate permanently. Table 4.D shows that the percentage of respondents with a non-native partner among those with intentions to migrate permanently (9.3 %) is larger than among those with intentions to migrate temporarily (7.5 %) in the GI sample, as well as in the student sample (7.4 % compared to 4.2 %). Welch two-sample t-tests reveal that those differences are significant at the 10 % level. This result is driven by women in both samples, while for men the association is positive but not statistically significant (see Tables 4.K and 4.L in the Appendix).

In conclusion, temporary migration seems to be considered primarily for educational and professional reasons. Family ties, on the contrary, are more strongly linked to intentions to migrate permanently than temporarily. Though the percentage of respondents who state partner or other family as the main motivation for potential migration is small among both those intending to migrate temporarily and permanently, it is consistently larger among the latter.

4.5.3 Conditionality and Differences

To investigate which factors explain differences between aspirations and intentions (as shown in Table 4.3), we estimate the multinomial probit models in Equation (4.4), conditional on respondents' aspirations to migrate. Results are shown in Table 4.6.

We take a closer look at those respondents with no migration aspirations (see Columns (1) and (2) of Table 4.6), since they have been largely neglected by the literature so far

Table 4.6: *Intentions by Aspirations*

	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
Panel A: GI sample				
Gender: woman	0.964	0.711*	1.124	0.877
	(0.122)	(0.186)	(0.118)	(0.103)
Age: under 18 years	1.165	1.303	0.557***	0.941
	(0.261)	(0.413)	(0.223)	(0.189)
Age: 25 to 34 years	0.922	1.323	1.285	1.721***
	(0.205)	(0.338)	(0.191)	(0.171)
Age: 35 to 49 years	0.442***	1.388	0.648*	0.764
	(0.254)	(0.391)	(0.254)	(0.228)
Age: 50 to 64 years	0.163***	0.214**	0.195***	0.461**
	(0.372)	(0.653)	(0.409)	(0.327)
Age: 65 years or above	0.090***	0.171**	0.093***	0.157***
	(0.537)	(0.856)	(0.724)	(0.575)
University degree	1.805***	1.161	1.165	0.935
	(0.172)	(0.281)	(0.154)	(0.136)
Student	1.207	1.100	0.838	0.716**
	(0.192)	(0.324)	(0.165)	(0.145)
Migration experience	1.106	1.213	1.468***	1.449***
	(0.139)	(0.214)	(0.137)	(0.123)
Partner: native	0.840	0.932	1.020	1.204
	(0.160)	(0.251)	(0.165)	(0.149)
Partner: non-native	1.154	1.662	1.069	2.057***
	(0.247)	(0.368)	(0.282)	(0.251)
Children	0.864	1.006	0.834	1.074
	(0.225)	(0.304)	(0.263)	(0.225)
Willingness to take risks	1.045	1.092*	1.094***	1.110***
	(0.030)	(0.047)	(0.028)	(0.024)
Patience	0.985	0.954	0.983	0.985
	(0.026)	(0.041)	(0.024)	(0.021)
Number of observations	1743		3367	
Country FE	✓		✓	
Correctly predicted values (%)	64.9		61.3	
McFadden Pseudo R2	0.08		0.08	
Panel B: Student sample				
Gender: woman	0.909	1.048	1.075	1.016
	(0.129)	(0.268)	(0.129)	(0.138)
Age: 25 to 34 years	1.093	0.909	1.430*	1.173
	(0.179)	(0.379)	(0.193)	(0.206)
Age: 35 to 49 years	0.471*	1.448	1.181	0.844
	(0.421)	(0.626)	(0.362)	(0.413)
University degree	1.246	1.451	1.607***	1.368**
	(0.135)	(0.276)	(0.141)	(0.150)
Migration experience	1.945***	0.822	1.449***	1.733***
	(0.134)	(0.310)	(0.136)	(0.146)
Partner: native	0.986	1.671*	0.910	1.028
	(0.134)	(0.284)	(0.139)	(0.149)
Partner: non-native	1.177	4.647***	1.892	4.256***
	(0.354)	(0.561)	(0.418)	(0.411)
Children	0.882	0.724	0.335***	0.231***
	(0.364)	(0.735)	(0.357)	(0.432)
Willingness to take risks	1.130***	0.931	1.098***	1.093***
	(0.031)	(0.060)	(0.031)	(0.033)
Patience	1.002	0.907*	1.037	1.068**
	(0.026)	(0.053)	(0.026)	(0.027)
Number of observations	1560		1682	
Country FE	✓		✓	
Correctly predicted values (%)	69.2		46.2	
McFadden Pseudo R2	0.06		0.05	

Note: This table shows risk ratios with standard errors in parentheses. Specifications (1) and (2) estimate the probability of having intentions to migrate temporarily or permanently for those with no aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Specifications (3) and (4) estimate the probability of having intentions to migrate temporarily or permanently for those with aspirations to migrate permanently; the reference category is 'no intentions to migrate'. See Table 4.3 for the number of observations for each response category. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

due to the conditionality in responses. As discussed before, highly educated individuals are likely to form intentions to migrate temporarily for career reasons. Consistently, respondents with a university degree are 1.8 times (GI sample) as likely to intend to migrate temporarily despite having no aspirations to migrate permanently as those without a university degree. In the student sample, this association is only significant for men underlying again that educational and career reasons seem influential on the migration choice primarily for men. The linkage between university degree and intentions to migrate temporarily is strongest among the younger age groups (see Tables 4.O and 4.P in the Appendix, where the GI sample is split by age and respondents under the age of 35 years and aged 35 or older are analyzed separately. Table 4.O analyzes the six countries, which are observed in both samples, and Table 4.P all countries in the GI sample).

Having a non-native partner increases the likelihood of intentions to migrate permanently among those who do not aspire to migrate permanently, though the result is significant only in the student sample. Subsampling by gender reveals that this result is driven by women. Generally, women are often found to support and follow their partner as “tied movers” (Geist and McManus 2011; Mincer 1978), and our results suggest that they might do so even when they do not aspire to migrate permanently.

In the student sample respondents with children are only 0.3 times as likely to intend to migrate temporarily and 0.2 times as likely to intend to migrate permanently as those without children, despite having migration aspirations. Again, this is driven by women in the sample (see Table 4.Q in the Appendix). Strong ties to their country of origin can therefore increase migration costs and be a factor which impedes aspirations to migrate from being fulfilled, especially for women.

For men, on the other hand, we find education and career to be the most influential in their migration decision and to explain differences between aspirations and intentions. Among men in both samples, holding a university degree is positively associated with intentions to migrate temporarily even when they do not aspire to migrate (see Table 4.R in the Appendix). In the GI sample, where we can differentiate between students and non-students, men who are studying are also less likely to intend to migrate permanently, despite aspirations to do so. These findings, too, go hand in hand with existing literature describing professional reasons as the main driver of migration decisions for men (Geist and McManus 2011; McKinnish 2008; Munk et al. 2022). Lastly, as discussed before, in the GI sample older age groups (35 years and above) show a much lower likelihood of intentions to migrate, and this holds even for those with migration aspirations. This is most robust among respondents from high-income (EU) countries (see Table 4.S and Table 4.T in the Appendix). In the student sample, interpretation of the results is not meaningful due to low variation in age among students.[29]

[29] In addition to the baseline results, we estimated a multinomial probit model with all 6 outcomes separately. Overall, results are comparable, but since the reference category here is always those with neither aspirations nor intentions, the model fit is inferior to our sample split in Table 4.6 and interpretation of results is not straightforward. Results are available from the authors upon request.

4.6 Conclusion

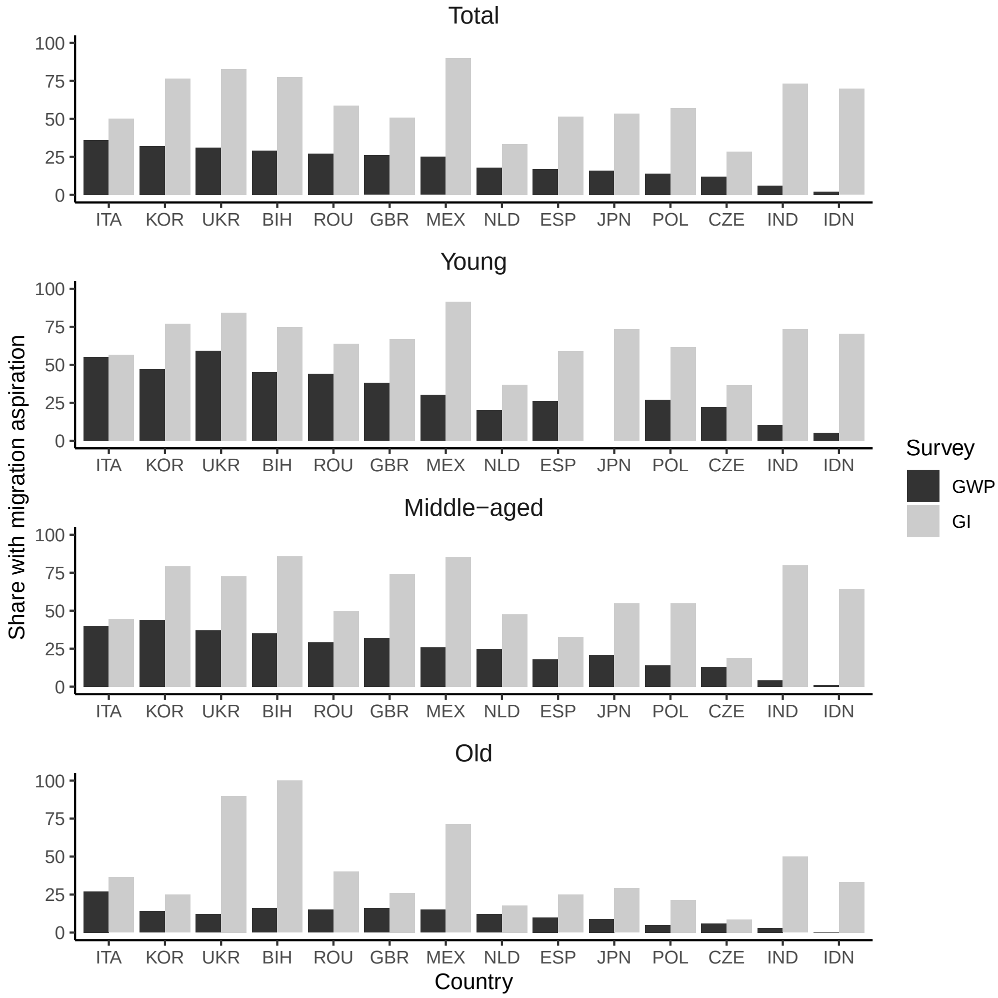
We conducted two multinational surveys – one among language course participants in 14 countries and one among university students in six countries – and use multinomial probit estimations to analyze individual level determinants of migration aspirations and intentions. Our analysis shows that migration considerations and plans often coincide with desires to migrate under ideal circumstances but there is a considerable share of respondents whose migration intentions differ from their migration aspirations. While the desire to leave one's country permanently might originate in pull factors abroad, push factors in the country of origin may result in considerations and plans to emigrate even if one would ideally like to stay. A better understanding of why aspirations and intentions sometimes differ could reveal such push and pull factors and thus be of great value for the design of targeted policy interventions.

To date, the GWP is the only globally representative survey available on migration aspirations and intentions, yet it suffers from the conditionality in responses and lacks a clear differentiation of potential future migration into temporary and permanent moves. Resolving those two limitations allows us to observe migration patterns which have received little attention in the literature on potential future migrants so far, and which play an essential role for migration research. Hence, while our data might be limited in its representativeness, our analysis provides a first step in this direction.

By introducing a clear time horizon to our measure of intentions, we highlight differences between those individuals who intend to migrate temporarily and those who intend to migrate permanently. Our results suggest that a temporary move abroad is considered or planned primarily for educational and career reasons, while family ties are of comparably larger importance regarding intentions to migrate permanently. By avoiding the conditionality in responses, we identify individual level determinants which explain differences between aspirations and intentions, and find them to vary between genders. The percentage of respondents who state educational or professional reasons as their main motivation is consistently larger among men, and the percentage of those who state partner or other family as their main motivation is consistently larger among women.

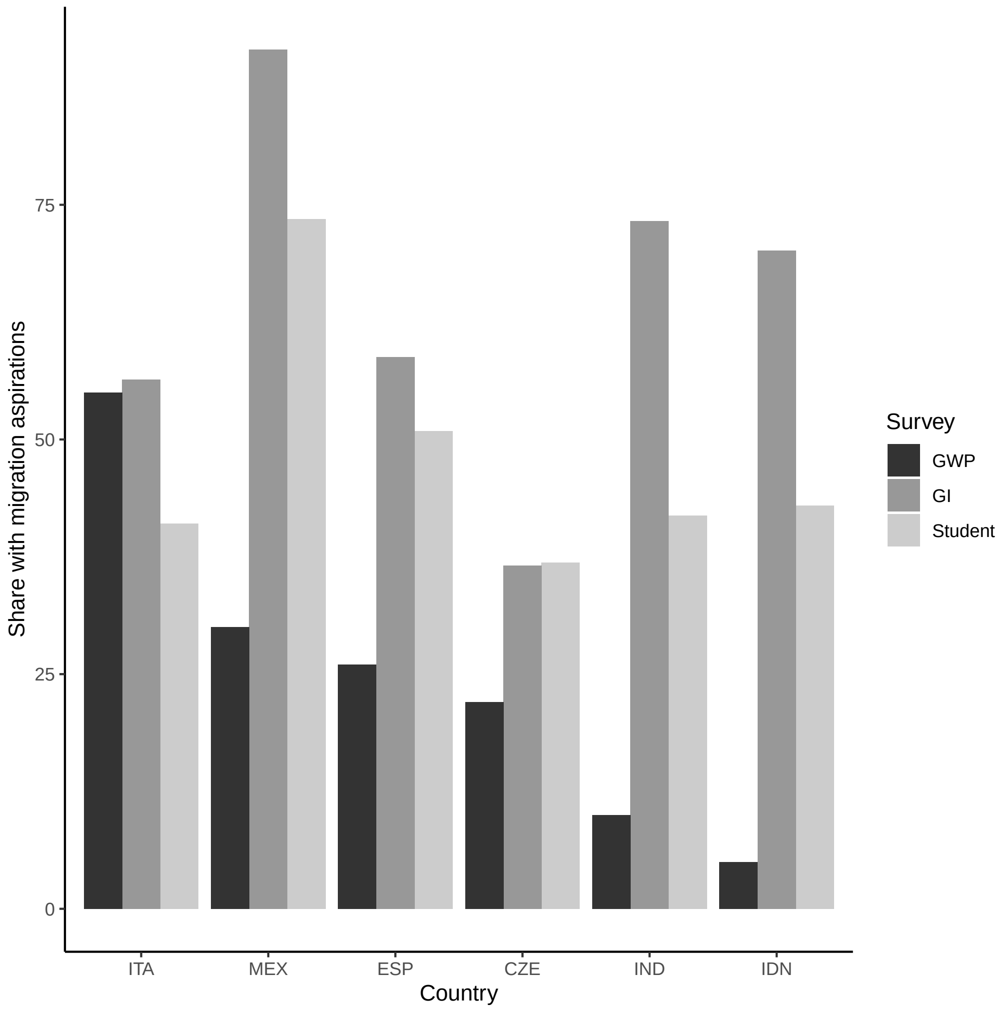
4.A Appendix: Descriptive Statistics and Variable Descriptions

Figure 4.A: *Percentage of Respondents with Migration Aspirations by Age Groups — GWP and GI Sample*



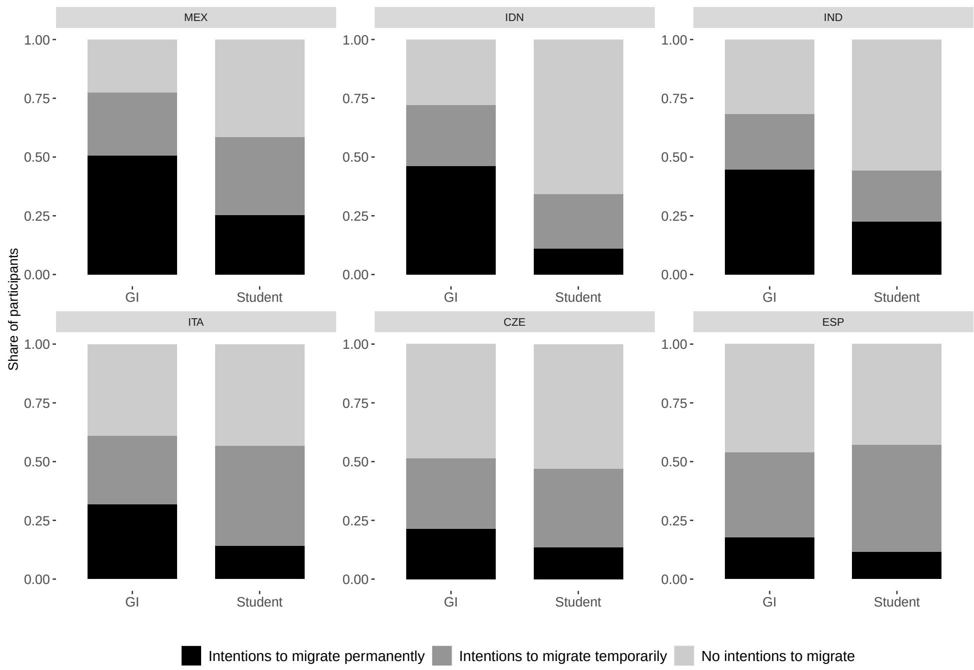
Note: This figure compares the percentages of respondents with migration aspirations in the GWP data and the GI sample by age group. Age groups do not perfectly match across surveys due to discrepancies in the categories and are therefore defined as follows: Younger (less than 35 (GI sample); less than 30 (GWP Data)), middle aged (35-49 (GI sample); 30-49 (GWP)) and older (50 or more (GI sample and GWP)). Data from the GWP refers to the 2018 wave, apart from data for Spain, Italy, the Netherlands, and the United Kingdom that refers to the 2017 wave, as data for 2018 was not available. Entry for Japanese of young age group (less than 30 years) is missing in the GWP data.

Figure 4.B: *Percentage of Respondents with Migration Aspirations — GWP, GI Sample and Student Sample, Younger Age Groups*



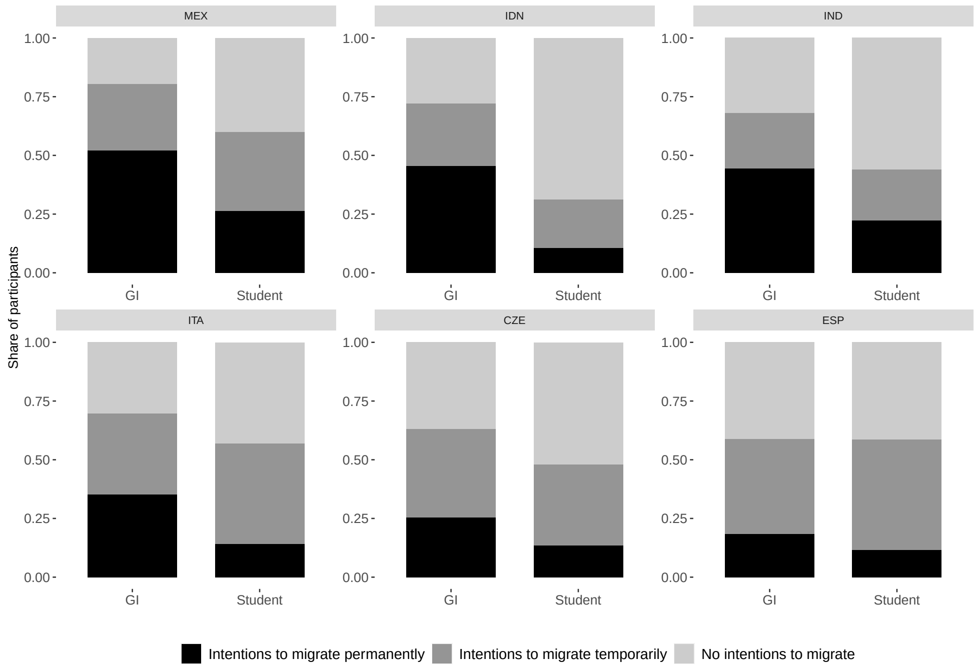
Note: This figure compares the percentages of respondents with migration aspirations among the younger age groups in the GWP data with the GI and the student samples for the 6 countries which are covered in both surveys. Age groups do not perfectly match across surveys due to discrepancies in the categories and are therefore defined as follows: Younger than 35 (GI and student sample) or younger than 30 (GWP Data). Data from the GWP refers to the 2018 wave, apart from data for Spain and Italy that refers to the 2017 wave, as data for 2018 was not available.

Figure 4.C: *Share of Respondents with Migration Intentions — GI and Student Sample*



Note: This figure compares the shares of migration intentions between the GI and the student samples (including all age groups), for the 6 countries which are covered in both surveys.

Figure 4.D: *Share of Respondents with Migration Intentions — GI and Student Sample, Younger Age Groups*



Note: This figure compares the shares of migration intentions between the GI and the student samples, both restricted to the younger age groups (under 35 years of age), for the 6 countries which are covered in both surveys. For the equivalent figure for the GI and the student samples including all age groups, see Figure 4.C in the Appendix.

Table 4.A: *Variables Description*

Variable	Type	Description
Migration aspirations	Binary	Indicates respondent's aspirations to migrate permanently: *"Ideally, if you had the opportunity, would you like to move permanently to another country or would you prefer to continue living in [country in which the survey took place]?"* - *"Like to permanently move to another country"* (migration aspirations) - *"Like to stay in [country in which the survey took place]."* (no migration aspirations) Reference category is 'no migration aspirations'.
Migration intentions	Categorical	Indicates respondent's intentions to migrate: *"Tick the statement that applies to your current situation"* - *"I would not move to another country under any circumstances"* (no intentions to migrate) - *"In principle, I would move to another country, but I have not thought about it in the last 12 months"* (no intentions to migrate) - *"I have been thinking about moving to another country in the last 12 months, but have no specific plans."* (intentions to migrate) - *"I am planning a move to another country."* (intentions to migrate) - *"I already have a date for my planned move to another country."* (intentions to migrate) Those respondents with migration intentions are further asked for their preferred destination country and their preferred length of stay (*"How long would you most likely stay in [preferred destination country]"?*) and likelihood of return (*"How likely is it that you will return to [country in which the survey took place] after a temporary stay in [preferred destination country]?"*). Those who state that they would most likely stay in their preferred destination country for more than 5 years or state that their return to [country in which the survey took place] after a temporary stay in their preferred destination country is unlikely are classified as having permanent migration intentions; the rest is classified as having temporary migration intentions. Those who state no migration intentions are not asked those questions. Reference category is 'no migration intentions'.
Gender	Categorical	Indicates respondent's gender. Takes a value of 0 if respondent indicated to be a man, a value of 1 if respondent indicated to be a woman, and a value of 2 if respondent indicated *"No answer/prefer not to say"* or if response is missing. The last category is not reported in the result tables.
Age	Categorical	Indicates respondent's age group according to the ranges: *under 18, 18 to 24, 25 to 34, 35 to 49, 50 to 64, 65 or above*. Reference category is 18 to 24 years for whole samples and samples restricted to younger age groups and 35 to 49 years for samples restricted to older age groups.
University degree	Binary	Indicates whether respondent has a university degree. Reference category is 'no university degree'.
Student	Binary	Indicates whether respondent is a student. Reference category is 'not a student'.
Migration experience	Binary	Indicates whether respondent has stayed abroad for at least three consecutive months in the past. Reference category is 'no migration experience'.

Table 4.B: *Variables Description (Continued)*

Variable	Type	Description
Partner: native	Binary	Indicates whether respondent is in a long-term relationship with or married to a partner whose native language is an official or officially recognized (minority) language in the country in which the survey took place. Reference category is 'no native partner'.
Partner: non-native	Binary	Indicates whether respondent is in a long-term relationship with or married to a partner whose native language is different from the official or officially recognized (minority) language(s) in the country in which the survey took place. Reference category is 'no non-native partner'.
Children	Binary	Indicates whether respondent has any children. Reference category is 'no children'. Missing responses are interpreted as 'no children'.
Willingness to take risks	Numerical (0-10 / 1-10)	Measures respondent's willingness to take risks (*"Would you describe yourself as someone who tries to avoid risks (risk-averse) or as someone who is willing to take risks (risk-prone)?"*) on a 11-point scale from 0 for *"risk-averse"* to 10 for *"risk-prone"* in the GI sample; and on a 10-point scale from 1 for *"risk-averse"* to 10 for *"risk-prone"* in the student sample.
Patience	Numerical (0-10 / 1-10)	Measures respondent's self-reported patience (*"Would you describe yourself as an impatient or a patient person in general?"*) on a 11-point scale from 0 for *"very impatient"* to 10 for *"very patient"* in the GI sample; and on a 10-point scale from 1 for *"very impatient"* to 10 for *"very patient"* in the student sample.
Main motivation for potential migration	Categorical	Indicates respondent's main reason for a potential move to their preferred destination country. Educational reasons include study/education/PhD. Professional reasons include work experience/(unpaid) traineeship, own higher income, more interesting job, poor job prospects in origin country, transfer by employer and other own professional reasons. Reasons related to partner or other family include professional reasons/studies of partner, partner lives in the destination country, other family/partner related reasons, friends/relatives live in destination country (South Korea only). Other reasons include interest in the country and culture, adventure, environmental reasons, higher quality of life, and all other reasons. Those who state no migration intentions are not asked this question.

Table 4.C: *Joint Distribution of Aspirations and Intentions — Foreign-Born Respondents Only*

Panel A: GI sample	No intentions to migrate		Intentions to migrate temporarily		Intentions to migrate permanently		*Total*	
No aspirations to migrate permanently	121	(19.5)	37	(6.0)	27	(4.4)	185	(29.8)
Aspirations to migrate permanently	52	(8.4)	99	(15.9)	284	(45.8)	435	(70.2)
Total	173	(27.9)	136	(21.9)	311	(50.2)	620	(100.0)
Panel B: Student sample								
No aspirations to migrate permanently	104	(21.9)	34	(7.2)	16	(3.4)	154	(32.5)
Aspirations to migrate permanently	95	(20.0)	92	(19.4)	133	(28.1)	320	(67.5)
Total	199	(41.9)	126	(26.6)	149	(31.5)	474	(100.0)

Note: This table shows the numbers of observation with percentages in parentheses. Row and column *Total(s)* show row and column totals; percentages of total sample size in parentheses.

Table 4.D: *Descriptive Statistics*

Panel A: GI sample	No intentions to migrate	Intentions to migrate temporarily	Intentions to migrate permanently	*Total*
Gender: woman	58.6	59.5	55.6	57.6
Gender: man	38.9	37.6	41.3	39.5
Age: under 18 years	11.0	11.2	17.4	13.7
Age: 18 to 24 years	36.1	43.9	37.8	38.7
Age: 25 to 34 years	19.6	29.7	29.0	26.0
Age: 35 to 49 years	17.2	12.3	12.8	14.2
Age: 50 to 64 years	10.2	2.3	2.6	5.1
Age: 65 years or above	6.0	0.6	0.4	2.4
University degree	62.5	63.5	56.4	60.2
Student	31.1	36.3	31.9	32.7
Migration experience	33.4	36.8	31.2	33.3
Partner: native	33.8	25.6	24.9	28.1
Partner: non-native	6.6	7.5	9.3	8.0
Children	19.9	8.1	11.2	13.4
Willingness to take risks (0=risk averse ... 10=risk prone)	6.0	6.5	6.7	6.4
Patience (0=very impatient ... 10=very patient)	6.2	6.1	6.3	6.2
Number of observations	1729	1256	2125	5110

Panel B: Student sample				
Gender: woman	52.8	56.9	53.3	54.3
Gender: man	46.9	42.6	46.3	45.3
Age: under 18 years	0.3	0.2	0.2	0.2
Age: 18 to 24 years	77.4	75.9	76.7	76.8
Age: 25 to 34 years	16.0	20.0	19.0	17.9
Age: 35 to 49 years	5.0	3.5	3.6	4.3
Age: 50 to 64 years	1.3	0.4	0.5	0.8
Age: 65 years or above	0.1	0.0	0.0	0.0
University degree	45.6	54.1	50.4	49.4
Migration experience	26.2	43.5	40.1	34.6
Partner: native	39.0	35.3	32.8	36.6
Partner: non-native	2.2	4.2	7.4	3.8
Children	6.6	3.5	2.9	4.9
Willingness to take risks (1=risk averse ... 10=risk prone)	5.7	6.2	6.2	6.0
Patience (1=very impatient ... 10=very patient)	5.9	5.8	5.9	5.9
Number of observations	1554	1105	583	3242

Note: This table shows the percentages of observations; except for willingness to take risks and patience which show means.

Table 4.E: *Answered the Question on Migration Aspirations: Selection on Observable Characteristics — GI and Student Sample*

Panel A: GI sample	Answered question on migration aspirations (1)
Age: 25 years or above	0.804^{***}
	(0.082)
Gender: woman	0.886^{*}
	(0.069)
Gender: n.a.	0.777
	(0.475)
Partner: in relationship or married	0.907
	(0.077)
Willingness to take risks	1.012
	(0.016)
Patience	1.005
	(0.014)
Number of observations	6327
Country FE	✓
McFadden Pseudo R2	0.15

Panel B: Student sample	
Age: 25 years or above	1.009
	(0.057)
Gender: woman	0.986
	(0.045)
Gender: n.a.	1.103
	(0.351)
Partner: in relationship or married	1.048
	(0.046)
Willingness to take risks	1.029
	(0.089)
Patience	1.019^{**}
	(0.009)
Number of observations	5176
University FE	✓
McFadden Pseudo R2	0.27

Note: This table shows risk ratios with standard errors in parentheses. Specification (1) estimates the probability of having answered the question on migration aspirations via binomial probit. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.F: *Joint Distribution of Aspirations and Intentions by Gender*

	No intentions to migrate		Intentions to migrate temporarily		Intentions to migrate permanently		*Total*	
Panel A: GI sample, women								
No aspirations to migrate permanently	651	(22.1)	254	(9.0)	75	(2.6)	990	(33.7)
Aspirations to migrate permanently	363	(12.3)	483	(16.4)	1107	(37.6)	1953	(66.3)
Total	1014	(34.4)	747	(25.4)	1182	(40.2)	2943	(100.0)
Panel B: GI sample, men								
No aspirations to migrate permanently	447	(22.1)	187	(9.3)	71	(3.5)	705	(34.9)
Aspirations to migrate permanently	225	(11.1)	285	(14.1)	806	(39.9)	1316	(65.1)
Total	672	(33.2)	472	(23.4)	877	(43.4)	2021	(100.0)
Panel C: Student sample, women								
No aspirations to migrate permanently	559	(31.7)	217	(12.3)	40	(2.3)	816	(46.3)
Aspirations to migrate permanently	262	(14.9)	412	(23.4)	271	(15.4)	945	(53.7)
Total	821	(46.6)	629	(35.7)	311	(17.7)	1761	(100.0)
Panel D: Student sample, men								
No aspirations to migrate permanently	509	(34.6)	197	(13.4)	34	(2.3)	740	(50.3)
Aspirations to migrate permanently	220	(15.0)	274	(18.6)	236	(16.1)	730	(49.7)
Total	729	(49.6)	471	(32.0)	270	(18.4)	1470	(100.0)

Note: This table shows the numbers of observation with percentages in parentheses. Row and column *Total(s)* show row and column totals; percentages of total sample size in parentheses.

Table 4.G: *Preferred Destination Countries for Potential Migration by Origin Country*

Origin country	Destination country				
	Country 1	Country 2	Country 3	Country 4	Country 5
Panel A: GI sample					
Bosnia	Germany (79.6)	Austria (12.0)	United Kingdom (2.1)	Switzerland (1.6)	Italy (1.0)
Czechia	Germany (60.2)	Austria (10.5)	United Kingdom (7.4)	Switzerland (5.9)	United States (3.5)
United Kingdom	Germany (65.1)	France (8.9)	Switzerland (6.8)	United States (5.5)	Austria (4.1)
India	Germany (84.1)	United States (4.6)	Canada (4.1)	United Kingdom (2.0)	Switzerland (1.9)
Indonesia	Germany (83.5)	United States (3.3)	United Kingdom (3.0)	Singapore (2.0)	Austria (1.4)
Italy	Germany (46.1)	Switzerland (13.7)	United Kingdom (12.9)	United States (8.3)	France (5.4)
Japan	Germany (80.0)	Austria (9.0)	Switzerland (3.4)	United States (2.1)	United Kingdom (1.4)
Mexico	Germany (68.5)	United States (11.4)	Canada (6.1)	Switzerland (2.5)	France (2.3)
Netherlands	Germany (58.1)	Austria (8.1)	United Kingdom (6.5)	Switzerland (4.8)	United States (4.8)
Poland	Germany (58.9)	United Kingdom (11.7)	Switzerland (8.0)	United States (4.3)	Austria (3.1)
Romania	Germany (52.1)	Austria (17.5)	Switzerland (7.8)	United Kingdom (6.0)	Netherlands (3.7)
South Korea	Germany (85.7)	United States (6.0)	United Kingdom (2.2)	Canada (1.9)	Austria (0.8)
Spain	Germany (58.7)	United Kingdom (14.2)	United States (9.2)	Switzerland (4.5)	France (4.2)
Ukraine	Germany (73.7)	Austria (11.6)	United States (4.9)	Switzerland (2.1)	United Kingdom (1.2)
EU/EEA countries	Germany (56.5)	United Kingdom (9.6)	Switzerland (7.4)	Austria (6.7)	United States (5.9)
Non-EU/EEA countries	Germany (79.5)	United States (5.1)	Austria (4.0)	Canada (2.4)	United Kingdom (2.1)
Panel B: Student sample					
Czechia	United States (19.3)	United Kingdom (15.9)	Germany (14.8)	France (6.4)	Canada (6.0)
India	United States (21.8)	United Kingdom (13.7)	Germany (12.6)	Canada (7.2)	France (7.2)
Indonesia	United States (25.2)	United Kingdom (16.5)	Germany (13.0)	Spain (8.7)	France (6.1)
Italy	United States (30.8)	United Kingdom (15.7)	Germany (15.1)	Canada (6.4)	France (4.7)
Mexico	United States (23.0)	United Kingdom (15.8)	Germany (10.6)	Canada (5.5)	Spain (4.8)
Spain	United States (24.4)	United Kingdom (12.8)	Germany (11.9)	Canada (6.2)	France (4.6)
EU/EEA countries	United States (23.0)	United Kingdom (14.7)	Germany (13.7)	Canada (6.1)	France (5.4)
Non-EU/EEA countries	United States (22.9)	United Kingdom (15.3)	Germany (11.3)	Canada (5.9)	France (5.4)

Note: This table shows the top 5 preferred destination countries for potential migration by origin country. Percentages of respondents per origin country in parentheses. Numbers of observation differ from sample totals since respondents who stated *"I would not move to another country under any circumstances"* did not answer the question on their preferred destination and not all of those who intend to migrate indicated their preferred destination either. We observe the preferred destination country for 4355 respondents (GI sample) and 2338 respondents (student sample).

Table 4.H: *Main Motivations for Potential Migration by Aspirations and Intentions*

	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
Panel A: GI sample, women				
	(n = 167)	*(n = 57)*	*(n = 319)*	*(n = 705)*
Educational reasons	32.3	19.3	48.9	39.4
Professional reasons	29.3	31.6	24.8	22.2
Partner or other family	12.6	22.8	7.8	14.6
Other reasons	25.8	26.3	18.5	23.8
Panel B: GI sample, men				
	(n = 116)	*(n = 49)*	*(n = 189)*	*(n = 473)*
Educational reasons	44.0	20.4	36.0	36.4
Professional reasons	37.9	38.8	32.2	27.9
Partner or other family	2.6	14.3	5.3	8.9
Other reasons	15.5	26.5	26.5	26.8
Panel C: Student sample, women				
	(n = 217)	*(n = 40)*	*(n = 412)*	*(n = 271)*
Educational reasons	37.3	25.0	37.4	28.4
Professional reasons	29.5	37.5	29.6	25.5
Partner or other family	6.0	17.5	5.1	9.2
Other reasons	27.2	20.0	27.9	36.9
Panel D: Student sample, men				
	(n = 195)	*(n = 34)*	*(n = 274)*	*(n = 236)*
Educational reasons	44.6	44.1	43.1	35.2
Professional reasons	30.3	41.2	25.6	26.3
Partner or other family	4.6	2.9	3.6	4.7
Other reasons	20.5	11.8	27.7	33.9

Note: Each panel of this table shows column percentages, n show column totals. Numbers of observation differ from sample totals since only respondents who stated intentions to migrate answered the question on main motivations for potential migration and not all of those who intend to migrate indicated such main reason. Hence, in the GI sample we observe main reasons for potential migration for 1248 of 1929 women with intentions to migrate, and for 827 of 1349 men with intentions to migrate. In the student sample we observe main reasons for potential migration for 940 of 940 women with intentions to migrate, and for 739 of 741 men with intentions to migrate.

4.B Appendix: Further Estimations and Robustness Checks

Table 4.I: *Aspirations and Intentions — High-Income Countries*

	Binomial probit	Multinomial probit		Multinomial probit	
	(1) Aspirations to migrate permanently	(2) Intentions to migrate temporarily	(3) Intentions to migrate permanently	(4) Intentions to migrate temporarily	(5) Intentions to migrate permanently
Panel A: GI sample					
Migration aspiration				3.538***	20.390***
				(0.121)	(0.148)
Gender: woman	1.001	0.989	0.692***	0.935	0.616***
	(0.056)	(0.113)	(0.110)	(0.117)	(0.126)
Age: under 18 years	1.231	0.497**	0.878	0.419***	0.651
	(0.149)	(0.295)	(0.270)	(0.307)	(0.304)
Age: 25 to 34 years	0.959	0.931	1.323	0.980	1.520**
	(0.095)	(0.185)	(0.183)	(0.192)	(0.208)
Age: 35 to 49 years	0.746***	0.422***	0.638**	0.467***	0.805
	(0.111)	(0.216)	(0.219)	(0.224)	(0.250)
Age: 50 to 64 years	0.472***	0.115***	0.205***	0.141***	0.312***
	(0.140)	(0.300)	(0.295)	(0.308)	(0.339)
University degree	0.905	1.186	0.906	1.251	0.962
	(0.081)	(0.156)	(0.159)	(0.162)	(0.179)
Student	0.979	0.745*	0.841	0.740*	0.817
	(0.090)	(0.175)	(0.171)	(0.183)	(0.194)
Migration experience	1.011	1.139	1.150	1.162	1.187
	(0.058)	(0.114)	(0.113)	(0.118)	(0.129)
Partner: native	1.002	0.908	0.991	0.899	0.989
	(0.069)	(0.135)	(0.138)	(0.139)	(0.159)
Partner: non-native	0.998	1.237	1.806***	1.341	2.248***
	(0.108)	(0.214)	(0.215)	(0.223)	(0.246)
Children	0.958	0.865	1.020	0.860	1.023
	(0.091)	(0.190)	(0.182)	(0.195)	(0.211)
Willingness to take risks	1.107***	1.149***	1.200***	1.116***	1.131***
	(0.013)	(0.027)	(0.026)	(0.028)	(0.029)
Patience	0.991	0.956**	0.977	0.956*	0.978
	(0.012)	(0.023)	(0.023)	(0.024)	(0.026)
Number of observations	2433	2433		2433	
Country FE	✓	✓		✓	
Correctly predicted values (%)	67.4	53.0		61.7	
McFadden Pseudo R2	0.12	0.11		0.22	
Panel B: Student sample					
Migration aspiration				4.151***	15.348***
				(0.127)	(0.206)
Gender: woman	1.126*	1.081	1.112	1.018	1.012
	(0.069)	(0.120)	(0.174)	(0.126)	(0.188)
Age: 25 to 34 years	0.940	1.009	0.842	1.030	0.883
	(0.105)	(0.180)	(0.259)	(0.187)	(0.282)
Age: 35 to 49 years	1.024	0.194***	0.719	0.172***	0.662
	(0.243)	(0.550)	(0.616)	(0.570)	(0.683)
University degree	0.838**	1.220	1.013	1.386**	1.251
	(0.071)	(0.123)	(0.179)	(0.130)	(0.194)
Migration experience	1.378***	2.019***	2.386***	1.856***	2.027***
	(0.070)	(0.121)	(0.174)	(0.127)	(0.189)
Partner: native	0.751***	0.998	0.935	1.165	1.247
	(0.070)	(0.121)	(0.178)	(0.128)	(0.193)
Partner: non-native	1.309*	1.624	3.704***	1.602	3.594***
	(0.152)	(0.298)	(0.329)	(0.316)	(0.367)
Children	0.949	0.755	0.323*	0.734	0.237*
	(0.229)	(0.419)	(0.669)	(0.436)	(0.755)
Willingness to take risks	1.091***	1.175***	1.111**	1.136***	1.043
	(0.017)	(0.029)	(0.042)	(0.031)	(0.045)
Patience	1.007	1.008	1.053	1.011	1.048
	(0.014)	(0.025)	(0.036)	(0.026)	(0.038)
Number of observations	1594	1594		1594	
Country FE	✓	✓		✓	
Correctly predicted values (%)	63.9	54.5		60.5	
McFadden Pseudo R2	0.06	0.06		0.14	

Note: This table shows risk ratios with standard errors in parentheses. Specification (1) estimates the probability of aspirations to migrate permanently via binomial probit; the reference category is 'no aspirations to migrate permanently'. Specifications (2) and (4) estimate the probability of intentions to migrate temporarily, and specifications (3) and (5) the probability of intentions to migrate permanently via multinomial probit; the reference category is 'no intentions to migrate'. High-income countries include countries which have a GNI per capita larger than \$12,535 in current US-Dollars, as of 2020 (Czechia, Italy, Japan, Netherlands, Poland, Romania, South Korea, Spain, and the United Kingdom). All EU member states in the sample are high-income countries. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.J: *Aspirations and Intentions — Middle-Income Countries*

Panel A: GI sample

	Binomial probit	*Multinomial probit*		*Multinomial probit*	
	(1) Aspirations to migrate permanently	(2) Intentions to migrate temporarily	(3) Intentions to migrate permanently	(4) Intentions to migrate temporarily	(5) Intentions to migrate permanently
Migration aspiration				2.513^{***}	26.610^{***}
				(0.123)	(0.159)
Gender: woman	1.115^{*}	1.143	1.116	1.109	1.025
	(0.058)	(0.119)	(0.101)	(0.122)	(0.114)
Age: under 18 years	1.277^{**}	1.077	1.570^{***}	1.042	1.415^{*}
	(0.099)	(0.205)	(0.174)	(0.210)	(0.198)
Age: 25 to 34 years	1.146	1.070	1.398^{**}	1.041	1.326
	(0.096)	(0.197)	(0.171)	(0.205)	(0.194)
Age: 35 to 49 years	1.060	0.647	0.893	0.625	0.803
	(0.142)	(0.307)	(0.245)	(0.311)	(0.280)
Age: 50 to 64 years	0.994	0.158^{**}	0.790	0.158^{**}	0.740
	(0.256)	(0.800)	(0.404)	(0.806)	(0.486)
University degree	0.807^{***}	1.469^{**}	0.930	1.627^{***}	1.145
	(0.075)	(0.154)	(0.131)	(0.159)	(0.151)
Student	0.919	1.306	0.836	1.344^{*}	0.865
	(0.081)	(0.166)	(0.143)	(0.172)	(0.163)
Migration experience	1.104	1.428^{**}	1.491^{***}	1.456^{**}	1.510^{***}
	(0.077)	(0.162)	(0.138)	(0.167)	(0.156)
Partner: native	1.075	0.980	1.200	0.963	1.189
	(0.088)	(0.184)	(0.155)	(0.190)	(0.175)
Partner: non-native	1.232	0.783	1.590^{*}	0.749	1.486
	(0.132)	(0.300)	(0.237)	(0.309)	(0.271)
Children	0.764^{**}	0.708	0.839	0.782	1.130
	(0.134)	(0.312)	(0.234)	(0.314)	(0.271)
Willingness to take risks	1.019	1.026	1.076^{***}	1.028	1.077^{***}
	(0.015)	(0.030)	(0.025)	(0.030)	(0.028)
Patience	0.996	1.016	0.991	1.018	0.993
	(0.012)	(0.026)	(0.022)	(0.026)	(0.024)
Number of observations	2677	2677		2677	
Country FE	✓	✓		✓	
Correctly predicted values (%)	77.3	52.0		62.9	
McFadden Pseudo R2	0.04	0.04		0.16	

Panel B: Student sample

	(1)	(2)	(3)	(4)	(5)
Migration aspiration				3.541^{***}	15.622^{***}
				(0.133)	(0.201)
Gender: woman	1.195^{***}	1.012	1.125	0.929	0.972
	(0.069)	(0.126)	(0.136)	(0.131)	(0.148)
Age: 25 to 34 years	0.937	1.461^{**}	1.066	1.537^{**}	1.165
	(0.098)	(0.175)	(0.192)	(0.183)	(0.208)
Age: 35 to 49 years	0.774	1.358	0.812	1.525	1.006
	(0.172)	(0.305)	(0.364)	(0.316)	(0.396)
University degree	1.051	1.338^{**}	1.251	1.351^{**}	1.311^{*}
	(0.078)	(0.144)	(0.151)	(0.150)	(0.165)
Migration experience	1.083	1.389^{**}	1.411^{**}	1.388^{**}	1.382^{**}
	(0.078)	(0.140)	(0.151)	(0.146)	(0.164)
Partner: native	0.991	0.737^{**}	1.046	0.737^{**}	1.034
	(0.078)	(0.143)	(0.152)	(0.148)	(0.165)
Partner: non-native	1.048	1.112	3.050^{***}	1.224	3.693^{***}
	(0.214)	(0.434)	(0.399)	(0.448)	(0.443)
Children	1.085	0.618	0.503^{*}	0.553^{*}	0.418^{**}
	(0.168)	(0.303)	(0.376)	(0.314)	(0.410)
Willingness to take risks	1.003	1.063^{**}	1.065^{**}	1.071^{**}	1.079^{**}
	(0.016)	(0.030)	(0.032)	(0.031)	(0.035)
Patience	0.987	1.007	1.006	1.015	1.017
	(0.014)	(0.025)	(0.027)	(0.026)	(0.029)
Number of observations	1648	1648		1648	
Country FE	✓	✓		✓	
Correctly predicted values (%)	67.1	50.2		54.3	
McFadden Pseudo R2	0.07	0.04		0.12	

Note: This table shows risk ratios with standard errors in parentheses. Specification (1) estimates the probability of aspirations to migrate permanently via binomial probit; the reference category is 'no aspirations to migrate permanently'. Specifications (2) and (4) estimate the probability of intentions to migrate temporarily, and specifications (3) and (5) estimate the probability of intentions to migrate permanently via multinomial probit; the reference category is 'no intentions to migrate'. Middle-income countries are countries which have a GNI per capita of $1,036 to $12,535 in current US-Dollars, as of 2020 (Bosnia, India, Indonesia, Mexico, and Ukraine). None of the middle-income countries are EU member states. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.K: *Aspirations and Intentions — Women*

Panel A: GI sample					
	Binomial probit	*Multinomial probit*		*Multinomial probit*	
	(1) Aspirations to migrate permanently	(2) Intentions to migrate temporarily	(3) Intentions to migrate permanently	(4) Intentions to migrate temporarily	(5) Intentions to migrate permanently
Migration aspiration				3.054***	23.125***
				(0.113)	(0.147)
Age: under 18 years	1.387***	0.822	1.764***	0.742	1.464*
	(0.117)	(0.232)	(0.204)	(0.240)	(0.227)
Age: 25 to 34 years	1.093	0.992	1.512**	0.996	1.538**
	(0.089)	(0.172)	(0.163)	(0.180)	(0.184)
Age: 35 to 49 years	0.804**	0.567***	0.755	0.610**	0.876
	(0.110)	(0.214)	(0.206)	(0.221)	(0.235)
Age: 50 to 64 years	0.540***	0.110***	0.333***	0.128***	0.440**
	(0.155)	(0.388)	(0.302)	(0.396)	(0.353)
Age: 65 years or above	0.375***	0.111***	0.107***	0.137***	0.190**
	(0.229)	(0.520)	(0.643)	(0.528)	(0.696)
University degree	0.838**	1.229	0.896	1.335**	1.027
	(0.072)	(0.141)	(0.133)	(0.146)	(0.150)
Student	0.975	1.137	0.965	1.157	0.967
	(0.083)	(0.158)	(0.149)	(0.165)	(0.167)
Migration experience	1.050	1.475***	1.474***	1.510***	1.546***
	(0.061)	(0.120)	(0.114)	(0.124)	(0.129)
Partner: native	0.960	0.936	0.912	0.941	0.931
	(0.071)	(0.137)	(0.133)	(0.141)	(0.151)
Partner: non-native	1.116	1.081	1.867***	1.136	2.050***
	(0.106)	(0.218)	(0.202)	(0.227)	(0.233)
Children	0.990	0.854	1.110	0.863	1.161
	(0.099)	(0.208)	(0.187)	(0.212)	(0.218)
Willingness to take risks	1.083***	1.080***	1.150***	1.057**	1.102***
	(0.013)	(0.025)	(0.024)	(0.026)	(0.027)
Patience	1.001	0.991	0.991	0.988	0.988
	(0.011)	(0.022)	(0.020)	(0.023)	(0.023)
Number of observations	2943	2943		2943	
Country FE	✓	✓		✓	
Correctly predicted values (%)	73.3	52.2		61.4	
McFadden Pseudo R2	0.15	0.10		0.20	
Panel B: Student sample					
Migration aspiration				4.609***	14.187***
				(0.126)	(0.198)
Age: 25 to 34 years	0.915	1.260	0.946	1.357*	1.053
	(0.095)	(0.167)	(0.209)	(0.176)	(0.225)
Age: 35 to 49 years	0.767	0.591	0.502	0.643	0.606
	(0.188)	(0.346)	(0.432)	(0.357)	(0.458)
University degree	0.876*	1.155	1.148	1.300**	1.349*
	(0.070)	(0.123)	(0.156)	(0.130)	(0.168)
Migration experience	1.311***	2.063***	2.060***	1.927***	1.868***
	(0.069)	(0.120)	(0.151)	(0.126)	(0.163)
Partner: native	0.791***	0.959	0.800	1.086	0.958
	(0.068)	(0.119)	(0.155)	(0.126)	(0.166)
Partner: non-native	1.131	1.834**	4.063***	2.100**	4.911***
	(0.153)	(0.308)	(0.318)	(0.328)	(0.354)
Children	1.145	0.513**	0.591	0.429**	0.452*
	(0.180)	(0.332)	(0.411)	(0.349)	(0.449)
Willingness to take risks	1.068***	1.111***	1.115***	1.085***	1.071*
	(0.016)	(0.029)	(0.036)	(0.030)	(0.038)
Patience	0.984	1.023	0.989	1.036	1.006
	(0.013)	(0.023)	(0.029)	(0.024)	(0.031)
Number of observations	1761	1761		1761	
Country FE	✓	✓		✓	
Correctly predicted values (%)	66.8	50.9		57.2	
McFadden Pseudo R2	0.10	0.06		0.14	

Note: This table shows risk ratios with standard errors in parentheses. Specification (1) estimates the probability of aspirations to migrate permanently via binomial probit; the reference category is 'no aspirations to migrate permanently'. Specifications (2) and (4) estimate the probability of intentions to migrate temporarily, and specifications (3) and (5) estimate the probability of intentions to migrate permanently via multinomial probit; the reference category is 'no intentions to migrate'. Respondents who gave no answer regarding their gender or indicated *"No answer/prefer not to say"* are excluded from the gender subsamples. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.L: *Aspirations and Intentions — Men*

	Binomial probit	Multinomial probit		Multinomial probit	
	(1) Aspirations to migrate permanently	(2) Intentions to migrate temporarily	(3) Intentions to migrate permanently	(4) Intentions to migrate temporarily	(5) Intentions to migrate permanently
Panel A: GI sample					
Migration aspiration				2.899***	23.192***
				(0.140)	(0.164)
Age: under 18 years	1.153	1.055	1.025	0.986	0.885
	(0.122)	(0.252)	(0.217)	(0.259)	(0.251)
Age: 25 to 34 years	1.038	1.229	1.317	1.272	1.387
	(0.104)	(0.217)	(0.193)	(0.224)	(0.222)
Age: 35 to 49 years	0.956	0.427***	0.697	0.433***	0.668
	(0.135)	(0.288)	(0.250)	(0.296)	(0.291)
Age: 50 to 64 years	0.564***	0.155***	0.248***	0.177***	0.311***
	(0.180)	(0.399)	(0.349)	(0.407)	(0.413)
Age: 65 years or above	0.359***	0.041***	0.071***	0.051***	0.102***
	(0.244)	(0.691)	(0.603)	(0.698)	(0.688)
University degree	0.823**	1.452**	0.912	1.560**	1.110
	(0.089)	(0.186)	(0.163)	(0.192)	(0.187)
Student	0.932	0.914	0.672**	0.901	0.670**
	(0.092)	(0.191)	(0.168)	(0.199)	(0.193)
Migration experience	1.073	1.036	1.138	1.050	1.124
	(0.072)	(0.151)	(0.135)	(0.157)	(0.156)
Partner: native	1.149	0.971	1.455**	0.940	1.406*
	(0.087)	(0.184)	(0.164)	(0.190)	(0.189)
Partner: non-native	1.044	1.006	1.570*	1.018	1.732*
	(0.140)	(0.304)	(0.271)	(0.314)	(0.315)
Children	0.804*	0.771	0.856	0.822	1.072
	(0.119)	(0.271)	(0.225)	(0.276)	(0.265)
Willingness to take risks	1.039**	1.110***	1.114***	1.102***	1.101***
	(0.015)	(0.033)	(0.028)	(0.034)	(0.032)
Patience	0.981	0.960	0.951**	0.968	0.960
	(0.013)	(0.028)	(0.025)	(0.029)	(0.029)
Number of observations	2021	2021		2021	
Country FE	✓	✓		✓	
Correctly predicted values (%)	71.6	53.1		64.5	
McFadden Pseudo R2	0.11	0.10		0.22	
Panel B: Student sample					
Migration aspiration				3.139***	16.427***
				(0.134)	(0.209)
Age: 25 to 34 years	1.018	1.106	0.999	1.088	0.967
	(0.104)	(0.185)	(0.224)	(0.190)	(0.245)
Age: 35 to 49 years	1.010	0.930	1.083	0.913	1.069
	(0.208)	(0.376)	(0.454)	(0.385)	(0.502)
University degree	1.004	1.558***	1.197	1.610***	1.291
	(0.078)	(0.142)	(0.170)	(0.146)	(0.188)
Migration experience	1.178**	1.331**	1.466**	1.278*	1.297
	(0.079)	(0.140)	(0.171)	(0.145)	(0.188)
Partner: native	0.931	0.770*	1.362*	0.811	1.462**
	(0.080)	(0.147)	(0.174)	(0.151)	(0.192)
Partner: non-native	1.471*	0.857	2.565**	0.752	1.995
	(0.220)	(0.416)	(0.435)	(0.426)	(0.475)
Children	0.942	0.823	0.289**	0.789	0.260**
	(0.204)	(0.366)	(0.552)	(0.375)	(0.600)
Willingness to take risks	1.018	1.141***	1.069*	1.139***	1.082**
	(0.016)	(0.031)	(0.036)	(0.032)	(0.040)
Patience	1.006	0.981	1.062*	0.982	1.062*
	(0.015)	(0.027)	(0.032)	(0.028)	(0.035)
Number of observations	1470	1470		1470	
Country FE	✓	✓		✓	
Correctly predicted values (%)	63.7	53.8		59.3	
McFadden Pseudo R2	0.06	0.05		0.14	

Note: This table shows risk ratios with standard errors in parentheses. Specification (1) estimates the probability of aspirations to migrate permanently via binomial probit; the reference category is 'no aspirations to migrate permanently'. Specifications (2) and (4) estimate the probability of intentions to migrate temporarily, and specifications (3) and (5) estimate the probability of intentions to migrate permanently via multinomial probit; the reference category is 'no intentions to migrate'. Respondents who gave no answer regarding their gender or indicated *"No answer/prefer not to say"* are excluded from the gender subsamples. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.M: *Aspirations and Intentions — GI Sample, 6 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above)*

GI sample, younger age groups (under 35 years of age)					
	Binomial probit	*Multinomial probit*		*Multinomial probit*	
	(1) Aspirations to migrate permanently	(2) Intentions to migrate temporarily	(3) Intentions to migrate permanently	(4) Intentions to migrate temporarily	(5) Intentions to migrate permanently
Migration aspiration				2.366^{***}	20.462^{***}
				(0.116)	(0.159)
Gender: woman	1.095	1.056	0.999	1.024	0.917
	(0.058)	(0.112)	(0.105)	(0.114)	(0.117)
Age: under 18 years	1.270^{**}	0.780	1.208	0.728	1.006
	(0.103)	(0.204)	(0.184)	(0.208)	(0.207)
Age: 25 to 34 years	1.081	1.092	1.350^{*}	1.093	1.320
	(0.091)	(0.173)	(0.169)	(0.177)	(0.188)
University degree	0.882^{*}	1.604^{***}	1.011	1.687^{***}	1.135
	(0.075)	(0.143)	(0.137)	(0.146)	(0.153)
Student	0.999	1.163	0.896	1.161	0.879
	(0.081)	(0.154)	(0.149)	(0.159)	(0.165)
Migration experience	1.105	1.577^{***}	1.375^{**}	1.547^{***}	1.332^{*}
	(0.075)	(0.140)	(0.141)	(0.143)	(0.156)
Partner: native	1.012	1.121	1.245	1.116	1.242
	(0.086)	(0.162)	(0.164)	(0.166)	(0.182)
Partner: non-native	1.054	1.037	1.796^{**}	1.098	2.028^{**}
	(0.132)	(0.271)	(0.262)	(0.282)	(0.301)
Children	0.717^{*}	0.565	0.644	0.620	0.933
	(0.189)	(0.374)	(0.348)	(0.379)	(0.403)
Willingness to take risks	1.059^{***}	1.069^{**}	1.113^{***}	1.057^{*}	1.085^{***}
	(0.015)	(0.028)	(0.026)	(0.029)	(0.029)
Patience	0.982	0.987	0.972	0.992	0.982
	(0.012)	(0.024)	(0.022)	(0.024)	(0.025)
Number of observations	2443	2443		2443	
Country FE	✓	✓		✓	
Correctly predicted values (%)	71.4	46.8		56.8	
McFadden Pseudo R2	0.09	0.04		0.14	
GI sample, older age groups (35 years and above)					
Migration aspiration				7.872^{***}	19.646^{***}
				(0.332)	(0.337)
Gender: woman	0.788^{*}	1.061	0.616^{*}	1.149	0.707
	(0.134)	(0.269)	(0.255)	(0.284)	(0.288)
Age: 50 to 64 years	0.838	0.387^{**}	0.564^{*}	0.363^{**}	0.546
	(0.167)	(0.373)	(0.320)	(0.398)	(0.370)
Age: 65 years or above	0.488^{***}	0.193^{***}	0.196^{***}	0.243^{**}	0.281^{*}
	(0.252)	(0.572)	(0.590)	(0.584)	(0.657)
University degree	0.785	1.408	0.779	1.576	1.023
	(0.203)	(0.436)	(0.387)	(0.469)	(0.462)
Migration experience	0.805	1.208	0.856	1.383	0.987
	(0.134)	(0.271)	(0.257)	(0.289)	(0.296)
Partner: native	1.161	0.980	0.941	0.856	0.827
	(0.175)	(0.333)	(0.330)	(0.355)	(0.375)
Partner: non-native	1.088	1.189	1.322	1.174	1.305
	(0.214)	(0.425)	(0.411)	(0.449)	(0.473)
Children	0.898	0.396^{***}	0.738	0.388^{***}	0.686
	(0.152)	(0.306)	(0.287)	(0.325)	(0.326)
Willingness to take risks	1.060^{*}	0.949	1.273^{***}	0.951	1.252^{***}
	(0.034)	(0.068)	(0.071)	(0.072)	(0.078)
Patience	0.937^{**}	0.858^{***}	0.806^{***}	0.860^{**}	0.814^{***}
	(0.029)	(0.058)	(0.055)	(0.061)	(0.063)
Number of observations	478	478		478	
Country FE	✓	✓		✓	
Correctly predicted values (%)	73.2	61.1		68.8	
McFadden Pseudo R2	0.21	0.15		0.26	

Note: This table shows risk ratios with standard errors in parentheses. The GI sample is restricted to younger age groups (under 35 years of age) or older age groups (35 years and above) and to the 6 countries which are also observed in the student sample. Specification (1) estimates the probability of aspirations to migrate permanently via binomial probit; the reference category is 'no aspirations to migrate permanently'. Specifications (2) and (4) estimate the probability of intentions to migrate temporarily, and specifications (3) and (5) estimate the probability of intentions to migrate permanently via multinomial probit; the reference category is 'no intentions to migrate'. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.N: *Aspirations and Intentions — GI Sample, 14 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above)*

GI sample, younger age groups (under 35 years of age)

	Binomial probit	*Multinomial probit*		*Multinomial probit*	
	(1) Aspirations to migrate permanently	(2) Intentions to migrate temporarily	(3) Intentions to migrate permanently	(4) Intentions to migrate temporarily	(5) Intentions to migrate permanently
Migration aspiration				2.628^{***}	23.622^{***}
				(0.094)	(0.125)
Gender: woman	1.106^{**}	1.057	0.963	0.998	0.848^{*}
	(0.046)	(0.091)	(0.082)	(0.093)	(0.094)
Age: under 18 years	1.259^{***}	0.839	1.355^{**}	0.788	1.180
	(0.083)	(0.167)	(0.146)	(0.171)	(0.166)
Age: 25 to 34 years	1.046	0.993	1.345^{**}	1.011	1.404^{**}
	(0.068)	(0.135)	(0.125)	(0.139)	(0.142)
University degree	0.856^{***}	1.301^{**}	0.903	1.375^{***}	1.019
	(0.060)	(0.117)	(0.108)	(0.121)	(0.123)
Student	0.940	1.020	0.829^{*}	1.035	0.844
	(0.061)	(0.121)	(0.110)	(0.125)	(0.125)
Migration experience	1.165^{***}	1.326^{***}	1.462^{***}	1.309^{**}	1.411^{***}
	(0.055)	(0.108)	(0.101)	(0.111)	(0.114)
Partner: native	0.943	0.944	1.084	0.980	1.176
	(0.063)	(0.126)	(0.119)	(0.130)	(0.136)
Partner: non-native	1.059	1.155	1.667^{**}	1.197	1.784^{**}
	(0.103)	(0.215)	(0.202)	(0.223)	(0.230)
Children	0.805	0.736	0.768	0.778	0.969
	(0.136)	(0.293)	(0.254)	(0.298)	(0.298)
Willingness to take risks	1.060^{***}	1.088^{***}	1.121^{***}	1.075^{***}	1.093^{***}
	(0.011)	(0.022)	(0.020)	(0.023)	(0.023)
Patience	0.993	0.993	0.994	0.996	0.999
	(0.010)	(0.019)	(0.017)	(0.019)	(0.020)
Number of observations	4005	4005		4005	
Country FE	✓	✓		✓	
Correctly predicted values (%)	73.1	50.0		59.9	
McFadden Pseudo R2	0.08	0.06		0.17	

GI sample, older age groups (35 years and above)

	(1)	(2)	(3)	(4)	(5)
Migration aspiration				6.960^{***}	26.717^{***}
				(0.218)	(0.228)
Gender: woman	0.913	1.085	0.688^{**}	1.103	0.682^{*}
	(0.087)	(0.190)	(0.169)	(0.202)	(0.199)
Age: 50 to 64 years	0.685^{***}	0.289^{***}	0.481^{***}	0.317^{***}	0.532^{**}
	(0.106)	(0.249)	(0.208)	(0.262)	(0.250)
Age: 65 years or above	0.424^{***}	0.126^{***}	0.149^{***}	0.173^{***}	0.236^{***}
	(0.160)	(0.412)	(0.412)	(0.422)	(0.462)
University degree	0.737^{**}	1.245	0.791	1.472	1.114
	(0.145)	(0.331)	(0.277)	(0.349)	(0.334)
Migration experience	0.849^{*}	1.263	1.028	1.471^{*}	1.268
	(0.086)	(0.188)	(0.169)	(0.201)	(0.200)
Partner: native	1.277^{**}	0.796	1.027	0.617^{**}	0.758
	(0.109)	(0.225)	(0.216)	(0.241)	(0.255)
Partner: non-native	1.209	0.720	1.889^{**}	0.673	1.902^{*}
	(0.148)	(0.325)	(0.286)	(0.340)	(0.335)
Children	0.882	0.829	0.950	0.903	1.015
	(0.096)	(0.209)	(0.185)	(0.223)	(0.218)
Willingness to take risks	1.092^{***}	1.095^{**}	1.189^{***}	1.058	1.132^{***}
	(0.020)	(0.045)	(0.040)	(0.048)	(0.047)
Patience	0.987	0.924^{*}	0.909^{***}	0.913^{**}	0.897^{**}
	(0.018)	(0.040)	(0.036)	(0.043)	(0.042)
Number of observations	1105	1105		1105	
Country FE	✓	✓		✓	
Correctly predicted values (%)	70.8	62.8		71.4	
McFadden Pseudo R2	0.18	0.17		0.30	

Note: This table shows risk ratios with standard errors in parentheses. The GI sample is restricted to younger age groups (under 35 years of age) or older age groups (35 years and above). Specification (1) estimates the probability of aspirations to migrate permanently via binomial probit; the reference category is 'no aspirations to migrate permanently'. Specifications (2) and (4) estimate the probability of intentions to migrate temporarily, and specifications (3) and (5) estimate the probability of intentions to migrate permanently via multinomial probit; the reference category is 'no intentions to migrate'. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.O: *Intentions by Aspirations — GI Sample, 6 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above)*

	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
GI sample, younger age groups (under 35 years of age)				
Gender: woman	0.908	0.943	1.138	0.969
	(0.169)	(0.297)	(0.157)	(0.138)
Age: under 18 years	1.191	1.021	0.524**	0.903
	(0.311)	(0.492)	(0.282)	(0.240)
Age: 25 to 34 years	1.045	1.393	1.253	1.422
	(0.263)	(0.479)	(0.254)	(0.235)
University degree	2.018***	0.944	1.478*	1.093
	(0.222)	(0.402)	(0.199)	(0.181)
Student	1.492	1.118	0.927	0.744
	(0.244)	(0.431)	(0.216)	(0.196)
Migration experience	1.276	1.024	1.970***	1.655**
	(0.209)	(0.391)	(0.209)	(0.198)
Partner: native	1.049	0.824	1.351	1.508*
	(0.236)	(0.448)	(0.252)	(0.236)
Partner: non-native	1.026	1.823	1.270	2.497*
	(0.368)	(0.605)	(0.509)	(0.468)
Children	0.690	0.289	0.795	1.671
	(0.452)	(1.099)	(0.869)	(0.770)
Willingness to take risks	0.995	1.093	1.110***	1.109***
	(0.043)	(0.077)	(0.040)	(0.035)
Patience	0.992	0.967	0.995	0.987
	(0.036)	(0.064)	(0.033)	(0.030)
Number of observations	763		1680	
Country FE	✓		✓	
Correctly predicted values (%)	56.7		56.8	
McFadden Pseudo R2	0.03		0.05	
GI sample, older age groups (35 years and above)				
Gender: woman	0.887	1.357	1.945	0.510
	(0.375)	(0.483)	(0.519)	(0.431)
Age: 50 to 64 years	0.616	0.241*	0.151***	0.703
	(0.484)	(0.810)	(0.710)	(0.501)
Age: 65 years or above	0.183**	0.159*	0.421	0.432
	(0.811)	(1.108)	(1.070)	(0.943)
University degree	1.241	/†	1.149	0.387
	(0.590)		(0.926)	(0.749)
Migration experience	1.208	1.701	1.598	0.650
	(0.379)	(0.521)	(0.523)	(0.444)
Partner: native	0.794	0.516	0.757	0.867
	(0.471)	(0.633)	(0.668)	(0.595)
Partner: non-native	2.196	0.932	0.378	0.836
	(0.587)	(0.727)	(0.798)	(0.703)
Children	0.521	1.310	0.224***	0.473
	(0.425)	(0.571)	(0.551)	(0.459)
Willingness to take risks	1.152	1.210	0.780**	1.172
	(0.107)	(0.131)	(0.123)	(0.110)
Patience	0.807**	0.805**	0.900	0.866*
	(0.084)	(0.107)	(0.102)	(0.086)
Number of observations	285		193	
Country FE	✓		✓	
Correctly predicted values (%)	76.1		66.8	
McFadden Pseudo R2	0.15		0.21	

Note: This table shows risk ratios with standard errors in parentheses. The GI sample is restricted to younger age groups (under 35 years of age) or older age groups (35 years and above), and to the 6 countries which are also observed in the student sample. Specifications (1) and (2) estimate the probability of having intentions to migrate temporarily or permanently for those with no aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Specifications (3) and (4) estimate the probability of having intentions to migrate temporarily or permanently for those with aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Since only 1 of 478 respondents in the GI older age group sample is a student, we control for it, but do not report the coefficient. †Since all respondents in the older age group sample who have no migration aspirations but who have intentions to migrate permanently ($n = 26$) possess a university degree, the coefficient is not meaningful. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.P: *Intentions by Aspirations — GI Sample, 14 Countries, Younger Age Groups (Under 35 Years of Age) and Older Age Groups (35 Years and Above)*

GI sample, younger age groups (under 35 years of age)	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
Gender: woman	0.968	0.723	1.075	0.899
	(0.138)	(0.225)	(0.129)	(0.112)
Age: under 18 years	1.187	1.140	0.600**	1.050
	(0.265)	(0.418)	(0.227)	(0.193)
Age: 25 to 34 years	0.913	1.529	1.201	1.549**
	(0.210)	(0.353)	(0.195)	(0.175)
University degree	1.708***	0.819	1.192	0.979
	(0.186)	(0.319)	(0.163)	(0.146)
Student	1.199	0.993	0.890	0.759*
	(0.194)	(0.330)	(0.167)	(0.147)
Migration experience	1.060	1.123	1.552***	1.626***
	(0.165)	(0.274)	(0.159)	(0.145)
Partner: native	0.905	1.067	1.139	1.324
	(0.186)	(0.307)	(0.194)	(0.178)
Partner: non-native	1.053	1.762	1.462	2.207**
	(0.307)	(0.529)	(0.372)	(0.344)
Children	0.923	0.594	0.771	1.197
	(0.370)	(0.601)	(0.576)	(0.500)
Willingness to take risks	1.018	1.093	1.126***	1.118***
	(0.033)	(0.056)	(0.031)	(0.027)
Patience	0.996	0.994	0.996	0.999
	(0.029)	(0.049)	(0.027)	(0.024)
Number of observations	1160		2845	
Country FE	✓		✓	
Correctly predicted values (%)	58.5		60.9	
McFadden Pseudo R2	0.04		0.07	
GI sample, older age groups (35 years and above)				
Gender: woman	0.994	0.738	1.481	0.673
	(0.280)	(0.357)	(0.326)	(0.273)
Age: 50 to 64 years	0.409**	0.182***	0.286***	0.708
	(0.359)	(0.605)	(0.409)	(0.326)
Age: 65 years or above	0.226***	0.118***	0.153**	0.266**
	(0.541)	(0.823)	(0.732)	(0.589)
University degree	1.736	/†	0.955	0.551
	(0.509)		(0.541)	(0.451)
Migration experience	1.282	1.592	1.756*	1.232
	(0.280)	(0.379)	(0.313)	(0.267)
Partner: native	0.653	0.670	0.518*	0.770
	(0.340)	(0.476)	(0.386)	(0.351)
Partner: non-native	1.107	1.345	0.349*	1.914
	(0.444)	(0.575)	(0.553)	(0.465)
Children	0.892	1.329	0.933	0.899
	(0.319)	(0.416)	(0.339)	(0.290)
Willingness to take risks	1.163**	1.058	0.960	1.114*
	(0.073)	(0.092)	(0.071)	(0.060)
Patience	0.920	0.854*	0.927	0.919
	(0.061)	(0.081)	(0.064)	(0.055)
Number of observations	583		522	
Country FE	✓		✓	
Correctly predicted values (%)	79.1		64.8	
McFadden Pseudo R2	0.16		0.20	

Note: This table shows risk ratios with standard errors in parentheses. The GI sample is restricted to younger age groups (under 35 years of age) or older age groups (35 years and above). Specifications (1) and (2) estimate the probability of having intentions to migrate temporarily or permanently for those with no aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Specifications (3) and (4) estimate the probability of having intentions to migrate temporarily or permanently for those with aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Since only 3 of 1105 respondents in the GI older age group sample are students, we control for it but do not report the coefficient. †Since all respondents who have no migration aspirations but who have intentions to migrate permanently ($n = 44$) possess a university degree, the coefficient is not meaningful. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.Q: *Intentions by Aspirations — Women*

Panel A: GI sample	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
Age: under 18 years	1.014	3.637**	0.621	1.228
	(0.395)	(0.601)	(0.302)	(0.255)
Age: 25 to 34 years	0.790	0.906	1.378	2.079***
	(0.285)	(0.529)	(0.246)	(0.224)
Age: 35 to 49 years	0.397***	1.296	0.895	0.905
	(0.342)	(0.575)	(0.321)	(0.297)
Age: 50 to 64 years	0.078***	0.284	0.202***	0.513
	(0.613)	(0.950)	(0.556)	(0.428)
Age: 65 years or above	0.105***	0.247	0.182**	0.153**
	(0.706)	(1.240)	(0.865)	(0.876)
University degree	1.457*	1.905	1.232	0.917
	(0.229)	(0.414)	(0.196)	(0.175)
Student	1.220	1.590	0.995	0.832
	(0.264)	(0.483)	(0.215)	(0.193)
Migration experience	1.323	1.366	1.802***	1.746***
	(0.188)	(0.320)	(0.177)	(0.162)
Partner: native	0.762	1.002	1.057	0.995
	(0.208)	(0.352)	(0.209)	(0.195)
Partner: non-native	1.308	2.176	1.010	1.997**
	(0.323)	(0.495)	(0.345)	(0.311)
Children	1.172	1.055	0.624	1.003
	(0.297)	(0.430)	(0.338)	(0.292)
Willingness to take risks	1.051	1.112	1.063*	1.105***
	(0.039)	(0.066)	(0.037)	(0.032)
Patience	0.962	0.902*	1.024	1.020
	(0.034)	(0.056)	(0.031)	(0.028)
Number of observations	990		1953	
Country FE	✓		✓	
Correctly predicted values (%)	66.4		59.4	
McFadden Pseudo R2	0.09		0.08	
Panel B: Student sample				
Age: 25 to 34 years	1.374	0.851	1.425	1.138
	(0.249)	(0.521)	(0.264)	(0.284)
Age: 35 to 49 years	0.276*	0.236	1.181	1.058
	(0.665)	(1.224)	(0.503)	(0.560)
University degree	0.988	1.329	1.569**	1.430*
	(0.191)	(0.382)	(0.186)	(0.204)
Migration experience	2.585***	1.071	1.544**	1.831***
	(0.182)	(0.410)	(0.181)	(0.196)
Partner: native	1.309	1.048	0.928	0.863
	(0.182)	(0.380)	(0.180)	(0.199)
Partner: non-native	1.725	6.299***	3.330*	7.154***
	(0.439)	(0.643)	(0.636)	(0.629)
Children	0.492	2.161	0.302**	0.234***
	(0.570)	(0.919)	(0.475)	(0.555)
Willingness to take risks	1.088*	0.888	1.100**	1.127**
	(0.044)	(0.086)	(0.043)	(0.047)
Patience	0.995	0.894	1.080**	1.058
	(0.037)	(0.075)	(0.034)	(0.037)
Number of observations	816		945	
Country FE	✓		✓	
Correctly predicted values (%)	68.0		47.6	
McFadden Pseudo R2	0.08		0.05	

Note: This table shows risk ratios with standard errors in parentheses. Specifications (1) and (2) estimate the probability of having intentions to migrate temporarily or permanently for those with no aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Specifications (3) and (4) estimate the probability of having intentions to migrate temporarily or permanently for those with aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Respondents who gave no answer regarding their gender or indicated *"No answer/prefer not to say"* are excluded from the gender subsamples. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.R: *Intentions by Aspirations — Men*

Panel A: GI sample	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
Age: under 18 years	1.989*	0.516	0.604	0.746
	(0.391)	(0.637)	(0.360)	(0.306)
Age: 25 to 34 years	1.107	2.314*	1.375	1.295
	(0.323)	(0.459)	(0.320)	(0.279)
Age: 35 to 49 years	0.325***	1.861	0.466*	0.534*
	(0.430)	(0.571)	(0.444)	(0.381)
Age: 50 to 64 years	0.209***	0.214*	0.159***	0.309**
	(0.538)	(0.925)	(0.688)	(0.554)
Age: 65 years or above	0.069***	/†	0.025***	0.076***
	(0.863)		(1.319)	(0.872)
University degree	2.729***	0.669	1.096	1.025
	(0.302)	(0.411)	(0.265)	(0.230)
Student	1.059	0.765	0.786	0.600**
	(0.311)	(0.450)	(0.270)	(0.233)
Migration experience	1.009	1.178	1.106	1.170
	(0.230)	(0.315)	(0.231)	(0.202)
Partner: native	0.996	0.910	0.950	1.615*
	(0.271)	(0.382)	(0.290)	(0.251)
Partner: non-native	1.036	1.208	1.083	2.147*
	(0.413)	(0.596)	(0.529)	(0.459)
Children	0.681	0.945	1.119	1.271
	(0.374)	(0.463)	(0.467)	(0.387)
Willingness to take risks	1.054	1.073	1.148***	1.128***
	(0.051)	(0.074)	(0.047)	(0.038)
Patience	1.006	1.015	0.923*	0.926**
	(0.044)	(0.064)	(0.042)	(0.036)
Number of observations	705		1316	
Country FE	✓		✓	
Correctly predicted values (%)	65.4		64.1	
McFadden Pseudo R2	0.13		0.10	
Panel B: Student sample				
Age: 25 to 34 years	0.887	0.859	1.416	1.144
	(0.267)	(0.590)	(0.292)	(0.306)
Age: 35 to 49 years	0.729	5.510**	1.154	0.629
	(0.605)	(0.836)	(0.541)	(0.621)
University degree	1.469*	1.410	1.681**	1.290
	(0.198)	(0.424)	(0.220)	(0.226)
Migration experience	1.343	0.586	1.284	1.550*
	(0.206)	(0.509)	(0.215)	(0.226)
Partner: native	0.707	3.109***	0.854	1.264
	(0.214)	(0.429)	(0.225)	(0.233)
Partner: non-native	0.591	/††	0.992	2.405
	(0.634)		(0.615)	(0.593)
Children	1.492	0.164	0.390*	0.234**
	(0.516)	(1.245)	(0.557)	(0.699)
Willingness to take risks	1.174***	0.990	1.105**	1.082
	(0.044)	(0.091)	(0.048)	(0.048)
Patience	0.992	0.951	0.974	1.088**
	(0.039)	(0.078)	(0.041)	(0.043)
Number of observations	740		730	
Country FE	✓		✓	
Correctly predicted values (%)	69.6		47.3	
McFadden Pseudo R2	0.08		0.06	

Note: This table shows risk ratios with standard errors in parentheses. Specifications (1) and (2) estimate the probability of having intentions to migrate temporarily or permanently for those with no aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Specifications (3) and (4) estimate the probability of having intentions to migrate temporarily or permanently for those with aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Respondents who gave no answer regarding their gender or indicated *"No answer/prefer not to say"* are excluded from the gender subsamples. †Since none of the respondents who have no migration aspirations but intentions to migrate permanently ($n = 71$) are 65 years or above, the coefficient is not meaningful. ††Since none of the respondents who have no migration aspirations but intentions to migrate permanently ($n = 34$) have a non-native partner, the coefficient is not meaningful. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.S: *Intentions by Aspirations — High-Income Countries*

	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
Panel A: GI sample				
Gender: woman	0.919	0.516***	1.122	0.686**
	(0.158)	(0.242)	(0.183)	(0.163)
Age: under 18 years	1.037	0.507	0.206***	0.500**
	(0.439)	(0.857)	(0.422)	(0.345)
Age: 25 to 34 years	0.894	1.082	1.229	1.929**
	(0.268)	(0.452)	(0.284)	(0.261)
Age: 35 to 49 years	0.351***	1.142	0.646	0.813
	(0.315)	(0.503)	(0.348)	(0.328)
Age: 50 to 64 years	0.143***	0.121***	0.174***	0.417**
	(0.412)	(0.777)	(0.496)	(0.433)
Age: 65 years or above	0.074***	0.121**	0.051***	0.122***
	(0.567)	(0.910)	(0.887)	(0.701)
University degree	1.707**	1.063	0.896	0.773
	(0.228)	(0.373)	(0.239)	(0.223)
Student	0.863	0.934	0.589**	0.709
	(0.263)	(0.453)	(0.255)	(0.231)
Migration experience	1.047	1.250	1.286	1.234
	(0.163)	(0.255)	(0.183)	(0.167)
Partner: native	0.863	0.882	0.905	1.074
	(0.186)	(0.307)	(0.228)	(0.214)
Partner: non-native	1.460	1.493	1.238	2.610***
	(0.284)	(0.437)	(0.391)	(0.361)
Children	0.883	1.123	0.827	0.916
	(0.258)	(0.360)	(0.331)	(0.294)
Willingness to take risks	1.105***	1.115*	1.132***	1.145***
	(0.038)	(0.061)	(0.042)	(0.037)
Patience	0.955	0.963	0.960	0.982
	(0.033)	(0.053)	(0.036)	(0.033)
Number of observations	1137		1296	
Country FE	✓		✓	
Correctly predicted values (%)	69.6		57.6	
McFadden Pseudo R2	0.11		0.10	
Panel B: Student sample				
Gender: woman	0.877	0.954	1.428*	1.297
	(0.159)	(0.361)	(0.212)	(0.248)
Age: 25 to 34 years	1.197	0.973	0.728	0.736
	(0.226)	(0.538)	(0.352)	(0.399)
Age: 35 to 49 years	0.160**	/†	0.201*	1.044
	(0.799)		(0.924)	(0.873)
University degree	1.134	1.202	1.979***	1.504
	(0.166)	(0.365)	(0.223)	(0.264)
Migration experience	2.053***	0.904	1.633**	2.422***
	(0.161)	(0.391)	(0.213)	(0.252)
Partner: native	1.166	1.030	1.202	1.364
	(0.160)	(0.355)	(0.223)	(0.260)
Partner: non-native	1.423	2.050	2.180	5.213***
	(0.409)	(0.824)	(0.573)	(0.581)
Children	0.849	/†	0.478	0.110**
	(0.503)		(0.845)	(1.022)
Willingness to take risks	1.157***	0.854*	1.139**	1.102
	(0.039)	(0.088)	(0.052)	(0.061)
Patience	0.994	0.861**	1.060	1.147***
	(0.034)	(0.075)	(0.043)	(0.051)
Number of observations	920		674	
Country FE	✓		✓	
Correctly predicted values (%)	65.1		53.9	
McFadden Pseudo R2	0.07		0.07	

Note: This table shows risk ratios with standard errors in parentheses. Specifications (1) and (2) estimate the probability of having intentions to migrate temporarily or permanently for those with no aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Specifications (3) and (4) estimate the probability of having intentions to migrate temporarily or permanently for those with aspirations to migrate permanently; the reference category is 'no intentions to migrate'. High-income countries include countries which have a GNI per capita larger than $12,535 in current US-Dollars, as of 2020 (Czechia, Italy, Japan, Netherlands, Poland, Romania, South Korea, Spain, and the United Kingdom). †Since none of the respondents who have no migration aspirations but intentions to migrate permanently ($n = 39$) are between 35 and 49 years of age or have children, the coefficients are not meaningful. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.T: *Intentions by Aspirations — Middle-Income Countries*

Panel A: GI sample	No aspirations to migrate permanently		Aspirations to migrate permanently	
	(1) Intentions to migrate temporarily	(2) Intentions to migrate permanently	(3) Intentions to migrate temporarily	(4) Intentions to migrate permanently
Gender: woman	1.059	1.100	1.133	1.028
	(0.199)	(0.295)	(0.158)	(0.134)
Age: under 18 years	1.331	1.999	0.879	1.269
	(0.341)	(0.529)	(0.271)	(0.228)
Age: 25 to 34 years	0.742	1.595	1.283	1.503*
	(0.352)	(0.520)	(0.265)	(0.235)
Age: 35 to 49 years	0.650	1.416	0.529	0.660
	(0.519)	(0.691)	(0.406)	(0.337)
Age: 50 to 64 years	/†	1.621	0.190*	0.574
		(1.250)	(0.901)	(0.580)
University degree	2.192***	1.225	1.433*	1.071
	(0.270)	(0.430)	(0.203)	(0.175)
Student	1.897**	1.304	1.065	0.730*
	(0.292)	(0.469)	(0.219)	(0.190)
Migration experience	1.279	1.080	1.652**	1.707***
	(0.288)	(0.419)	(0.218)	(0.194)
Partner: native	0.783	1.069	1.131	1.340
	(0.339)	(0.445)	(0.245)	(0.214)
Partner: non-native	0.529	2.282	0.918	1.621
	(0.553)	(0.669)	(0.409)	(0.348)
Children	0.891	0.683	0.948	1.441
	(0.496)	(0.600)	(0.456)	(0.370)
Willingness to take risks	0.944	1.057	1.074*	1.102***
	(0.051)	(0.078)	(0.039)	(0.033)
Patience	1.051	0.945	1.007	0.992
	(0.045)	(0.065)	(0.034)	(0.029)
Number of observations	606		2017	
Country FE	✓		✓	
Correctly predicted values (%)	59.9		64.3	
McFadden Pseudo R2	0.06		0.05	
Panel B: Student sample				
Gender: woman	0.946	1.113	0.926	0.920
	(0.227)	(0.408)	(0.166)	(0.166)
Age: 25 to 34 years	1.015	0.805	2.103***	1.453
	(0.318)	(0.571)	(0.237)	(0.243)
Age: 35 to 49 years	0.976	2.079	2.162*	0.954
	(0.545)	(0.772)	(0.418)	(0.480)
University degree	1.390	1.655	1.294	1.230
	(0.246)	(0.448)	(0.190)	(0.185)
Migration experience	1.688**	0.819	1.324	1.444**
	(0.254)	(0.534)	(0.183)	(0.184)
Partner: native	0.652	3.278***	0.739	0.881
	(0.267)	(0.445)	(0.185)	(0.185)
Partner: non-native	0.662	14.024***	1.633	3.508**
	(0.818)	(0.835)	(0.630)	(0.594)
Children	1.080	1.224	0.334***	0.281***
	(0.540)	(0.802)	(0.404)	(0.478)
Willingness to take risks	1.074	1.019	1.066	1.090**
	(0.050)	(0.090)	(0.040)	(0.040)
Patience	1.016	0.950	1.023	1.030
	(0.044)	(0.077)	(0.033)	(0.033)
Number of observations	640		1008	
Country FE	✓		✓	
Correctly predicted values (%)	73.6		41.6	
McFadden Pseudo R2	0.06		0.05	

Note: This table shows risk ratios with standard errors in parentheses. Specifications (1) and (2) estimate the probability of having intentions to migrate temporarily or permanently for those with no aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Specifications (3) and (4) estimate the probability of having intentions to migrate temporarily or permanently for those with aspirations to migrate permanently; the reference category is 'no intentions to migrate'. Middle-income countries are countries which have a GNI per capita of $1,036 to $12,535 in current US-Dollars, as of 2020 (Bosnia, India, Indonesia, Mexico, and Ukraine). †Since none of the respondents who have no migration aspirations but intentions to migrate temporarily ($n = 186$) are between 50 and 64 years of age, the coefficient is not meaningful. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

5 | The Effect of Schooling on Parental Integration

Abstract. *Exploiting the age-at-enrollment policies in 16 German states as exogenous source of variation, I examine whether the schooling of the oldest child in a migrant household affects parents' integration. My analysis links administrative records on primary school enrollment cutoff dates with micro data from the German Socioeconomic Panel (GSOEP). I use a regression discontinuity design around the school enrollment cutoff and an instrumental variable approach to show that children's schooling improves the integration of parents along several dimensions, such as labor market outcomes, financial worries, and German language abilities – both on the extensive margin (i.e. school enrollment) and on the intensive margin (i.e. months of schooling). Labor market outcomes are most positively affected for parents who carry the main burden of childcare in the household, i.e. mothers. Additional analysis of underlying mechanisms suggests that the results are driven by gains in disposable time and exposure to the German language and culture.*

Keywords: *international migration, assimilation, integration, education, schooling, early age, family, regression discontinuity, instrumental variables*

5.1 Introduction

Immigration into developed countries has become an increasingly important topic in recent decades and is not going to subside anytime soon. By the end of 2020, Germany had a migrant population of over 10 million people[1], representing 13.7 % of the nation's total population (Destatis 2020). One million alone are Syrian refugees who entered Germany in the mid 2010s (BAMF 2016). Such inflow poses major challenges for public policy (Angelini et al. 2015), first and foremost the question of successful integration into the host country. The literature on labor market outcomes, cultural and social assimilation and well-being of migrants is vast and shows over and over again how migrants lack behind their native counterparts. They obtain lower wages and show higher unemployment rates (Algan et al. 2010; Borjas 2015) and suffer from cultural or political marginalization (Algan et al. 2012). Not only the migrants themselves suffer from their marginalization (Angelini et al. 2015) but so do the host countries, as it potentially

[1] Individuals who were not born in Germany but regularly reside in Germany.

leads to ethnic enclaves and social unrest (Gathmann and Keller 2018). However, the inflow of migrants constitutes a major chance for countries like Germany to address their demographic change and the subsequent shortage of skilled workers. Consequently, the most pressing question for host countries is how to facilitate successful integration of immigrants into the country – and specifically into the labor market.

As one step towards identifying a potential key factor for integration, I examine the effect children's primary schooling has on the integration of migrant parents into their host country. While the effect of school attendance and attendance of early childhood education and care facilities (ECEC, e.g. kindergarten) on migrant children themselves has been extensively studied (Bleakley and Chin 2008; Cornelissen et al. 2018; Felfe and Lalive 2018), relatively few studies have investigated how attendance of such facilities might impact migrant parents. Drange and Telle (2015) used data from Norway and found that ECEC attendance among migrant children did not have a significant impact on their parents' employment or educational outcomes. Gambaro et al. (2021), on the other hand, exploit regional differences in the availability of ECEC facilities in German states as exogenous sources of variation to estimate the effect of attendance of such facilities by refugee children on their parents' integration. They create an integration index from several measures of integration in the German Socioeconomic Panel (GSOEP) and find a significant positive effect on overall integration, with particularly strong effects on labor market outcomes and language proficiency. Instead of state dependent variation in access to early childhood education and care facilities like Gambaro et al. (2021), I use variation in school enrollment timing to analyze effects on migrant parents' outcomes. My results are consistent with those of Gambaro et al. (2021) and provide the first evidence regarding the effect of primary schooling on parents' integration.

Exploiting the German age-at-enrollment policy as exogenous source of variation in school entry timing, I investigate the effect of schooling of the oldest child in the household on parental integration outcomes.[2] For this, I link administrative records on primary school enrollment cutoff dates with micro data from the German Socioeconomic Panel. Using a regression discontinuity design around the school enrollment cutoff and an instrumental variable approach, I can estimate the effect of the oldest child's schooling on parents' integration both on the extensive margin (i.e. the effect of school enrollment) and on the intensive margin (i.e. the effect of one additional month of schooling).[3]

Integration is a complex, multidimensional process, spanning economic outcomes as well as social and cultural assimilation (Constant and Zimmermann 2008; Facchini et al. 2015). Lots of works focus on labor market outcomes (Aldashev et al. 2009;

[2] Using school enrollment cutoff dates as exogenous source of variation is frequently used in literature analyzing the effects of school entry timing on children's school and lifetime achievements (Angrist and Krueger 1991; Cascio and Schanzenbach 2016; DiPasquale et al. 1980). Like in Germany, school enrollment in many states is regulated by compulsory schooling cutoff dates; more detail on how school enrollment is designed in Germany and how this design is utilized in this paper can be found in Section 5.2.

[3] I focus on the oldest child in the household since they are the first ones to enter school. How the existence of younger children in the household could distort my identification strategy and how I address this will be discussed in detail in Section 5.2.

Bleakley and Chin 2004; Chiswick and Miller 1995; Dustmann 1994; Dustmann and Van Soest 2001), and the positive effects of native language skills on labor market outcomes specifically (Aldashev et al. 2009; Bleakley and Chin 2004; Chiswick and Miller 1995; Dustmann 1994; Dustmann and Fabbri 2003; Dustmann and Van Soest 2001). Some works also observe societal integration (Danzer and Yaman 2013; Gambaro et al. 2021), and well-being (Battisti et al. 2022). Hence, I focus on a host of integration outcomes – i.e. labor market outcomes, financial worries, health status, staying intentions, and German language skills.

My results show that both on the extensive and intensive margin the schooling of the oldest child in the household positively affects parental labor market outcomes. It increases labor market participation – i.e. parents' probability to be in regular employment and their weekly working hours – as well as monthly income and hourly wages. These effects are especially strong among the formerly unemployed and those who carry the main burden of childcare in the household, which are primarily the mothers. Apart from labor market outcomes, I find positive effects of the oldest child's schooling on parents' financial worries, health status, staying intentions and self-assessed German language skills in speaking, reading and writing.

Capturing different dimensions of integration allows the analysis of mechanisms underlying the integration process in some detail. I assess two potential channels: time and exposure. The first is based on the assumption that upon enrollment of their oldest child, parents have more disposable time on hand which they can then use to actively work on their integration (e.g. by attending language courses) or to participate in the labor market. The second relates to the idea that children's school attendance entails exposure to the German language and culture for the parents. This happens through personal contact to teachers, administrators and other children and parents, and indirect contact to the language since their children learn and speak the German language in school. Migrants are likely to benefit from this exposure. In my analysis I find evidence that both channels play a role in shaping integration outcomes. Disposable time clearly plays an important role and I find that upon school enrollment of the oldest child parents use their gained time to increase their labor market participation, which in turn increases their income. Yet, the observed increase in weekly working hours exceeds the decrease in hours spent with childcare daily. In addition, other outcomes like improved health status, staying intentions and German language skills cannot clearly be attributed to a time effect.

The rest of the paper is organized as follows. Section 5.2 provides institutional background on the German primary schooling system and describes the research design and data. Section 5.3 presents main results, Section 5.3.1 examines potential mechanisms and Section 5.3.2 discusses limitation of the analysis and provides robustness checks. Section 5.4 concludes.

5.2 Institutional Background, Estimation Strategy and Data

The aim of this paper is to estimate the effect schooling of the oldest child in the household has on the integration outcomes of the schooled child's parents. This can be challenging given that migrant parents differ in their ability and willingness to integrate, depending on their education and skills but also social embeddedness. To overcome those sources of potential bias, I exploit age-at-enrollment policies in the 16 German states as exogenous source of variation for school enrollment timing. This allows me to investigate the effect of schooling on parents' integration both on the extensive margin (i.e. the effect of early school enrollment) and on the intensive margin (i.e. the effect of one additional month of schooling).[4]

5.2.1 Institutional Background

In Germany schooling is free and compulsory. From the age of six up to the age of 18 (age of legal majority) children are officially obliged to attend school. This includes primary and secondary school and, after finishing secondary education, vocational school. Parents have to ensure that their child fulfills their obligation to attend school and otherwise face legal consequences – penalty fees and in some states even prison sentences up to 6 months. The exact length of compulsory schooling as well as its start is subject to states (*Bundesländer*) legislation.

In each of Germany's 16 states, the start of compulsory schooling is defined relative to a cutoff date. While these cutoff dates differ between states, they all follow the same general rule: children who turn six on or before the cutoff date of the state they regularly reside in are admitted to primary school in the respective school year. Children who turn six after the cutoff date are admitted to primary school one year later. The start of the school year itself differs between states, too, but is usually between the end of July and the middle of September.

In addition, there is some basic maturity test administered to all children who are about to enter school. Based on this test, school enrollment can be postponed by one year even if children are born before the cutoff date. Postponement can also happen upon the parents' specific wish that their child be enrolled a year later. According to the parents' wish, children can also be admitted prematurely if they are born after the cutoff date. Parents might bring forward enrollment if their child shows signs of learning potential that exceeds their age cohort average (Angrist and Krueger 1992). On the other hand, they might postpone enrollment because they feel their child lacks the necessary maturity for enrollment (*absolute age effect*, see e.g. DiPasquale et al. 1980; Fredriksson and Öckert 2014) or to give their child a comparative advantage over their (then younger) classmates

[4]"Early" school enrollment here refers to children who are eligible for school enrollment the year they turn six; not enrollment which has been brought forward despite the child only being eligible for enrollment the following year. I.e. it refers to children who were enrolled one year earlier than their peers of comparable age who were not eligible for enrollment the year they turned six; not children who were enrolled early in their lifetime.

(*relative age effect*, see e.g. Deming and Dynarski 2008).[5] This might lead to distortion in the identification strategy and will be addressed in Section 5.2.3.[6].

5.2.2 Data

For this analysis I match administrative records on primary school enrollment cutoff dates with micro data from the German Socioeconomic Panel (GSOEP). The enrollment cutoff dates are obtained from the Journals of Laws and Ordinances (*Gesetz- und Verordnungs-blätter*) of the respective states and are available for all 16 states from the year 1992 on. The cutoff dates vary, depending on state and year, between the last day of June and the last day of December. The German Socioeconomic Panel is a longitudinal household survey across all 16 German states (Goebel et al. 2019). It provides yearly information on households and all individual household members since 1984. In addition to information on migration background and state of residence as well as birth dates and enrollment years of children, it offers a wide variety of questions on sociodemographic status and integration outcomes. For my sample I utilize the GSOEP waves of 1992 to 2020 (the first year for which I can provide complete records on school enrollment cutoffs for all 16 states up to the last currently available year). I identify all adult migrants (i.e. individuals who are not born in Germany and have migrated to Germany at age 16 or above) for whom data on their oldest child's birth date and actual school enrollment is available.

I focus on the oldest child in the household since they are naturally the first ones to enter school. The existence of children who had entered school earlier in the household would distort the identification of the effect of school enrollment (extensive margin) on parental integration.[7] The existence of younger children in the household, who potentially visit early childhood education and care (ECEC) facilities, could similarly distort the identification. Hence in the following estimations I control for whether younger children are present in the households and whether they visit ECEC facilities, as well as whether the oldest child has visited ECEC facilities before school enrollment.

Additionally, I run all main estimations on a subset of households with only one child, and find that results are not substantially different to those with (multiple) younger children. To assess the child's enrollment eligibility I compare their birth month and year with the enrollment cutoff date at the state their family resides in the year they turn six.[8] If they were born before or on the respective cutoff date, they are assumed eligible for

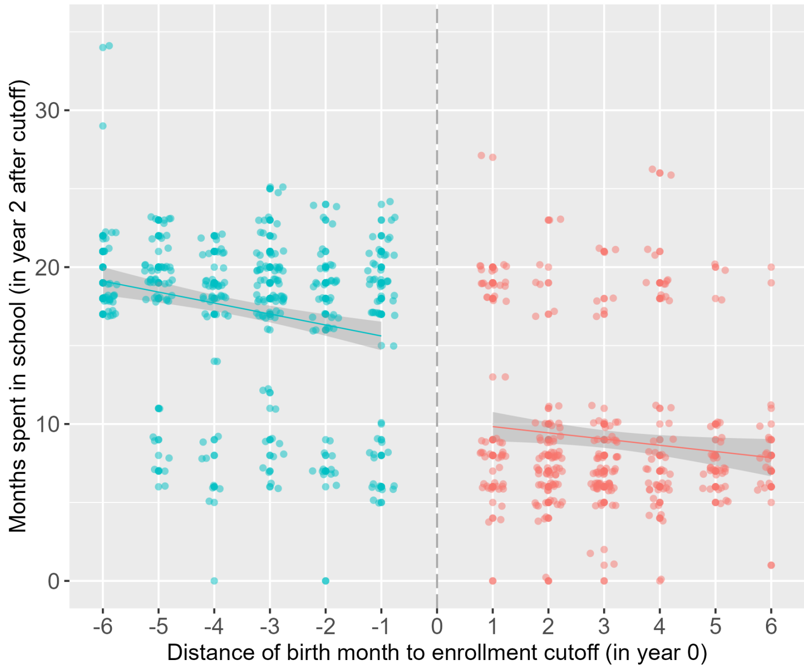
[5] Research on age-at-enrollment has overtly shown that later school entry can raise academic achievement (Black et al. 2011; Cascio and Schanzenbach 2016; McEwan and Shapiro 2008; Puhani and Weber 2008), and even positively affect long-term life outcomes (Bauer and Riphahn 2009; Bedard and Dhuey 2006, 2012; McAdams 2016). Other works, however, show how early school entrance can be beneficial due to longer total schooling (in states where compulsory schooling legally ends after a certain age is reached) or due to peer effects (Currie 2001). Both is of special benefit for disadvantaged pupils who would otherwise have not spend so much time in a positive learning environment, like migrants (Schneeweis 2006).

[6] For a detailed description of the German school system see Lohmar and Eckhardt (2014).

[7] There are no households in the sample for which a younger child is enrolled earlier than the oldest child.

[8] Since birth month and year of children is most commonly provided while exact birth day is not, I utilize monthly cutoffs. This does not reduce the precision of treatment assignment since all enrollment cutoff dates in all states and years observed relate to the first or last day of a respective month.

Figure 5.1: *Discontinuity in Months Spent in School (in Year 2 After Cutoff)*



Note: Number of months spent in school in year 2 after initial cutoff by eligibility for enrollment in year 0 (no = red, yes = blue). Vertical and horizontal noise added to avoid overplotting.

enrollment in this year, and if they were born after the cutoff they are assumed eligible for enrollment in the following year. Lastly, I eliminate all individuals with missings in relevant variables. This yields 678 individuals in 473 households.[9]

Figure 5.1 plots the average months spent in school by the oldest child two years after they turned six (*initial cutoff*) against their birth month distance to the enrollment cutoff. Negative distance means they are born before the cutoff and hence were eligible for enrollment the year they turned six (year 0), positive distance means they were born after the cutoff and hence were not eligible for enrollment the year they turned six (year 0), but only one year later (year 1). E.g. if an oldest child in a given household turned six on the June 15th and the cutoff date in their state if residence was June 30th, they would be eligible for enrollment in this year (year 0), and had a distance to the cutoff of -1 in the graph. Had they been born on July 1st of the respective year in the same state, they would not have been eligible for enrollment in the same year (year 0) but only one year later (year 1) and had a distance to the cutoff of +1 in the graph.[10]

The focus on outcomes in the second year after the oldest child turns six (year 2) is explained by one major limitation in the GSOEP data. The yearly surveys of the GSOEP are done during all 12 months of each year, but actual school enrollment, depending

[9]For a step-by-step explanation of sample construction and shrinkage see Table 5.A in the Appendix.

[10]Due to the aforementioned limitations in the data regarding precise birth days the cutoff distance can only relate to full months. I.e. I cannot differentiate whether the child is born on June 1st or June 30th, their cutoff distance will in both cases be -1 if the cutoff date is June 30th. Hence the cutoff distance can take any integer value between -6 and 6, except for 0.

on state, only happens between July and September. Focusing on outcomes two years after the earliest enrollment ensures that the households which children were eligible for enrollment in year 0 had at least one full year of schooling before their outcome is measured. The potential threat to identification the differences in survey months poses and how I deal with it is discussed in Section 5.2.3.

As can be seen in Figure 5.1, there is a considerable discontinuity in months spent in school by the oldest child in year 0 between those eligible for enrollment in year 0 (i.e. born before the cutoff, pictured in blue), and those not eligible (i.e. born after the cutoff, pictured in red). On average, those eligible in year 0 have spent 8.7 more months in school in year 2 than those not eligible in year 0 (17.7 months compared to 9.0 months).

To obtain reliable estimates of effects this difference in oldest child's school enrollment timing and months spent in school has on parental integration outcomes, several assumptions regarding the data need to hold. First, birth dates of the oldest children and consequently their enrollment eligibility should be exogenously given. While there has been some discussion on potential correlation between children's birth month and parental characteristics, the assumption that children's birth dates are exogenously given is widely used in economic literature, especially regarding the effects of age at school entry (Angrist and Krueger 1992) and policy changes that affect only children born after a certain cutoff and their parents (Cygan-Rehm et al. 2018; Danzer and Lavy 2018; Dustmann and Schönberg 2012).

Second, assuming that the children's birth dates are exogenously given, parents should not differ in their characteristics except for their oldest child's enrollment eligibility. Hence, I compare the averages in parental sociodemographic characteristics and observed integration outcomes between both groups prior to any school enrollment of the oldest child in the household. Due to the data structure of the GSOEP, households are potentially surveyed after the enrollment of their oldest child in year 0, which could distort the results if I used year 0 as control year. Therefore I use one year before the initial enrollment cutoff as control year. This ensures that all observed households are surveyed strictly before the enrollment of their oldest child. As Table 5.1 shows, their differences are mostly negligible, except for a slightly higher average years of education among the group assigned treatment. This is to be expected given that children who are eligible for enrollment at age six tend to be slightly olden than those who are not. Also, the percentage of children being in ECEC facilities, such as kindergartens, before enrollment is slightly higher among the eligible. This differences in age of children and their earlier ECEC attendance could pose a threat to identification, which will be addressed in Section 5.2.3.

Third, the *common trend assumption* should hold, i.e. in the absence of treatment (here: eligibility for enrollment in year 0), the difference between the treated and not treated should be constant over time. Though I cannot statistically test this assumption, I can examine time points before the initial enrollment cutoff and extrapolate that the outcomes would have followed the same trajectory if it weren't for the treatment. Hence, prior to any effects of school enrollment of the oldest child in the household, the observed parental integration outcomes should be comparable between groups (eligible and not

Table 5.1: *Differences in Parental Characteristics and Integration Outcomes between Parents whose Oldest Child was Eligible and Not Eligible for Enrollment the Year They Turned Six*

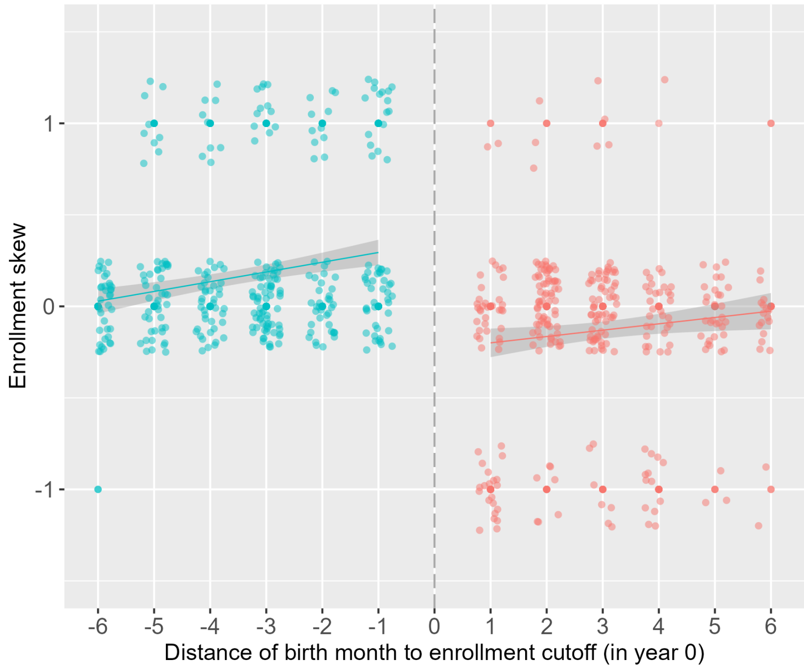
	Not eligible *Control*	Eligible *Treatment*	Difference
Parental characteristics			
Age	32.5	32.8	0.3
Female (%)	52.5	55.6	3.1
Years of education	10.6	10.9	0.3*
Currently in parental leave (%)	9.3	10.7	1.4
Owner of housing (%)	24.8	23.6	−1.2
Refugee (%)	5.0	6.5	1.5
Years since migration	15.3	15.1	−0.2
Number of children	1.7	1.7	0.0
Younger children in ECEC (%)	12.4	16.6	4.2
Oldest in ECEC before enrollment (%)	84.8	91.0	5.2**
Parental integration outcomes			
Monthly parental income (Euro)	1,128.8	1,079.3	−49.5
Currently employed (%)	62.1	56.5	−5.6
Working hours per week	23.9	22.3	−1.6
Hourly net wage (Euro)	7.0	6.7	−0.3
Childcare hours per day	4.7	5.0	0.3
Worried about own finances (scale 1-3)	2.1	2.2	0.1
Staying intention (%)	69.3	75.1	5.8
Number of observations	322	356	678

Note: Means 1 year before initial enrollment cutoff. Unpaired two-sample Wilcoxon tests with potential unequal variance in both samples for differences in variables between groups. For detailed variable description see Table 5.B in the Appendix. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

eligible). Figure 5.A in the Appendix plots all observed parental integration outcomes over several time points and shows that there are indeed no statistically significant differences between both groups.

Fourth, as established in Section 5.2.1, parents have the choice to bring forward or postpone the enrollment of their child, contrary to their state-mandated enrollment eligibility. This introduces non-compliance with the treatment assignment (eligibility for enrollment the year the child turns six) into the data. I check whether the oldest child in the household has actually been enrolled according to their eligibility or whether their enrollment has been brought forward or postponed – i.e. whether each specific household complied with the treatment assignment. Figure 5.B in the Appendix shows that there is non-compliance with the treatment assignment in the data. 8.3 % of parents bring forward the enrollment of their oldest child by one year, and 11.0 % postpone it. Altogether, 19.3 % of parents do not comply with the treatment assignment. There is a discontinuity in the actual enrollment around the enrollment cutoff. As Figure 5.2 shows, bringing

Figure 5.2: *Discontinuity in Actual Enrollment in Year 0 (Treatment Assignment)*



Note: Skew of actual enrollment by eligibility for enrollment in year 0 (no = red, yes = blue). Enrollment skew of -1 means enrollment has been brought forward by 1 year, enrollment skew of +1 means enrollment has been postponed by 1 year. Vertical and horizontal noise added to avoid overplotting.

forward the oldest child's enrollment (enrollment skew of -1) is more likely for children closer to the cutoff. This is not surprising, as children closer to the cutoff are older than those farther away from the cutoff. Hence their parents might feel that they are ready to enter school even though they were not born before the state-mandated birth date cutoff. Vice versa, children born before the state-mandated cutoff but quite close to it are more likely to be enrolled one year later. Their parents might feel that their children are not yet ready for school despite being eligible for enrollment and postpone their enrollment by one year.

Since parents have this choice regarding enrollment, students with brought forward or delayed school entry are not randomly selected. Naturally, this raises the question whether parents who decide to deviate from the state-mandated enrollment eligibility of their child, differ from parents who enroll their child in accordance with state-mandated eligibility in relevant characteristics. Within the migrant sample I compare parents who have brought forward or postponed their oldest' enrollment with those who have not made use of this option, using unpaired two-sample Wilcoxon tests. Table 5.C in the Appendix shows that there is not much difference between the groups. Migrant parents with more education and those who have come to Germany more recently seem to make use of the option to bring forward or postpone the enrollment of their child more often. Additionally, those who did not enroll their oldest child according to eligibility have less monthly net income. Non-compliance with the treatment assignment (eligibility for enrollment in year 0) poses a threat to identification of a treatment effect. Hence, in the estimation strategy in the following Section 5.2.3 this will be addressed.

5.2.3 Estimation Strategy

Utilizing those age-at-enrollment policies in the German states to introduce exogenous variation in the timing of school entry, I estimate several equations. On the intensive margin, I estimate the effect of one additional month of schooling. For this, I first run a naive linear regression via OLS method

$$\begin{aligned}\text{Integration Outcome}_{iht} = \alpha_1 + \beta_1 M_{ht} + \gamma_1 C_{ht} + \tau_1 T_t + \phi_1 P_i \\ + \omega_1 (S_{ht} \times W_{iht}) + \epsilon_{1,iht}\end{aligned} \tag{5.1}$$

where $\text{Integration Outcome}_{iht}$ denotes the integration outcome of parent i in household h in year t. Parent i is either of the parents of the oldest child in the household h. Integration is displayed in different aspects of individuals' lives, hence observed outcomes are parental monthly income, parental employment, working hours per week, hourly income, hours spent with childcare per day, worries about personal finances, health status, staying intentions and German language skills.[11] M_{ht} is the number of months spent in school by the oldest child of household h in year t. C_{ht} are time-variant controls at household h level at time t. T_t are time fixed effects to control for heterogeneity in observational years and P_i are time-invariant individual fixed effects. The latter also cover country of origin fixed effects to ensure that results are not driven by factors related to origin countries. $S_{ht} \times W_{iht}$ is an interaction between the state in which the household h resides in year t and the month in which parent i in household h was surveyed in year t.

The state fixed effect accounts for heterogeneity in institutional factors between German states. The interaction term is added to account for the heterogeneity in months spent in school between oldest children of different households and states. As mentioned earlier, in the GSOEP data survey months differ both between households h and between years t within households. As such, even if two households reside in the same state and their oldest child was enrolled in the same year, depending on survey month their oldest children might have spent different numbers of months in schooling at time t. For a detailed description of all variables see Table 5.B in the Appendix.

In this Equation 5.1 β_1 is the estimated effect of an additional month of schooling of the oldest child in the household on parental integration outcomes. However, the number of months the oldest child has spent in school is not exogenously given but driven by the enrollment timing. This, in turn, is determined by the exogenous variation in birth month distance to enrollment cutoff which predicts eligibility for enrollment, and unobservable factors driving parental discretion to enroll their children in accordance with eligibility or not. To exploit the exogenous variation in birth month distance to enrollment cutoff, I utilize a two stage least squares (2SLS) method. In the first stage months of schooling

[11]Contrary to the other outcomes are parental employment and staying intentions not linear but binary outcomes, yet are estimated via linear regression.

M_{ht} is instrumented by enrollment eligibility E_h

$$\widehat{M_{ht}} = \alpha_{21} + \zeta_{21} E_h + \gamma_{21} C_{ht} + \tau_{21} T_t + \phi_{21} P_i + \omega_{21}(S_{ht} \times W_{iht}) + \epsilon_{21,iht} \tag{5.2.1}$$

where E_h is a dummy variable which takes a value of 1 if the oldest child in household h was eligible for school enrollment the year they turned six (year 0), i.e. if they were assigned the treatment; and a value of 0 if the oldest child was not eligible for school enrollment the year they turned six (year 0), i.e. they were not assigned the treatment. Then the fitted values of $\widehat{M_{ht}}$ are plugged into the second stage of the 2SLS equation

$$\text{Integration Outcome}_{iht} = \alpha_{22} + \beta_{22}\widehat{M_{ht}} + \gamma_{22} C_{ht} + \tau_{22} T_t + \phi_{22} P_i + \omega_{22}(S_{ht} \times W_{iht}) + \epsilon_{22,iht} \tag{5.2.2}$$

where β_{22} is the estimate of the causal effect of one additional month of schooling of the oldest child in the household on parental integration outcomes.

On the extensive margin, I estimate the effect of early school enrollment of the oldest child in the household on parental integration. For this, I run a difference-in-differences (DID) regression

$$\text{Integration Outcome}_{iht} = \alpha_3 + \zeta_3 E_h + \tau_3 T_t + \delta_3(E_h \times T_t) + \gamma_3 C_{ht} + \phi_3 P_i + \omega_3(S_{ht} \times W_{iht}) + \epsilon_{3,iht} \tag{5.3}$$

where ζ_3 is the coefficient of the treatment fixed effect and τ_3 the coefficient of the time fixed effect. Since there are non-compliers in the data, the estimated treatment effect δ_3 only captures the average causal effect of the treatment (enrollment eligibility) for those who comply with the treatment assignment mechanisms, i.e. those who enroll their child according to state-mandated eligibility (compare Angrist and Pischke 2009; Imbens and Angrist 1994). Hence, δ_3 estimates the intention to treat effect (ITT).

In order to estimate an average treatment effect, the non-compliance with the treatment assignment in the data must be accounted for. As in the 2SLS approach, this is done via instrumental variable regression. Except now treatment status D_{ih} (being enrolled in year 0) is instrumented by treatment assignment E_h (being eligible for enrollment in year 0) in a fuzzy regression discontinuity design (RDD). RDD approaches generally exploit changes in treatment status at a certain observable cutoff point. Different from the sharp RDD approach, in which the treatment status is perfectly determined by a certain cutoff point (i.e. perfect compliance with treatment assignment), in the fuzzy RDD case treatment status D_{ih} is not deterministically related to the threshold-crossing of a certain cutoff. In the data, there is a jump in the probability of treatment D_{ih} (i.e. being enrolled in year 0) at the birth month cutoff $x_h = 0$. Before this cutoff $(x_h < x_h = 0)$ the oldest child in the household is eligible for enrollment in the given year, and after this cutoff

$(x_h > x_h = 0)$ they are eligible for enrollment only in the next year, such that

$$P(D_{ih} = 1|x_h) = \begin{cases} g_1(x_h) & \text{if } x_h < \ x_h = 0 \\ g_0(x_h) & \text{if } x_h > \ x_h = 0 \end{cases} \tag{5.4.0}$$

where $g_1(x_h = 0) \neq \ g_0(x_h = 0)$. Functions $g_0(x_h)$ and $g_1(x_h)$ differ. $g_1(x_h = 0) > g_0(x_h = 0)$ is assumed, such that $x_h < \ x_0$ makes treatment more likely. The dummy variable $E_h = 1$ for $x_h < \ x_h = 0$. It indicates the point where treatment D_{ih} dependent on x_h is discontinuous. Using only E_h as an instrument for treatment status D_{ih} leads to the first stage

$$\widehat{D_{ih}} = \alpha_{41} + \pi E_h + \epsilon_{41,ih} \tag{5.4.1}$$

where π is the first stage effect of E_h. The fitted values for D_{ih} from the first stage are then plugged into the second stage estimation

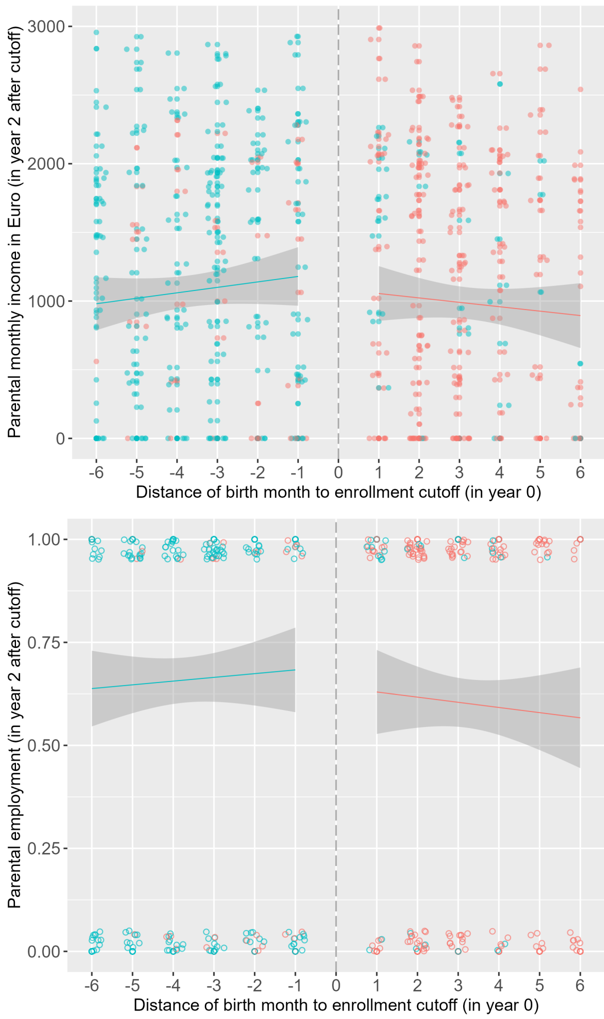
$$\text{Integration Outcome}_{iht} = \alpha_{42} + \eta x_h + \rho \widehat{D_{ih}} + \epsilon_{42,ih} \tag{5.4.2}$$

where ρ is the estimated local average treatment effect (LATE) for compliers in the bandwidth around the cutoff. This means it captures the causal effect of the treatment for those who comply with the treatment assignment mechanisms, i.e. those who enroll their child according to state-mandated eligibility, within the observed bandwidth (compare to Angrist and Pischke 2009; Imbens and Angrist 1994). Focusing on a narrow bandwidth around the treatment assigning cutoff has a certain advantage. In Section 5.2.2 I showed that the parents assigned treatment and those not assigned treatment did not differ much. Any last concerns regarding differences between both groups, i.e. differences in years of education and whether the oldest child was in ECEC before enrollment, can be ruled out once the bandwidth around the cutoff is below ± 4 (4 months left and right to the cutoff included, see Table 5.D in the Appendix). Concerns that the enrolled children differ in their age (given that children who are eligible for enrollment at age six are on average a bit older than those who are not) can also be ruled out by decreasing the observed bandwidth. Hence I will adjust the bandwidth around the enrollment cutoff accordingly in the following estimations in Section 5.3 to reduce the risk of potential confounders.

5.2.4 Descriptive Analysis

As established before, migrant parents are likely to profit from the exposure to the German language and culture the school attendance of their child entails. Direct personal contact to teachers, administrators as well as other children and parents offer the opportunity to build social networks and practice the local language. Further, indirect contact to the culture and language via community meetings (e.g. parents' evenings and school trips) and their children learning the language fosters cultural assimilation and develops language skills (Avitabile et al. 2013; Dustmann 1996). Lastly, school enrollment has a considerable effect on the disposable time of parents, who can use the gained time to actively work on their integration process (e.g. though language courses) or participate in the labor market (compare with Müller and Wrohlich (2020) who makes the same argument regarding early childhood education and care (ECEC)). The latter, labor market participation, in turn brings more exposure to German language and culture through direct contacts to coworkers and supervisors. And contacts to natives have been shown to foster assimilation (Facchini et al. 2015; Martinovic et al. 2009, 2015). Hence, schooling of the oldest child in the household can be expected to positively affect parents' integration outcomes.

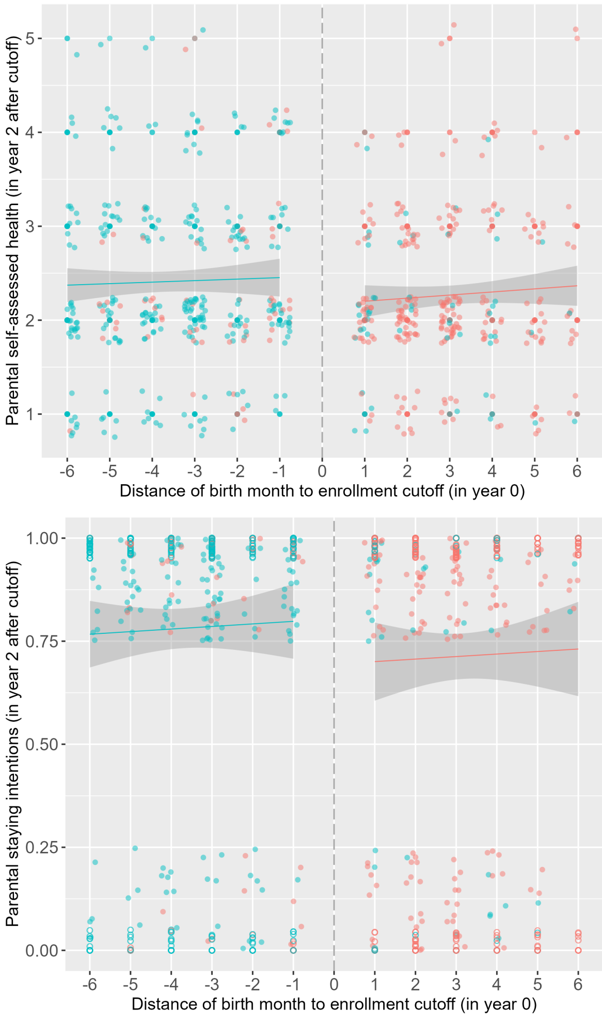
Figure 5.3: *Discontinuity in Parental Monthly Income and Parental Employment (in Year 2 After Cutoff)*



Note: Parental monthly income (top panel) and parental employment (bottom panel) in year 2 after initial cutoff by treatment status (not treated = red line, treated = blue line). Dots indicate whether the oldest child in the household was enrolled in year 0 (not enrolled in year 0 = red dots, enrolled in year 0 = blue dots). Vertical and horizontal noise added to avoid overplotting. Outliers with monthly parental income over 3000 Euro not pictured.

Indeed, in a first descriptive analysis on some selected integration outcomes I see that treated parents (parents whose oldest child was eligible for enrollment and enrolled in the year they turned six) have on average better outcomes than parents whose oldest child

Figure 5.4: *Discontinuity in Parental Self-Assessed Health and Parental Staying Intentions (in Year 2 After Cutoff)*



Note: Parental self-assessed health on a scale from 1 (bad) to 5 (very good) (top panel) and parental staying intentions (bottom panel) in year 2 after initial cutoff by treatment status (not treated = red line, treated = blue line). Dots indicate whether the oldest child in the household was enrolled in year 0 (not enrolled in year 0 = red dots, enrolled in year 0 = blue dots). Vertical and horizontal noise added to avoid overplotting.

was not eligible for enrollment and not enrolled in the year they turned six (control). E.g., Figure 5.3 plots a discontinuity at the birth date cutoff in the parental monthly income. There is no significant difference in monthly net income before the treatment between treatment and control (1126 Euro on average within the treatment group versus 1085 Euro on average within the control group one year before the initial cutoff), in year 2 after the treatment the income increases more strongly for the treated group (to 1304 Euro on average) compared to the control (to 1150 Euro on average). The difference between years within the control group is not significant (p-value of 0.334), yet within the treatment group it is highly significant (p-value of 0.008).[12] Also, there is substantial

[12]If not stated otherwise, p-values of tests of differences in means refer to p-values obtained from unpaired two-sample Wilcoxon tests with two-sided alternative and potentially unequal variance in both samples.

differences between treatment and control regarding regular employment. The probability to be in regular employment increases within the treatment group by 7 percentage points (from 59 % to 66 %), while in the control this is only 2 percentage points (from 59 % to 61 %).

Figure 5.4 plots discontinuity at the birth date cutoff in parental self-assessed health on a scale from 1 (bad) to 5 (very good). Again, pre-treatment there is no significant difference between treatment and control group. Post-treatment the average health status in the treatment group is 2.40 compared to 2.28 in the control group (and paired two-sample t-tests reveal that this difference is significant). Also, the probability to intend to stay in Germany indefinitely increases within the treatment group by 4 percentage points (from 74 % to 78 %), while in the control this is only 1 percentage points (from 71 % to 72 %).

Since there is non-compliance with the treatment assignment in the data, those purely descriptive results could be partially driven by self-selection into treatment (e.g. parents who integrate more easily to begin with enroll their children according to their eligibility). In the following results in Section 5.3 those concerns will be addresses by applying the estimation strategies introduced earlier in Section 5.2.3. Afterwards, a channel analysis in Section 5.3.1 sheds some more light on potential channels through which parents' integration is affected by schooling.

5.3 Results

First, estimation results regarding parents' labor market outcomes will be presented in the following. In Section 5.3.1 follows an analysis of potential channels through which schooling of the oldest child in the household affects parents' integration outcomes. In addition, a comparison to outcomes among a sample of parents born in Germany is drawn. Lastly, Section 5.3.2 discusses limitations of the data and analysis and the external validity of the results.

The estimated effects on parental monthly net income are shown in Table 5.2. On the intensive margin, the coefficient of the OLS regression in Column (1) shows that one additional month of schooling of the oldest child is positively associated with parental monthly income. However, the OLS regression does not account for potential selection effects (e.g. that parents who are better equipped to integrate to begin with enroll their children into school earlier). To account for selection effects, I estimate a 2SLS regression where in the first stage the months of schooling are instrumented by the enrollment eligibility in year 0 (see Equation 5.2.1 in Section 5.2.3). Column (2) shows that in the first stage enrollment eligibility in year 0 strongly predicts the months of schooling in year 2. A Wald test comparing the model including and excluding the instrument proves instrument relevance (F-statistic of the first stage is highly significant with p-value $< 2.2e^{-16}$). In the second stage the fitted values of stage 1 are plugged into a regression which estimates the causal effect of one additional month of schooling of the oldest child on parental monthly income (see Equation 5.2.2 in Section 5.2.3). As Column (3) shows,

Table 5.2: *Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		$\pm$5 months	$\pm$4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.81^{***}		129.60^{*}		
		(0.46)		(72.78)		
Months in school	7.01^{**}		15.99^{***}			
	(2.89)		(2.80)			
Treated (compliers)					303.26^{**}	472.81^{***}
					(148.74)	(148.91)
R^2	0.91	0.65	0.90	0.92	–	–
Adj. R^2	0.87	0.53	0.87	0.84	–	–
Num. obs.	2712	2712	2712	1356	577	491
Num. individuals	678	678	678	678	577	491
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State $\times$ interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental monthly income in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental monthly income in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

one additional month of schooling of the oldest child increases the parental monthly net income by around 16 Euro and the effect is significant on the 1 % level. On the extensive margin, I estimate the effect of school enrollment of the oldest child in the household on parental monthly income. First, I estimate an intention to treat (ITT) effect of school enrollment eligibility in year 0 (drawn from the difference-in-differences (DID) regression in Equation 5.3 in Section 5.2.3). As Column (4) shows, parents of children who were eligible for enrollment in year 0 have on average around 130 Euro more monthly income in year 2 than those whose children were not eligible for enrollment in year 0. This effect is only slightly significant on the 10 % level and is independent of whether the eligible children were actually enrolled in year 0, i.e. whether the parents complied with the treatment assignment and enrolled their child accordingly. The fuzzy RDD approach (see Equation 5.4.1 and 5.4.2 in Section 5.2.3) accounts for the non-compliance with the treatment assignment in the data. The LATE for compliers with the treatment assignment within a bandwidth of $\pm$5 months around the enrollment cutoff is roughly an additional 303 Euro monthly parental income. Within a smaller bandwidth of $\pm$4 months the LATE is even larger with around 473 Euro additional parental monthly income and significant on the 1 % level.[13]

One additional month of schooling for the oldest child increases the parents' prob-

[13] All following result tables will be build and interpreted like Table 5.2, showing the estimation results for naive OLS, 2SLS, ITT from DID, and Fuzzy RDD approach.

Table 5.3: *Estimates of Months of Schooling and Enrollment Timing on Parental Employment*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.81^{***}		0.06		
		(0.46)		(0.05)		
Months in school	0.00^{*}		0.01^{***}			
	(0.00)		(0.00)			
Treated (compliers)					0.21^{***}	0.31^{***}
					(0.08)	(0.09)
R^2	0.78	0.65	0.77	0.81	–	–
Adj. R^2	0.69	0.53	0.69	0.61	–	–
Num. obs.	2712	2712	2712	1356	577	491
Num. individuals	678	678	678	678	577	491
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental employment in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental employment in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

ability to be regularly employed, as Table 5.3 shows. On the intensive margin, each additional month brings 0.01 higher probability to be in regular employment. On the extensive margin, parents whose oldest child was enrolled one year earlier have a 0.21 to 0.31 higher probability to be employed in year 2 (within a bandwidth of ± 5 and ± 4 around the enrollment cutoff, respectively). All coefficients are significant at the 1 % level.

Both parental employment and monthly income are positively affected by children's schooling. With the improved employment and income situation also comes a better outlook on personal finances. As the fuzzy RDD results of Table 5.4 show, parents whose oldest child was enrolled one year earlier are 0.21 to 0.29 points less worried about their personal financial situation on a 3-point scale (within a bandwidth of ± 5 and ± 4 around the enrollment cutoff, respectively).

5.3.1 Channel Analysis

Given the observed positive effects on the labor market outcomes, the question arises which mechanisms underlie them. Specifically, I want to shed light on two potential channels which might drive my results – a *time effect* and *exposure* to the German language and culture. For the first the assumption is that upon enrollment the hours the oldest child spends at school every weekday becomes disposable to the parent(s), increasing their daily number of disposable hours. They can use this gained time to

Table 5.4: *Estimates of Months of Schooling and Enrollment Timing on Parental Worries about Personal Finances*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.76^{***}		-0.08		
		(0.47)		(0.10)		
Months in school	-0.00		0.00			
	(0.00)		(0.00)			
Treated					-0.21^{**}	-0.29^{***}
					(0.09)	(0.09)
R^2	0.62	0.65	0.51	0.65	–	–
Adj. R^2	0.48	0.53	0.34	0.26	–	–
Num. obs.	2596	2596	2596	1298	549	469
Num. individuals	649	649	649	649	549	469
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental worries about personal finances in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental worries about personal finances in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

actively work on their integration (e.g. through language courses) or participate in the labor market. This channel will be assessed in detail in the first part of this Section. The latter relates to the idea that migrant parents are likely to profit from the exposure to the German language and culture the school attendance of their child entails. This will be assessed in the second part of this Section. Additionally, effects of children's schooling on parental labor market outcomes for German parents are discussed.

Time Effect

School enrollment considerably increases the daily disposable time of parents who can use this gained time to actively work on their integration process (e.g. through language courses) or participate in the labor market.

Indeed, the hours per weekday spent with childcare decrease from an average of 4.86 to 4.42 after school enrollment of the oldest child, which is a significant difference with p-value of 0.011.[14] This decrease in childcare hours differs considerably between genders. For the mothers in the sample the average number of childcare hours per weekday decreases from 6.97 to 6.33 and the difference is statistically significant at

[14]Since all children, independent of their enrollment eligibility in year 0, will not be enrolled in year -1 and be enrolled in school in year 2, I will here and in the following analysis compare year -1 and year 2 averages.

Table 5.5: *Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		± 5 months	± 4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.81^{***}		0.09		
		(0.46)		(0.40)		
Months in school	-0.02		-0.05^{***}			
	(0.02)		(0.01)			
Treated (compliers)					-3.05^{***}	-3.43^{***}
					(0.47)	(0.60)
R^2	0.75	0.65	0.74	0.80	–	–
Adj. R^2	0.65	0.53	0.65	0.59	–	–
Num. obs.	2712	2712	2712	1356	577	491
Num. individuals	678	678	678	678	577	491
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental childcare hours spent per day in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental childcare hours spent per day in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

the 1 % level (p-value of 0.006). On the contrary, the difference for fathers with only 0.22 (decrease from 2.38 to 2.16 hours on average) is much smaller and statistically not significant.[15]

Clearly, the increase in disposable time for both parents is not nearly proportional to the child's time spent at school – assuming a 5 to 6 hours school day during the first and second grades. Parents have to bring and pick up their children from school, prepare lunches, help with homework, and keep contact to teachers and administrators (e.g. via parents' evenings). Also, parents often have more than one child – in the migrant sample 80 % of households have 2 children or more. Since I analyze the school enrollment of the oldest child, potential younger children in the household are not in school, yet. Even though the oldest child spends a lot of time at school daily, younger siblings still need childcare. Apart from this, parents whose children have attended early childhood education and care facilities before primary school enrollment gain less or no additional

[15] In general mothers carry the main burden of childcare. Over all observed years they spend an average of 6.79 hours per weekday with childcare (compared to only 2.27 hours the fathers spend). Since the regularly employed individual in the migrant sample spends an average of 7.95 hours per weekday at work, this is almost equivalent to full-time employment. Do their childcare responsibilities mean that mothers spend more time with children and less time at work overall? In the sample roughly 39 % of mothers are regularly employed and spend an average of 12.52 hours per week at work (or 2.50 hours per day if I assume a 5 day work week). Fathers, on the other hand, have a regular employment share of 86 % and spend an average of 36.20 hours per week at work (7.24 hours per day).

time at all upon school enrollment.[16] All of these factors lead to the rather limited decrease in childcare hours upon enrollment. Nonetheless, the hours spent with childcare daily decrease, and they decrease more strongly for the treated parents. As the 2SLS estimate in Table 5.5 shows, one additional month of schooling reduces the daily time spent with childcare by 0.05 hours on average. On the extensive margin, the LATE is even more pronounced. Parents whose oldest child was enrolled one year earlier spent 3.05 to 3.43 hours less with childcare per day (within a bandwidth of ± 5 and ± 4 around the enrollment cutoff, respectively).

Table 5.6: *Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		± 5 months	± 4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.81^{***}		2.06		
		(0.46)		(1.77)		
Months in school	0.15^{**}		0.22^{***}			
	(0.07)		(0.05)			
Treated (compliers)					5.12^{*}	7.75^{**}
					(2.86)	(3.44)
R^2	0.83	0.65	0.82	0.85	–	–
Adj. R^2	0.76	0.53	0.76	0.68	–	–
Num. obs.	2712	2712	2712	1356	577	491
Num. individuals	678	678	678	678	577	491
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental working hours in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental working hours in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

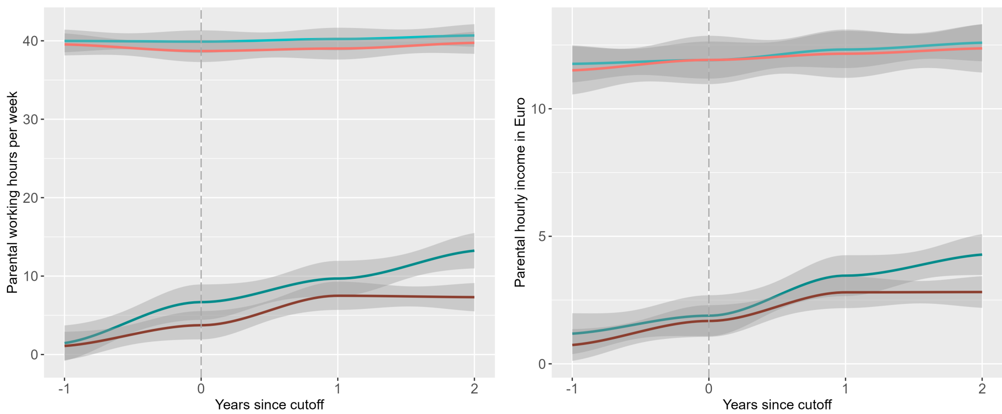
Parents can use this additional disposable time during the weekday to enter the labor market by taking up regular employment or increase their working hours if they were already employed. If the time gained by reduced childcare hours would perfectly translate into increased working hours, the effect of schooling of the oldest child on labor market outcomes would be driven fully by a time effect. Yet, among the treated parents the increase in working hours exceeds the hours gained through less childcare, as Table 5.6 shows. One additional month of schooling increases the parental working hours per week by 0.22 hours, far exceeding the reduce in childcare hours of 0.05 hours shown in Table 5.5. On the extensive margin, parents whose oldest child was enrolled one year

[16]Hence in all regressions I control for whether the oldest child in the household has attended early childhood education and care facilities before enrollment, and whether any younger children attend early childhood education and care facilities.

earlier had on average 5.12 to 7.75 more weekly working hours (within a bandwidth of ± 5 and ± 4 around the enrollment cutoff, respectively). This also exceeds the disposable time gained.

As seen in Table 5.3, there is a significant surge in employment probability among the treated parents. Does this mean that the increase in working hours is mainly driven by job uptake? While working hours among the formerly employed stay largely constant, for the formerly unemployed they increase, which is stronger among the treated parents (see Figure 5.5). This could be because parents who were regularly employed before school enrollment of their oldest child have already hit full-time employment (on average 39.8 hours per week). As a result, there is not much possibility to increase labor market participation for them. As such, the increase in working hours is clearly due to job uptakes after school enrollment, and there is vast difference between treated and control. Prior to treatment, shares of individuals who are not in regular employment are comparable between treated and control (41 % in each). Post-treatment, 2.1 % of the complete control group and 6.8 % of the treated have taken up regular employment. This is also underlined by the positive effect of the oldest child's schooling on parental employment probability among the formerly unemployed shown in Table 5.E in the Appendix.

Figure 5.5: *Means in Parental Working Hours per Week and Parental Hourly Income Over Years*



Note: Parental parental working hours per week (left panel) and parental hourly income (right panel) over the years by treatment status and former employment (not treated = red lines, treated = blue lines, formerly employed = lighter lines, formerly unemployed = darker lines). 95 % confidence intervals.

But can those increased working hours fully explain the increase in individual income that was shown in Table 5.2? The average parental monthly net income increases by approximately 65 Euro (from around 1085 to 1150 Euro) among the not treated parents and the difference is not significant. Among the treated parents the monthly net income increases from around 1126 to 1304 Euro and the difference is highly significant ($p < 0.01$). This depicts an increase of roundly 178 Euro per month, while the average working hours per month (assuming a month with 4.35 working weeks) have increased by 12.18 hours among the treated parents. Given the pre-treatment average hourly income among the treated of 7.00 Euro, the increase in monthly working hours should have been 25.36 to explain the increase in income; a value which is more than double the actual increase.

The increase in individual income cannot be explained by increase in working hours alone. Hence, in the next step I analyze the change in hourly income.

Among the treated parents net income increases by 1.13 Euro per hour on average (from 7.00 to 8.13 Euro) and the difference is highly significant (p-value < 0.01). This is almost double the increase in hourly income among the not treated parents (by 0.72 Euro from 6.70 to 7.42 with a p-value of 0.083). The estimation in Table 5.7 shows that one additional month of schooling of the oldest in the household brings a net income increase of 0.11 Euro per parental working hour. On the extensive margin, parents whose oldest child was enrolled one year earlier have on average 3.62 to 4.35 Euro more hourly income (within a bandwidth of ±5 and ±4 around the enrollment cutoff, respectively). Subsampling reveals that for those who were in regular employment prior to enrollment of their oldest child, hourly income stays mostly constant and there is no significant difference between treated and not treated (see lighter lines in right panel of Figure 5.5). For the formerly unemployed, however, hourly income increases and this is stronger among the treated parents (see darker lines in right panel of Figure 5.5). Hence, like the increase in working hours, the increase in hourly income seems mainly driven by job uptake of the formerly unemployed.

Table 5.7: *Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.81^{***}		0.48		
		(0.46)		(0.65)		
Months in school	0.09^{***}		0.11^{***}			
	(0.02)		(0.02)			
Treated (compliers)					3.62^{***}	4.35^{***}
					(1.37)	(1.22)
R^2	0.78	0.65	0.78	0.85	–	–
Adj. R^2	0.71	0.53	0.71	0.70	0–	–
Num. obs.	2712	2712	2712	1356	577	491
Num. individuals	678	678	678	678	577	491
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State $\times$ interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental hourly income in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental hourly income in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

While the gained disposable time upon school enrollment of the oldest child in the household is limited, the allocation of disposable time throughout the day could be of importance. School attendance frees up time between morning and midday, a time slot that offers good working opportunities as many jobs require attendance during school hours, especially traditional part-time positions. This gives the part of the household which carries the majority of the childcare burden the opportunity to enter the labor market.[17] Not coincidentally, these are mostly the women in the sample.

Women make up the largest share of those not in regular employment (85.0 %) and have only an average of 11.3 weekly working hours prior to the enrollment of their oldest child (compared to 37.0 hours among men). Taking a closer look at gender subsamples, I find that most positive effects of the oldest child's schooling on parental outcomes seem to be largely driven by the women in the sample. As Table 5.G in the Appendix shows, one additional month of schooling of the oldest child brings on average around 17 Euro more monthly income for the mothers. On the extensive margin, mothers whose oldest child was enrolled one year earlier had on average around 484 to 581 Euro more monthly income (within a bandwidth of ± 5 and ± 4 around the enrollment cutoff, respectively).[18] The enrollment probability of mothers increases significantly with their oldest child's schooling (see Table 5.H in the Appendix). As Table 5.I in the Appendix shows, womens' working hours per week increase strongly among the treated. One additional month of schooling of their oldest brings 0.28 more working hours per week. Mothers whose oldest child was enrolled one year earlier had on average 8.64 to 11.72 more weekly working hours (within a bandwidth of ± 5 and ± 4 around the enrollment cutoff, respectively). These increases in working hours are considerably larger in size compared to the whole sample, which is not surprising given the difference of daily childcare hours for women upon enrollment of their oldest child. Table 5.K shows that mothers whose oldest child was enrolled one year earlier spent on average 4.49 to 5.05 less hours per day with childcare (within a bandwidth of ± 5 and ± 4 around the enrollment cutoff, respectively). Each additional month of schooling of their oldest child brings a decrease of 0.08 hours per day spent with childcare for mothers.

Those results are in favor of the argument that for the formerly unemployed time has freed up in the mornings till midday, allowing them to take up jobs (see the substantial effect of early enrollment on treated parents among the formerly unemployed in Table 5.F in the Appendix) – and this applies majorly to the mothers in the sample who are the main carriers of the childcare burden in the household. Consequently, there is strong

[17] Parents have to allocate their time between childcare at home and work. Before primary school enrollment childcare options outside the home are either quite limited in their availability and in the time they free up (e.g. communal kindergartens) or come with additional costs (e.g. privately paid kindergartens, day care centers, nannies). So for parents who carry the main childcare burden in the household (i.e. who are not the breadwinners), their income generated due to the time freed up by outside childcare options has to exceed the amount they spend on these outside options to make employment a financially feasible option. This changes upon school entry, which basically constitutes an outside childcare option that is free of charge (since there are no school fees at German public schools). Hence, it can then be a financially feasible option to work during school times even for parents with low hourly incomes. Also, more flexibility regarding time and place of work offers opportunities for better positions and higher pay.

[18] Tables 5.G to 5.K in the Appendix show the estimation results for labor market outcomes (monthly income, employment, working hours per week, hourly income) as well as childcare hours per day among women.

evidence for a *time effect* of school enrollment. However, since the increase in labor market participation exceeds the gained disposable time, my findings cannot be driven solely by such time effect. In addition, if the outcomes were fully explainable by gained disposable time, parents in households with only one child should have considerably stronger decreases in childcare hours and increases in labor market participation compared to those with several children. Yet, looking at a subsample of parents with only one child in the household, the effects among parents are largely comparable with the whole sample (see Table 5.L to Table 5.P in the Appendix for regression results for parents with only one child). Even with regard to the change in childcare hours, there is not much difference. Though they were less to begin with compared to parents with several children, the daily hours spent with childcare on average decrease by 0.44 hours a day (from 3.68 to 3.28), which is exactly the same amount as the decrease for parents with several children.

Exposure Channel

School enrollment has a considerable effect on the disposable time of parents who can use the additional time to actively work on their integration process (e.g. through language courses) or to participate in the labor market. Besides such gains in disposable time – that can only to some extent explain the outcomes – which mechanisms lead to improved integration outcomes for parents? When it comes to drivers of language acquisition, some authors, like Chiswick and Miller (2005) and Isphording and Otten (2014), differentiate between three ones: economic incentives, exposure, and individual ability. Those can also be applied more broadly to other assimilation measures. I will not focus on the first or last, since I do not expect differences in economic incentives between parents of children depending on different birth dates (treatment assignment). Similarly, differences in individual ability should be controlled for by the identification strategy and individual fixed effects. This leaves me with exposure as a potential channel through which integration happens.

Upon the enrollment of their oldest child into primary school, parents enter a higher level of exposure to the German language and culture. Direct personal contact to teachers, administrators as well as other children and parents offers the opportunity to build social networks and practice the local language. Further, indirect contact to the culture and language via community meetings (e.g. parents' evenings and school trips) and their children's language acquisition can promote cultural assimilation and language skills (Avitabile et al. 2013; Dustmann 1996). Regular contacts to natives overall have been shown to support assimilation (Facchini et al. 2015; Martinovic et al. 2009, 2015). As such, migrant parents are likely to profit from their children's school enrollment in their whole integration process – and not only regarding labor market outcomes. Integration is a complex, multidimensional process, spanning economic outcomes (Aldashev et al. 2009; Bleakley and Chin 2004; Chiswick and Miller 1995; Dustmann 1994; Dustmann and Van Soest 2001) as well as social and cultural assimilation (Constant and Zimmermann 2008; Danzer and Yaman 2013; Facchini et al. 2015; Gambaro et al. 2021) and well-being (Battisti et al. 2022). Hence, in the following I analyze the effect of exposure to the German language and culture via the oldest child's schooling on several parental integration measures, namely their self-assessed German language knowledge, self-assessed health status and staying intentions.

While health might not be understood as a traditional integration outcome in itself, it very well can be a proxy for overall well-being. Table 5.8 shows that on a self-assessed health scale which runs from 1 (very bad) to 5 (very good), parents whose oldest child was enrolled one year earlier report a 0.70 to 0.86 points higher health status (within a bandwidth of ± 5 and ± 4, respectively).

Table 5.8: *Estimates of Months of Schooling and Enrollment Timing on Parental Self-Assessed Health*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		± 5 months	± 4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.16^{***}		0.02		
		(0.48)		(0.12)		
Months in school	0.00		0.00			
	(0.00)		(0.00)			
Treated (compliers)					0.70^{***}	0.86^{***}
					(0.21)	(0.17)
R^2	0.60	0.66	0.59	0.70	–	–
Adj. R^2	0.45	0.54	0.45	0.36	–	–
Num. obs.	2304	2304	2304	1152	488	419
Num. individuals	576	576	576	576	488	419
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental self-assessed health in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental self-assessed health in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

When it comes to the intention to stay in Germany indefinitely, parents whose oldest child was enrolled one year earlier have on average a 0.15 to 0.22 higher probability to have permanent staying intentions (within a bandwidth of ± 5 and ± 4, respectively). However, the results are only significant within a bandwidth of ± 4 around the enrollment cutoff (see Table 5.9).[19]

Migrant parents can profit from exposure to the German language in their own German skills. In addition to the direct contact to teachers, administrators and other parents, schooling also brings indirect contact to the German language. In their everyday lives migrant parents might not need their German language skills too often. But once their oldest child enters school, their child not only learns to write and read the German

[19] I find no significant effect of the enrollment of the oldest child in the household on the parents' probability of living in government subsidized housing and their overall life satisfaction (not reported here).

Table 5.9: *Estimates of Months of Schooling and Enrollment Timing on Parental Staying Intentions*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.69***		0.00		
		(0.46)		(0.06)		
Months in school	0.00*		0.00			
	(0.00)		(0.00)			
Treated (compliers)					0.15	0.22**
					(0.09)	(0.11)
R^2	0.74	0.65	0.73	0.75	–	–
Adj. R^2	0.64	0.57	0.64	0.49	–	–
Num. obs.	2552	2552	2552	1276	544	460
Num. individuals	638	638	638	638	544	460
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental staying intentions in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental staying intentions in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

language, but also learns all other subjects in German, and might need help with homework assignments. Indeed, Table 5.10 shows that treated parents report better German speaking, reading and writing skills on a scale from 0 (no knowledge) to 4 (very good knowledge). Parents whose oldest child was enrolled one year earlier report significantly higher knowledge: 0.53 to 0.66 points higher speaking skills, 0.78 to 0.85 points higher reading skills, and 0.77 to 0.89 points higher writing skills (within a bandwidth of ±5 and ±4, respectively). Additionally, I build an index for German language ability in which I combine the values on speaking, reading and writing by equal weights. Treated parents report on average 0.69 to 0.80 points higher German skills on this index on a scale from 0 to 4 (within a bandwidth of ±5 and ±4, respectively).[20]

Overall, there is some evidence on the role of exposure in everyday life on parents' integration outcomes. Exposure to Germans through personal contact to teachers, administrators, other children and parents offers the opportunity to practice the German language and build social networks. Those are particularly important for job search and promotion opportunities (Aldashev et al. 2009; Bleakley and Chin 2004; Chiswick and

[20]It has to be noted that the number of observations is limited due to the periodicity of the language skills questions, which were only included in alternate survey years until 2007 and were not asked in 2012. Thus, the analysis is based on a sample of 437 individuals from 310 households who provided information on their German language skills for both time periods and I only observe year 2 after the initial enrollment cutoff. This number further decreases by narrowing the bandwidth to ±5 and ±4 months around the enrollment cutoff.

Table 5.10: *Estimates of Enrollment Timing on Parental Self-Assessed German Language Abilities*

	Fuzzy RDD							
	German index		German speaking		German reading		German writing	
	±5 months	±4 months	±5 months	±4 months	±5 months	±4 months	±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated (compliers)	0.69^{***}	0.80^{***}	0.53^{***}	0.66^{***}	0.78^{***}	0.85^{***}	0.77^{***}	0.89^{***}
	(0.19)	(0.21)	(0.19)	(0.21)	(0.18)	(0.19)	(0.24)	(0.25)
Num. obs.	383	320	383	320	383	320	383	320
Num. individuals	383	320	383	320	383	320	383	320

Note: Fuzzy RDD LATE estimates of enrollment in year 0 on parental self-assessed German language knowledge in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on level of cutoff distance for discontinuity samples. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Miller 1995; Dustmann 1994; Dustmann and Fabbri 2003; Dustmann and Van Soest 2001). Increased labor market participation in turn creates more exposure to German language and culture through direct contacts to coworkers, supervisors and customers. Thus, the child's school entrance can foster a circle of exposure for the parents, which in turn fosters assimilation (Facchini et al. 2015; Martinovic et al. 2009, 2015). This is especially important for mothers, who have much lower labor market participation rates before school enrollment of their oldest child. Since they weren't in regular employment prior to school enrollment, they have not been subject to the German language and culture – at least not through their job. If they do not maintain regular contact to natives outside of work, they establish regular contact to Germans only upon school enrollment. This can potentially explain the strong labor market effects of the child's school enrollment among formerly unemployed parents and mothers.

Comparison to German Parents

Another interesting aspect to shed light on is whether schooling of the oldest child also has effects on German parents' labor market outcomes and other aspects. For this, I identify German individuals[21] residing in households with an oldest child for whom I have information on their enrollment eligibility and actual enrollment year. As in the migrant sample, I eliminate all individuals with missings in relevant variables. This yields 3137 individuals in 1996 households. Tables 5.Q to 5.U in the Appendix show the same estimations on labor market outcomes and childcare hours as estimated for the migrant sample.

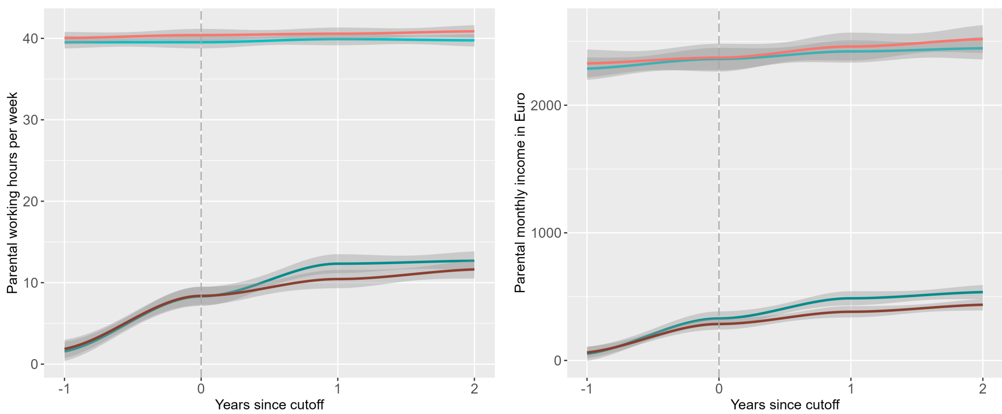
The first stage of the 2SLS regression shows that enrollment eligibility in year 0 strongly predicts the months of schooling in year 2 (see Column (2) of Table 5.Q in the Appendix). One additional month of schooling of the oldest child increases parental monthly income by around 12 Euro. Additionally, it increases parents' working hours by 0.15 hours per week and the hourly income by 0.08 Euro (see Table 5.S and Table 5.T in the Appendix, respectively). Therefore, on the intensive margin the results among Germans are largely comparable to the migrant sample, though smaller in magnitude. A

[21] Individuals who were born in Germany.

possible explanation for that is that native Germans, contrary to migrants, have easier access to ECEC. Indeed, the share of parents whose oldest child was in ECEC before school enrollment is higher among Germans (94.2 %) than among migrants (87.7 %). As such, school entrance of the oldest child in the household is a large positive shock in disposable time for fewer Germans than migrant parents.

Still, like among migrants, the increased labor market participation and returns are largely driven by those who were not in regular employment prior to the school enrollment of their oldest child. And just like among migrants, the vast majority of parents who were not in regular employment before enrollment of their oldest child are women (91.8 % of the formerly unemployed). Among the formerly employed German parents there is not much possibility to increase labor market participation, as they worked on average 39.8 hours per week before school enrollment of their oldest child – which is equivalent to full-time employment. Therefore, the increase in working hours upon school enrollment is due to job uptakes among the formerly unemployed, i.e. mothers. As the lighter lines in the left panel of Figure 5.6 show, working hours among those who were employed before stay constant. The darker lines show that working hours have increased for those who were not in regular employment in year -1. The right panel of Figure 5.6 shows that the same pattern applies to monthly income.

Figure 5.6: *Means in Parental Working Hours per Week and Parental Monthly Net Income Over Years (Germans)*



Note: German sample. Parental working hours per week (left panel) and parental monthly net income (right panel) over the years by treatment status and former employment (not treated = red lines, treated = blue lines, formerly employed = lighter lines, formerly unemployed = darker lines). 95 % confidence intervals.

The gender split of childcare is not substantially different in the German sample compared to the migrant sample. Carrying the main burden of childcare in the household, German mothers spend an average of 7.48 hours per weekday with childcare (compared to the 2.13 hours the fathers spend) – almost equivalent to the 8.01 hours the regularly employed individual in the German sample spends at work per weekday. They reduce their childcare time upon enrollment of their oldest child by 1.04 hours per day, a highly significant difference (p-value $< 2.2e^{-16}$). For German fathers, the reduction in childcare hours is almost negligible (0.03 hours on average) and not significant.

Interestingly, Figure 5.6 also shows that the effects on income and working hours among the treated parents, though small, already realize in the first year after the initial enrollment. Parents whose oldest child was enrolled one year earlier see no significant effects regarding their labor market outcomes in the second year (see LATE from RDD regressions in Tables 5.Q to 5.T in the Appendix, which are based on year 2 outcomes and are all insignificant). It seems that labor market returns upon school enrollment of the oldest child in the household realize faster among German parents compared to migrant parents. Formerly unemployed German parents seem to get into employment rather quickly after the surge in disposable time, while for migrant parents job search seems to take longer. The latter experience labor market returns of increased disposable time only in the second year after the initial enrollment. This is not surprising, as migrant parents can use their gained disposable time to improve their own value on the German labor market and hence employment opportunities, e.g. by attending German language courses and building networks. Then they can enter the labor market with increased employment chances and higher potential wages. In conclusion, since integration is a gradual process, it is expected to take some time to fully manifest in the outcome variables. In addition, in the German sample – though it is much larger and thus even smaller effect sizes should be identified – I cannot identify effects of schooling on parental financial worries or health status (not reported here). All of this evidence suggests that schooling of the oldest child in the household drives integration among migrant parents, which relies on mechanisms besides the effect of disposable time.

5.3.2 Threats to Identification and Sensitivity Analysis

In the following some major threats to identification and how they are dealt with, as well as additional robustness checks, will be discussed.

Identification of causal effects in DID and fuzzy RDD approaches relies on several assumptions. First, there should be no manipulation, i.e. individuals should not display sorting on the enrollment cutoff distance. McCrary density test is used to check whether there is bunching of units at the cutoff (McCrary 2008). Under the null hypothesis, the density should be continuous at the cutoff point and under the alternative hypothesis, the density should increase at the cutoff point. The null is rejected (at a p-value of 0.026), so there is some evidence for manipulation. Yet, since children's birth dates are randomly distributed, there is no reason to assume manipulation around the cutoff. A more likely explanation for the observation of bunching at the cutoff is that there are too little observations in the sample to distinguish a discontinuity in the density from noise.

Second, individuals and households should have parallel trends in outcomes under the absence of treatment. In Section 5.2.2, I demonstrate that the pre-treatment means of both controls and outcomes are similar between the treatment and control groups, indicating that the pre-treatment trends are parallel.

Third, pre-treatment characteristics that are in expectation not qualitatively affected by the treatment should be invariant to change in treatment assignment. A covariat balance test reveals that there is no observable discontinuous change around the cutoff in the average values of covariates that should not be affected by the treatment assignment, i.e. parents' gender, years since migration, whether they own the house they reside in, whether they are a refugee, and whether they have a permanent residence permit or German citizenship.[22]

Fourth, just as there should not be any effects on those covariates, there should also not be effects on the outcomes of interest at arbitrarily chosen cutoffs. In following Imbens and Lemieux (2008), I look at one side of the discontinuity and take the median value of the running variable (distance to the enrollment cutoff) in this selection. Looking at the right side of the discontinuity and using the median of 3 as an arbitrarily chosen cutoff, I find no sign of discontinuity at this point in any of the outcomes of interest (parental monthly income, employment probability, working hours per week, hourly income as well as childcare hours per day, worries about personal finances, health status, staying intention or German language skills). The same holds when I look at the left side of the discontinuity and use the median of -3 as an arbitrarily chosen cutoff.[23]

Fifth, I run an additional placebo test in which I assign the treatment randomly, given the same probability to be assigned the treatment as under the birthdate cutoff rule. To check whether this random treatment assignment can predict the months spent in school by the oldest child, I regress the random treatment assignment on the months spent in school by the oldest child as an outcome (this is the same set-up as the first stage of the 2SLS instrumental variable approach (Equation 5.2.1), except now the treatment is not assigned based on the birthday of the oldest child with respect to the enrollment cutoff, but randomly). Figure 5.7 plots the distribution of the random treatment assignment coefficient for $N = 10000$ repetitions. The coefficient is normally distributed around 0, which means in most cases of random treatment assignment the null hypothesis that the randomly assigned treatment has no effect on the months the oldest child spent in school cannot be rejected. This underlines the validity of the actual treatment assignment as an instrument for months of schooling of the oldest child.

The use of self-reported measures (e.g. worries about personal finances and self-assessed language skills) might introduce unobserved heterogeneity between individuals. However, the panel data structure and the introduction of individual fixed effects in the estimations should account for varying reporting styles and personality traits across respondents (Angelini et al. 2015).[24]

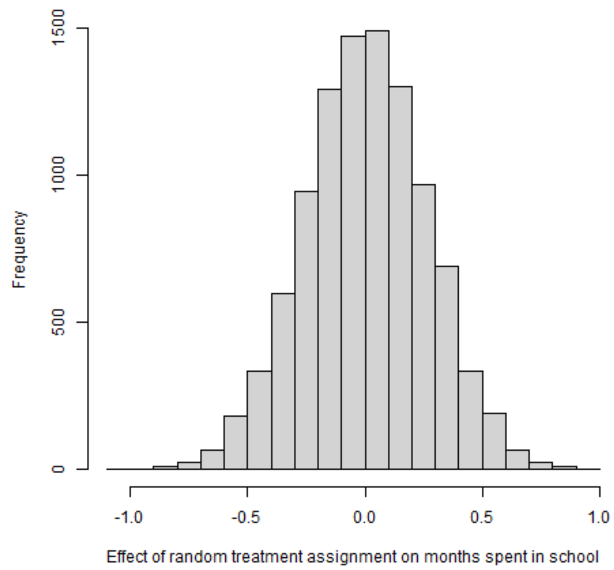
Another potential threat to identification is the timing of the survey interviews. As mentioned in Section 5.2.3, the yearly GSOEP interviews are conducted during all 12

[22]In the long run, residence permits and naturalization can be an integration output which is potentially affected by children's schooling. However, in the short run I expect no effects on these outcomes since changes in residence status and acquiring citizenship take a lot of time.

[23]The only exception is hourly income for which I find marginally significant results on the left side of the cutoff, but not the right.

[24]Also, most studies on the effect of school enrollment timing on children's outcomes, like test results and lifetime earnings, are potentially biased by age-at-test effects since children who were enrolled earlier are of younger age when they are tested for their academic achievements, and vice versa. This, however, is not a concern in this study as it focuses on the outcomes of parents in the household rather than those of the children.

Figure 5.7: *Distribution of Random Treatment Assignment Coefficient on Months Spent in School (N=10000)*



Note: Distribution of linear regression coefficient of random treatment assignment on oldest child's months spent in school. $N = 10000$ repetitions of random treatment assignment, given same assignment probability as under birthdate cutoff rule.

months of the year. The actual school enrollment, depending on the federal state, only happens between July and September, though. Hence, I have to ensure that when I measure results post-treatment, the treated have been subject to at least one year of schooling of their oldest child. In addition, it is possible that households are surveyed in the cutoff year 0 after the actual school entry of their child (for example if the oldest child was eligible for enrollment in year 0 and resided in a state where school started in August but the household was only interviewed in November). To address this, I define the year before the enrollment cutoff (year -1) as the pre-treatment period, and use the second year after enrollment cutoff (year 2) as the post-treatment period. In addition, I introduce an interaction term between the state of residence and interview month (see Section 5.2.3).

Heterogeneity in effects between migrants who live in households with German individuals and migrants who live in migrant-only households is another concern. All previous observations include all migrants, independent of whether they live in a household with other migrants or Germans. Now I run the main regressions on a subsample of migrants living in migrant-only households. Results are shown in Table 5.V to Table 5.Z in the Appendix. Since they are largely comparable to the whole sample results there is no evidence for considerable heterogeneity in effects.

Lastly, the external validity of the results is limited. Though the GSOEP is a German-wide representative survey, the data only observes migrants in Germany, and the sample shrinkage leaves only a rather limited number of observations, especially in subsample analyses. Also, only two years after school enrollment are observed. Hence, estimated effects within those years cannot easily be generalized to a larger time frame. I.e. on the intensive margin the estimated causal effect of one additional month of schooling of the oldest child in the household on parental integration cannot be extrapolated to an arbitrary number of years after initial enrollment exceeding the observed 2 years. On the extensive margin, the LATE estimated via RDD approach can only be applicable to parents whose children are born close to the enrollment cutoff. This limits the extent in how far the presented results can be generalized to other migrants, more years of observation and other countries. Despite those limitations, the analysis produces interesting first insights on the effect of schooling and school enrollment timing on migrant parents' integration outcomes.

5.4 Conclusion

In this paper, I exploit age-at-enrollment policies in 16 German states as exogenous source of variation to examine the effect of schooling of the oldest child in the household on parental integration. For this, I link administrative records on primary school enrollment cutoff dates with micro data from the German Socioeconomic Panel. Via a fuzzy regression discontinuity design around the school enrollment cutoff, I estimate the effect the early school entry of the oldest child in the household has on parental integration outcomes (extensive margin). Via an instrumental variable approach, I estimate the effect one additional month of schooling of the oldest child has on the parents' integration outcomes (intensive margin).

Both on the extensive and intensive margin I find that the schooling of the oldest child in the household positively affects parental labor market outcomes. It increases labor market participation, parental monthly income and hourly wages. These effects are especially strong among the formerly unemployed and those who carry the main burden of childcare in the household, i.e. the mothers. Apart from labor market outcomes, I find positive effects of the oldest child's schooling on parental health status, staying intentions and self-assessed German language skills in speaking, reading and writing. My results are robust to various robustness checks, and not driven by self-selection into school entry due to parental choice to deviate from the state-mandated enrollment eligibility.

An analysis of potential channels reveals that both gained disposable time and exposure to the German language and culture play a role in shaping integration outcomes. Schooling not only opens up time for migrant parents to spend at work, but also boosts their overall labor market outcomes, and language skills. Those results contribute to our understanding in how far direct and indirect exposure to the German language and culture via compulsory schooling hold the potential to enhance the integration of migrant parents.

In conclusion, primary schooling is one important factor in the complex integration process. Yet, assimilation processes in the family context are not well understood and warrants further research.

5.A Appendix

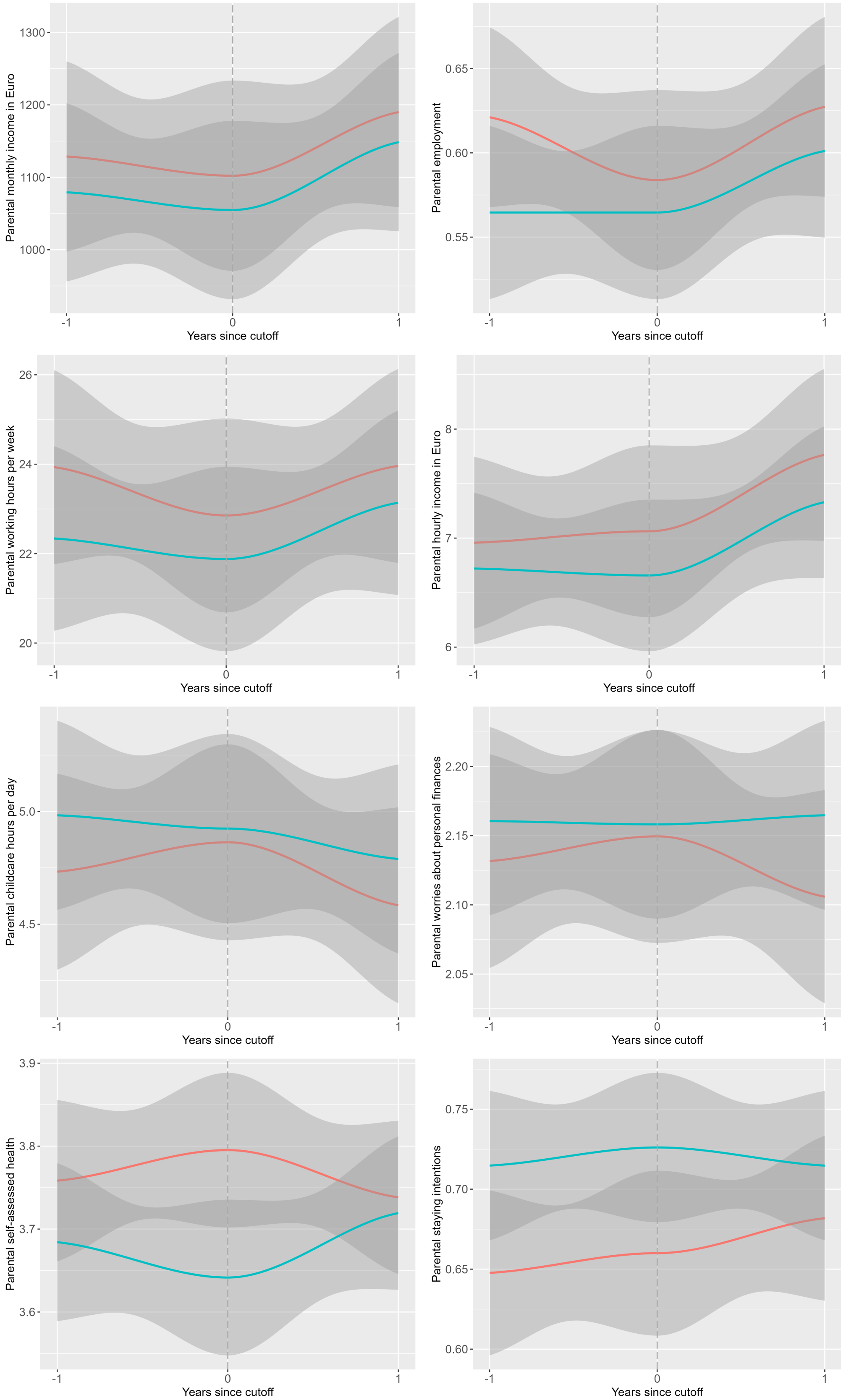
Table 5.A: *Sample Shrinkage due to Missings*

Step	Action	Observations	Individuals	Households
1	Identify adult migrants with children of enrollment age	21573	2758	1643
2	Identify observations with complete records on enrollment	15313	1897	1137
3	Remove observations with $<$ 4 years of interviews around cutoff	4244	1062	682
4	Remove missings: Interview month	3556	889	592
5	Remove missings: Parental characteristics	3192	798	543
5	Remove missings: Parental employment	3192	798	543
5	Remove missings: Parental monthly income	2972	743	503
6	Remove missings: Parental working hours per week	2852	713	494
7	Remove missings: Parental childcare hours per week	2712	678	473

Table 5.B: *Variables Description*

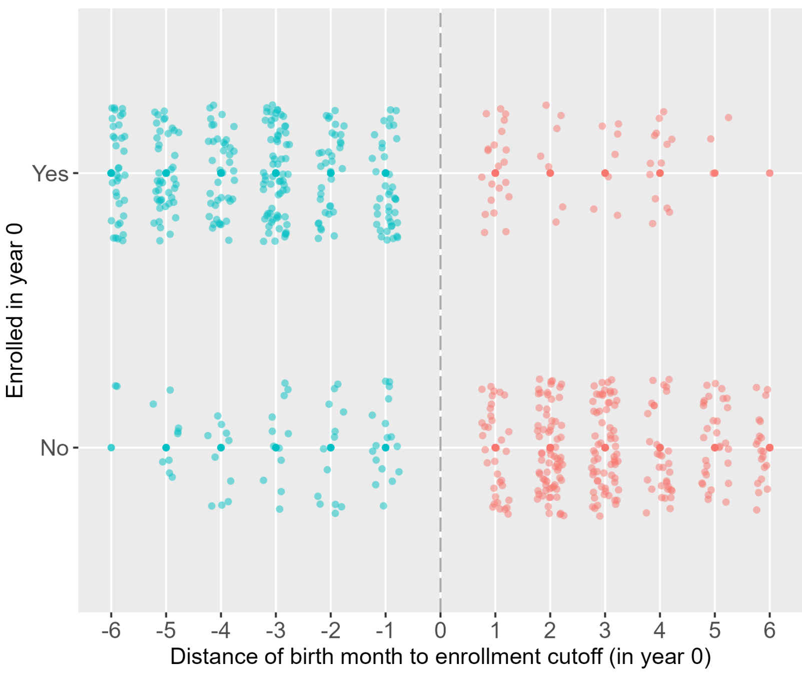
Variable	Type	Description
Eligible for enrollment	Binary	Oldest child was eligible for school enrollment in the year they turned six years old (*treatment group*) (*Ref = not eligible (control group)*).
Months in school	Numerical	Number of months spent in school by oldest child since enrollment.
Age	Numerical	Age in years.
Female	Binary	Female gender (*Ref = male*).
Years of education	Numerical	Number of years spent in formal education.
Currently in parental leave	Binary	Currently in maternity or paternity leave (*Ref = not in parental leave*).
Owner of housing	Binary	Owner of current dwelling (*Ref = not owner*).
Refugee	Binary	Status as a refugee (*Ref = no refugee*).
Years since migration	Numerical	Years spent in Germany since initial migration.
Number of children	Numerical	Number of children (under 18 years old) in household.
Younger children in ECEC	Binary	Younger children have spent time in any kind of early childhood education and care facilities (e.g. kindergarten) (*Ref = younger children not in ECEC*).
Oldest in ECEC before enrollment	Binary	Oldest child has spent time in any kind of early childhood education and care facilities (e.g. kindergarten) before school enrollment (*Ref = oldest child not in ECEC before enrollment*).
State	Categorical	One of 16 current states of living in Germany.
Interview month	Categorical	Month in which interview was conducted in given year.
German language skills: speaking / writing / reading	Numerical	Self-assessed ability in speaking / writing / reading German on a scale from 0 (no knowledge) to 4 (very good).
Parental monthly income	Numerical	Monthly income in Euro earned by individual after taxes and social security contributions, adjusted for inflation.
Parental employment	Binary	Current regular employment in paid occupation or in education (*Ref = not regularly employed*).
Parental working hours per week	Numerical	Average actual working hours in paid employment per week.
Parental hourly income	Numerical	Monthly income in Euro earned by individual after taxes and social security contributions, adjusted for inflation, divided by the actual average working hours per month (assuming a month with 4.35 working weeks).
Parental worries about personal finances / economic development / job security	Numerical	Worries about personal finances / the overall economic development of the country / own job security on a scale from 1 (not worried) to 3 (strongly worried).
Parental staying intentions	Binary	Intention to stay in Germany indefinitely (*Ref = intention to stay only for several years*).
Parental health	Numerical	Parental self-assessed current health status on a scale from 1 (bad) to 5 (very good).
German language index	Numerical	Self-assessed ability in speaking / writing / reading German combined to an index with equal weight on a scale from 0 (no knowledge) to 4 (very good).

Figure 5.A: *Means in Parental Integration Outcomes Over Years*



Note: Plot of outcome means in years -1, 0 and 1 from cutoff by eligibility for enrollment in year 0 (no = red, yes = blue). 95% confidence intervals.

Figure 5.B: *Actual Enrollment in Year 0 (Treatment Compliance)*



Note: Actual enrollment in year 0 by eligibility for enrollment in year 0 (no = red, yes = blue). Vertical and horizontal noise added to avoid overplotting.

Table 5.C: *Pre-Treatment Differences in Parental Characteristics and Integration Outcomes between Parents who Enrolled their Children According to Eligibility and Parents who Brought Forward or Postponed Enrollment*

	Enrolled according to eligibility		Difference
	Yes	*No*	
Parental characteristics			
Age	32.8	31.8	-1.0
Female (%)	53.8	55.5	1.7
Years of education	10.7	11.0	0.3^{*}
Currently in parental leave (%)	10.4	8.6	-1.8
Owner of housing (%)	25.1	20.3	-4.8
Refugee (%)	5.5	7.0	1.5
Years since migration	15.6	13.5	-2.1^{**}
Number of children	1.7	1.7	0.0
Younger children in ECEC (%)	14.7	14.1	-0.6
Oldest in ECEC before enrollment (%)	88.5	85.9	2.6
Parental integration outcomes			
Monthly parental income (Euro)	1,140.5	940.8	-200.5^{**}
Currently employed (%)	50.8	61.1	10.3
Working hours per week	23.6	20.8	-2.8
Hourly net wage (Euro)	7.0	6.0	-1.0
Childcare hours per day	4.8	5.3	0.5
Worried about own finances (scale 1-3)	2.2	2.1	-0.1
Staying intention (0/1)	73.8	72.1	-1.7
Number of observations	550	128	678

Note: Means 1 year before initial enrollment cutoff. Unpaired two-sample Wilcoxon tests for differences in variables between groups. For detailed variable description see Table 5.B. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 5.D: *Differences in Parental Characteristics and Integration Outcomes between Parents whose Oldest Child was Eligible and Not Eligible for Enrollment the Year They Turned Six (Bandwidth* ± 4*)*

	Not eligible *Control*	Eligible *Treatment*	Difference
Parental characteristics			
Age	32.6	32.2	−0.4
Female (%)	52.4	52.3	−0.1
Years of education	10.7	11.0	0.3
Currently in parental leave (%)	9.1	10.1	1.0
Owner of housing (%)	24.8	20.3	−4.5
Refugee (%)	5.1	7.6	2.5
Years since migration	15.4	15.1	−0.3
Number of children	1.6	1.7	0.1
Younger children in ECEC (%)	13.0	14.8	1.8
Oldest in ECEC before enrollment (%)	89.0	91.1	2.1
Parental integration outcomes			
Monthly parental income (Euro)	1,172.8	1,094.2	−78.6
Currently employed (%)	63.8	57.0	−6.8
Working hours per week	24.7	22.9	−2.8
Hourly net wage (Euro)	7.3	6.7	−0.6
Childcare hours per day	4.7	4.9	0.2
Worried about own finances (scale 1-3)	2.2	2.1	−0.1
Staying intention (%)	76.4	71.1	−5.3
Number of observations	254	237	491

Note: Means 1 year before initial enrollment cutoff. Unpaired two-sample Wilcoxon tests with potential unequal variance in both samples for differences in variables between groups. For detailed variable description see Table 5.B. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 5.E: *Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Formerly Unemployed*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.46^{***}		0.09		
		(0.66)		(0.08)		
Months in school	-0.00		0.02^{***}			
	(0.00)		(0.00)			
Treated (compliers)					0.15^{*}	0.24^{**}
					(0.09)	(0.10)
R^2	0.63	0.65	0.54	0.65	–	–
Adj. R^2	0.47	0.52	0.37	0.23	–	–
Num. obs.	1108	1108	1108	554	232	194
Num. individuals	277	277	277	277	232	194
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental employment for formerly unemployed (unemployed in year -1) in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental employment for formerly unemployed (unemployed in year -1) in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.F: *Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Formerly Unemployed*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.46^{***}		-0.33		
		(0.66)		(0.78)		
Months in school	-0.01		-0.08^{***}			
	(0.03)		(0.02)			
Treated (compliers)					-3.51^{***}	-4.21^{***}
					(0.81)	(0.87)
R^2	0.68	0.65	0.67	0.75	–	–
Adj. R^2	0.55	0.52	0.54	0.45	–	–
Num. obs.	1108	1108	1108	554	232	194
Num. individuals	277	277	277	277	232	194
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental childcare hours spent per day for formerly unemployed (unemployed in year -1) in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental childcare hours spent per day for formerly unemployed (unemployed in year -1) in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.G: *Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Women*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.90^{***}		63.24		
		(0.49)		(88.24)		
Months in school	5.67		16.64^{***}			
	(3.53)		(2.99)			
Treated (compliers)					484.40^{***}	580.88^{***}
					(153.03)	(147.04)
R^2	0.85	0.66	0.84	0.86	–	–
Adj. R^2	0.79	0.53	0.79	0.71	–	–
Num. obs.	1468	1468	1468	734	309	257
Num. individuals	367	367	367	367	309	257
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental monthly income for women in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental monthly income for women in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.H: *Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Women*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.90^{***}		0.07		
		(0.49)		(0.07)		
Months in school	0.00		0.01^{***}			
	(0.00)		(0.00)			
Treated (compliers)					0.35^{***}	0.48^{***}
					(0.13)	(0.11)
R^2	0.73	0.66	0.72	0.77	–	–
Adj. R^2	0.62	0.53	0.62	0.50	–	–
Num. obs.	1468	1468	1468	734	309	257
Num. individuals	367	367	367	367	309	257
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental employment for women in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental employment for women in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.I: *Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Women*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.90^{***}		2.17		
		(0.49)		(2.44)		
Months in school	0.09		0.29^{***}			
	(0.09)		(0.08)			
Treated (compliers)					8.64^{***}	11.72^{***}
					(3.31)	(3.65)
R^2	0.78	0.66	0.77	0.78	–	–
Adj. R^2	0.69	0.53	0.69	0.52	–	–
Num. obs.	1468	1468	1468	734	309	257
Num. individuals	367	367	367	367	309	257
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental working hours for women in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental working hours for women in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.J: *Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Women*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.90^{***}		0.43		
		(0.49)		(0.71)		
Months in school	0.05^{*}		0.14^{***}			
	(0.03)		(0.03)			
Treated (compliers)					2.94^{***}	2.69^{***}
					(1.03)	(0.97)
R^2	0.73	0.66	0.71	0.81	–	–
Adj. R^2	0.62	0.53	0.61	0.60	–	–
Num. obs.	1468	1468	1468	734	309	257
Num. individuals	367	367	367	367	309	257
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental hourly income for women in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental hourly income for women in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.K: *Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Women*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		$\pm$5 months	$\pm$4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.90^{***}		-0.10		
		(0.49)		(0.60)		
Months in school	-0.03		-0.08^{***}			
	(0.02)		(0.02)			
Treated (compliers)					-4.49^{***}	-5.05^{***}
					(0.41)	(0.43)
R^2	0.69	0.66	0.68	0.77	–	–
Adj. R^2	0.57	0.53	0.56	0.51	–	–
Num. obs.	1468	1468	1468	734	309	257
Num. individuals	367	367	367	367	309	257
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State $\times$ interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental childcare hours spent per day for women in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental childcare hours spent per day for women in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.L: *Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Parents with only One Child*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.67^{***}		117.45		
		(1.01)		(147.85)		
Months in school	2.51		8.33^{*}			
	(5.38)		(4.35)			
Treated (compliers)					708.51^{**}	585.93^{**}
					(310.82)	(240.41)
R^2	0.92	0.65	0.91	0.93	–	–
Adj. R^2	0.88	0.52	0.88	0.84	–	–
Num. obs.	532	532	532	266	114	95
Num. individuals	133	133	133	133	114	95
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental monthly income for parents with only one child in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental monthly income for parents with only one child in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.M: *Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Parents with only One Child*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		± 5 months	± 4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.67^{***}		0.07		
		(1.01)		(0.11)		
Months in school	0.00		0.00			
	(0.00)		(0.00)			
Treated (compliers)					0.24^{**}	0.28^{***}
					(0.10)	(0.11)
R^2	0.78	0.65	0.76	0.82	–	–
Adj. R^2	0.67	0.52	0.67	0.59	–	–
Num. obs.	532	532	532	266	114	95
Num. individuals	133	133	133	133	114	95
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental employment for parents with only one child in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental employment for parents with only one child in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.N: *Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Parents with only One Child*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.67^{***}		0.48		
		(1.01)		(3.46)		
Months in school	0.05		0.07			
	(0.14)		(0.11)			
Treated (compliers)					4.46	3.21
					(7.29)	(7.10)
R^2	0.87	0.65	0.85	0.89	–	–
Adj. R^2	0.80	0.52	0.79	0.75	–	–
Num. obs.	532	532	532	266	114	95
Num. individuals	133	133	133	133	114	95
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental working hours for parents with only one child in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental working hours for parents with only one child in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.O: *Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Parents with only One Child*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.67^{***}		-0.78		
		(1.01)		(2.39)		
Months in school	0.03		0.08^{*}			
	(0.04)		(0.05)			
Treated (compliers)					11.13^{***}	10.79^{***}
					(3.31)	(2.16)
R^2	0.76	0.65	0.73	0.78	–	–
Adj. R^2	0.63	0.52	0.63	0.50	–	–
Num. obs.	532	532	532	266	114	95
Num. individuals	133	133	133	133	114	95
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental hourly income for parents with only one child in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental hourly income for parents with only one child in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.P: *Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Parents with only One Child*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		6.67^{***}		-0.05		
		(1.01)		(0.71)		
Months in school	-0.01		-0.05^{**}			
	(0.02)		(0.02)			
Treated (compliers)					-1.45^{**}	-1.33^{*}
					(0.76)	(0.76)
R^2	0.77	0.65	0.76	0.82	–	–
Adj. R^2	0.66	0.52	0.66	0.59	–	–
Num. obs.	532	532	532	266	114	95
Num. individuals	133	133	133	133	114	95
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental childcare hours spent per day for parents with only one child in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental childcare hours spent per day for parents with only one child in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.Q: *Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Germans*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.29^{***}		19.56		
		(0.20)		(42.65)		
Months in school	−0.38		12.07^{***}			
	(1.37)		(1.17)			
Treated (compliers)					128.63	158.74
					(101.25)	(113.43)
R^2	0.93	0.67	0.92	0.93	–	–
Adj. R^2	0.90	0.55	0.90	0.86	–	–
Num. obs.	13112	13112	13112	6556	2978	2495
Num. individuals	3278	3278	3278	3278	2978	2495
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental monthly income for Germans in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental monthly income for Germans in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.R: *Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Germans*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.29^{***}		-0.00		
		(0.20)		(0.02)		
Months in school	-0.00		0.00^{***}			
	(0.00)		(0.00)			
Treated (compliers)					0.01	0.02
					(0.02)	(0.02)
R^2	0.76	0.67	0.76	0.80	–	–
Adj. R^2	0.68	0.55	0.68	0.59	–	–
Num. obs.	13112	13112	13112	6556	2978	2495
Num. individuals	3278	3278	3278	3278	2978	2495
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental employment for Germans in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental employment for Germans in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.S: *Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Germans*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.29^{***}		-0.04		
		(0.20)		(0.70)		
Months in school	-0.01		0.15^{***}			
	(0.03)		(0.02)			
Treated (compliers)					0.46	0.15
					(0.91)	(1.02)
R^2	0.85	0.67	0.85	0.87	–	–
Adj. R^2	0.80	0.55	0.80	0.74	–	–
Num. obs.	13112	13112	13112	6556	2978	2495
Num. individuals	3278	3278	3278	3278	2978	2495
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental working hours for Germans in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental working hours for Germans in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.T: *Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Germans*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		$\pm$5 months	$\pm$4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.29^{***}		0.38		
		(0.20)		(0.34)		
Months in school	-0.00		0.08^{***}			
	(0.01)		(0.01)			
Treated (compliers)					0.06	0.09
					(0.50)	(0.54)
R^2	0.78	0.67	0.78	0.82	–	–
Adj. R^2	0.70	0.55	0.70	0.64	–	–
Num. obs.	13112	13112	13112	6556	2978	2495
Num. individuals	3278	3278	3278	3278	2978	2495
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State $\times$ interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental hourly income for Germans in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental hourly income for Germans in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.U: *Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Germans*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.29^{***}		-0.11		
		(0.20)		(0.17)		
Months in school	-0.00		-0.05^{***}			
	(0.01)		(0.01)			
Treated (compliers)					-0.31	-0.33
					(0.23)	(0.24)
R^2	0.80	0.67	0.80	0.85	–	–
Adj. R^2	0.74	0.55	0.73	0.69	–	–
Num. obs.	13112	13112	13112	6556	2978	2495
Num. individuals	3278	3278	3278	3278	2978	2495
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental childcare hours spent per day for Germans in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental childcare hours spent per day for Germans in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.V: *Estimates of Months of Schooling and Enrollment Timing on Parental Monthly Income for Households with only Migrants*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.24^{***}		146.05		
		(0.68)		(113.81)		
Months in school	8.41^{**}		15.61^{***}			
	(4.12)		(3.85)			
Treated (compliers)					247.67^{**}	318.00^{***}
					(122.92)	(108.17)
R^2	0.91	0.67	0.91	0.92	–	–
Adj. R^2	0.88	0.55	0.88	0.83	–	–
Num. obs.	1480	1480	1480	740	323	282
Num. individuals	370	370	370	370	323	282
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental monthly income for households with only migrants in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental monthly income for households with only migrants in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.W: *Estimates of Months of Schooling and Enrollment Timing on Parental Employment for Households with only Migrants*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.24^{***}		0.08		
		(0.68)		(0.06)		
Months in school	0.00^{*}		0.01^{***}			
	(0.00)		(0.00)			
Treated (compliers)					0.27^{**}	0.39^{***}
					(0.11)	(108.17)
R^2	0.78	0.67	0.77	0.83	–	–
Adj. R^2	0.69	0.55	0.69	0.63	–	–
Num. obs.	1480	1480	1480	740	323	282
Num. individuals	370	370	370	370	323	282
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental employment for households with only migrants in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental employment for households with only migrants in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.X: *Estimates of Months of Schooling and Enrollment Timing on Parental Working Hours for Households with only Migrants*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.24^{***}		3.85^{*}	3.85^{*}	3.85^{*}
		(0.68)		(2.31)	(2.31)	(2.31)
Months in school	0.21^{**}		0.30^{***}			
	(0.10)		(0.08)			
Treated (compliers)					4.04	6.21^{**}
					(3.11)	(2.99)
R^2	0.84	0.67	0.83	0.85	–	–
Adj. R^2	0.77	0.55	0.77	0.69	–	–
Num. obs.	1480	1480	1480	740	323	282
Num. individuals	370	370	370	370	323	282
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental working hours for households with only migrants in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental working hours for households with only migrants in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.Y: *Estimates of Months of Schooling and Enrollment Timing on Parental Hourly Income for Households with only Migrants*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.24^{***}		-0.08		
		(0.68)		(1.23)		
Months in school	0.01		0.11^{***}			
	(0.03)		(0.03)			
Treated (compliers)					6.10^{***}	7.51^{***}
					(1.12)	(0.60)
R^2	0.82	0.67	0.81	0.86	–	–
Adj. R^2	0.74	0.55	0.74	0.71	–	–
Num. obs.	1480	1480	1480	740	323	282
Num. individuals	370	370	370	370	323	282
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental hourly income for households with only migrants in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental hourly income for households with only migrants in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table 5.Z: *Estimates of Months of Schooling and Enrollment Timing on Parental Childcare Hours per Day for Households with only Migrants*

	OLS	2SLS		DID	Fuzzy RDD	
		1. stage	2. stage		±5 months	±4 months
	(1)	(2)	(3)	(4)	(5)	(6)
Eligible for enrollment		7.24^{***}		-0.82		
		(0.68)		(0.58)		
Months in school	-0.05^{***}		-0.06^{***}			
	(0.02)		(0.01)			
Treated (compliers)					-2.56^{***}	-2.88^{***}
					(0.65)	(0.72)
R^2	0.76	0.67	0.75	0.81	–	–
Adj. R^2	0.66	0.55	0.66	0.60	–	–
Num. obs.	1480	1480	1480	740	323	282
Num. individuals	370	370	370	370	323	282
Controls	✓	✓	✓	✓	–	–
Time FE	✓	✓	✓	✓	–	–
Individual FE	✓	✓	✓	✓	–	–
State × interview month	✓	✓	✓	✓	–	–

Note: OLS (Column 1) and 2SLS (Column 2-3) estimate coefficients of months of schooling on parental childcare hours per day for households with only migrants in all years. First stage of 2SLS instruments months of schooling via enrollment eligibility, and Wald test comparing the model including instrument and excluding instrument proves instrument relevance. DID ITT effect (Column 4) and Fuzzy RDD LATE (Column 5-6) estimates of enrollment in year 0 on parental childcare hours per day for households with only migrants in year 2. Fuzzy RDD estimates use triangular kernel weighting and all first stage F-statistics are highly significant ($p < 2.2e^{-16}$). Standard errors (in parentheses) are clustered on household level for OLS, 2SLS and DID model, and clustered on level of cutoff distance for discontinuity samples. Controls include number of children, a dummy for younger child(ren) in household in ECEC, and a dummy for oldest child of household in ECEC before enrollment. $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

6 | General Conclusions

Although there remains much to be learned about the complex migration process, this thesis provides insights into the role of language learning, education and social context in migration. It contributes to three strands of literature. Chapter 2 and Chapter 3 highlight the motives of language learning and its importance for success in both origin and host country labor markets. Chapter 4 enhances the understanding of migration motives and intentions of potential migrants before actual migration occurs. Lastly, Chapter 5 offers first evidence of primary schooling of children in the household being a potential integration driver for migrant parents.

In this final Chapter, policy implications from the presented findings will be derived and the limitations of the analyses discussed, highlighting potential for further research. Since all analyses revolve around Germany and the German language, the policy implications mostly focus on Germany.

6.1 Policy Implications

Germany faces massive demographic changes due to its rapidly aging population and a considerable shortage of skilled workers. Against this background, migration constitutes a huge opportunity for Germany to address these problems. In the following it will be discussed what policy makers can learn from this thesis in order to attract young and highly skilled immigrants to Germany and foster their integration into the German labor market.

In the first part of this thesis (Chapter 2 and Chapter 3), it is established that language learning is an important factor for migration and integration. Analyzing motivations for and timing of foreign language learning can help identify policy measures that may result in more foreign language acquisition, e.g. of the German language. Language skills can be acquired in many different contexts, but are mainly acquired through two channels: First, at school at an early age, which is subject to schools' curricula and parents' preferences. While early age language learning is unlikely affected by migration decisions that are only made later in life, it is quite plausible that early age language learning builds up relevant foreign language skills. Migrants may very well sort into the destination countries where these skills are most useful, as evidence suggests that language learning at school can affect migration decisions (Aparicio Fenoll and Kuehn 2016). Second, at adult age language learning occurs mostly through specific language courses, offered by public or private providers. Such adult age language learning is more directly related to individuals' human capital and consumption motives, especially in the context of migration decisions. Hence, while linguistic properties may be beyond the control of policy makers, policy interventions can effectively address incentives for both early age and adult age language learning, recognizing its impact on migration decisions. One crucial institution which promotes adult age language learning is the Goethe Institute. Goethe Institutes act in countries around the world and prepare migrants for their stay in Germany by offering German language and cultural courses and counseling, hence they already play a significant role in German foreign cultural policy. Expanding the role of institutes like the GI can provide valuable support for language learning and subsequent integration. E.g. Jaschke and Keita (2021) provide evidence for the positive effect the existence of Goethe Institutes in a country has on the (self-) selection of migrants with regards to their education and integration outcomes after arrival in Germany.

Chapter 3 of this thesis also shows that there exists considerable difference in the timing of language learning, depending on the certainty of access to the destination country. Language learning in Germany is strongly associated with restricted access to the German labor market (i.e. migration from countries outside of the EU or Schengen Area), while language learning in origin countries is positively associated with secure access to the German labor market (i.e. migration from countries inside the EU or Schengen Area). Policy intervention should address such uncertainty of investment due to access barriers by facilitating access to the labor market for skilled workers. The Skilled Immigration Act from 2020 – aimed at facilitating migration of skilled workers from third countries to Germany – is a first step in removing uncertainty of investment, with potential positive effects on preparatory language learning abroad and thus economic and social integration. These results can also inform institutions like the Goethe Institute regarding their choice of location and courses offered. E.g. in countries with restricted access to the German labor market German language courses could be combined with in-depth information and consultation on German immigration regulations and how to acquire work and residence permits.

Chapter 4 of this thesis establishes that a considerable amount of potential migrants actually intends to migrate only temporarily, i.e. stay in the host country only for a limited amount of time before returning to their home country or leaving for another country. Understanding the intended duration of stay can assist origin countries in differentiating which share of planned emigration is about brain drain and which is about brain circulation. With regards to the receiving country – in this case Germany – it allows for the design of optimal integration measures, which should target individuals who intend to stay long-term rather than short-term. Integration policies in general should recognize the significance of family- and gender-related aspects in migration to utilize the full potential of migrating families. This can be done e.g. by specifically targeting women who can potentially generate spillover effects from a consumption motive of language learning to professional use in the local labor market (i.e. women who immigrated in a "tied mover" context without current labor market participation).

Chapter 5 of this thesis emphasizes the importance of family context in the integration process. Upon school enrollment of the oldest child in the household, migrant parents not only gain disposable time, but also are exposed to the German language and culture.

Hence, migrant parents benefit from their oldest child's schooling with regards to their labor market outcomes, health, staying intentions and German language skills. These effects are most prominent among those who carry the main burden of childcare, i.e. mothers. In conclusion, schooling of migrant children is one crucial factor in the complex integration process of migrant parents – especially for women whose labor market potential is underutilized.

6.2 Limitations and Future Research

While this thesis sheds some light on the complex processes of language learning and migration, the presented analyses are subject to some limitations.

The analysis on migration intentions investigates potential migrants in their home country, prior to any move. While this allows to identify all potential migrants, not only those who follow through with their migration plans, and examine migration motives and barriers ex-ante, intentions are naturally not a perfect predictor of actual behavior. Migration considerations and plans might change or be discarded, and the cross-country structure of the data utilized in the analysis on migration intentions does not allow to follow up with surveyed individuals in whether they have realized their plans. This is to some extend mitigated by evidence that migration intentions are indeed a strong predictor of subsequent behavior, e.g. by Jacob and Linkow (2011), Tjaden et al. (2019), and Wanner (2021).

Furthermore, all presented analyses are limited with respect to their representativeness and generalizability. The analysis of the effect of schooling on parental labor market outcomes (Chapter 5) is based on data from the German Socioeconomic Panel. While the GSOEP itself is a representative survey among the German population, the results from its data might be only applicable to Germany. In addition, the variation in school enrollment dates is driven by specific German federal states' policies. Replications of the study in countries with comparable policies regarding school enrollment could provide some insights on the generalizability of the results. E.g. with data from the British Household Panel Survey in the UK, where enrollment policies are quite similiar to the German case but the overall share of immigrants is higher.

The analyses of language learning (Chapter 2 and Chapter 3) focus primarily on the German language and rely on data from the Goethe Institute. While the GI is a large institute which provides language courses to many countries worldwide, it is only one supplier of language services. Naturally, there are a number of other language learning opportunities, including universities, private language schools, and internet platforms. Comparisons to competitors and conversations with GI employees do not indicate that the GI is the most expensive provider of language courses on a given market. Still, the price of courses are above average in comparison to competitors and most GIs are located in large cities; hence language learners at GIs might be a self-selected group with regards to income and location. Further, recent developments, especially the COVID-19 pandemic, might have brought a shift in language learning to increased digital learning via online platforms. The data utilized in this thesis cannot account for such alternative sources of language learning. This may limit the replicability of the results beyond the context of the GI and the German language. Hence, the studies should be replicated, including different languages and various forms of language acquisition.

The analysis of migration intentions in Chapter 4 also utilizes data from surveys conducted at GIs. Beyond their focus on Germany and the German language, those surveys are not representative for the general population of the respective country they are conducted in. The surveys are selective with regards to education and age, as they over-represent young and highly skilled individuals. While this group of potential migrants is of special interest as Germany is in need of skilled workers, medium and low skilled individuals are also of importance with regard to the rapidly over-aging German labor force. A better representation of the overall population of interest would be necessary to understand migration decisions and make Germany an attractive destination for potential migrants of all ages and skill levels. It is tried to broaden the observed population by including surveys among university students into the analysis. Those do not relate to the German language or Germany as a potential destination in general. Yet, they are limited, too, as they sample young and highly skilled individuals in universities in 6 selected countries. Expanding such studies across wider population groups and more countries would allow the identification of country level push and pull factors that lead to differences between migration desires and plans. A potential way of generating representative data on this question would be to eliminate the conditionality in questions in the Gallup World Poll, such that all respondents are asked for their migration desires as well as considerations and plans consequently, independent of their desires, and including questions on the intended length of stay abroad.

While these limitations constitute possible areas of future research – there is need to validate the findings with more representative data and with regards to different countries – this thesis provides valid recommendations for policies to improve the access to foreign language learning and international mobility. Overall, the results of the presented thesis and subsequent policy implications highlight the importance of addressing access barriers and opportunities of language learning, and considering family and gender context in migration and integration. Implementing such policies could contribute to improved international mobility, economic integration, and overall welfare gains for both origin and destination countries.

Bibliography

Abel, Martin, Rulof Burger, Eliana Carranza, and Patrizio Piraino (2019). “Bridging the Intention-Behavior Gap? The Effect of Plan-Making Prompts on Job Search and Employment”. In: *American Economic Journal: Applied Economics* 11.2, 284–301. DOI: 10.1257/app.20170566.

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson (2012). “Europe’s Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration”. In: *American Economic Review* 102.5, 1832–1856. DOI: 10.1257/aer.102.5.1832.

Adserà, Alícia and Mariola Pytliková (2015). “The Role of Language in Shaping International Migration”. In: *Economic Journal* 125.586, 49–81. DOI: 10.1111/ecoj.12231.

Ajzen, Icek (1991). “The Theory of Planned Behavior”. In: *Organizational Behavior and Human Decision Processes* 50.2, 179–211. DOI: 10.1016/0749-5978(91)90020-T.

Aldashev, Alisher, Johannes Gernandt, and Stephan L. Thomsen (2009). “Language Usage, Participation, Employment and Earnings: Evidence for Foreigners in West Germany with Multiple Sources of Selection”. In: *Labour Economics* 16.3, 330–341. DOI: 10.1016/j.labeco.2008.11.004.

Algan, Yann, Alberto Bisin, Alan Manning, and Thierry Verdier, eds. (2012). *Cultural Integration of Immigrants in Europe*. Oxford University Press. DOI: 10.1093/acprof:oso/9780199660094.001.0001.

Algan, Yann, Christian Dustmann, Albrecht Glitz, and Alan Manning (2010). “The Economic Situation of First and Second-Generation Immigrants in France, Germany and the United Kingdom”. In: *The Economic Journal* 120.542, 4–30. DOI: 10.1111/j.1468-0297.2009.02338.x.

Alstadsaeter, Annette (2011). “Measuring the Consumption Value of Higher Education”. In: *CESifo Economic Studies* 57.3, 458–479. DOI: 10.1093/cesifo/ifq009.

Angelini, Viola, Laura Casi, and Luca Corazzini (2015). “Life Satisfaction of Immigrants: Does Cultural Assimilation Matter?” In: *Journal of Population Economics* 28.3, 817–844. DOI: 10.1007/s00148-015-0552-1.

Angrist, Joshua D. and Alan B. Krueger (1991). “Does Compulsory School Attendance Affect Schooling and Earnings?” In: *The Quarterly Journal of Economics* 106.4, 979–1014. DOI: 10.2307/2937954.

— (1992). “The Effect of Age at School Entry on Educational Attainment: An Application of Instrumental Variables with Moments from Two Samples”. In: *Journal of the American Statistical Association* 87.418, 328–336. DOI: 10.1080/01621459.1992.10475212.

Angrist, Joshua D. and Jörn-Steffen Pischke (2009). *Mostly Harmless Econometrics: An Empiricist’s Companion*. Princeton University Press. ISBN: 978-0-691-12035-5.

Aoki, Yu and Lualhati Santiago (2018). “Speak Better, Do Better? Education and Health of Migrants in the UK”. In: *Labour Economics* 52, 1–17. DOI: 10.1016/j.labeco.2018.03.003.

Aparicio Fenoll, Ainhoa and Zoë Kuehn (2016). “Does Foreign Language Proficiency Foster Migration of Young Individuals within the European Union?” In: *The Economics of Language Policy*. Ed. by Bengt-Arne Wickstroem and Michele Gazzola. Cambridge, MA: MIT Press, 331–355.

Arcidiacono, Peter (2004). “Ability Sorting and the Returns to College Major”. In: *Journal of Econometrics* 121.1-2, 343–375. DOI: 10.1016/j.jeconom.2003.10.010.

Aslany, Maryam, Jørgen Carling, Mathilde Bålsrud Mjelva, and Tone Sommerfelt (2021). *Systematic Review of Determinants of Migration Aspirations*. QuantMig Project Deliverable 2.2.

Auswärtiges Amt and Goethe-Institut (2004). *Rahmenvertrag*. Munich and Berlin.

Avitabile, Ciro, Irma Clots-Figueras, and Paolo Masella (2013). “The Effect of Birthright Citizenship on Parental Integration Outcomes”. In: *The Journal of Law and Economics* 56.3, 777–810. DOI: 10.1086/673266.

Azam, Mehtabul, Aimee Chin, and Nishith Prakash (2013). “The Returns to English-Language Skills in India”. In: *Economic Development and Cultural Change* 61.2, 335–367. DOI: 10.1086/668277.

BAMF (2016). *Migrationsbericht 2015*.

Bassett, William F. and Robin L. Lumsdaine (2001). “Probability Limits: Are Subjective Assessments Adequately Accurate?” In: *Journal of Human Resources* 36.2, 327–363. DOI: 10.2307/3069662.

Battisti, Michele, Giovanni Peri, and Agnese Romiti (2022). “Dynamic Effects of Co-Ethnic Networks on Immigrants’ Economic Success”. In: *The Economic Journal* 132.641, 58–88. DOI: 10.1093/ej/ueab036.

Bauer, Philipp C. and Regina T. Riphahn (2009). “Age at School Entry and Intergenerational Educational Mobility”. In: *Economics Letters* 103.2, 87–90. DOI: 10.1016/j.econlet.2009.01.032.

Becker, Gary S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. National Bureau of Economic Research.

Bedard, Kelly and Elizabeth Dhuey (2006). “The Persistence of Early Childhood Maturity: International Evidence of Long-Run Effects”. In: *Quarterly Journal of Economics* 21.4, 1437–1472. DOI: 10.1093/qje/121.4.1437.

— (2012). “School-Entry Policies and Skill Accumulation Across Directly and Indirectly Affected Individuals”. In: *Journal of Human Resources* 47.3, 643–683. DOI: 10.3368/jhr.47.3.643.

Beine, Michel, Frédéric Docquier, and Çaglar Özden (2011). “Diasporas”. In: *Journal of Development Economics* 95.1, 30–41. DOI: 10.1016/j.jdeveco.2009.11.004.

Bell, Andrew and Kelvyn Jones (2014). “Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data”. In: *Political Science Research and Methods* FirstView, 1–21. DOI: 10.1017/psrm.2014.7.

Belot, Michèle and Sjef Ederveen (2012). "Cultural Barriers in Migration between OECD Countries". In: *Journal of Population Economics* 25.3, 1077–1105. DOI: 10.1007/s00148-011-0356-x.

Belot, Michèle and Timothy J. Hatton (2012). "Immigrant Selection in the OECD". In: *The Scandinavian Journal of Economics* 114.4, 1105–1128. DOI: 10.1111/j.1467-9442.2012.01721.x.

Bertoli, Simone, Frédéric Docquier, Hillel Rapoport, and Ilse Ruyssen (2022). "Weather Shocks and Migration Intentions in Western Africa: Insights from a Multilevel Analysis". In: *Journal of Economic Geography* 22.2, 289–323. DOI: 10.1093/jeg/lbab043.

Bertoli, Simone and Jesús Fernández-Huertas Moraga (2015). "The Size of the Cliff at the Border". In: *Regional Science and Urban Economics* 51, 1–6. DOI: 10.1016/j.regsciurbeco.2014.12.002.

Bertoli, Simone and Ilse Ruyssen (2018). "Networks and Migrants' Intended Destination". In: *Journal of Economic Geography* 18, 705–728. DOI: 10.1093/jeg/lby012.

Bertrand, Marianne and Sendhil Mullainathan (2001). "Do People Mean What They Say? Implications for Subjective Survey Data". In: *American Economic Review* 91.2, 67–72. DOI: 10.1257/aer.91.2.67.

Black, Sandra E., Paul J. Devereux, and Kjell G. Salvanes (2011). "Too Young to Leave the Nest? The Effects of School Starting Age". In: *The Review of Economics and Statistics* 93.2, 455–467. DOI: 10.1162/REST_a_00081.

Bleakley, Hoyt and Aimee Chin (2004). "Language Skills and Earnings: Evidence from Childhood Immigrants". In: *The Review of Economics and Statistics* 86.2, 481–496. DOI: 10.1162/003465304323031067.

— (2008). "What Holds Back the Second Generation?: The Intergenerational Transmission of Language Human Capital among Immigrants". In: *Journal of Human Resources* 43.2, 267–298. DOI: 10.3368/jhr.43.2.267.

— (2010). "Age at Arrival, English Proficiency, and Social Assimilation among US Immigrants". In: *American Economic Journal: Applied Economics* 2.1, 165–192. DOI: 10.1257/app.2.1.165.

Borjas, George J. (1987). "Self-Selection and the Earnings of Immigrants". In: *American Economic Review* 77.4, 531–553. DOI: 10.3386/w2248.

— (2015). "The Slowdown in the Economic Assimilation of Immigrants: Aging and Cohort Effects Revisited Again". In: *Journal of Human Capital* 9.4, 483–517. DOI: 10.1086/676461.

Borjas, George J., Ilpo Kauppinen, and Panu Poutvaara (2019). "Self-Selection of Emigrants: Theory and Evidence on Stochastic Dominance in Observable and Unobservable Characteristics". In: *Economic Journal* 129.617, 143–171. DOI: 10.1111/ecoj.12585.

Brown, Meta (2006). "Informal Care and the Division of End-of-Life Transfers". In: *Journal of Human Resources* 41.1, 191–219. DOI: 10.3368/jhr.XLI.1.191.

Budría, Santiago, Alberto Colino, and Carlos Martínez de Ibarreta (2019). “The Impact of Host Language Proficiency on Employment Outcomes among Immigrants in Spain”. In: *Empirica* 46.4, 625–652. DOI: 10.1007/s10663-018-9414-x.

Burda, Michael C., Wolfgang Härdle, Marlene Müller, and Axel Werwatz (1998). “Semiparametric Analysis of German East-West Migration Intentions: Facts and Theory”. In: *Journal of Applied Econometrics* 13.5, 525–541.

Bureau, US Census (2021). *Current Population Survey Datasets*. https://www.census.gov/programs-surveys/cps/data/datasets.html.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011). “Robust Inference with Multiway Clustering”. In: *Journal of Business & Economic Statistics* 29.2, 238–249. DOI: 10.1198/jbes.2010.07136.

Canton, Erik and Frank de Jong (2005). “The Demand for Higher Education in The Netherlands, 1950–1999”. In: *Economics of Education Review* 24.6, 651–663. DOI: 10.1016/j.econedurev.2004.09.006.

Cascio, Elizabeth U. and Diane W. Schanzenbach (2016). “First in the Class? Age and the Education Production Function”. In: *Education Finance and Policy* 11.3, 225–250. DOI: 10.1162/EDFP_a_00191.

Chandon, Pierre, Vicki G. Morwitz, and Werner J. Reinartz (2005). “Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research”. In: *Journal of Marketing* 69.2, 1–14. DOI: 10.1509/jmkg.69.2.1.60755.

Chiswick, Barry R. and Paul W. Miller (1995). “The Endogeneity between Language and Earnings: International Analyses”. In: *Journal of Labor Economics* 13.2, 246–288. DOI: 10.1086/298374.

— (2005). “Linguistic Distance: A Quantitative Measure of the Distance Between English and Other Languages”. In: *Journal of Multilingual and Multicultural Development* 26.1, 1–11. DOI: 10.1080/14790710508668395.

— (2007a). “Computer Usage, Destination Language Proficiency and the Earnings of Natives and Immigrants”. In: *Review of Economics of the Household* 5.2, 129–157. DOI: 10.1007/s11150-007-9007-0.

— (2007b). *The Economics of Language*. Routledge.

— (2015). “International Migration and the Economics of Language”. In: *Handbook of the Economics of International Migration*. Ed. by Barry R. Chiswick and Paul W. Miller. Vol. 1, 211–269.

Constant, Amelie F. and Klaus F. Zimmermann (2008). “Measuring Ethnic Identity and Its Impact on Economic Behavior”. In: *Journal of the European Economic Association* 6.2-3, 424–433. DOI: 10.1162/JEEA.2008.6.2-3.424.

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg (2018). “Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance”. In: *Journal of Political Economy* 126.6, 2356–2409. DOI: 10.1086/699979.

Currie, Janet (2001). “Early Childhood Education Programs”. In: *Journal of Economic Perspectives* 15.2, 213–238. DOI: 10.1257/jep.15.2.213.

Cygan-Rehm, Kamila, Daniel Kuehnle, and Regina T. Riphahn (2018). "Paid Parental Leave and Families' Living Arrangements". In: *Labour Economics* 53, 182–197. DOI: 10.1016/j.labeco.2018.05.008.

Danzer, Alexander M. and Firat Yaman (2013). "Do Ethnic Enclaves Impede Immigrants' Integration? Evidence from a Quasi-experimental Social-interaction Approach". In: *Review of International Economics* 21.2, 311–325. DOI: 10.1111/roie.12038.

Danzer, Natalia and Victor Lavy (2018). "Paid Parental Leave and Children's Schooling Outcomes". In: *The Economic Journal* 128.608, 81–117. DOI: 10.1111/ecoj.12493.

DaVanzo, Julie (1983). "Repeat Migration in the United States: Who Moves Back and Who Moves On?" In: *Review of Economics and Statistics* 65.4, 552–559. DOI: 10.2307/1935923.

De Jong, Gordon F. (2000). "Expectations, Gender, and Norms in Migration Decision-Making". In: *Population Studies* 54.3, 307–319. DOI: 10.1080/713779089.

Delogu, Marco, Frédéric Docquier, and Joël Machado (2018). "Globalizing Labor and the World Economy: The Role of Human Capital". In: *Journal of Economic Growth* 23.2, 223–258. DOI: 10.1007/s10887-017-9153-z.

Deming, David and Susan Dynarski (2008). "The Lengthening of Childhood". In: *Journal of Economic Perspectives* 22.3, 71–92. DOI: 10.1257/jep.22.3.71.

Destatis (2020). *Foreign Population by Place of Birth and Selected Citizenships*.

Dhar, Diva, Tarun Jain, and Seema Jayachandran (2022). "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India". In: *American Economic Review* 112.3, 899–927. DOI: 10.1257/aer.20201112.

DiPasquale, Glenn W., Allan D. Moule, and Robert W. Flewelling (1980). "The Birthdate Effect". In: *Journal of Learning Disabilities* 13.5, 4–8. DOI: 10.1177/002221948001300501.

Docquier, Frédéric, Joël Machado, and Khalid Sekkat (2015). "Efficiency Gains from Liberalizing Labor Mobility". In: *Scandinavian Journal of Economics* 117.2, 303–346. DOI: 10.1111/sjoe.12097.

Docquier, Frédéric, Çağlar Ozden, and Giovanni Peri (2014). "The Labour Market Effects of Immigration and Emigration in OECD Countries". In: *Economic Journal* 124.579, 1106–1145. DOI: 10.1111/ecoj.12077.

Docquier, Frédéric, Aysit Tansel, and Riccardo Turati (2020). "Do Emigrants Self-Select Along Cultural Traits? Evidence from the MENA Countries". In: *International Migration Review* 54.2, 388–422. DOI: 10.1177/0197918319849011.

Drange, Nina and Kjetil Telle (2015). "Promoting Integration of Immigrants: Effects of Free Child Care on Child Enrollment and Parental Employment". In: *Labour Economics* 34, 26–38. DOI: 10.1016/j.labeco.2015.03.006.

Dustmann, Christian (1994). "Speaking Fluency, Writing Fluency and Earnings of Migrants". In: *Journal of Population Economics* 7.2, 133–156.

— (1996). "The Social Assimilation of Immigrants". In: *Journal of Population Economics* 9.1, 37–54. DOI: 10.1007/PL00003825.

Dustmann, Christian (2003). “Return Migration, Wage Differentials, and the Optimal Migration Duration”. In: *European Economic Review* 47.2, 353–369. DOI: 10.1016/S0014-2921(01)00184-2.

— (2008). “Return Migration, Investment in Children, and Intergenerational Mobility Comparing Sons of Foreign- and Native-Born Fathers”. In: *Journal of Human Resources* 43.2, 299–324. DOI: 10.3368/jhr.43.2.299.

Dustmann, Christian and Francesca Fabbri (2003). “Language Proficiency and Labour Market Performance of Immigrants in the UK”. In: *The Economic Journal* 113.489, 695–717. DOI: 10.1111/1468-0297.t01-1-00151.

Dustmann, Christian, Francesco Fasani, Xin Meng, and Luigi Minale (2020). “Risk Attitudes and Household Migration Decisions”. In: *Journal of Human Resources*, 1019–10513R1. DOI: 10.3368/jhr.58.3.1019-10513R1.

Dustmann, Christian and Joseph-Simon Görlach (2016). “The Economics of Temporary Migrations”. In: *Journal of Economic Literature* 54.1, 98–136. DOI: 10.1257/jel.54.1.98.

Dustmann, Christian and Uta Schönberg (2012). “Expansions in Maternity Leave Coverage and Children’s Long-Term Outcomes”. In: *American Economic Journal: Applied Economics* 4.3, 190–224. DOI: 10.1257/app.4.3.190.

Dustmann, Christian and Arthur Van Soest (2001). “Language Fluency and Earnings: Estimation with Misclassified Language Indicators”. In: *Review of Economics and Statistics* 83.4, 663–674. DOI: 10.1162/003465301753237740.

— (2002). “Language and the Earnings of Immigrants”. In: *Industrial & Labor Relations Review* 55.3, 473–492. DOI: 10.1177/001979390205500305.

Egger, Peter H. and Andrea Lassmann (2012). “The Language Effect in International Trade: A Meta-Analysis”. In: *Economics Letters* 116.2, 221–224. DOI: 10.1016/j.econlet.2012.02.018.

Egger, Peter H. and Farid Toubal (2016). “Common Spoken Languages and International Trade”. In: *The Palgrave Handbook of Economics and Language*. Ed. by Victor Ginsburgh and Shlomo Weber, 263–289. ISBN: 978-1-137-32505-1. DOI: 10.1007/978-1-137-32505-1_10.

Espenshade, Thomas J. and Haishan Fu (1997). “An Analysis of English-Language Proficiency among U.S. Immigrants”. In: *American Sociological Review* 62.2, 288–305. DOI: 10.2307/2657305.

Eurostat (2021). *Migration and Migrant Population Statistics*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Migration_and_migrant_population_statistics.

Facchini, Giovanni, Eleonora Patacchini, and Max F. Steinhardt (2015). “Migration, Friendship Ties, and Cultural Assimilation”. In: *The Scandinavian Journal of Economics* 117.2, 619–649. DOI: 10.1111/sjoe.12096.

Falck, Oliver, Robert Gold, and Stephan Heblich (2017). “Lifting the Iron Curtain: School-Age Education and Entrepreneurial Intentions”. In: *Journal of Economic Geography* 17.5, 1111–1148. DOI: 10.1093/jeg/lbw026.

Felfe, Christina and Rafael Lalive (2018). "Does Early Child Care Affect Children's Development?" In: *Journal of Public Economics* 159, 33–53. DOI: 10.1016/j.jpubeco.2018.01.014.

Fortin, Nicole M., Philip Oreopoulos, and Shelley Phipps (2015). "Leaving Boys Behind Gender Disparities in High Academic Achievement". In: *Journal of Human Resources* 50.3, 549–579. DOI: 10.3368/jhr.50.3.549.

Fredriksson, Peter and Björn Öckert (2014). "Life-Cycle Effects of Age at School Start". In: *The Economic Journal* 124.579, 977–1004. DOI: 10.1111/ecoj.12047.

Friebel, Guido, Juan Miguel Gallego, and Mariapia Mendola (2013). "Xenophobic Attacks, Migration Intentions, and Networks: Evidence from the South of Africa". In: *Journal of Population Economics* 26.2, 555–591. DOI: 10.1007/s00148-012-0455-3.

Fry, Richard and B. Lindsay Lowel (2003). "The Value of Bilingualism in the U.S. Labor Market". In: *ILR Review* 57.1, 128–140. DOI: 10.1177/001979390305700107.

Gallup (2018). *Gallup World Poll.*

— (2021). *Worldwide Research Methodology and Codebook.*

Gambaro, Ludovica, Guido Neidhöfer, and C. Katharina Spiess (2021). "The Effect of Early Childhood Education and Care Services on the Integration of Refugee Families". In: *Labour Economics* 72, 102053. DOI: 10.1016/j.labeco.2021.102053.

Gathmann, Christina and Nicolas Keller (2018). "Access to Citizenship and the Economic Assimilation of Immigrants". In: *The Economic Journal* 128.616, 3141–3181. DOI: 10.1111/ecoj.12546.

Geist, Claudia and Patricia A. McManus (2011). "Different Reasons, Different Results: Implications of Migration by Gender and Family Status". In: *Demography* 49.1, 197–217. DOI: 10.1007/s13524-011-0074-8.

Ginsburgh, Victor, Jacques Melitz, and Farid Toubal (2017). "Foreign Language Learning and Trade". In: *Review of International Economics* 25.2, 320–361. DOI: 10.1111/roie.12268.

Ginsburgh, Victor A. and Juan Prieto-Rodriguez (2011). "Returns to Foreign Languages of Native Workers in the European Union". In: *ILR Review* 64.3, 599–618. DOI: 10.1177/001979391106400309.

Goda, Gopi Shah, John B. Shoven, and Sita Nataraj Slavov (2012). "Does Stock Market Performance Influence Retirement Intentions?" In: *Journal of Human Resources* 47.4, 1055–1081. DOI: 10.3368/jhr.47.4.1055.

Goebel, Jan et al. (2019). "The German Socio-Economic Panel (SOEP)". In: *Jahrbücher für Nationalökonomie und Statistik* 239.2, 345–360. DOI: 10.1515/jbnst-2018-0022.

Goethe Institute (2011). *Jahrbuch 2010/11*. Munich.

— (2014). *Jahrbuch 2013/14*. Munich.

— (2019a). *Annual Report 2018/19.*

— (2019b). *Jahrbuch 2018/2019.*

— (2021). *Jahrbuch 2020/21*. Munich.

Graham, Carol and Julia R. Pozuelo (2022). "Do High Aspirations Lead to Better Outcomes? Evidence from a Longitudinal Survey of Adolescents in Peru". In: *Journal of Population Economics*, 1–39. DOI: 10.1007/s00148-021-00881-y.

Grogger, Jeffrey and Gordon H. Hanson (2011). "Income Maximization and the Selection and Sorting of International Migrants". In: *Journal of Development Economics* 95.1, 42–57. DOI: 10.1016/j.jdeveco.2010.06.003.

Head, Keith, Thierry Mayer, and John Ries (2010). "The Erosion of Colonial Trade Linkages after Independence". In: *Journal of International Economics* 81.1, 1–14. DOI: 10.1016/j.jinteco.2010.01.002.

Hofstede, Geert and Michael Minkov (2013). *Values Survey Module*. Tech. rep.

Huber, Matthias, Ann-Marie Sommerfeld, and Silke Uebelmesser (2022). "Language Learning: Human Capital Investment or Consumption?" In: *Empirica* 49.4, 897–948. DOI: 10.1007/s10663-022-09548-7.

Huber, Matthias and Silke Uebelmesser (2023). "Presence of Language-Learning Opportunities and Migration". In: *Labour Economics* 84, 102409. DOI: 10.1016/j.labeco.2023.102409.

Imbens, Guido W. and Joshua D. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects". In: *Econometrica* 62.2, 467–475. DOI: 10.2307/2951620.

Imbens, Guido W. and Thomas Lemieux (2008). "Regression Discontinuity Designs: A Guide to Practice". In: *Journal of Econometrics*. The Regression Discontinuity Design: Theory and Applications 142.2, 615–635. DOI: 10.1016/j.jeconom.2007.05.001.

Isphording, Ingo E. (2013). "Returns to Foreign Language Skills of Immigrants in Spain". In: *Labour (Committee on Canadian Labour History)* 27.4, 443–461. DOI: 10.1111/labr.12020.

Isphording, Ingo E. and Sebastian Otten (2013). "The Costs of Babylon—Linguistic Distance in Applied Economics". In: *Review of International Economics* 21.2, 354–369. DOI: 10.1111/roie.12041.

— (2014). "Linguistic Barriers in the Destination Language Acquisition of Immigrants". In: *Journal of Economic Behavior & Organization* 105, 30–50. DOI: 10.1016/j.jebo.2014.03.027.

Jacob, Brian A. and Tamara W. Linkow (2011). "Educational Expectations and Attainment." In: *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances.* Vol. Chapter 7. Ed. by Greg J. Duncan and Richard J. Murnane.

Jaeger, David A. et al. (2010). "Direct Evidence on Risk Attitudes and Migration". In: *Review of Economics and Statistics* 92.3, 684–689. DOI: 10.1162/REST_a_00020.

Jaschke, Philipp and Sekou Keita (2021). "Say It Like Goethe: Language Learning Facilities Abroad and the Self-Selection of Immigrants". In: *Journal of Development Economics* 149, 102597. DOI: 10.1016/j.jdeveco.2020.102597.

Kodde, David A. and Jozef M. M. Ritzen (1984). "Integrating Consumption and Investment Motives in a Neoclassical Model of Demand for Education". In: *Kyklos* 37.4, 598–608. DOI: 10.1111/j.1467-6435.1984.tb00760.x.

Lazear, Edward P. (1977). “Education: Consumption or Production?” In: *Journal of Political Economy* 85.3, 569–597.

— (1999). “Culture and Language”. In: *Journal of Political Economy* 107.6, S95.

Lohmann, Johannes (2011). “Do Language Barriers Affect Trade?” In: *Economics Letters* 110.2, 159–162. DOI: 10.1016/j.econlet.2010.10.023.

Lohmar, Bridgette and Thomas Eckhardt (2014). *The Education System in the Federal Republic of Germany 2012/2013: A Description of the Responsibilities, Structures and Developments in Education Policy for the Exchange of Information in Europe*. Tech. rep. Secretariat of the Standing Conference of the Ministers of Education and Cultural Affairs of the Lander in the Federal Republic of Germany.

Manchin, Miriam and Sultan Orazbayev (2018). “Social Networks and the Intention to Migrate”. In: *World Development* 109, 360–374. DOI: 10.1016/j.worlddev.2018.05.011.

Manski, Charles F. (1990). “The Use of Intentions Data to Predict Behavior: A Best-Case Analysis”. In: *Journal of the American Statistical Association* 85.412, 934–940. DOI: 10.1080/01621459.1990.10474964.

Martinovic, Borja, Frank Van Tubergen, and Ineke Maas (2009). “Dynamics of Interethnic Contact: A Panel Study of Immigrants in the Netherlands”. In: *European Sociological Review* 25.3, 303–318. DOI: 10.1093/esr/jcn049.

— (2015). “A Longitudinal Study of Interethnic Contacts in Germany: Estimates from a Multilevel Growth Curve Model”. In: *Journal of Ethnic and Migration Studies* 41.1, 83–100. DOI: 10.1080/1369183X.2013.869475.

Mayda, Anna (2010). “International Migration: A Panel Data Analysis of the Determinants of Bilateral Flows”. In: *Journal of Population Economics* 23.4, 1249–1274. DOI: 10.1007/s00148-009-0251-x.

McAdams, John M. (2016). “The Effect of School Starting Age Policy on Crime: Evidence from U.S. Microdata”. In: *Economics of Education Review* 54, 227–241. DOI: 10.1016/j.econedurev.2014.12.001.

McCrary, Justin (2008). “Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test”. In: *Journal of Econometrics*. The Regression Discontinuity Design: Theory and Applications 142.2, 698–714. DOI: 10.1016/j.jeconom.2007.05.005.

McEwan, Patrick J. and Joseph S. Shapiro (2008). “The Benefits of Delayed Primary School Enrollment Discontinuity Estimates Using Exact Birth Dates”. In: *Journal of Human Resources* 43.1, 1–29. DOI: 10.3368/jhr.43.1.1.

McKenzie, David and Hillel Rapoport (2010). “Self-Selection Patterns in Mexico-U.S. Migration: The Role of Migration Networks”. In: *Review of Economics and Statistics* 92.4, 811–821. DOI: 10.1162/REST_a_00032.

McKinnish, Terra (2008). “Spousal Mobility and Earnings”. In: *Demography* 45.4, 829–849. DOI: 10.1353/dem.0.0028.

Melitz, Jacques and Farid Toubal (2014). “Native Language, Spoken Language, Translation and Trade”. In: *Journal of International Economics* 93.2, 351–363. DOI: 10.1016/j.jinteco.2014.04.004.

Mincer, Jacob (1978). “Family Migration Decisions”. In: *Journal of Political Economy* 86.5, 749–773. DOI: 10.1086/260710.

Müller, Kai-Uwe and Katharina Wrohlich (2020). “Does Subsidized Care for Toddlers Increase Maternal Labor Supply? Evidence from a Large-Scale Expansion of Early Childcare”. In: *Labour Economics* 62, 101776. DOI: 10.1016/j.labeco.2019.101776.

Mundlak, Yair (1978). “On the Pooling of Time Series and Cross Section Data”. In: *Econometrica : journal of the Econometric Society* 46.1, 69–85. DOI: 10.2307/1913646.

Munk, Martin D, Till Nikolka, and Panu Poutvaara (2022). “International Family Migration and the Dual-Earner Model”. In: *Journal of Economic Geography* 22.2, 263–287. DOI: 10.1093/jeg/lbab050.

Oosterbeek, Hessel and Hans Van Ophem (2000). “Schooling Choices: Preferences, Discount Rates, and Rates of Return”. In: *Empirical Economics* 25.1, 15–34. DOI: 10.1007/s001810050002.

Oosterbeek, Hessel and Dinand Webbink (1995). “Enrolment in Higher Education in The Netherlands”. In: *De Economist* 143.3, 367–380. DOI: 10.1007/BF01434014.

Ortega, Francesc and Giovanni Peri (2013). “The Effect of Income and Immigration Policies on International Migration”. In: *Migration Studies* 1.1, 47–74. DOI: 10.1093/migration/mns004.

Papapanagos, Harry and Peter Sanfey (2001). “Intention to Emigrate in Transition Countries: The Case of Albania”. In: *Journal of Population Economics* 14.3, 491–504. DOI: 10.1007/s001480000037.

Puhani, Patrick A. and Andrea M. Weber (2008). “Does the Early Bird Catch the Worm?” In: *The Economics of Education and Training*. Ed. by Christian Dustmann, Bernd Fitzenberger, and Stephen Machin. Studies in Empirical Economics, 105–132. DOI: 10.1007/978-3-7908-2022-5_6.

Rendon, Silvio (2007). “The Catalan Premium: Language and Employment in Catalonia”. In: *Journal of Population Economics* 20.3, 669–686. DOI: 10.1007/s00148-005-0048-5.

Ruyssen, Ilse and Sara Salomone (2018). “Female Migration: A Way out of Discrimination?” In: *Journal of Development Economics* 130, 224–241. DOI: 10.1016/j.jdeveco.2017.10.010.

Saiz, Albert and Elena Zoido (2005). “Listening to What the World Says: Bilingualism and Earnings in the United States”. In: *Review of Economics and Statistics* 87.3, 523–538. DOI: 10.1162/0034653054638256.

Schaafsma, Joseph (1976). “The Consumption and Investment Aspects of the Demand for Education”. In: *The Journal of Human Resources* 11.2, 233. DOI: 10.2307/145456.

Schneeweis, Nicole (2006). *How Should We Organize Schooling to Further Children with Migration Background?* Economics Working Papers 20.

Schultz, Theodore W. (1960). "Capital Formation by Education". In: *Journal of Political Economy* 68.6, 571–583. DOI: 10.1086/258393.

Shastry, Gauri Kartini (2012). "Human Capital Response to Globalization: Education and Information Technology in India". In: *Journal of Human Resources* 47.2, 287–330. DOI: 10.3368/jhr.47.2.287.

Sjaastad, Larry A. (1962). "The Costs and Returns of Human Migration". In: *Journal of Political Economy* 70.5, Part 2, 80–93. DOI: 10.1086/258726.

Spolaore, Enrico and Romain Wacziarg (2016). "Ancestry, Language and Culture". In: *The Palgrave Handbook of Economics and Language*. Ed. by Victor Ginsburgh and Shlomo Weber, 174–211. ISBN: 978-1-137-32505-1. DOI: 10.1007/978-1-137-32505-1_7.

Stark, Oded and David E. Bloom (1985). "The New Economics of Labor Migration". In: *American Economic Review* 75.2, 173–178.

Stöhr, Tobias (2015). "The Returns to Occupational Foreign Language Use: Evidence from Germany". In: *Labour Economics* 32, 86–98. DOI: 10.1016/j.labeco.2015.01.004.

Tjaden, Jasper, Daniel Auer, and Frank Laczko (2019). "Linking Migration Intentions with Flows: Evidence and Potential Use". In: *International Migration* 57.1, 36–57. DOI: 10.1111/imig.12502.

Toomet, Ott (2011). "Learn English, Not the Local Language! Ethnic Russians in the Baltic States". In: *American Economic Review* 101.3, 526–531. DOI: 10.1257/aer.101.3.526.

Uebelmesser, Silke (2006). "To Go or Not to Go: Emigration from Germany". In: *German Economic Review* 7.2, 211–231. DOI: 10.1111/j.1468-0475.2006.00153.x.

Uebelmesser, Silke, Matthias Huber, and Severin Weingarten (2018a). "The German Language Worldwide: A New Data Set on Language Learning". In: *CESifo Economic Studies* 64.1, 103–121. DOI: 10.1093/cesifo/ify007.

— (2018b). "The German Language Worldwide: Data - Course Participation Abroad". In: Harvard Dataverse, V1. DOI: 10.7910/DVN/XVNUY8.

— (2018c). "The German Language Worldwide: Data - Course Participation in Germany". In: Harvard Dataverse, V1. DOI: 10.7910/DVN/V5BWMT.

Uebelmesser, Silke, Ann-Marie Sommerfeld, and Severin Weingarten (2022). "A Macro-Level Analysis of Language Learning and Migration". In: *German Economic Review* 23.2, 181–232. DOI: 10.1515/ger-2020-0067.

Van Dalen, Hendrik P. and Kène Henkens (2012). "Explaining Low International Labour Mobility: The Role of Networks, Personality, and Perceived Labour Market Opportunities". In: *Population, Space and Place* 18.1, 31–44. DOI: 10.1002/psp.642.

Wanner, Philippe (2021). "Can Migrants' Emigration Intentions Predict Their Actual Behaviors? Evidence from a Swiss Survey". In: *Journal of International Migration and Integration* 22.3, 1151–1179. DOI: 10.1007/s12134-020-00798-7.

World Bank (2021). *World Development Indicator*.

Yao, Yuxin and Jan C. van Ours (2015). “Language Skills and Labor Market Performance of Immigrants in the Netherlands”. In: *Labour Economics* 34, 76–85. DOI: 10.1016/j.labeco.2015.03.005.

Declaration for Admission to the Doctoral Examination

I confirm

1. that the dissertation “Empirical Studies on Migration – From Language Learning to Integration Processes” that I submitted was produced independently without assistance from external parties, and not contrary to high scientific standards and integrity,

2. that I have adhered to the examination regulations, including upholding a high degree of scientific integrity, which includes the strict and proper use of citations so that the inclusion of other ideas in the dissertation are clearly distinguished,

3. that in the process of completing this doctoral thesis, no intermediaries were compensated to assist me neither with the admissions or preparation processes, and in this process,

    No remuneration or equivalent compensation were provided

    No services were engaged that may contradict the purpose of producing a doctoral thesis

4. that I have not submitted this dissertation or parts of this dissertation elsewhere.

I am aware that false claims (and the discovery of those false claims now, and in the future) with regards to the declaration for admission to the doctoral examination can lead to the invalidation or revoking of the doctoral degree.

Date: ____________________________ Signature: ____________________________

Author Contributions

The main part of the thesis builds on four research papers, three of which are joint work with co-authors. My contributions to the papers are as follows:

- Chapter 2 is based on the paper titled *Language Learning: Human Capital Investment or Consumption?*, which is joint work with Matthias Huber (Federal Office for Migration and Refugees and Friedrich Schiller University Jena) and Silke Uebelmesser (Friedrich Schiller University Jena and CESifo). It is published in *Empirica* 49.4, 897–948 (DOI:10.1007/s10663-022-09548-7). Matthias Huber and Silke Uebelmesser developed the research question and the draft of the manuscript. Matthias Huber provided the data collection and preparation as well as the first implementation of the empirical analysis. I substantially contributed to the empirical analysis and the writing of the manuscript.

- Chapter 3 is based on the paper titled *A Macro-Level Analysis of Language Learning and Migration*, which is joint work with Silke Uebelmesser and Severin Weingarten (Friedrich Schiller University Jena). It is published in *German Economic Review* 23.2, 181–232 (DOI:10.1515/ger-2020-0067). Silke Uebelmesser and Severin Weingarten developed the research question and the draft of the manuscript. Severin Weingarten provided the data collection and preparation as well as the first implementation of the empirical analysis. I substantially contributed to the empirical analysis and the writing of the manuscript.

- Chapter 4 is based on the paper titled *Migration Aspirations and Intentions*, which is joint work with Matthias Huber, Till Nikolka (German Youth Institute and CESifo), Panu Poutvaara (University of Munich, ifo Institute, CESifo, CReAM and IZA), and Silke Uebelmesser. It is pre-published as CESifo Working Paper No. 9708 (Link). Matthias Huber and Silke Uebelmesser mainly developed the research question. The collection and preparation of the Goethe Institute data was provided by Matthias Huber, and the data for the survey among university students was provided by Till Nikolka and Panu Poutvaara. I implemented the empirical analysis and drafted the largest part of the manuscript with Matthias Huber's and Silke Uebelmesser's support. All co-authors contributed to the collaborative development of the paper.

- Chapter 5 is based on the yet unpublished paper *The Effect of Schooling on Parental Integration*, which is single-authored.

Date: ______________________ Signature: ______________________