Language-based discrimination and its role for ethnic inequalities in the educational system and the labor market

Miriam Schmaus

Doctoral thesis, submitted 09/2018

First supervisor: Prof. Dr. Cornelia Kristen

Second supervisor: Prof. Dr. Lucinda Platt

Third supervisor: Prof. Dr. Steffen Schindler

Date of Defence: Dezember 19, 2018

Table of contents

* *Notes are presented as chapter endnotes.*

URN: urn:nbn:de:bvb:473-irb-469492
DOI: https://doi.org/10.20378/irb-46949

Preface

Proficiency in a host country language (L2) is crucial to immigrants' structural integration into 'core' societal institutions, such as the education system and the labor market (Esser 2006). It is widely accepted that educational achievement and trajectories notably improve with L2 proficiency (e.g. Stanat and Edele, 2016). Similarly, considerable research demonstrates that income, occupational attainment, and employment opportunities increase with rising skills in the relevant national language (e.g. Chiswick and Miller, 2001; Dustmann and van Soest, 2002; Lindemann and Kogan, 2013; Kalter, 2006; van Tubergen et al., 2004).

The importance of (L2) language for immigrants' structural integration is attributable to its multiple functionality (Esser, 2006). Language, in general, is assumed to fulfil three different functions. Firstly, (L2) language proficiency is seen as a crucial *resource*, which, as part of an actor's human capital, increases workers' productivity on the labor market, or students' learning efficiency within schools. Secondly, it serves as a *medium,* which facilities the transfer of information in firms or schools, thus reducing transaction costs. Thirdly, it can also have *symbolic* effects, as the linguistic patterns associated with lower (L2) language proficiency, such as deviant vocabulary, grammar, or pronunciation, saliently signal the speaker's ethnic origin and emphasize foreignness, which can be associated with different types of discrimination (ibid.).

Studies analyzing large-scale survey data, which are commonly called upon to demonstrate the significance of L2 language for immigrants' structural integration, predominantly apply arguments based on the first two language functions[1] and often disregard assumptions about language-based forms of discrimination. For instance, 74 % of examined SSCI ranked studies (1990-2017) that indicate effects of L2 language proficiency for immigrants' labor market integration based on large-scale survey data, consider the respective proficiency solely as a (human capital) resource, that facilitates communication or apply related assumptions (Figure A, Table A in Appendix). Conversely, 26 % of these studies also mention arguments on language based forms of discrimination and less than half of them specifically address such arguments. Patterns for educational outcomes seem similar, however here research is less ample, arguments more diverse and some studies have a different research focus, so that L2 effects are only very briefly discussed (Figure B, Table B in Appendix).

Figure A: Studies examining immigrants' labor market outcomes

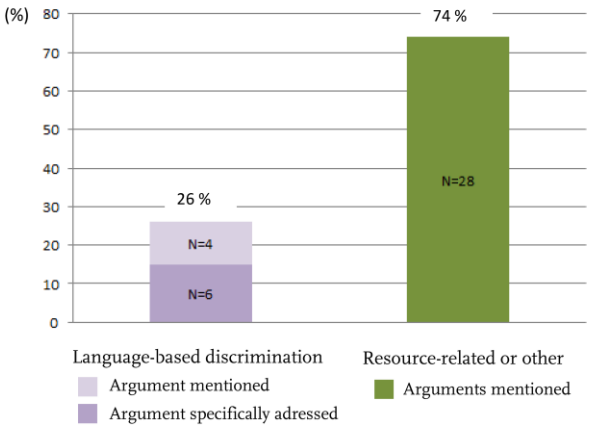
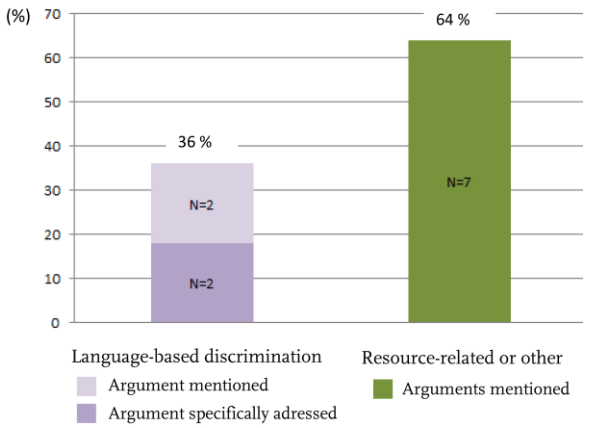


Figure B: Studies examining immigrants' educational outcomes



Note: Author's illustrations. Summary statistics are based on the information in Table A (Figure A) and Table B (Figure B) (Appendix A). Considered studies are SSCI ranked, were conducted between 1990 and 2017 and examine the relationship between L2 proficiency and immigrants' educational or labor market outcomes with large-scale survey data. Tables include all studies that were located by the author and meet these criteria (with a small number of marked exceptions). However, completeness is not claimed.

Those large-scale studies that specifically examine arguments on language-based discrimination often seem to find such effects. Certain immigrant groups seem to get particularly penalized for lower L2 language skills and deviant linguistic patterns (Stolzenberg and Tienda, 1997; Kossoudji, 1988; Davila et al., 1993; Pendakur and Pendakur, 2002. For one exception see Kratzmann and Schneider, 2009), which the authors interpret as indicative of language-based discrimination. In addition to that, there is experimental evidence which suggests that certain linguistic features, such as first and second language accents, are associated with unfavorable evaluations and differential treatment of applicants (Table C in Appendix) and students (Table D in Appendix).

While providing valuable indications that language-based forms of discrimination may matter for immigrants' structural integration, the generalizability of these findings may be limited. Studies that examine such processes on the labor market by analyzing large-scale survey data generally employ self-reported measures of L2 proficiency (Kossoudji, 1988; Stolzenberg and Tienda, 1997; Pendakur and Pendakur, 2002; Davila et al., 1993), which might make it difficult to adequately assess arguments on language-based discrimination. The patterns of interest mainly relate to lower L2 proficiency, yet most immigrant groups seem to overestimate their own language abilities and scarcely report low proficiencies (Edele et al., 2015). Moreover, some studies examine intricate arguments on interrelations between L2 proficiency and other factors that may shape the extent of ethnic discrimination, such as educational attainment (Stolzenberg and Tienda, 1997), yet they bypass other, more direct forms of language-based discrimination. Other studies examine specific penalties resulting from minority language proficiency (L1) (Pendakur and Pendakur, 2002). The abilities in question also enable access to ethnic enclave economies, which might mitigate or mask effects of language-based discrimination (ibid). *Lastly,* all (four) of the respective studies examined one singular labor market outcome, namely earnings, which might paint the most conservative picture of the potential role of language-based discrimination. One the one hand, discriminatory tendencies seem to reduce over the course of the hiring process (Cedie and Foronie, 2008, 107), so that employed immigrants and their earnings may be less affected by any kind of discrimination than those in search of (new) work. On the other hand, as employed immigrants have already passed several hurdles and convinced employers of their qualifications, they may be positively selected on unobservable traits, such as motivation or drive. This might additionally reduce or cancel out discriminatory inclinations. In addition to that, the processes of language-based discrimination that affect earnings within labor markets may be (vastly) different from those within schools. Studies that examine arguments on language-based discrimination with regard to educational outcomes by analyzing large-scale survey data are also either based on self-reported measures (Barrett et al., 2012), or minority language exposure (Shifer et al., 2011) and examine very specific arguments (Kratzmann and Schneider, 2009) or outcomes (ibid.; Shifer et al., 2011).

Experimental studies might sometimes struggle with threats to internal validity. For instance, stimulus materials of certain studies provide additional markers of ethnic origin, such as the name, or pictures, alongside the language cue of interest, which exacerbates inferences on purely language-related effects (e.g. Hosoda and Stone-Romero, 2010). Studies that examine the role of language cues for immigrants' labor market outcomes within such experimental frameworks sometimes also struggle with threats to external validity. They by and large employ student samples, which have been shown to exhibit lower levels of ethnic discrimination in fictitious hiring situations than other parts of the population (Blommaert et al., 2013) and might pay attention to different criteria than employers when evaluating potential applicants. Moreover, most studies were (more or less) conducted in laboratory settings, which generally do not involve

future interactions with (fictitious) applicants, thus reducing discriminatory tendencies that are based on the wish to avoid personal contact with members of certain ethnic groups. Studies that experimentally examine the role of language cues for students' educational outcomes, in general, were conducted several decades ago and might not always meet standards of current experimental research, such as unambiguous stimulus materials (e.g. Williams et al., 1972) or clear-cut randomization (e.g. Choy and Dodd, 1976). Moreover, in contrast to studies that analyze survey data, experimental research sometimes disregards resource-related arguments on language-related productivity or efficiency when interpreting the respective findings (e.g. Crowl and MacGinitie, 1974).

This dissertation attempts to disentangle the effects of language-related productivity and efficiency from those of language-based discrimination. In more general terms it aims to assess the concrete role of language-based discrimination for immigrants' structural integration into the educational system and the labor market.

Two broad questions guide this endeavor. One asks about the *extent to which language-based discrimination affects immigrants' educational or labor market outcomes beyond aspects of language-related productivity or efficiency*. The second question relates to *the motives and mechanisms behind such forms of discrimination and asks about the main drivers of language-based discrimination within the educational system or on the labor market*.

I examine these issues within three different research papers. While the scope that is dedicated to each aspect varies, all papers loosely address both of these questions. The first two papers examine the respective processes on the labor market, the third one within the educational system.

Paper 1

In paper one, I aim to disentangle the effect of (L2) language related productivity from that of language-based discrimination for immigrants' labor market integration in Germany. Exploiting data from the National Educational Panel Study (NEPS), I examine whether immigrant groups associated with unfavorable attitudes receive lower returns to lower German language skills than groups facing more positive attitudes or the native population, and how large the respective discounts may be. I also investigated whether the size of these discounts varies with the amount of information available to employers to distinguish between different types of discrimination. Findings indicate that 'unpopular' immigrant groups receive discounts to lower L2 proficiency, irrespective of the amount of information available to employers. This could suggest that, in addition to affecting employees' productivity, lower language proficiency might also be associated with taste-based forms of discrimination.

Paper 2

As the first paper allows only indirect inferences on the occurrence and roots of language-based discrimination, Paper 2 takes a more direct approach. It focuses on a salient language cue, namely a foreign accent, and attempts to assess how this affects the labor market prospects of immigrants in Germany within a field experimental framework. Here again, of primary interest is whether the respective language cue affects employers' evaluations predominantly in terms of productivity or whether it also represents a source of bias, which triggers discrimination. The field experiment focuses on the initial contact between employers and applicants who call to inquire about an advertised position. The experimental audit thus allows assessments of whether foreign-accented

speech affects the opportunity to partake in the hiring process. In addition to varying applicant characteristics, such as the job seeker's accent and name, information on job and firm characteristics was collected. This included information on the communicative demands or the extent of customer contact in the advertised job, as well as information on company orientation or the structure of the recruitment process. Moreover, from the Federal Employment Agency and the German General Survey, context information was taken and merged to the experimental set. This included information on regional labor supply, job-specific vacancy rates, or regional levels of anti-immigrant attitudes. The aim behind this procedure was to allow for tentative inferences on the reasons behind any accent effects. Results demonstrated that foreign-accented applicants were turned down more often than German jobseekers and accent-free applicants of the same origin. This disadvantage was not conditional on the communicative demands of the job in question, which might suggest that an accent may not be primarily considered in terms of productivity. Moreover, the adverse effects of foreign-accented speech were less pronounced in firms with a more standardized recruitment processes, where employers might have less room to factor personal tastes into their decision-making. In addition to that, negative accent effects were more pronounced in regions with higher levels of anti-immigrant attitudes, which might also hint at the potential role of employers' tastes.

Paper 3

The third paper addresses processes of language-based discrimination within the educational system. It examines how lower second language proficiency affects the identification of special educational needs (SEN) for language minority children in England, about which there has been little research to date. Drawing upon risk-aversion and discrimination theories, I investigate whether the language-related ambiguities of language minority children are associated with diagnosing hesitancy, or if they give rise to different types of discrimination. Building upon competence development models, I also assess how inaccurate SEN judgements affect students' educational achievement. Employing data from the Millennium Cohort Study (MCS), I examine discrepancies between children's performance-based predicted probability of being identified with SEN and teachers' actual SEN identification. Using logistic regression models I assess whether language minority children with low levels of second language proficiency are more likely to experience diagnosing hesitancy or biased judgements upon school entry. Next, I employ linear regression models to assess how inaccurate SEN identification upon school entry affects children's Maths and English competences at the end of primary school. Results indicate group-specific effects. For black children, limited English proficiency was associated with over-identification, particularly of behavioral needs, which might reflect biased judgements. All other groups were less likely to be over-identified, which might reflect a greater (mis)diagnosing hesitancy. Inaccurate SEN identification was negatively related to educational achievement.

These studies go beyond the scope of prior research in the following aspects: *Firstly*, when analyzing large-scale survey data, I draw upon test-based L2 language proficiencies, which should result in more adequate assessments of the respective abilities and might thus allow for more clear-cut tests of the relevant arguments (Paper 1 & Paper 3). *Secondly,* paper 3 is one of the few studies thus far that examine the effects of language-based discrimination for school-related outcomes by analyzing large-scale survey data. Moreover, it is the first investigation (to the author's knowledge) of how inaccurate SEN identification, in this case as a consequence of language-based discrimination, affects student competence acquisition. *Furthermore,* when

examining the role of language-based discrimination for immigrants' labor market integration by analyzing large-scale survey data, paper 1 is the first (ibid.) to consider different indicators of labor market success that relate to employment decisions, as well as to subsequent outcomes such as job status, job type or earnings. *Thirdly,* in contrast to prior research, which was mainly conducted in (more or less) laboratory-like settings, paper 2 was based on a field experiment, which might be associated with a higher degree of external validity, as subjects were employers expecting future interactions with the respective applicants. Moreover, the field experimental design included systematic variations of language-based cues and other indicators of ethnic origin, as well as the additional collection of job, firm and context data, which enabled comparatively broad assessments of the relevant processes. *Lastly,* I examined direct forms of language-based discrimination as well as potential interrelations between L2 proficiency and other factors, such as the availability of additional information, which allows for additional and complementary inferences on how language-based discrimination affects the respective outcomes (paper 1).

Together, these studies hope to contribute to the literature that examines the reasons behind ethnic inequalities and might call for a sensitive interpretation of language effects on structural integration outcomes. By examining influences on immigrants' structural integration, this dissertation, in a broader sense, ties in with research that addresses issues of inequality and stratification within society more generally and relates to aspects that shape social cohesion as a whole (Esser 2006). Policy implications could relate to earlier and more widespread language training, as well as to specific anti-discrimination measures, such as more standardized recruitment or student evaluation processes.

Endnotes to chapter

1 In general, arguments on the L2 language's resource and medium function should be closely related. As a resource, L2 proficiency increases an actor's productivity; as a medium, it reduces transaction costs and facilitates communication, which ultimately should again affect the total productivity in firms or schools. I thus sum these two functions in the provided overview of the literature and the following argumentation. However, when explaining L2 effects on the labor market, it has sometimes also been argued that employers' *anticipation* of transaction costs affects outcomes for less proficient speakers (Esser, 2006). This line of arguments does not directly relate to the medium function, but, in essence, reflects a type of statistical discrimination, which I consider as such in the corresponding part of my theoretical considerations.

Theoretical considerations – a note on the applied arguments

When trying to explain the reasons behind any kind of discrimination, literature in general distinguishes between arguments on the individual and on the institutional level. Individual-level discrimination includes all forms of differential treatment that are carried out by individual actors, solely based on certain group memberships, resp. ascriptive characteristics of their respective counterparts (Blank et al., 2004, 39f; Diehl and Fick, 2012, 5). Institutional-level discrimination, on the other hand, goes beyond the actions of individual actors and proposes that discrimination is "woven into the very fabric of society" (e.g. Radtke, 1996, 122), for instance in prevailing norms, laws, and both formal and informal regulations (Feagin and Feagin, 1986; Kristen, 2006, 5). These rules and regulations can have a differential and adverse impact on members of certain groups (Blank et al., 2004, 39f; Diehl and Fick, 2012, 39), "behind the backs" of individual decision makers and irrespective of their personal intent (Radtke, 1996, 122; Kristen, 2006, 5).

When explaining individual level forms of discrimination, the lion's share of arguments as to why actors discriminate is brought forth by economic and socio-psychological works. Economic approaches indicate two main reasons for discriminatory behavior. Models of taste discrimination consider elements of prejudice and group-specific distaste as the main motive[2] for discriminatory behavior (Becker, 1971). Theories of statistical discrimination, on the other hand, regard a lack of information as critical (Arrow, 1973; Aigner and Cain, 1977; Phelps, 1972). In lieu of information on a certain individual, actors are assumed to draw upon group-specific beliefs, which can be more or less accurate (England, 1992). In turn they can over-, or underestimate this individual (Arrow, 1973; Aigner and Cain, 1977; Phelps, 1972) or all members (England, 1992) of the respective groups.

Socio-psychological approaches can be seen as complementary to these economic models. One the one hand, they provide more nuanced descriptions of the two core concepts of prejudice and group-specific beliefs and what these actually entail. On the other hand, they give ample arguments as to how these concepts come to exist. Group-specific beliefs, for example, can be understood as stereotypes, meaning socially shared knowledge structures, about attributes and characteristics that members of certain groups typically have (e.g. McRae et al., 1996). These assumptions can be more or less accurate and more or less accepted, yet most members of society are assumed to know the respective content (ibid.). The origins of such stereotypic beliefs are for instance traced back to illusory correlations, or a distorted "grain of truth", which are then "conveyed and reproduced in all the usual socio-cultural ways", such as families, schools, or the media (Brown, 1995: 83). Stereotypic beliefs are also seen as the base for prejudice. However, here, they are paired with a negative affective response, acceptance of the respective assumptions, and a tendency to act upon these beliefs and feelings in a negative and harmful way (Allport, 1979; Werth and Mayer, 2007, 379). Origins of prejudice, for example, are seen in conflict over actual or perceived competition over scarce resources (Campbell, 1965; Sherif, 1966), feelings of relative deprivation (Gurr, 1970; Runciman, 1966), or threats to social identity (Tajfel and Turner, 1986). In addition to that, socio-psychological theories provide many arguments on the conditions that prevent or foster reliance on prejudice and group-specific beliefs, such as opposing or consenting external norms, personal motivation, or situational factors, such as time pressure or alcohol use (Werth and Mayer, 2007).

Institutional forms of discrimination can be the result of prior and far-reaching prejudice and group-specific beliefs on the individual level, yet they may also occur as by-products of norms, rules, or regulations with neutral or even positive intent (e.g. Gomolla, 2010; Feagin and Feagin,

1986). Literature thus distinguishes between direct and indirect forms of institutional discrimination[3]. Direct forms of institutional discrimination can stem from prejudice or historic beliefs about the inferiority of certain groups, or may relate to the wish of a dominant group to secure their privileges and protect economic and political interests (Feagin and Feagin, 1986). These direct forms comprise all informal or highly formalized rules on an organizational, community, or societal level, which have "an *intentionally* differential and negative impact on members of subordinate groups" (Feagin and Feagin, 1986, 30). Indirect forms of institutional discrimination on the other hand refer to all norms, rules, or regulations that have an *unintended* adverse impact on members of certain groups and the underlying rationale "appears fair or at least neutral" (ibid.).

L2 language proficiency and discrimination

L2 language proficiency, or different linguistic features, may relate to each of these types of discrimination. On the individual level, L2 language proficiency may firstly *shape the level of prejudice or 'distaste'* that it potentially triggered. Limited L2 proficiency and the associated linguistic cues are very salient reminders of the speaker's ethnic origin (Rakić et al., 2011; Rödin and Özcan, 2011) and additionally signal 'foreignness' (Kossoudji, 1988, 211) or an 'unintegrated' immigration status (Horr et al., 2010). Together, this might trigger stronger aversions than other group-specific indicators. Moreover, individuals may react particularly strongly to language, as "natural selection may have favored attention to language cues" (cf. Kinzler et. al, 2009, 625), as social groups have not always been unambiguously identifiable by physiognomy or appearance, but by speech patterns (ibid.). In addition to that, ethnolinguistic identity theory proposes a strong bond between language and ethnic identity (Giles and Johnson, 1987). As prominent indicators of outgroup-status, limited L2 proficiency and the corresponding patterns may thus trigger a notably negative affective response within members of the majority group (Rakić and Stoeßel, 2013, 15; Bresnahan et al., 2002). Moreover, considerable evidence suggests that language cues may be dominant in impression formation (Rakić et al., 2011; Rödin and Özcan, 2011), and can convert ethnic 'distaste', in that a rejected 'foreign-looking' person was chosen as a friend when she spoke in a 'native-like' manner (Kinzler et. al, 2009).

Secondly, limited L2 language proficiency can exacerbate evaluations of personal attributes, such as competences, skill level, or attitudes, thus *limiting access to relevant individual information*. In other words, lower L2 proficiency can leave actors guessing as to whether their less proficient counterparts actually possess certain traits, or if the lower language skills simply do not permit their identification. The resulting insecurity can *foster the application of group-specific beliefs* (Arrow, 1973; Aigner and Cain, 1977; Phelps, 1972), which for some groups may be systematically biased (England, 1992). The application of such beliefs should be additionally enhanced when lower L2 language proficiency is combined with other indicators of ethnic origin, such as one's name or appearance, as the presence of multiple consistent group stimuli is assumed to facilitate stereotype application (Fiske et al., 1999). Yet, lower L2 language proficiency by itself can be enough to trigger ethnic beliefs and their application, and has been shown to overwrite associations that opposing cues trigger (Rakić et al. 2011; Rödin and Özcan 2011).

On labor markets, this may mean that employers not only value language skills in terms of productivity, but also *factor in strong(er) group-specific distaste* when evaluating immigrant jobseekers with lower L2 language skills or deviant language patterns. 'Unpopular' ethnic groups might thus receive discounts to lower (L2) language proficiency or deviant linguistic patterns. When language skills increase, however, and deviances disappear, the speaker's ethnic origin

should become less salient and distaste should reduce, which might allow non-preferred groups to 'catch up' with other ethnic groups or the native population.

In addition to that, lower L2 language proficiency may hamper employers' assessments of jobseekers' skills and competencies, *limiting information on applicants' 'true' productivity*. On the one hand, lower L2 skills may obscure the observability of specific or soft skills, leaving employers guessing whether applicants actually 'lack' certain abilities, or if the missing language skills simply 'block' their identification. On the other hand, employers may expect a greater variance in unobservable abilities for less proficient speakers (cf. Kossoudji, 1988, 209), as in some circumstances both less and highly productive immigrants can exhibit lower L2 proficiency, for instance shortly after their arrival. In addition to that, employers may be unsure about transaction costs, as they might feel unable to assess the 'true' extent of training costs and upheaval within teams that can be associated with hiring a less proficient speaker (Esser, 2006). If employers thus lack information about a (prospective) worker's 'true' value, they may try to reduce this uncertainty by drawing upon their assumption about the (average) productivity of the worker's ethnic group (Arrow, 1973; Aigner and Cain, 1977; Phelps, 1972). Lower L2 language proficiency may thus *foster the application of ethnic beliefs*, which again should have particularly adverse consequences for groups associated with predominantly negative beliefs (England, 1992). Higher L2 skills, on the other hand, may prevent such applications, as when the necessary information is available employers should have no need to draw upon their beliefs to begin with.

Similar processes might also relate to students and teachers within schools. Here too, lower L2 language skills should exacerbate the evaluation of students' competences, as they obscure the observability of children's actual abilities. This holds especially true when trying to assess students' special educational needs. Difficulties associated with lower proficiency in the language of instruction overlap with indications of special learning needs, *limiting information on children's 'true' abilities*. For instance, problems with the correct production of sounds, or with understanding and conveying information through language, associated with second language learning, overlap with indications of 'communication and interaction needs', or could be interpreted as specific learning difficulties such as dyslexia (e.g. Case and Taylor, 2005). Moreover, still developing L2 language proficiency might prevent children from fully following lessons, causing demeanors that can be interpreted as signs of 'behavioral difficulties', such as short attention spans, temper outbursts, low self-confidence, a general lack of interest, or apathy (Ortiz and Maldonado-Colón, 1986). It might also be associated with general underachievement, which could be interpreted as (moderate) 'cognitive difficulties'. The resulting insecurities due to limited information on children's 'true' abilities can, again, be associated with *application of ethnic beliefs*. Consequently, groups associated with negative beliefs about intelligence, educability, or deviance may be more likely to be misidentified with different educational needs than groups associated with more positive beliefs about educational or behavioral attributes. Good command of the host country language should, again, reduce the risk of misidentification, as all necessary information is available[4].

In addition to that, lower L2 language proficiency might also enhance *group-specific distaste* within teachers. As models of taste discrimination centre around the idea that actors strive to compensate or – when possible – avoid the psychological costs that they face when they interact with members of their non-preferred groups, implications might again particularly revolve around the identification of special educational needs. The identification of such needs can function as a strategy to decrease interactions with students of the respective groups and thus

reduce the associated psychological costs, as it makes them eligible for external help and more likely to spend time outside the regular classroom setting. Consequently, here again, especially less proficient children of 'unpopular' ethnic origins might be more likely to be misidentified with special educational needs. When the speaker's origin becomes less salient with rising L2 language skills, here, too, such patterns should reduce.

Limited L2 language proficiency might also trigger certain types of discrimination that affect all L2 learning children. As limited L2 proficiency saliently signals students' out-group membership, it may be associated with *(perceived) interethnic competition and conflicting group interests*. In this context it has been argued that adverse evaluations or grading, unfavorable track recommendations, or the identification of special educational needs might be employed as some sort of "early stratification measures" to keep out-group children from reaching valued resources, such as spots in distinguished colleges, or (good) employment positions (Sullivan and Artiles, 2011, 1531). The ambiguities that children with limited L2 language proficiency pose might be an easy handle to justify each of these measures. Moreover, the salient out-group signal might foster "ultimate attribution errors" (Pettigrew, 1979), suggesting that negative events are likely attributed to unfavorable dispositions within out-group members themselves. Majority teachers might thus be more prone to (mis)attribute school-related difficulties to unfavorable personality or ability dispositions within less proficient students themselves, rather than to the external factor of being a second language learner, which again might be associated with adverse evaluations or grading, unfavorable track recommendations, or the misidentification of special educational needs.

In addition to these processes on the individual level, limited L2 language proficiency and the associated linguistic patterns might also be associated with institutional-level discrimination. One the one hand, this may engender strategies or regulations *that aim to disadvantage* the respective speakers; on the other hand, neutral rules or regimentations may *have unintended negative consequences* for less L2 proficient actors.

On the labor market, regulations that aim to keep less proficient speakers out of professions or organizations may relate to the *requirement of 'good L2 language command' for positions where such skills aren't necessary to adequately carry out the relevant job tasks*. For instance, the rejection of an African applicant for a job as a mailman on grounds of "not being able to communicate clearly in the German language" was ruled unlawful by a German court. It was deemed "neither appropriate nor necessary" to require L2 language skills to be a mailman, and thus seen as a form of institutional or mediate discrimination (Antidiskriminierungsstelle des Bundes, 2012, 7). Such direct strategies might also relate to requesting additional language certificates for second language learners, while language skills of the majority population do not need to be certified beyond the common credentials.

Indirect forms of institutional discrimination, i.e. rules that have unintended negative consequences for less L2 proficient speakers, may relate to *the requirement of good language command for positions that do require such skills to adequately carry out the job tasks*. The rationale is fair or at least justified; however, the regulation still adversely affects less proficient L2 speakers, as the respective positions are often associated with particularly favorable outcomes, such as high income, tenure, or private health insurance.

Within the educational system, discussions of such indirect forms of institutional discrimination, for instance, revolved around unreliable test instruments (e.g. Mercer and Brown, 1973). *Commonly used IQ tests and other standardized measures can be less accurate in the case of less L2 proficient students*, as they often rely on (cultural) knowledge that is specific to the majority

population and – when administered in the language of instruction – also function as a test of second language proficiency (Burr et al., 2015; Duran, 1988). Reliance on such measures might thus be associated with test-inherent misevaluations of L2 language learning children (e.g. MacSwan and Rolstad, 2006). Moreover, it has been noted that within schools, children's general *linguistic abilities are measured by their skills in the L2 language* (Gomolla and Radtke, 2000, 326; Kristen, 2006, 7). This, again, may inadequately capture the respective abilities for second language learning children, and underestimate their 'true' linguistic potential. In addition to that, it has been criticized that *schools reproduce and reward attributes or behaviors of the educated middle-class, which typically includes proficiency in the standard speech variety* (Gomolla, 2003, 105; Diehl and Fick, 2012, 13). Though being fair in setting equal expectations for all children within school, these standards are disproportionally harder to reach for lower class or immigrant children, who often are second language learners or speak non-standard speech varieties, such as regional dialects (ibid.). In addition to that, lower proficiency in the language of instruction might make it disproportionally difficult for less L2 proficient students to follow the lessons, which can adversely affect educational achievement (Diehl and Fick, 2012, 30). *L2 language-based schooling* per se might thus also be seen as indirect institutional discrimination.

Other contributions to this discussion also address more direct forms of institutional discrimination that relate to the organizational processes within classrooms and schools. It has been suggested that schools might feel ill-prepared to deal with linguistically diverse students or to provide teaching for multilingual classrooms (Gandara et al., 2005; Gomolla and Radtke, 2003). They might thus *develop informal strategies to reduce such difficulties and relieve some of the burdens of inter-lingual teaching*. These informal strategies may include referring less L2 proficient children to 'school kindergartens' or preparatory classes (Gomolla and Radtke, 2000, 329; Kristen, 2005, 6). They may also include the identification of special educational needs and referrals to suitable school types, particularly when resources for specific second language learning programs are scarce (Gomolla and Radtke, 2003; Wagner and Powell, 2003; Diehl and Fick, 2012, 37). Authors who propose such arguments suggest that schools rely on "intricate arguments that are built on second language difficulties" (Gomolla and Radtke, 2003, 208) to justify their respective routines. These "intricate arguments" can, for instance, include the assertion that lower L2 proficiency makes it more difficult to follow the lessons, which might result in a loss of motivation and subsequent learning disadvantages in all subjects (Gomolla and Radtke, 2003, 208), making mainstream school attendance the least preferable option. In general, such strategies may reflect a pragmatic, yet intentional and biased routine to facilitate knowledge transfer in linguistically diverse classes and to tackle the associated organizational difficulties.

I draw upon all of the mentioned arguments within my three research papers, yet to a different extent. When examining the role of language-based discrimination for immigrants' labor market integration, I primarily focus on individual-level forms of discrimination. When assessing its role for educational outcomes, I consider arguments on both levels. The main reason for this is that institutional discrimination seems to be of greater relevance for processes within the educational system, and its role has been heavily debated for the corresponding outcomes. Nevertheless, with benching institutional-level arguments when examining immigrants' labor market outcomes, I might be excluding potential influences of language-based discrimination. Results may thus paint a conservative picture of the full extent of language-based discrimination on the labor market.

Language and productivity

For both contexts, I aim to investigate whether the mentioned types of language-based discrimination affect outcomes over and above resource-related influences of L2 proficiency. Resource-based approaches primarily regard L2 language proficiency as an element of human capital. The basic idea here is that, like other aspects of human capital, language skills are embodied in the person in question. They are costly to gain, but productive, as they can raise all kinds of benefits (Chiswick and Miller, 1995, 248; Chiswick, 2008, 4; Esser, 2006).

On the labor market, L2 proficiency may for example facilitate communication with supervisors, co-workers, suppliers, or customers. This allows workers with better language skills to be more efficient on a given job (Chiswick and Miller, 1995, 248; Chiswick, 2008, 4). The associated increase in productivity should translate into higher earnings and other more favorable labor market outcomes. This mechanism should generally work for all ethnic groups alike. Within schools, L2 proficiency should improve students' learning efficiency. Information within classes is generally transmitted in the host country language and proficiency in that language thus facilities students' knowledge acquisition (Esser, 2006). Here, too children of all ethnic groups should profit alike.

An integrative framework?

While such resource-based arguments, in general, compete with assumptions about language-based discrimination, both types of arguments could be integrated into a general theory of action. The model of frame selection (e.g. Kroneberg, 2005), for instance, allows explicit modeling of how and when decision-makers consciously optimize and evaluate more rationally, e.g. in terms of productivity or efficiency, and when they make more subconscious, potentially biased judgements. This should, for example, depend on (1) their framing of the situation, e.g. either as context-specific, such as hiring or teaching scenarios, or as an intergroup setting, (2) the known set of actions on how to behave within such situations, e.g. fair or antagonistic, and (3) the evaluation of each of these actions. For each of these steps, the selection depends on the salience of certain situational cues, the accessibility of the respective frames and scripts, the expected payoffs, and more generally whether sufficient opportunities for reflection exist (ibid.). Individuals with strong unfavorable group attitudes, with clear and internalized antagonistic strategies, or in cognitively demanding situations with a low degree of external control might thus be more likely perform discriminatory actions on grounds of lower language proficiency, either on the labor market or within schools. Moreover, it could be argued that, as linguistic cues are a very salient signal of ethnic origin, they might generally foster intergroup definitions of situations and the associated sets of action.

This dissertation has a slightly subsequent focus, however. It is primarily interested in examining whether language-based discriminatory behaviors are chosen to such an extent that they affect labor market or educational outcomes beyond the resource-related influences of L2 proficiency, and what the main motives behind such types of discrimination are. Investigating the processes of the respective decision-making more closely could be an avenue for future research.

Endnotes to chapter

2 Becker (1971) touches only briefly upon the origin of group-specific tastes. It is suggested that a taste might incorporate both, elements of prejudice and ignorance, yet prejudice is seen as the more pervasive driver (Becker, 1917, 17f).

3 Certain authors additionally distinguish between institutional-level and structural-level discrimination (e.g. Pincus, 1996). What they term structural discrimination, in essence, captures indirect forms of institutional discrimination (ibid, 186). In line with other works (e.g. Feagin and Feagin, 1986) I thus distinguish between different forms of institutional discrimination and do not refer to 'structural' types.

4 Here, the condition I am primarily interested in relates to incomplete information. Only with lower L2 proficiency, students exhibit ambiguities that foster insecurities about their 'true' abilities. The main arguments and hypotheses thus relate to this condition (paper 3). However, I also assess how evaluations change when language skills increase and information deficits reduce (ibid.).

Empirical approach - a note on methodology

As with theoretical approaches, there are many different methods for measuring discrimination (e.g. Blank et al., 2004, 72).

Large scale survey data, for instance, enables inferences on the possible extent to which discrimination may shape certain outcomes. For that, outcomes of interest are regressed on an indicator of group-membership and other factors that may shape the respective outcome, but are unrelated to discrimination. The main idea here is that if all relevant 'non-discriminatory' factors are taken into account, adverse effects of group membership should be the result of discrimination (Blank et al., 2004, 118ff; Holzer and Ludwig, 2003, 1151ff). The occurrence of discrimination is thus indirectly inferred (ibid.). The main advantage of such approaches is that they allow for inferences on whether and to what extent discrimination could affect outcomes beyond other relevant factors, i.e. on the *relative importance of discrimination*. Moreover, inferences are seen as *comparatively generalizable*, as they are generally based on large, representative samples (ibid.). The main drawback of such methods is that discrimination can never actually be 'proven', as it is only indirectly inferred. This is particularly precarious when not all relevant 'non-discriminatory' factors are taken into account, as differences in such 'omitted variables' can also underlie the adverse effects of group membership (ibid.). Moreover, it has been noted that explanatory factors might not always be unrelated to discrimination. One the one hand, certain groups may receive discounts to some of these factors, due to certain forms of discrimination. Standard statistical models cannot pick up on such processes, as estimated functions, in general, do not allow the respective slope coefficients to differ between groups[5] (Blank et al., 2004, 122). On the other hand, these factors may include effects of prior discrimination[6] (Blank et al., 2004, 118ff; Holzer and Ludwig, 2003, 1151ff). Parental social background or actors' work experience, for instance, may be the result of past labor market discrimination. Moreover, as analyzed sets in general do not contain data on the potential 'discriminators', inferences on the reasons behind discrimination are limited.

By contrast, when 'well-designed and executed', experimental studies allow for more direct inferences on whether a certain group membership actually causes discrimination and on the main motives behind it (Blank et al., 2004, 93). Laboratory experiments, in particular, are seen as the *'royal road' to causality*, as they provide a high degree of control over treatments and settings and – like other types of experiments – should face limited omitted variable bias, due to the 'magic' of randomization (Blank et al., 2004, 90ff). Such laboratory experiments have been predominately used to assess subtle, subconscious, or nonverbal forms of discrimination (ibid.), yet they are also applied to study more manifest forms of differential treatment, such as applicant selections (Blommaert et al., 2013) or student evaluations (e.g. Crowl and MacGinitie, 1974). When combined with additional tests, measures, or manipulations, laboratory experiments also allow for comprehensive *assessments of the mechanisms behind any observed discrimination*. When 'well-designed', drawbacks primarily relate to generalization. Settings, tasks, and samples of laboratory experiments often differ from the real-word conditions of interest (Jackson and Cox, 2014, 38). Moreover, it is not possible to assess the relative importance of discrimination, as other explanatory factors are always fixed at certain levels. Laboratory experiments are thus often seen as less appropriate for addressing the role of discrimination for outcomes in the aggregate (Blank et al., 2004, 72).

In contrast, field experimental findings are more directly generalizable, as they are obtained under 'natural' conditions and, when 'well-executed', can indicate the extent of discrimination in

the examined domain, e.g. housing or labor markets (Blank et al., 2004, 72). However, this comes at the cost of losing control over settings and the conditions under which treatments are applied (Jackson and Cox, 2014, 39). Moreover, here too, the relative importance of discrimination is not straightforward to assess, as other explanatory factors are also set at certain levels.

Additionally, researchers may draw upon qualitative methods such as in-depth interviews. Such approaches generate 'richly detailed' information on individuals' discriminatory experiences and can uncover novel or highly specific forms of discrimination (Blank et al., 2004, 175). Moreover, such methods may be the only way to identify certain forms of institutional discrimination, such as specific regulations with neutral intent that adversely affect certain groups. Here, too, the main drawback relates to generalizability. In-depth interviews, for instance, are generally based on "small and unrepresentative samples that are often biased, because participants are of higher status, more articulate, and more politically aware than most of the subject population" (Blank et al., 2004, 175). Moreover, such approaches are based on subjective reports of experienced discrimination and allow only limited considerations of other factors that may be of relevance. The relative importance of discrimination for outcomes in the aggregate can thus not be assessed.

Methods within this dissertation

I draw upon the first two approaches and employ large-scale survey data as well as an experimental design. Large-scale survey data allows assessments of the relative importance of language-based discrimination and enables first, tentative inferences on the reasons behind such processes[7]. The implemented field experimental design additionally enables more direct examinations of whether language cues truly elicit differential treatment, and provides supplementary insights into the corresponding motives.

When analyzing large-scale survey data, I use information from the Millennium Cohort Study (MCS) to examine the cited issues within the educational system, and information from the National Educational Panel Study (NEPS) to address these matters on the labor market. With both of these sets, I examine whether and to what extent L2 language proficiency affects immigrants' outcomes differently when groups are associated with negative attitudes and beliefs. I examine this in different specifications and with regard to different arguments (Paper 1 & Paper 3)

The implemented field experiment is based on a telephonic audit study, in which trained individuals (testers) called employers and responded to advertised positions. Language accents and ethnicity of name were randomized (2 x 2 design); employer reactions, job-, and firm characteristics were logged; context data was merged to the main data set. It was examined whether foreign-accented applicants are turned down more often than comparable German jobseekers and accent-free applicants of the same origin. Moreover, it was assessed whether the effect of language accent is conditional on job, firm, and context characteristics, to gain insights into the potential reasons behind language-based forms of differential treatment (Paper 2)

Endnotes to chapter

5 To consider discounts to certain explanatory variables, researchers for instance add interactions to their specifications, estimate separate regressions, and employ decompositions (e.g. Blinder, 1973; Oxaca, 1973), or apply alternative techniques such as propensity score matching (Holzer and Ludwig, 2003, 1151ff)

6 In general, however, forms of prior discrimination are not disregarded altogether, but seen as a different phenomenon that precedes the outcome of interest.

7 When examining the role of language-based discrimination for immigrants' educational outcomes, I also address institutional-level forms of discrimination within a large-scale survey framework. However, I can only apply very indirect tests of the relevant arguments and cannot fully disentangle their role from individual-level forms of discrimination. It could thus be fruitful to additionally assess the institutional-level arguments with in more qualitative designs.

Paper 1

Ethnic differences in labour market outcomes: The role of language-based discrimination

*Submitted to: European Sociological Review**

Full citation:
Schmaus, M. (2020). Ethnic Differences in Labour Market Outcomes—The Role of Language-Based Discrimination. European Sociological Review,36 (1), 82-103 (https://doi.org/10.1093/esr/jcz044)

Host country language proficiency has been shown to account for ethnic differences in labour market outcomes. Prior studies generally assume that language skills represent a form of human capital, affecting employees' productivity. However, language proficiency may also be associated with discrimination. Lower language proficiency may elicit distaste for certain ethnic groups, as it is a prominent reminder of the respective origin. When this reminder vanishes as language skills rise, group-specific distaste should also reduce. Employers may thus not only value language skills in terms of productivity, but also factor in less group-specific distaste when evaluating immigrant jobseekers with high language skills. Moreover, if employers lack information on competences that are hard to observe, high language proficiency may prevent the application of adverse ethnic stereotypes.

Using data from the National Educational Panel Study (NEPS), I examine whether language skills affect crucial indicators of labour market success differently for groups that vary with respect to the levels of distaste they face. I also investigate whether this effect is conditional on the amount of information available to employers. Findings indicate group-specific returns to higher language proficiency, irrespective of the available information, which may point to language-based forms of taste discrimination.

Introduction

Proficiency in a host country language (L2) is crucial for the economic assimilation of immigrants. Considerable research demonstrates that income, occupational attainment as well as employment opportunities increase with rising skills in the respective national language (e.g. Chiswick and Miller, 2010; Dustmann and van Soest, 2002; Lindemann and Kogan, 2013; Kalter, 2006; van Tubergen et. al, 2004).

In most of the previous studies it is assumed that (L2) language skills represent a form of human capital (Chiswick and Miller, 2001) that impact labour market outcomes via an increase in the employee's productivity. However, another mechanism could co-produce the general findings: The linguistic patterns associated with lower language proficiency can trigger discrimination (Esser, 2006; Kossoudji, 1988; Stolzenberg and Tienda, 1997). As these patterns vanish with improving language skills, related forms of discrimination should disappear as well. In this case, labour market outcomes would also increase with higher (L2) language skills. On the one hand, lower language proficiency may elicit distaste for certain ethnic groups (Becker, 1971), as it is a particularly prominent reminder of the respective origin that "emphasizes foreignness" (c.f. Kossoudji, 1988:211). With improving L2 language skills this reminder becomes less salient and the associated distaste should reduce as well. Employers may thus not only value language skills in terms of productivity, but also factor in less group-specific distaste when evaluating immigrant jobseekers with higher language skills. Moreover, if employers lack relevant information on skills and competences that are hard to observe, high language proficiency may prevent the application of adverse ethnic stereotypes (England, 1992). As a consequence, the effect generally interpreted as stemming from an (L2) language related productivity increase could co-capture the effect of language-based discrimination. If that were the case, previous studies could systematically overestimate the influence of (L2) language competence in terms of productivity.

In this contribution, I aim to disentangle the effect of (L2) language related productivity increments from the effect of language-based discrimination for immigrants in Germany. Exploiting employment biography data from the "National Educational Panel Study (NEPS)", I examine whether or not (L2) language skills affect crucial indicators of labour market success

differently for immigrant groups that vary with respect to the level of distaste they face (Becker, 1971). I also investigate whether or not this effect is conditional on the amount of information available to employers (England, 1992). Analyses indicate group-specific returns to higher language proficiency. As this is not conditional on other relevant information available to employers, this might indicate language-related taste discrimination (Becker, 1971).

Theoretical framework

To disentangle the effect of language related productivity from that of language-based discrimination, I test the assumptions implied by the two competing approaches. According to the "classical" approach (Chiswick and Miller, 1995; 2001), L2 proficiency should affect the labour market outcomes for all ethnic groups alike. The basic idea behind this assumption is that, as with human capital, a person is embodied with language skills, which are costly to gain but efficient in facilitating communication, for example with supervisors, suppliers or customers. This allows workers with better language skills to be more productive (Chiswick and Miller, 1995:248; Chiswick, 2008:4). The associated increase in productivity should (directly) translate into higher earnings and other more favourable labour market outcomes. This mechanism should generally work for all ethnic groups alike, as the same increase in L2 skills should bring about the same increase in productivity and thus translate into the same increase in returns (Figure I, a).

The methodological approach generally applied to test such arguments exemplifies this assumption. The relationship between language skills and labour market outcomes is usually examined by estimating a certain function in which the respective outcome is regressed on a linear combination of explanatory variables (Borjas, 1982; Chiswick, 2008; Chiswick and Miller, 2002; Kalter, 2006; Kogan and Lindemann, 2013). Among crucial characteristics typically associated with successful labour market integration, these sets include indicators of language proficiency and ethnic origin (ibid.). For language proficiency a difference in the slope coefficient by ethnic group, or an alternative procedure testing for different returns, is generally not implemented.

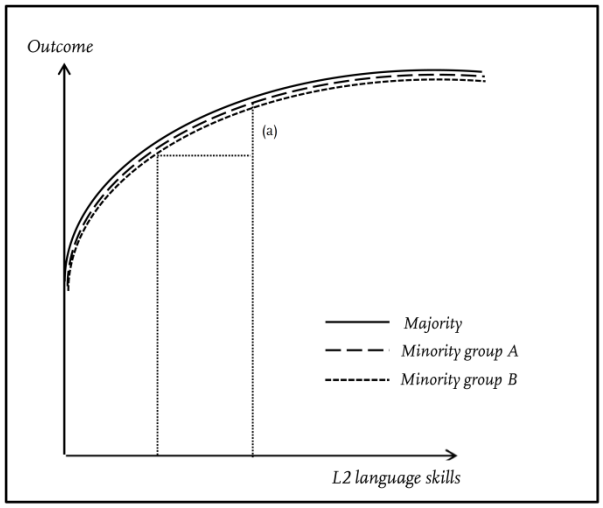


Figure I: Returns to L2 language proficiency for different ethnic groups according to productivity arguments.

Building on these theoretical and methodological considerations, I assume *that if language proficiency primarily works through an increase in productivity, (L2) language skills should (c.p.) affect the labour market prospects for all ethnic groups alike (hypothesis 1)*.

However, according to (economic) discrimination theories one would expect group-specific returns to rising L2 proficiency. First, (L2) language proficiency may be linked with taste-based discrimination (Becker ,1971). The main assumption of such models is that actors, such as employers, prefer members of some groups over members of others. When making hiring or payment decisions, they may factor this (dis)taste into their decision-making, which can then lead to lower labour market outcomes for members of the non-preferred group. (L2) language proficiency may moderate the strength of this distaste: lower language proficiency may elicit a (stronger) distaste for certain ethnic groups, as it highlights the respective origin and "emphasizes foreignness" (c.f. Kossoudji, 1988:211). Native-like language command on the other hand can reduce that foreignness and override the effect of other ethnic cues such as name or looks in

impression formation (Rakić et al., 2011; Rödin and Özcan, 2011). This has been shown to convert ethnic preferences in experimental settings, in that a previously rejected foreign looking person was chosen as a friend when she spoke in a 'native-like' manner (Kinzler et. al, 2009).

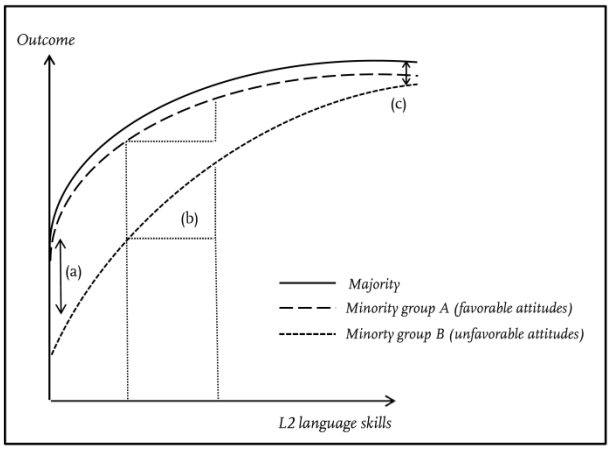


Figure II: Returns to L2 language proficiency for different ethnic groups according to discrimination arguments (distaste).

Following this logic, ethnic groups that are subjected to prejudice and aversions should be doubly disadvantaged with lower (L2) language skills: Here, employers may not only consider limited language proficiency in terms of productivity, but also factor in (greater) group-specific distaste when evaluating the respective jobseekers. Consequently, these groups should receive greater discounts to lower (L2) proficiency. When language skills increase however, the respective ethnic origin should become less salient and ethnic distaste should reduce which allows non-preferred groups to catch up with other immigrants and natives. In more technical terms, the slope of the language coefficient should be steeper for non-preferred ethnicities than for members of other immigrant groups. Figure II above illustrates these assumptions. Disliked groups receive lower returns to lower (L2) language skills as here, employers factor in any distaste they may harbour against this ethnic group; this results in a discount from the potentially achievable return for the respective group of origin (a). However, when language skills increase the respective group membership becomes less salient and distaste is reduced; compared with non-proficient speakers of the respective origin, this generates a greater increase in returns for unpopular ethnicities (b, c) (for related considerations see Esser, 2006:431ff.). Consequently, if *employers factored group-specific distaste into their decision-making, ethnic groups that are subjected to strong(er) distaste and aversions should profit more from rising (L2) language skills than other ethnic groups (hypothesis 2).*

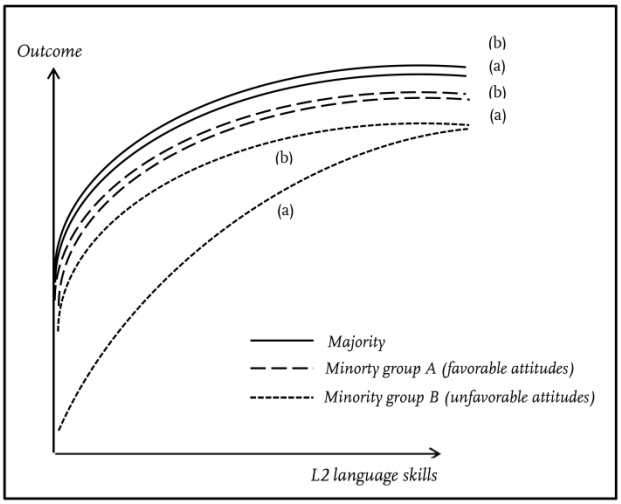


Figure III: Returns to L2 language proficiency for different ethnic groups according to discrimination arguments (adverse beliefs, resp. error discrimination).

Additionally, lower (L2) language skills may also trigger statistical or error discrimination (England, 1992). Such models are based on the idea that employers substitute information they lack about an individual worker with information they hold about his group (Phelps, 1972). When evaluating new workers, employers in general lack information on applicants' prospective productivity and potential value to their firm. Employers may be particularly uncertain about prospective productivity when immigrants have limited (L2) language proficiency (c.f Kossoudji, 1988:209). On the one hand limited language proficiency can hamper the assessment of other skills and competencies, thus exacerbating unambiguous evaluations of the applicant's 'true' productivity. Lower L2 skills may, for instance, obscure the observability of specific or soft skills, leaving employers guessing whether applicants 'lack' certain abilities, or if the missing language skills simply 'block' their identification. On the other hand, employers may expect a greater variance in unobservable abilities for less proficient speakers

(ibid.), as in some circumstances both less and highly productive immigrants can exhibit lower L2 proficiency, for instance shortly after their arrival. In addition to that, employers may be unsure about transaction costs, as they might feel unable to assess the 'true' extent of training costs and upheaval within teams that can be associated with hiring a less proficient speaker (Esser, 2006).

Thus, if employers thus lack information about a (prospective) worker's 'true' value, they may try to reduce this uncertainty by drawing upon their assumption about the (average) productivity of the worker's ethnic group. For some origins, assumptions about productivity may be systematically downwardly-biased (error discrimination, England, 1992). In this case, employers may transfer their own negative beliefs into their evaluations and erroneously expect the respective workers to cluster on the lower end of the ability distribution. Consequently, groups confronted with strong negative beliefs should receive greater discounts to lower language proficiency. However, when language skills increase these discounts should reduce, as higher L2 skills send out a positive signal about ability, drive and initiative which should deter the application of negative beliefs. Additionally, higher L2 proficiency may reduce insecurity per se, as it allows for more clear-cut skill evaluations (c.f Kossoudji, 1988:211).

As this type of discrimination should occur only if employers lack relevant information (say, on specific or soft skills), this implies that *for ethnic groups associated with negative beliefs the influence of (L2) language proficiency on labour market outcomes is more (negatively) conditional on the amount of information available to employers (hypothesis 3*). Expressed less technically, if necessary skill-specific information is unavailable to employers, workers from ethnic groups associated with negative beliefs receive greater discounts to lower language proficiency, as employers then draw upon these group-specific assumptions. If language skills improve, these discounts reduce, because higher L2 skills prevent the application of the negative beliefs (a); however, if the necessary information is available, then employers have no need to draw upon their beliefs to begin with. Consequently, in this case, there should be no ethnic differences in the effect of language proficiency (b). Figure III above sums up these considerations.

The German context

As in most (western) societies, the German labour market is characterised by ethnic inequalities. Many immigrant groups fare worse than their native counterparts in terms of employment, occupational levels, job prospects and earnings (Algan et al., 2010; Kalter, 2006, 2008; Kogan, 2011a, 2011b; Luthra, 2013). German language proficiency (L2) has been shown to shape immigrants' labour market outcomes (Dustmann and vanSoest, 2001; Kalter, 2006). For some ethnic groups, German language skills seem to be more relevant for their success on the labour market than for others. Turkish immigrants, for instance, appear to have a higher need for German language skills than immigrant groups from other origins (Kalter, 2006).

In the German context, immigrant groups vary substantially with regard to the distaste they face, as well as with the beliefs that are associated with them. Distaste is particularly pronounced for the Turkish group (e.g. Blohm and Wasmer, 2008; Degner and Wentura, 2010; Steinbach, 2004), as well as asylum seekers (Steinbach, 2004), in recent years particularly from Arab or African countries (Cymara, 2017). Similarly, beliefs about these groups have been shown to be mostly negative (Karaham and Knoblich, 2010; Kotzur et al., 2017) and include assumptions about violence (ibid.) as well as lower (professional) competence (Asbrock, 2010).
To a lesser extent, distaste and adverse beliefs also seem to be associated with immigrants from the former communist states of Eastern Europe, particularly Bulgaria and Romania (Wagner, 2018), but also extending further to other neighbouring states. Many of these origins face social

distance, especially with regard to close personal interaction and distrust (Keita and Valette, 2016:12; Kucharczyk et. al, 2013:11; Steinbach, 2004:145), sometimes comparable to those faced, for instance, by the Turkish group (ibid). However, here ethnic boundaries are generally assumed to be less bright (Diehl et. al, 2016: 243; Schulz and Leszczensky 2016:172) and most groups of these origins report lower levels of perceived discrimination (André and Dronkers, 2017:114f; Ganter, 2003). Similarly, while for many of these groups beliefs contain distinct adverse assumptions about criminality (Gatzke 2010; Szczęk, 2006; Kucharczyk et. al, 2013:8; Wagner, 2018), corruption and untrustworthiness (Gatzke 2010:66; Keita and Valette, 2016:12; Kucharczyk et. al, 2013:8), they also include multiple positive attributes (Kucharczyk et. al, 2013:8). By and large for these origins, attitudes thus seem to be moderately unfavourable, yet comprehensive research is scarce.
By contrast, immigrants from other European countries (EU 15), such as Italy, are comparatively well-liked (Steinbach 2004) and beliefs seem to be predominantly positive, particularly with regard to social behaviour, parenting style or culture (Froehlich et al., 2015:77). While beliefs about competence sometimes appear slightly lower than those applying to the native population, such assumptions do not seem to be very salient when people envision immigrants with these origins (Froehlich et al., 2016:78).

It may thus be fruitful to test the main theoretical assumptions within the German country context: L2 language proficiency is shaping immigrants' labour market outcomes, some ethnic groups seem to have a higher need for L2 language skills than others and these ethnic groups are distinct with regard to the distaste they face and their associated beliefs. Yet to date, processes of language-based discrimination remain largely unexamined.

Data & methods

To test the above mentioned assumptions empirically, I use data from the "National Educational Panel Study (NEPS)". The NEPS is a nationwide study conducted in Germany that is based on a multi-cohort sequence design and aims to capture educational trajectories across the entire life span (Blossfeld et al., 2011). I focus on the adult population sample (SC6)*, as the NEPS offers detailed information on employment history as well as other relevant (background) characteristics for this cohort (Skopek, 2013). Moreover, it provides test-based information on (L2) language proficiency, which is crucial to test the proposed hypotheses adequately (for potential bias associated with self-reported language proficiency see Edele et.al., 2015).

Thus far, information is available from two different L2 language assessments. The first language tests took place either in sweep three (2010/2011) or sweep five (2012/2013), depending on when participants entered the panel (Fuß et al., 2016). In these initial assessments, reading competence and reading speed were captured. The second assessments took place in sweep seven (2014/2015) and captured listening comprehension. To date, no repeated measurements of L2 language skills are available. Consequently, longitudinal analyses cannot (yet) be undertaken.

I focus on the first assessment of L2 language skills and include information only up to sweep five in the analyses. There are two main reasons for this approach: *Firstly*, as sweeps progress, participants become increasingly positively selected, particularly with regard to their language proficiency. About 37% of participants who scored within the lowest decile in their first language test (sweep 3 or sweep 5) dropped out before the additional language assessment in sweep seven (supplementary materials, Table III). For ethnic groups that face strong distaste and negative beliefs, this percentage is even higher (approx. 51%), which might limit comprehensive examination of discounts to lower language proficiency when using information of the later

language competence tests (sweep 7). *Secondly,* while the construct measured in sweep 7 (listening comprehension) represents a widely-used indicator, the constructs captured during first assessment also take cognitive requirements, test functions and task formats into account (Gehrer et al., 2013), which might allow for a more comprehensive test of L2 abilities.

Moreover, I apply one sample restriction: self-employed participants and respondents working for family members were excluded from the analyses, as here (language-based) employer discrimination should not be relevant (n=295). As I consider different indicators of labour market success and employ an 'imputation then deletion' strategy (von Hippel, 2007), final sample sizes vary between 10,708 and 6958[1] (Appendix, Table II).

Dependent Variables

All hypotheses propose a (differential) relationship between (L2) language proficiency and the labour market success of different immigrant groups. To capture the labour market success, I employ five indicators that are frequently used in integration research: employment, occupational status, occupational type, occupational prestige and earnings. I consider information on these outcomes after all participants faced their first L2 language assessment (sweep 5)[2].

Employment is captured by a dummy variable, indicating if participants were employed (1) or unemployed (0) at the time of the interview. For employed respondents, I then consider *occupational status*, as indicated by the "International Socio-Economic Index of Occupational Status" (ISEI-08; Ganzeboom et.al., 1992); *Occupational type*, distinguishing between blue (0) and white collar jobs (1), based on a condensed version of the EGP class category scheme[3] (Erikson et.al., 1979; Kalter, 2006:150); *Occupational prestige*, as indicated by the magnitude prestige scale (MPS) to capture idiosyncratic prestige hierarchies that are specific to the German context (Bernhard, 2005), as well as *earnings,* which are indicated by the gross (pretax) monthly earned income[4].

Independent variables

To examine whether or not different ethnic groups receive disparate returns to L2 language proficiency, I categorise respondents' countries of origin with regard to the distaste and beliefs associated with them.

As with many other (western) countries, non-western immigrants face notable distaste in Germany, particularly those originating from Turkey, or more recently from Arab or African countries (e.g. Steinbach, 2004; Cymara, 2017). Similarly, beliefs about these groups have been shown to be mostly negative (Karaham and Knoblich, 2010; Kotzur et al., 2017) and include assumptions about violence (ibid.) as well as lower (professional) competence (Asbrock, 2010). To a lesser extent this should also apply to immigrants from former communist states of Eastern Europe (Gatzke 2010; Szczęk, 2006; Keita and Valette, 2016; Kucharczyk et. al, 2013; Wagner, 2018). By contrast, for immigrants from other European countries (EU 15), such as Italy, distaste and beliefs seem to be considerably more favourable (Froehlich et al., 2015:77; Steinbach, 2004).

I thus distinguish between groups facing *strongly unfavourable attitudes* (1), *moderately unfavourable attitudes* (2) and those facing *favourable attitudes* (3). I use *natives* as a point of reference (0), as here attitudes should in general be the most positive (Steinbach, 2004:145). Groups facing *strongly unfavourable attitudes* include respondents who either had two grandparents, one or two parent(s) or were themselves born in Turkey or in countries in the near East or Africa, such as Algeria, Morocco, Iran, Iraq, Lebanon, Nigeria, Ghana, or Kenya. Groups facing *moderately unfavourable attitudes* include respondents who either had two grandparents, one or two parent(s) or were themselves born in former communist states of Eastern Europe, such as Russia, Bulgaria,

Romania, former Yugoslavia, Poland or the Czech republic. Groups facing *favourable attitudes* include respondents whose (two) grandparents, parent(s) or they themselves were born in another EU 15 country, such as Italy, Spain, France or Greece. Respondents whose parents, more than two grandparents and they themselves were born in Germany were categorised as *German natives*. On the (rare) occasion, that respondents belong to an ethnic group with unclear associated attitudes, such as immigrants from South America, they are categorised as *others* (4).
The NEPS provides a comparatively detailed measure of *respondents' generational status* (1st, 1.5th, 2nd, 2.25th, 2.5th, 2.75th, 3rd, 3.25th, 3.5th); however, as this fine-grained distinction stretches the case numbers a little thin, I differentiate only between the first (born abroad) and the second or later generation/s (born in Germany). Information on migration background is always captured when participants first enter the panel (sweep 1, 2, or 4)

(L2) language proficiency is indicated by the score of a reading competence test[5], focusing on the appropriate handling of written texts in different everyday situations (Gehrer et al., 2012:2). It is based on a functional understanding of reading competence, similar to the Anglo-Saxon literacy concept (ibid.). As this kind of competence testing requires integrative processing of multiple bottom-up and top-down text handling abilities, it is assumed to represent a comparatively broad indicator of (L2) language skills (Edele et al., 2015:106). For adults, the implemented test consists of 5 different texts and 32 items that refer to each of these texts (Koller et al., 2014:4). Every text serves a different purpose (text function), which can either be informative, argumentative, literary, instructional or advertising something (Gehrer et al., 2012). Moreover, each text requires one of three cognitive tasks (cognitive requirements): finding information in the text, drawing text-related conclusions or reflecting on the reading material (ibid.). Additionally, item format varied between multiple-choice, evaluating statements and assigning titles to text sections, broadening the assessment scheme further. For each respondent, weighted maximum likelihood scores (WLEs) are provided, indicating their most likely competence scores. I rescale this indicator so that 0 represents the lowest observable value, in other words the lowest observable competence. Higher scores are indicative of better competences. Tests took place either in sweep 3 or sweep 5, depending on when participants entered the panel (Fuß et al., 2016).

Hypothesis three proposes that for non-western immigrants the influence of (L2) language proficiency on labour market outcomes is (more negatively) conditional on the *amount of information available to employers*. Following Altonji and Pierrett (2001), I assume that *work experience* provides pivotal information on productivity for employers. It should not only be indicative of prior achievements per se, but should also be associated with (more) reference letters that are available to potential employers. Such letters should provide additional information and have been shown to reduce statistical discrimination for negatively viewed groups in the German context (Kaas/Manger 2010). I thus use work experience, meaning the number of years participants (successfully) held positions within the German labour market (sweep 5), as an indicator for the amount of information that is available to employers.

Controls

Alongside those primary independent variables, I include other relevant characteristics typically associated with successful labour market integration in a single country setting; that is, I include central aspects of the respondent's human, cultural, social, economic, symbolic and personal capital. To indicate *human capital*, I *use participants' highest level of education* (sweep 5, CASMIN). To indicate *cultural capital* I consider the *highest level of education of the respondent's parents* (panel entry, CASMIN). To capture *social capital*, I draw upon information from the NEPS position generator (sweep2, Schulz et al., 2017). Within this instrument respondents are asked whether

they personally know people with different occupations. These occupations range from low-status positions, such as car-mechanics to high-status professions, such as engineers or doctors (ibid.). Following Schulz et al, 2017, I assign status values to each of these occupations using the International Socio-Economic Index of Occupational Status (ISEI 08). If respondents do not know anyone with the respective profession, they are assigned a value of 0. I then calculate the mean occupational status across all occupations for each respondent. This procedure has the advantage of producing final scores that are not only indicative of *respondents' network status composition,* as professions with a higher occupational status increases the respective indicator, but *also of their network extent,* as not knowing somebody with a certain profession will always decrease respondents' scores due to the value 0 (ibid.). Additionally, I included two variables indicating how likely it is that somebody from the respondent's *network will pass on information regarding job vacancies* (sweep 2, 4 point scale) or *would put in effort to help the respondent obtain a new job* (ibid.) *Economic capital* is indicated by the highest *occupational status of the respondent's parents* (panel entry, ISEI-08) and the *question of how likely the respondent could borrow an amount larger than 250 Euro* from somebody in his social environment (sweep 2, 4 point scale). Holding a *German citizenship* is seen as an aspect of *symbolic capital* (panel entry, no vs. yes). To indicate respondents' *personal capital,* I included a measure of *personal drive* that captures the subjective importance of professional success (sweep2, 4 point scale), as well as a measure of mental health (sweep 4, SF 12 mental component score). Additionally, in all models *respondents' age, gender* (sweep5), as well as the *time of language testing* (sweep3 vs. sweep5) was considered. When examining respondents' earnings, the level of the respective job (EGP), the numbers of hours worked as well as the job sector (public vs. private) were also included.

Method

To investigate whether or not groups associated with distaste and adverse beliefs profit more strongly from their (L2) language proficiency than other ethnic groups and if this relation is conditional on the amount of information available to employers, I take a gradual approach: *Firstly*, I apply (logistic and linear) regression models to replicate the findings of prior research, demonstrating that alongside other crucial characteristics typically associated with successful labour market integration in a single country setting, (L2) language proficiency accounts for ethnic differences in labour market outcomes (Model 1, Model 2, Table A-D). *Secondly*, I then estimate two-way interactions, to examine whether or not the (L2) language skills affect crucial indicators of labour market success differently for groups that face different distaste and beliefs (Model 3, Table A-D). *Lastly,* I investigate whether or not this effect is conditional on the amount of information available to employers (England 1992), by estimating a three-way interaction (Model 4, Table A-D). To account for missing values, I used multiple imputations and estimated 25 datasets with complete information (e.g. Allison 2001). Regression analyses were performed on each dataset and afterwards combined following Rubin's (1978) approach. Furthermore, I followed an 'imputation then deletion' strategy (MID) (van Hippel 2007). Here all cases are used for imputation, but thereafter cases with imputed Ys values are deleted to increase precision[6]. All models were estimated both with and without design weights to ensure results remain comparable.

Results

Table I indicates results from the regression analyses for each of the five indicators of labour market success.

In line with prior research findings, immigrants fare worse than the native population on the examined indicators of labour market success, particularly when originating from non-European countries (M1, Table I). Resource endowment, such as respondents' human, cultural, social, economic and personal capital, by and large accounts for these disadvantages, yet some ethnic gaps remain unexplained (M2, Table I). For groups associated with unfavourable attitudes the size of these gaps seems to vary with L2 language proficiency (M3, Table I).

Groups associated with distaste and adverse beliefs seem to receive lower returns to lower L2 language proficiency, so that here disadvantages are particularly pronounced. Conditional main effects indicate that the respective groups are approximately 20 % less likely to be employed than members of the native population when German skills are poor (M3, Table I). Similarly, groups associated with unfavourable attitudes are about 36 % less likely to obtain white collar positions (M3, Table I) or have about 17 points lower job status (ISEI) scores than their native counterparts, when German skills are low (M3, Table I). However, when language skills improve, these disadvantages reduce and the respective groups can catch up with the native population (Figure I-IV). With high German language skills, the likelihood of being employed and obtaining a 'good' position does not seem to differ considerably between groups associated with unfavourable attitudes and the majority population (Figure I-IV)[7]. This relationship seems not to be conditional on the amount of information available to employers (M4, Table I). Results are comparable across all indicators of labour market success; however, in the case of earnings, results are not significant and patterns are less unequivocal (M3, Table I).

In sum, results seem to align with the assumptions of language-moderated employer tastes (hypothesis 2). For groups associated with unfavourable attitudes, the linguistic patterns associated with lower language skills might elicit (stronger) distaste, which employers may factor into their decision-making, thus generating lower returns. However, when language skills increase, group membership becomes less salient and distaste is reduced; compared with non-proficient speakers of the respective origin, this generates a greater increase in returns for unpopular groups. The amount of information available to the employer does not seem to matter, which might limit the applicability of arguments concerning adverse beliefs and error discrimination (hypothesis 3)

[Table 1]

[Figures I - IV]

Robustness of these findings

The most profound limitation of this study may be the use of cross-sectional data, as this allows for effects of *reverse causality*. Groups that are faced with unfavourable attitudes may be more segregated and have more homogeneous networks in terms of L2 language proficiency and use; thus for this group, exposure to German language proficiency in the work environment may be more beneficial in bringing about improvements to their language skills than for other ethnic groups, who are more frequently exposed to the German language in other contexts as well.

To examine such arguments, I re-conducted all analyses, while including indicators on respondents' German language exposure, reflecting their number of German friends and

acquaintances[8]. The main idea here is that potential effects of reverse causality might then be captured by the respective network variables and should not affect the theoretically relevant interaction effects as gravely. Results remained comparable (supplementary materials, Table IV). Moreover, I estimated group-specific models, to examine whether or not German language exposure in networks (differentially) mediates the effect of German language proficiency for different ethnic groups[9]. In all models, the respective exposure did not reduce the effect of language proficiency to a substantial degree. (supplementary materials, Table V).

In addition to that the presumed role of reverse causality might primarily pertain to the models that examine (un)employment, as here the respective outcome should be (more) congruent with German language exposure. Respondents from groups facing unfavourable attitudes should have greater exposure to the German language when they hold positions in the German labour market compared with those who do not and might thus experience greater benefits than other groups. However, for other labour market outcomes, such as job quality (prestige, status), these congruencies should be less pronounced. (Customer) service or sales positions for instance score comparatively low on the respective indicators, but generally provide considerable exposure to L2 language. Conversely, IT professions score comparatively highly on these scales, but do not necessarily guarantee vast exposure to German language. However, groups that are faced with unfavourable attitudes also profit differently when examining these outcomes, which might be another hint that interactions are not (solely) driven by reverse causality. While I cannot rule out issues of reverse causality altogether, in sum all of these results might suggest that the main findings might not (solely) be due to such effects and there may be something else generating different returns to L2 language proficiency for different ethnic groups, which might be language-based discrimination.

A second criticism may relate to *omitted variables*. Firstly, groups that are faced with different attitudes may also be distinct with regard to certain resources that are not included in the main models, specifically proficiency in compensatory languages, such as English. Particularly in larger firms *English proficiency* may make up for a lack of German language ability and groups facing stronger distaste and adverse beliefs may be less likely to have acquired the respective skill (Kirgöz 2009). Consequently German language skills may be of greater relevance for these groups. Furthermore, groups faced with unfavourable attitudes, may be more inclined to work in ethnic enclave economies, as means to shield against (anticipated) discrimination. With lower levels of L2 proficiency, they may thus more frequently rely on their first language, resp. *minority language skills*, as this grants access to the respective ethnic economy (Pendakur and Pendakur, 2002). As mainstream labour markets may provide more favourable outcomes, particularly when ethnic enclaves are small (ibid.), L2 proficient members of 'unpopular' groups that have entered the mainstream market may also reap greater returns.

Secondly, groups that differ with regard to the distaste and beliefs that they face, may also select *jobs with different L2 language demands*. On the one hand, groups facing stronger distaste and beliefs more often work in unskilled occupations (e.g. Kogan, 2004), amongst which positions in the sales or service industry, in hotels or restaurants, say, (e.g. Governatori et al., 2008: 73), where (at least basic) skills in the German language may be advantageous. On the other hand, when reaching high-skill positions, these groups may target jobs associated with particularly high income or prestige, to protect their children from perceived discrimination, for instance by sustaining longer stays in the education system or to permit greater leverage with teachers. Many of these may be managerial positions, which seem to require extensive use of the respective national language (Gonzales, 2005:789). In turn, this might again create a greater need for German language skills with the respective groups. While such controls are not available for the

whole sample, for respondents who participated in the first version of initial language assessments (sweep 3), relevant information is at hand. Hence, I re-conducted all analyses for this subsample[10], while including information on whether or not respondents use a compensatory language, meaning English, at work to communicate with their co-workers, or if they use a minority language to interact with their colleagues, which might indicate access to ethnic enclaves. In addition to that I also considered the extent to which they stated to use language (for reading or writing) on the job. Results remained comparable (supplementary materials, Table VI).

Summary & Discussion

Prior research has regularly demonstrated that proficiency in the host country language (L2) is crucial for the labour market success of immigrants. In most of these studies it is assumed that (L2) language skills represent a form of human capital and impact the labour market outcomes via an increase in productivity. However, the effect generally interpreted as an L2 language related productivity increase could co-capture the effect of language-based discrimination. In this contribution, my aim has been to disentangle the effect of (L2) language related productivity increments from the effect of language-based discrimination for immigrants in Germany. Accordingly, I have investigated whether or not (L2) language skills affect crucial indicators of labour market success differently for groups that vary with respect to the distaste they face (Becker, 1971). I have also examined whether or not this effect is conditional on the amount of information available to employers (England, 1992). Analyses indicate group-specific returns to higher language proficiency, irrespective of the amount of information available to employers. This could suggest that, in addition to affecting employees' productivity, L2 language proficiency may also be associated with taste-based discrimination (Becker, 1971).

In the overall trend, groups facing distaste and adverse beliefs seemed to profit more strongly from their L2 language skills on all examined indicators of labour market success. However, for the generation of earnings, this relationship was insignificant. This could be due to a large number of (potentially selective) missing values on the earnings variable, or the assumed mechanisms may apply more strongly to hiring, than to payment decisions. Employers may initially include their preferences in their hiring decisions, but once members of certain groups have passed that hurdle, they might not face additional discounts. This could be in line with results of prior studies suggesting that discriminatory tendencies might reduce over the course of the hiring process (Cedie and Foronie, 2008:107) and once positions have been obtained, earnings should not be affected (Constant and Massey, 2005). However, I could not directly test such assumptions.

Moreover, the amount of information available to employers did not seem to matter. I used respondents' work experience to indicate the amount of information, as it might not only be indicative of prior achievements per se, but should also be associated with (more) reference letters, which provide additional information. However, this specification does not represent a direct indicator of the available information. A study that varied the amount of information more directly within an experimental setting did find indications of statistical discrimination in the German labour market (Kaas and Manger, 2010). Yet all fictitious applicants within this design grew up in Germany and had good command of the German language, which might cue different types of discrimination than those examined here. It may thus be an avenue for future research to examine the proposed mechanisms more directly, for instance within tailored experimental settings. Additionally it may be fruitful to further validate results – when possible – within more longitudinal designs.

Despite its limitations, this study might contribute to our understanding of the mechanisms that produce ethnic differences in labour market outcomes and call for a sensitive interpretation of language effects on labour market integration.

1 Final samples are slightly positively selected on social origin as well as L2 language skills. While this might limit generalisability as a whole, tests of hypotheses should remain unaffected and might be more conservative: If language-based discrimination already affects the labour market outcomes of better endowed immigrants, the same mechanisms should affect less advantaged foreigners even more severely.

2 Information on labour market outcomes was taken from sweep 5, irrespective of when participants faced their first L2 language assessments (sweep3 or sweep 5). The main reason behind this is to ensure comparability and to rule out the effects of economy fluctuations. To account for potential differences related to divergent test dates, timing of the first L2 language assessment is controlled for in all analyses.

3 EGP classes I, II, and III were recoded into white collar jobs (1) and EGP classes V and IV into blue collar jobs (0). General trends remained comparable when trying alternative categorisations.

4 If respondents reported to work in more than one job at the time of interview, for each outcome, the job with the highest respective level was considered (e.g. highest ISEI), as this should allow for the strictest test of arguments on (language-based) discrimination. If participants reported multiple jobs with identical levels, the one for which they claimed to work the most hours, or deemed their main activity was considered.

5 NEPS provides two measurements of (L2) language reading proficiency. The first reflects a reading competence test (Gehrer et al. 2012), based on the literacy concept (OECD 2009). The second one captures reading speed and draws upon the principles of the Salzburg reading screenings (Zimmermann et al., 2012).The reading competence test requires integrative processing of multiple bottom-up and top-down text handling abilities, while taking different text functions, cognitive requirements and item formats into account. Assessments of reading speed primarily capture basic reading processes such as decoding (ibid.) and do not consider additional aspects. Thus, I focus on reading competence scores as this represents the more comprehensive indicator.

6 To increase precision further, I did not impute migration background (n=63).

7 With high German language skills trends seem to reverse, meaning that groups associated with unfavourable attitudes – in tendency- get greater returns to their language skills than natives. This might be due to the empirical distribution of L2 proficiency for the respective groups. Observed values here do not go past "good" command (8 out of 10 scale points). As up to this point, "unpopular" groups profit more strongly from improving language skills, because they can regain initial (taste-based) discounts, the –in tendency-higher predicted probabilities for "very good" command (say a 10 on the scale) might simply be due to the linear modelling.

8 I include two indicators to capture respondents' German language exposure. I use the share of German, in regard to non-German friends (none, almost none, less than half, about half, more than half, almost all, all) (sweep 2), as well as the number of people in the respondent's professional network (NEPS position generator), that were stated to be of German origin (sweep 2)

9 I also tried to disassociate the model variables' temporal order and re-conducted all analyses with information on labour market outcomes from the sweep after all respondents had their language skills assessed (sweep 6). Results remained comparable, however, the majority of

respondents held the same job, or employment status in both sweeps (sweep 5, sweep 6) and for those that did not, changes may be path-dependent, thereby limiting the respective conclusions.

10 In this subsample, particularly for groups facing distaste and adverse beliefs, there are only few reading competence scores available (n=64), which do not span across the whole scale. As case numbers and distributions are slightly more favourable for the assessment of reading speed, I use this measure for the respective analyses. General results are comparable with those of the main models, which employ reading competence scores (main models: Table I, M1; additional models: supplementary materials, Table VI)

Data source

* This article uses data of Starting Cohort 6 (Adult sample) of the National Educational Panel Study (doi:10.5157/NEPS:SC6:10.0.1). From 2008 to 2013, NEPS data were collected as part of the Framework Program for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). Starting in 2014, NEPS has been carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg, in cooperation with a nationwide network.

References

Algan, Y., Dustmann, C., Glitz, A., & Manning, A. (2010). The economic situation of first and second-generation immigrants in France, Germany and the United Kingdom. *The Economic Journal*,20(542),4-30.

Altonji, JG. & Pierret, CR. (2001). Employer learning and statistical discrimination, *The Quarterly Journal of Economics*,116 (1),313-350.

André, S., & Dronkers, J. (2017). Perceived in-group discrimination by first and second generation immigrants from different countries of origin in 27 EU member-states. *International Sociology*,32(1),105-129.

Asbrock, F. (2010). Stereotypes of Social Groups in Germany in Terms of Warmth and Competence, *Social Psychology*,41,76–81.

Becker, G. (1971). *The economics of discrimination*. Chicago: UCP. 2nd ed.

Blohm, M., & Wasmer, M. (2008). Einstellungen und Kontakte zu Ausländern. Datenreport 2008: *Ein Sozialbericht für die Bundesrepublik*.

Cediey, E., & Foroni, F. (2008). Discrimination in access to employment on grounds of foreign origin in France,I*nternational Labour Office:*ILO.

Chiswick, B. & Miller, P. (2001). A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada,*Demography* ,38,391-409.

Chiswick, B. & Miller, P. (2001). Occupational language requirements and the value of English in the US labor market, *Journal of Population Economics*,23(1),353-372.

Constant, A., & Massey, DS. (2005). Labor market segmentation and the earnings of German guestworkers,*Population Research and Policy Review*,24(5),489-512.

Czymara, CS., & Schmidt-Catran, AW. (2017). Refugees Unwelcome? Changes in the Public Acceptance of Immigrants and Refugees in Germany in the Course of Europe's 'Immigration Crisis',*European Sociological Review*,33(6),735-751.

Degner, J., & Wentura, D. (2010). Automatic prejudice in childhood and early adolescence. *Journal of personality and social psychology*,98(3),356.

Diehl, C., Andorfer, VA., Khoudja, Y., & Krause, K. (2013). Not in my kitchen? Ethnic discrimination and discrimination intentions in shared housing among university students in Germany, *Journal of Ethnic and Migration Studies*,39(10),1679-1697.

Dustmann, C. & vanSoest, A. (2001). Language fluency and earnings: Estimation with misclassified language indicators, *Review of Economics and Statistics*,83(4),663-674.

Dustmann, C. & vanSoest, A. (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review* 55,(3),473-492.

Eckes, T. (2002). Paternalistic and Envious Gender Stereotypes: Testing Predictions from the Stereotype Content Model, *Sex Roles*,47,99–114.

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency, *Social Science Research*,52,99-123.

England, P. (1992). *Comparable worth: Theories and evidence*. NY:de Gruyter.

Esser, H. (2006). Migration, language and integration. WZB:AKI Research Review 4.

Froehlich, L., Martiny, SE., Deaux, K., & Mok, SY. (2015). It's their responsibility, not ours, *Social psychology*,47,74-86.

Fuß, D. Gnambs, T., Lockl, K., Attig, M. (2016). Competence Data in NEPS: Overview of Measures and Variable Naming Conventions (Starting Cohorts 1 to 6). *Bamberg: University of Bamberg,NEPS*.

Ganter, S. (2003). *Soziale Netzwerke und interethnische Distanz: Theoretische und empirische Analysen zum Verhältnis von Deutschen und Ausländern*. Wiesbaden: Westdt. Verlag.

Ganter, S. (2003). Soziale Netzwerke und interethnische Distanz: theoretische und empirische Analysen zum Verhältnis von Deutschen und Ausländern. Springer-Verlag.

Gatzke, N. (2010). Polenbild in Deutschland: Wandel und Kontinuität-von Polen als Gegenbild der Aufklärung bis zur EU-Reformdebatte. Potsdam:Universität Potsdam.

Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S. (2012). The assessment of reading competence (including sample items for grade 5 and 9). *Bamberg: University of Bamberg, NEPS*.

Horr, A., Hunkler, C., & Kroneberg, C. (2018). Ethnic Discrimination in the German Housing Market, *Zeitschrift für Soziologie,*47(2),134-146.

Kaas, L., & Manger, C. (2012). Ethnic discrimination in Germany's labour market: a field experiment, *German Economic Review,13*(1),1-20.

Kahraman, B. & Knoblich, G. (2000). Stechen statt Sprechen. Valenz und Aktivierbarkeit von Stereotypen über Türken*, Zeitschrift für Sozialpsychologie*,31,31–43.

Kahraman, B., & Knoblich, G. (2000). "Stechen statt Sprechen": Valenz und Aktivierbarkeit von Stereotypen über Türken. *Zeitschrift für Sozialpsychologie*.

Kalter, F. (2003). *Chancen, Fouls und Abseitsfallen: Migranten im deutschen Ligenfußball*. Wiesbaden: Westdt. Verlag.

Kalter, F. (2006). Auf der Suche nach einer Erklärung für die spezifischen Arbeitsmarktnachteile von Jugendlichen türkischer Herkunft*, Zeitschrift für Soziologie*,35,144-160.

Kalter, F. (2011). The Second Generation in the German Labor Market: Explaining the Turkish Exception, in Richard, A., Waters, M. (eds.). *The Next Generation – Immigrant Youth in a Comparative Perspective*. New York and London: NYUP,166-184.

Keita, S., & Valette, J. (2016). Natives' attitudes and immigrants' unemployment durations. Études et Documents, n° 23, CERDI.

Koller, I., Haberkorn, K., & Rohm, T. (2014). NEPS Technical Report for Reading: Scaling results of Starting Cohort 6 for adults in main study 2012, *NEPS Working Paper,*48,2-27.

Kossoudji, S. (1988). English language ability and the labor market opportunities of Hispanic and East Asian immigrant men, *Journal of Labor Economics*,6(2),205-228.

Kotzur, P., Forsbach, N., & Wagner, U. (2017). Choose Your Words Wisely. Stereotypes, Emotions, and Action Tendencies Toward Fled People as a Function of the Group Label, *Social Psychology,*48,226-241.

Kucharczyk, J., Łada, A., Ochmann,C.,Wenerski,L. (2013) Im Osten was Neues? Das Bild Polens und Russlands in Deutschland, Warschau:Bertelsmann.

Pendakur, K., & Pendakur, R. (2002). Language as both human capital and ethnicity, *International Migration Review*,36(1),147-177.

Rakić, T., Steffens, M. & Mummendey, A. (2011). Blinded by the accent! The minor role of looks in ethnic categorization, *Journal of Personality and Social Psychology*,100(1),16-29.

Rödin, M. & Özcan, G. (2011). Is it how you look or speak that matters? - An experimental study exploring the mechanisms of ethnic discrimination, *SULCIS working paper (Dep. of Economics, Stockholm University)*,11(3),1-51.

Schulz, B., & Leszczensky, L. (2016). Native friends and host country identification among adolescent immigrants in Germany: The role of ethnic boundaries. *International Migration Review*,50(1),163-196.

Steinbach, A. (2004). *Soziale Distanz: Ethnische Grenzziehung und die Eingliederung von Zuwanderern in Deutschland*. Wiesbaden: VS Verl. für Sozialwiss.

Szczęk, J. (2006). Man kann seine Nachbarn nicht wählen:Deutsche Polenwitze als Träger der Stereotype, *Studia Germanica Gedanensia 14*,169-179.

Van Tubergen, F., Maas, I., & Flap, H. (2004). The economic incorporation of immigrants in 18 Western societies: Origin, destination, and community effects, *American Sociological Review,* 69(5),704-727.

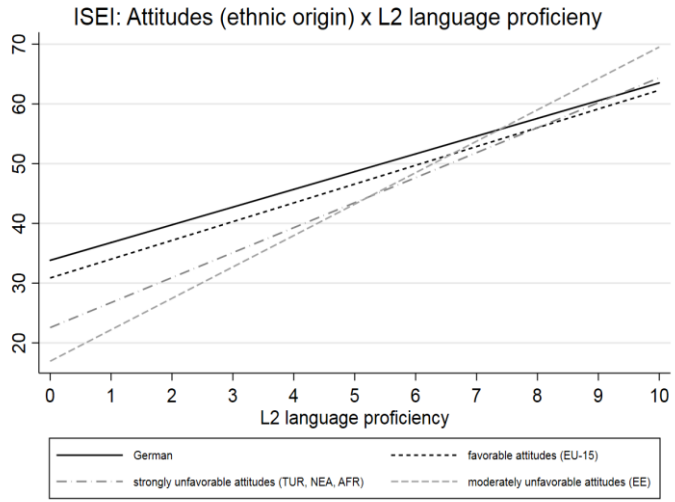
Wagner S. (2018). Die „Armutsmigranten" kommen – Zur Repräsentation von Migrantinnen und Migranten aus Bulgarien und Rumänien in der deutschen Tagespresse. In: Lünenborg M., Sell S. (eds) Politischer Journalismus im Fokus der Journalistik.Springer, Wiesbaden

Table I: Estimates from logistic and linear regression models examining the relationship between attitudes towards ethnic origin, L2 language proficiency and labor market outcomes

	Unemployment: Yes vs. No				*Job Type: Blue vs. White collar*				*Job Quality: ISEI*				*Job Prestige: MPS*				*Job Income*			
	M (1) AME (se)	M (2) AME (se)	M (3) AME (se)	M (4) ß (se)	M (1) AME (se)	M (2) AME (se)	M (3) AME (se)	M (4) ß (se)	M (1) ß (se)	M (2) ß (se)	M (3) ß (se)	M (4) ß (se)	M (1) ß (se)	M (2) ß (se)	M (3) ß (se)	M (4) ß (se)	M (1) ß (se)	M (2) ß (se)	M (3) ß (se)	M (4) ß (se)
Attitudes towards ethnic origin (Ref. Natives)																				
Favourable (EU-15)	-0.02	0.01	0.07	0.84	**-0.06+**	0.01	-0.11	-0.88	**-4.88****	**-2.40+**	-2.96	-2.65	-3.55	0.26	-8.15	-4.62	-59.01	127.73	-458.36	-240.76
	(0.01)	(0.01)	(0.05)	(1.23)	**(0.04)**	(0.03)	(0.15)	(1.54)	**(1.59)**	**(1.33)**	(3.53)	(5.12)	(3.01)	(2.42)	(7.58)	(10.38)	(200.17)	(164.89)	(628.79)	(1060.20)
Moderately unfavourable (Eastern Europe)	**-0.04*****	-0.01	**-0.19+**	**-1.61+**	**-0.22*****	**-0.10*****	**-0.36*****	**-3.83*****	**-10.84*****	**-5.13*****	**-11.27*****	**-11.51*****	**-17.97*****	**-8.22*****	**-15.44*****	**-15.62****	**-593.92*****	6.27	363.63	201.43
	(0.01)	(0.01)	**(0.10)**	**(0.86)**	**(0.02)**	**(0.02)**	**(0.04)**	**(0.83)**	**(0.98)**	**(0.82)**	**(1.90)**	**(3.24)**	**(1.69)**	**(1.45)**	**(3.45)**	**(4.91)**	**(95.67)**	(79.89)	(260.44)	(329.28)
Strongly unfavourable (Turkey, Near East, Africa)	**-0.14*****	-0.02	**-0.23+**	**-2.02***	**-0.28*****	**-0.08***	**-0.37*****	**-3.72***	**-14.55*****	**-5.76*****	**-16.89*****	**-18.22****	**-21.30*****	**-7.36***	**-30.26*****	**-27.78***	**-567.31****	-9.78	-73.01	-820.96
	(0.03)	(0.02)	**(0.14)**	**(0.99)**	**(0.04)**	**(0.04)**	(0.05)	(1.65)	**(1.90)**	**(1.60)**	**(4.74)**	**(6.94)**	**(3.32)**	**(2.97)**	**(8.88)**	**(11.62)**	**(215.31)**	(170.91)	(487.30)	(634.92)
Other	-0.02	0.01	-0.9	-0.99	-0.05	-0.05	-0.24	-0.91	**-4.43+**	**-3.81***	-6.96	**-2.44**	**-3.83**	**-4.03**	-13.64	-10.96	**-737.76****	-163.21	-670.35	-1450.58
	(0.02)	(0.02)	(0.19)	(1.85)	(0.05)	(0.05)	(0.16)	(2.67)	**(2.31)**	**(1.79)**	(5.63)	**(8.89)**	**(4.27)**	**(3.39)**	(10.55)	(16.94)	**(271.81)**	(189.33)	(873.26)	(1595.13)
Born in Germany (Ref. born abroad)	**0.02*****	0.00	-0.01	0.02	**0.15*****	**0.08*****	**0.58****	**0.56****	**10.31*****	**4.74*****	**3.99*****	**3.85*****	**16.62*****	**7.44*****	**6.54*****	**6.30*****	**504.55*****	-12.41	36.88	46.79
	(0.01)	(0.01)	(0.22)	(0.23)	**(0.12)**	**(0.02)**	**(0.19)**	**(0.20)**	**(1.13)**	**(0.93)**	**(0.95)**	**(0.96)**	**(2.02)**	**(1.70)**	**(1.73)**	**(1.76)**	**(121.52)**	(100.79)	(99.71)	(101.48)
(L2) Language proficiency		**0.01*****	**0.01****	0.12		**0.07*****	**0.05*****	**0.46*****		**3.26*****	**2.97*****	**2.99*****		**5.16*****	**4.70*****	**4.72*****		**180.45*****	**187.61*****	**122.89***
		(0.00)	**(0.00)**	(0.08)		**(0.00)**	**(0.00)**	**(0.05)**		**(0.16)**	**(0.18)**	**(0.28)**		**(0.32)**	**(0.37)**	**(0.50)**		**(35.30)**	**(42.47)**	**(53.24)**
Additional Information		**0.00*****	**0.00*****	0.00		**0.00**	**0.00**	0.00		**0.01****	**0.01****	0.01		**0.01***	**0.01***	0.01		**0.88+**	**0.88+**	-2.13
		(0.00)	**(0.00)**	(0.00)		**(0.00)**	**(0.00)**	(0.00)		**(0.00)**	**(0.00)**	(0.01)		**(0.00)**	**(0.00)**	(0.02)		**(0.52)**	**(0.52)**	(2.09)
Interaction Effects																				
Attitudes towards ethnic origin x L2 proficiency																				
Favourable x L2 proficiency			-0.01	-0.08			0.00	0.17			0.17	0.02			1.49	0.45			89.91	39.62
			(0.01)	(0.21)			(0.02)	(0.29)			(0.51)	(0.76)			(1.22)	(1.73)			(102.55)	(170.14)
Moderately unfavourable x L2 proficiency			**0.02****	**0.34+**			**0.06*****	**0.62*****			**1.21*****	**1.14+**			**1.42***	**1.21**			-71.82	-32.46
			(0.01)	**(0.18)**			**(0.01)**	**(0.17)**			**(0.34)**	**(0.58)**			**(0.64)**	**(0.90)**			(48.64)	(64.49)
Strongly unfavourable x L2 proficiency			**0.04****	**0.35**			**0.06*****	**0.64+**			**2.29***	**2.51+**			**4.67****	**4.03**			7.47	137.74
			(0.01)	**(0.21)**			**(0.02)**	**(0.34)**			**(0.90)**	**(1.38)**			**(1.75)**	**(2.44)**			(96.23)	(126.48)
Other x L2 proficiency			0.01	0.22			0.01	0.14			0.62	0.08			1.70	1.55			80.66	201.55
			(0.01)	(0.35)			(0.02)	(0.50)			(0.91)	(1.44)			(1.77)	(2.86)			(151.04)	(282.16)
Attitudes towards ethnic origin x L2 proficiency x Information																				
Favourable x information x L2				-0.00				0.00				-0.00				-0.03				-2.50
				(0.00)				(0.00)				(0.03)				(0.08)				(8.02)
Moderately unfavourable x information x L2				-0.00				0.00				0.00				0.00				1.09
				(0.00)				(0.00)				(0.02)				(0.03)				(2.64)
Strongly unfavourable x information x L2				-0.00				0.00				0.02				-0.04				8.49
				(0.00)				(0.00)				(0.09)				(0.13)				(6.74)
Other x information x L2				0.00				0.00				-0.03				-0.02				6.50
				(0.00)				(0.00)				(0.06)				(0.11)				(10.25)
Attitudes towards ethnic origin x Information																				
Favourable x information				0.01				0.00				0.00				0.01				0.52
				(0.01)				(0.01)				(0.00)				(0.01)				(1.29)
Moderately unfavourable x information				0.00				0.00				0.00				0.00				-0.30
				(0.01)				(0.01)				(0.00)				(0.01)				(0.52)
Strongly unfavourable x information				0.01				-0.01				0.00				0.01				-1.49
				(0.01)				(0.02)				(0.00)				(0.03)				(1.29)
Other x information				0.00				-0.01				-0.00				0.00				-1.01
				(0.01)				(0.02)				(0.02)				(0.02)				(1.89)
L2 proficiency x Information				0.00				-0.00				-0.00				-0.00				0.55
				(0.00)				(0.00)				(0.00)				(0.00)				(0.42)
Human Capital	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
Cultural Capital	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
Social Capital	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
Economic Capital	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
Symbolic Capital	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
Personal Capital	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES	NO	YES	YES	YES
N	10,708	10,708	10,708	10,708	8,557	8,557	8,557	8,557	8,694	8,694	8,694	8,694	8,683	8,683	8,683	8,683	6,958	6,958	6,958	6,958
R2	.02	.20	.20	.27	.03	.45	.45	.45	.02	.36	.36	.36	.02	.35	.35	.35	.003	.31	.31	.31

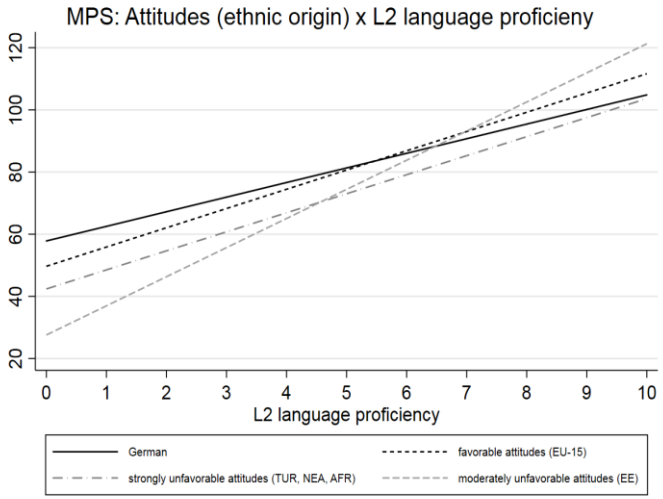
Note: ***p<.001; p<.01;*p<0.05; +p<.01; additionally controlled for age, gender and time of first L2 assessment; significant coefficients are **bold**, in general AMEs are estimated to report effects for logistic models, when conceptually not feasible (three-way interaction, inc. two continuous variables) logit coefficients are reported

Figure III. Interaction effects, margin plots (Job quality: ISEI)



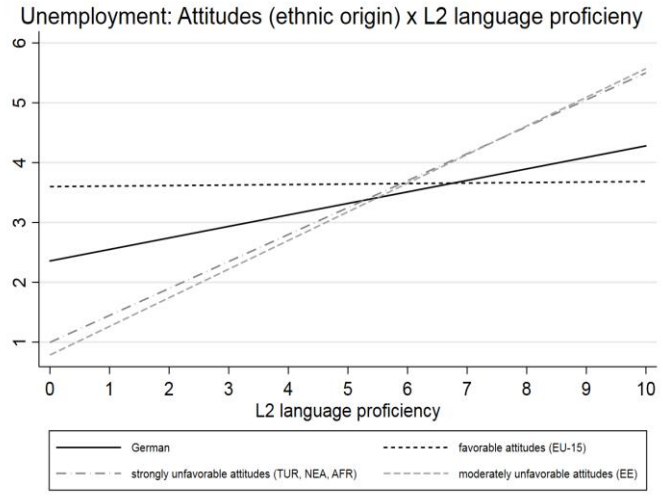
Note: Excluding others origins to facilitate ease of reading.

Figure IV. Interaction effects, margin plots (Job prestige: MPS)



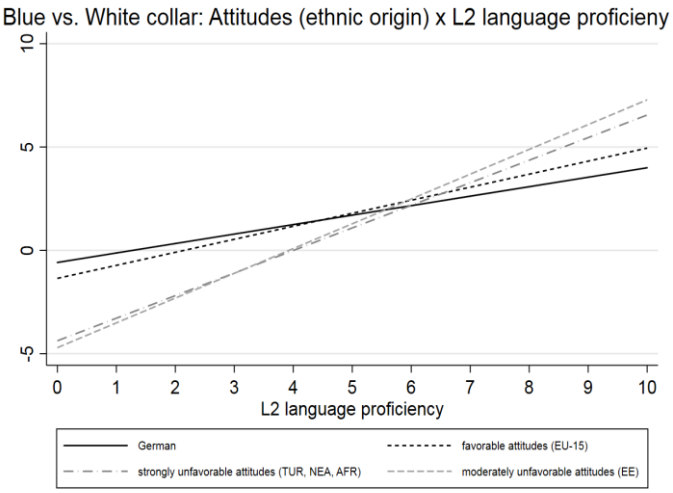
Note: Excluding others origins to facilitate ease of reading.

Figure I. Interaction effects, margin plots (Unemployment)



Note: Excluding others origins to facilitate ease of reading.

Figure II. Interaction effects, margin plots (Blue vs. White collar position)



Note: Excluding others origins to facilitate ease of reading.

Appendix: Table II. Sample characteristics.

	Job Type: Blue vs. White collar (n=8577)				Job Quality: ISEI (n=8694)				Job Prestige: MPS (n=8683)				Job Income (n=6958)				Unemployment: Yes/No (n=10708)			
	N	Mean/%	SD	% MI	N	Mean/%	SD	% MI	N	Mean/%	SD	% MI	N	Mean/%	SD	% MI	N	Mean/%	SD	% MI
Job type	(8557)			0.00																
Blue collar	2043	23.88																		
White collar	6514	76.12																		
Job quality - ISEI					8694	49.88	21.26	0.00												
Job prestige - MPS									(8683)	83.35	39.31	0.00								
Job Income													(6958)	2910.08	2756.08	0.00				
Unemployment																	(10708)			0.00
Employed																	10244	95.67		
Unemployed																	464	4.33		
Unemployment duration																				
Ethnic origin	(8557)			0.00	(8694)			0.00	(8683)			0.00	(6958)			0.00	(10708)			0.00
German	7040	82.27			7152	82.26			7144	82.28			5727	82.31			8794	82.13		
Western Immigrants (EU15)	260	3.04			265	3.05			265	3.05			198	2.85			346	3.23		
Non-Western Immigrants	130	1.52			133	1.53			133	1.53			102	1.47			172	1.61		
eastern europe	991	11.58			1004	11.55			1001	11.53			817	11.74			1221	11.40		
Other	136	1.59			140	1.61			140	1.61			114	1.64			175	1.63		
Born in Germany	(8557)			0.00	(8694)			0.00	(8683)			0.00	(6958)			0.00	(10708)			0.00
Born in Germany	7628	89.14			7748	89.12			7740	89.14			6204	89.16			9545	89.14		
Born abroad	929	10.86			946	10.88			943	10.86			754	10.84			1163	10.86		
(L2) Language proficiency	(5543)	5.70	1.34	35.22	(5623)	5.69	1.33	35.32	(5619)	5.70	1.33	35.29	(4485)	5.69	1.32	35.54	(6955)	5.59	1.35	35.05
Additional Information	(3057)	107.76	109.32	64.27	(3096)	107.97	109.61	64.39	(3092)	108.06	109.66	64.39	(2455)	115.07	111.33	64.72	(3708)	116.18	120.58	65.37
Controls																				
Human Capital																				
Education (CASMIN)	(8551)			0.07	(8688)			0.07	(8677)			0.07	(6950)			0.11	(10697)			0.10
1a. No education (below general)	54	0.63			55	0.63			55	0.63			45	0.65			91	0.85		
1b. General elementary education	178	2.08			179	2.06			179	2.06			134	1.93			263	2.46		
1c. Basic vocational qualification	1331	15.57			1375	15.83			1374	15.83			1085	15.61			1949	18.22		
2a. Intermediate vocational qualification	2657	31.07			2708	31.17			2703	31.15			2234	32.14			3239	30.28		
2b. Intermediate general qualification	155	1.81			159	1.83			159	1.83			126	1.81			222	2.08		
2c-gen General maturity certificate	216	2.53			219	2.52			219	2.52			176	2.53			315	2.94		
2c-voc Vocational maturity certificate	1294	15.13			1314	15.12			1311	15.11			1074	15.45			1517	14.18		
3a. Lower tertiary certificate	910	10.64			915	10.53			918	10.58			726	10.45			1067	9.97		
3b. Higher tertiary certificate	1756	20.54			1764	20.30			1759	20.27			1350	19.42			2034	19.01		
Cultural Capital																				
Highest parental education (years)	(8316)	13.17	2.44	2.82	(8446)	13.16	2.44	2.85	(8435)	13.16	2.44	2.86	(6758)	13.11	2.40	2.87	(10376)	13.08	2.41	3.10
Social Capital																				
Network status composition	(5748)	32.64	12.73	32.83	(5866)	32.69	12.73	32.53	(5867)	32.69	12.73	32.43	(4675)	32.73	12.73	32.81	(7065)	32.46	12.90	34.02
Likelihood of reference for new job	(5623)	2.24	.96	34.29	(5731)	2.24	.96	34.08	(5733)	2.24	.96	33.97	(4580)	2.24	.95	34.18	(6668)	2.28	.98	37.73
Likelihood of information on new job	(5637)	2.43	1.01	34.12	(5748)	2.44	1.01	33.89	(5750)	2.44	1.01	33.78	(4589)	2.45	1.01	34.05	(6695)	2.48	1.02	37.48
Economic Capital																				
Highest parental ISEI	(7909)	45.56	21.40	7.57	(8033)	45.55	21.38	7.60	(8025)	45.53	21.38	7.58	(6437)	45.16	21.37	7.49	(9878)	44.76	21.25	7.75
Borrowing money (>250 EUR)	(5711)	1.49	.84	33.26	(5827)	1.49	.84	32.98	(5828)	1.49	.84	32.88	(4649)	1.49	.83	33.18	(7013)	1.52	.87	34.51
Symbolic Capital																				
German citizenship	(8556)			0.01	(8693)			0.01	(8682)			0.01	(6957)			0.01	(10707)			0.01
yes	8297	96.97			8428	96.95			8417	96.95			6758	97.14			10353	96.69		
no	259	3.03			265	3.05			265	3.05			199	2.86			354	3.31		
Personal Capital																				
Importance career	(5701)	2.13	.97	33.38	(5815)	2.31	.97	33.11	(5816)	2.31	.97	33.02	(4635)	2.34	.97	33.39	(6844)	2.29	.99	36.09
Importance additional qualifications	(5742)	1.65	.77	32.90	(5859)	1.65	.77	32.61	(5860)	1.65	.77	32.51	(4670)	1.67	.77	32.88	(7035)	1.66	.78	34.30
Mental health	(5489)	50.23	9.97	35.85	(5603)	50.22	9.97	35.55	(5602)	50.22	9.97	35.48	(4473)	50.34	9.90	35.71	(6753)	50.37	10.22	36.94
Age	(8557)			0.00	(8694)			0.00	(8683)			0.00	(6958)			0.00	(10708)			0.00
28-44	2313	27.03			2336	26.87			2333	26.87			1730	24.86			2673	24.96		
45-52	2469	28.85			2859	32.88			2502	28.81			2442	35.10			3075	28.72		
53-59	2478	28.96			2151	24.74			2502	28.81			1866	26.82			2376	22.19		
60-70	1297	15.16			1348	15.50			1346	15.50			920	13.22			2584	24.13		
Gender	(8557)			0.00	(8694)			0.00	(8683)			0.00	(6958)			0.00	(10708)			0.00
male	4380	51.19			4464	51.35			4458	51.34			3599	51.72			5285	49.36		
female	4177	48.81			4230	48.65			4225	48.66			3359	48.28			5423	50.64		
Family status	(8303)			2.97	(8438)			2.94	(8427)			2.95	(6778)			2.59	(10425)			2.64
married	5675	68.35			5771	68.39			5764	68.40			4770	70.37			7179	68.86		
registered partnership	23	0.28			24	0.28			24	0.28			18	0.27			30	0.29		
divorced	729	8.78			745	8.83			742	8.81			560	8.26			890	8.54		
widowed	134	1.61			136	1.61			136	1.61			101	1.49			273	2.62		
single	1742	20.98			1762	20.88			1761	20.90			1329	19.61			2053	19.69		
subsample	(8557)			0.00	(8694)			0.00	(8683)			0.00	(6958)			0.00	(10708)			0.00
ALWA	3966	46.35			3997	45.97			3999	46.06			3248	46.68			4348	40.61		
Supplement	996	11.64			1052	12.10			1051	12.10			742	10.66			1811	16.91		
Refreshment	924	10.80			957	11.01			956	11.01			790	11.35			1059	9.89		
NEPS	2671	31.21			2688	30.92			2677	30.83			2178	31.30			3490	32.59		

Supplementary materials: Table III. Drop-outs after the first L2 assessment. Breakdown by deciles of L2 proficiency (Total sample and separated by different ethnic groups)

Total sample

L2 percentile	N first L2 test	N scnd L2 test	% dropout
1	848	531	0,37
2	851	568	0,33
3	850	632	0,26
4	850	626	0,26
5	842	627	0,26
6	848	633	0,25
7	850	653	0,23
8	845	656	0,22
9	848	680	0,20
10	848	690	0,19
Total	8,480	6,296	0,26

Dropouts by L2 proficiency (deciles) for ethnic groups associated with different attitudes

L2 percentile	German			Favorable (EU-15)			Moderately unfavorable			Strongly unfavorable		
	N first L2 test	N scnd L2 test	% dropout	N first L2 test	N scnd L2 test	% dropout	N first L2 test	N scnd L2 test	% dropout	N first L2 test	N scnd L2 test	% dropout
1	631	414	0,34	25	10	0,60	145	81	0,44	39	19	0,51
2	672	453	0,33	24	12	0,50	122	85	0,30	22	11	0,50
3	677	513	0,24	26	18	0,31	112	76	0,32	17	10	0,41
4	693	521	0,25	22	16	0,27	111	74	0,33	12	6	0,50
5	704	529	0,25	29	22	0,24	80	60	0,25	11	6	0,45
6	693	530	0,24	31	25	0,19	96	61	0,36	10	6	0,40
7	715	550	0,23	26	19	0,27	79	64	0,19	7	3	0,57
8	741	576	0,22	19	16	0,16	65	49	0,25	7	6	0,14
9	715	580	0,19	29	20	0,31	69	52	0,25	7	5	0,29
10	723	587	0,19	28	23	0,18	73	64	0,12	4	3	0,25
Total	6,964	5,253	0,25	259	181	0,30	952	666	0,30	136	75	0,45

Supplementary materials: Table IV. Robustness of main models, additional indicators on German language exposure

	Unemployment: Yes vs. No	*Job Type: Blue vs. White collar*	*Job Quality: ISEI*	*Job Prestige: MPS*	*Job Income*
	Model (3)	Model (3)	Model (3)	Model (3)	Model (3)
	AME (SE)	AME (SE)	ß (SE)	ß (SE)	ß (SE)
Attitudes associated with ethnic origin (Ref. German natives)					
Favorable (EU-15)	**0.11****	-0.10	-2.44	-9.13	-450.26
	(0.04)	(0.15)	(3.46)	(7.24)	(657.77)
Moderately unfavorable (Eastern Europe)	**-0.25***	**-0.36*****	**-11.56*****	**-16.31*****	93.32
	(0.10)	**(0.04)**	**(2.00)**	**(3.54)**	(227.82)
Strongly unfavorable (Turkey, Near East, Africa)	**-0.35***	**-0.37*****	**-17.50*****	**-31.82*****	-195.28
	(0.16)	(0.05)	(4.42)	(8.81)	(459.32)
Other	-0.39	**-0.26+**	-6.86	-15.41	-329.14
	(0.31)	**(0.14)**	(5.31)	(11.45)	(952.89)
Born in Germany (Ref. born abroad)	**-0.02+**	**0.06****	**4.13*****	**7.20*****	**53.40**
	(0.01)	**(0.02)**	**(1.01)**	**(1.91)**	**(113.01**
(L2) Language proficiency	**0.01*****	**0.05*****	**3.02*****	**4.77*****	**131.76*****
	(0.00)	**(0.00)**	**(0.20)**	**(0.36)**	**(30.08)**
Additional Information	**0.00*****	0.00	**0.01****	**0.01****	**1.00+**
	(0.00)	(0.00)	**(0.00)**	**(0.00)**	**(0.56)**
Share of (non) German friends	**-0.00****	0.01	0.21	**0.86***	1.24
	(0.00)	(0.00)	(0.20)	**(0.36)**	(25.90)
Number of German acquaintances (professional network)	**0.00****	**0.01*****	**0.24****	**0.64*****	7.36
	(0.00)	**(0.00)**	**(0.07)**	**(0.15)**	(8.59)
Interaction Effects					
Attitudes associated with ethnic origin x L2 proficiency					
Favorable (EU-15) x L2 proficiency	-0.01	-0.00	0.07	1.53	88.51
	(0.01)	(0.02)	(0.51)	(1.15)	(106.41)
Moderately unfavorable (Eastern Europe) x L2 proficiency	**0.03*****	**0.05*****	**1.24*****	**1.45***	-22.64
	(0.01)	**(0.01)**	**(0.34)**	**(0.64)**	(41.65)
Strongly unfavorable (Turkey, Near East, Africa) x L2 proficiency	**0.06*****	**0.06*****	**2.39****	**4.81****	24.40
	(0.01)	**(0.02)**	**(0.82)**	**(1.74)**	(87.80)
Other x L2 proficiency	0.02	0.01	**0.61**	1.88	31.64
	(0.01)	(0.02)	(0.87)	(1.91)	(167.19)
Human Capital	YES	YES	YES	YES	YES
Cultural Capital	YES	YES	YES	YES	YES
Social Capital	YES	YES	YES	YES	YES
Economic Capital	YES	YES	YES	YES	YES
Symbolic Capital	YES	YES	YES	YES	YES
Personal Capital	YES	YES	YES	YES	YES
N	10708	8557	8694	8683	6526
R2	.20	.45	.36	.35	.30

Note: ***p<.001; p<.01;*p<0.05; +p<.01; additionally controlled for age, gender and time of first L2 assessment; significant coefficients are **bold.**

Supplementary materials: Table V. Robustness of main models, group-specific models, (differential) mediation effects examined

Attitudes associated with ethnic origin		*Unemployment: Yes/No*		*Job Type: Blue vs. White collar*		*Job Quality: ISEI*		*Job Prestige: MPS*		*Job Income*	
		Model (1) AME (SE)	Model (2) AME (SE)	Model (1) AME (SE)	Model (2) AME (SE)	Model (1) ß (SE)	Model (2) ß (SE)	Model (1) ß (SE)	Model (2) ß (SE)	Model (1) ß (SE)	Model (2) ß (SE)
German (natives)	(L2) Language proficiency	**0.01****	**0.01****	**0.05*****	**0.05*****	**2.79*****	**2.78*****	**4.36*****	**4.34*****	**121.98*****	**122.04*****
		(0.00)	**(0.00)**	**(0.00)**	**(0.00)**	**(0.20)**	**(0.20)**	**(0.36)**	**(0.36)**	**(31.00)**	**(31.15)**
	Share of (non) German friends		**-0.00***		**0.01***		**0.49***		**1.18****		**14.66**
			(0.00)		**(0.01)**		**(0.23)**		**(0.42)**		**(33.24)**
	Number of German acquaintances (prof.network)		**0.00****		**0.01*****		**0.20***		**0.58*****		9.43
			(0.00)		**(0.00)**		**(0.08)**		**(0.17)**		(9.08)
	N	8794	8794	7038	7038	7152	7152	7144	7144	5375	5375
Favorable (EU-15)	(L2) Language proficiency	**0.04*****	**0.03*****	**0.12*****	**0.12*****	**2.92*****	**2.92*****	**6.01*****	**6.05*****	**359.48***	**365.94***
		(0.01)	**(0.01)**	**(0.01)**	**(0.01)**	**(0.64)**	**(0.63)**	**(1.52)**	**(1.51)**	**(162.75)**	**(165.36)**
	Share of (non) German friends		-0.01		-0.01		0.22		**1.53**		13.95
			(0.00)		(0.01)		(0.83)		**(1.76)**		(141.03)
	Number of German acquaintances (prof. network)		**0.01***		0.01		0.25		0.84		3.25
			(0.00)		(0.00)		(0.41)		(0.75)		(82.33)
	N	1221	1221	991	991	265	265	265	265	191	191
Moderately unfavorable (Eastern Europe)	(L2) Language proficiency	**0.04*****	**0.03*****	**0.12*****	**0.12*****	**5.02*****	**4.95*****	**7.61*****	**7.49*****	**124.07****	**126.17****
		(0.01)	**(0.01)**	(0.01)	**(0.01)**	**(0.35)**	**(0.35)**	**(0.66)**	**(0.66)**	**(44.21)**	**(44.42)**
	Share of (non) German friends		-0.01		-0.01		-0.42		0.10		**-62.13**
			(0.00)		(0.01)		(0.42)		(0.80)		**(42.46)**
	Number of German acquaintances (prof.network)		**0.01***		0.01		0.32		**0.79+**		-6.59
			(0.00)		(0.00)		(0.22)		**(0.43)**		(23.12)
	N	1221	1221	991	991	1004	1004	1001	1001	762	762
Strongly unfavorable (Turkey, Near East, Africa)	(L2) Language proficiency	**0.08*****	**0.08*****	**0.13*****	**0.14*****	**7.32*****	**7.36*****	**11.83*****	**11.83*****	**402.92***	**419.89***
		(0.02)	**(0.02)**	**(0.03)**	**(0.04)**	**(1.27)**	**(1.29)**	**(2.32)**	**(2.36)**	**(172.55)**	**(178.59)**
	Share of (non) German friends		-0.02		-0.03		-0.80		-0.40		72.28
			(0.02)		(0.03)		(1.06)		(1.81)		(126.73)
	Number of German acquaintances (prof.network)		0.00		0.01		0.22		0.15		-33.66
			(0.01)		(0.02)		(0.62)		(1.00)		(80.17)
	N	215	215	152	152	162	162	162	162	117	117

Note: ***p<.001; p<.01;*p<0.05; +p<.01; additionally controlled for age and gender; significant coefficients are **bold.**

Supplementary materials: Table VI. Robustness of main models, additional indicators on compensatory languages (English, minority) and language use on the job included

	Job Type: Blue vs. White collar		*Job Quality: ISEI*		*Job Prestige: MPS*	
	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
	AME (SE)	AME (SE)	AME (SE)	AME (SE)	ß (SE)	ß (SE)
Attitudes towards ethnic origin (Ref. Natives)						
Favourable (EU-15)	-0.01	-0.06	3.66	3.20	3.29	2.38
	(0.14)	(0.14)	(4.49)	(4.22)	(8.18)	(8.13)
Moderately unfavourable (Eastern Europe)	**-0.37*** **	**-0.31*** **	**-10.79*** **	**-7.81*** **	**-16.36*** **	**-12.19** **
	(0.06)	**(0.06)**	**(2.14)**	**(2.05)**	**(3.77)**	**(3.72)**
Strongly unfavourable (Turkey, Near East, Africa	**-0.52*** **	**-0.49*** **	**-14.35*** **	**-10.77** **	**-28.19** **	**-23.59** **
	(0.07)	**(0.11)**	**(4.33)**	**(4.05)**	**(8.80)**	**(8.42)**
Other	0.22	0.19	5.38	**4.12**	**13.83**	**11.90**
	(0.19)	(0.15)	(6.29)	**(6.21)**	**(12.60)**	**(12.44)**
Born in Germany (Ref. born abroad)	0.04	0.02	**2.59***	**1.29**	**5.43***	**3.45**
	(0.03)	(0.03)	**(1.22)**	**(1.16)**	**(2.25)**	**(2.19)**
(L2) Language proficiency	**0.00*** **	**0.00*** **	**0.28*** **	**0.23*** **	**0.45*** **	**0.38*** **
	(0.00)	**(0.00)**	**(0.03)**	**(0.03)**	**(0.06)**	**(0.06)**
Additional Information	0.00	**-0.00**	**0.01***	(0.00)	0.01	0.00
	(0.00)	**(0.00)**	**(0.00)**	(0.00)	(0.01)	(0.01)
Language spoken with colleagues (Ref. German)						
English		**0.07*** **		**2.79*** **		**2.13**
		(0.02)		**(0.74)**		**(1.55)**
Minority language		**-0.04**		**-2.31+**		**-3.73**
		(0.03)		**(1.34)**		**(2.53)**
Extent reading on job		**0.03*** **		**1.22*** **		**1.27*** **
		(0.00)		**(0.11)**		**(0.22)**
Extent writing on job		**0.07*** **		**4.60*** **		**7.88*** **
		(0.01)		**(0.25)**		**(0.48)**
Interaction Effects						
Attitudes towards ethnic origin x L2 proficiency						
Favourable x L2 proficiency	-0.00	0.00	-0.12	-0.07	-0.08	0.00
	(0.00)	(0.00)	(0.11)	(0.11)	(0.21)	**(0.21)**
Moderately unfavourable x L2 proficiency	**0.01*** **	**0.01*** **	**0.20** **	**0.16** **	**0.25***	**0.19+**
	(0.00)	**(0.00)**	**(0.06)**	**(0.06)**	**(0.12)**	**(0.12)**
Strongly unfavourable x L2 proficiency	**0.01*** **	**0.01*** **	**0.30***	**0.25***	**0.71** **	**0.65***
	(0.00)	**(0.00)**	**(0.12)**	**(0.12)**	**(0.27)**	**(0.26)**
Other x L2 proficiency	-0.00	-0.00	-0.12	**-0.08**	-0.32	-0.26
	(0.00)	(0.00)	(0.17)	(0.17)	(0.34)	(0.33)
Human Capital	YES	YES	YES	YES	YES	YES
Cultural Capital	YES	YES	YES	YES	YES	YES
Social Capital	YES	YES	YES	YES	YES	YES
Economic Capital	YES	YES	YES	YES	YES	YES
Symbolic Capital	YES	YES	YES	YES	YES	YES
Personal Capital	YES	YES	YES	YES	YES	YES
N	5935	5935	6078	6078	6072	6072
R2	.46	.54	.35	.40	.33	.37

Note: ***p<.001; p<.01;*p<0.05; +p<.01; additionally controlled for age and gender; significant coefficients are **bold.**

Paper 2

Foreign accents in the early hiring process

Planned Submission: International Migration Review
Authors: Miriam Schmaus & Cornelia Kristen, Bamberg University

Foreign accents in the early hiring process

This study aims at assessing the ways in which foreign-accented speech influences the labor market prospects of immigrants in Germany. We are particularly interested in the question of how foreign accents affect the opportunities of being recruited. Building upon human capital- and discrimination theories, we investigate whether foreign accents affect the perceptions and evaluations of employers in terms of productivity or if they are a source of bias triggering discrimination. To address these questions we conducted a field experiment. The field experiment focuses on the initial contact between employers and applicants (with and without a Turkish accent) who call to inquire about an advertised position. The experimental audit thus allows studying whether foreign-accented speech affects the opportunity to partake in the hiring process. In addition to varying applicant characteristics, such as the job seeker's accent and name, information on job and firm characteristics was collected. This included information on the communicative demands or the extent of customer contact in the advertised job, as well as information on the structure of the recruitment process. Moreover, from the Federal Employment Agency and the German General Survey, context information was taken and merged to the experimental set. This included information on regional labor supply, job-specific vacancy rates, or regional levels of anti-immigrant attitudes. The aim behind this procedure was to allow for tentative inferences on the reasons behind any accent effects. Results demonstrated that foreign-accented applicants were turned down more often than German jobseekers and accent-free applicants of the same origin. This disadvantage was not conditional on the communicative demands of the job in question, which might suggest that an accent may not be primarily considered in terms of productivity. However, certain indications pointed at the potential role of accent-related taste discrimination. On the one hand, adverse effects of foreign-accented speech were less pronounced in firms with a more standardized recruitment process, where employers might have fewer opportunities to factor personal tastes into their decision-making. On the other hand, accented applicants face stronger disadvantages in regions, where ethnic aversions cluster and where acting upon such distaste should be perceived as less costly. Customers' tastes, however, seemed to be irrelevant for employers' decision making.

Introduction

Immigrants who acquire the language of their destination country as a second language frequently display a foreign accent. This means that phonology and intonation of the mother tongue are carried over into the destination language in question (Lippi-Green 1994). Such accents can persist many years after immigration, even when individuals achieve near-perfect control over other features of the majority language. They are very salient aspects of speech (Derwing and Munro 2009) and a strong cue of group membership (Rakić, Steffens and Mummendey 2011; Rödin and Özkan 2011). With this, they often serve as an important element of impression formation (Fuertes et al. 2012) and seem to be more influential in evaluation processes than other "ethnic cues", such as the name or looks (Rakić, Steffens and Mummendey 2011; Rödin and Özkan 2011).

Prior research has shown that speakers with a foreign accent are often perceived more negatively: they are rated as less pleasant to listen to, as less intelligent, less competent, and as being of lower social status compared to individuals who speak the majority language with a standard accent (Fuertes et al. 2012; Gluszek and Dovidio 2010). Findings from laboratory experiments on the effects of foreign-accented speech in the workplace likewise reveal that a foreign accent usually leads to less favorable assessments. Negative evaluations result for ratings of job suitability, assessments of job-relevant attributes, hiring recommendations and decisions, and the likelihood

of promotion (Carlson and McHenry 2006; Deprez-Sims and Morris 2013; Hansen et al. 2014; Hosoda and Stone-Romero 2010; Hosoda, Nguyen and Stone-Romero 2012; Huang, Pierce and Frideger 2013; Segrest Purkiss et al. 2006; Timming 2017).

Experimental studies are well suited for establishing whether accented speech yields such outcomes; it remains, however, a challenging endeavor to address the underlying processes. One strand of arguments suggests that immigrants with a foreign accent are more difficult to understand than individuals who speak with a standard accent (Deprez-Sims and Morris 2013; Munro and Derwing 1995). In these instances, productivity considerations could be responsible for a less favorable assessment. Another strand of arguments focuses on discrimination. In this line of reasoning, a foreign accent is seen as a salient characteristic, which can trigger adverse beliefs and distaste resulting in a penalty (e.g. Horr, Hunkler and Kroneberg 2018). Both processes, in principle, can be at work simultaneously.

This study contributes to current research by taking a closer look at these arguments. In addition to analyzing how foreign accented applicants fare in early stages on the hiring process, we examine whether these speech cues influence employers' decisions via productivity considerations or if they are a source of bias that lead to discrimination.

Based on a field experiment, we aim at providing insights in the ways in which a foreign accent alters the labor market prospects of immigrants in the initial stage of the hiring process when applicants get into first contact with employers. The initial contact frequently consists of a call in which applicants inquire about an advertised position. This interaction is entirely speech-based, so that a wide range of potentially confounding factors accompanying face-to-face interactions, such as appearance, are out of the picture. In addition, in this purely verbal form of contact, an easily perceptible language cue such as a foreign accent may serve as an especially prominent signal, which can influence initial assessments of applicants. As a result, a caller might be turned down immediately; or, if the candidate is encouraged to send an application, the initial evaluation may still have a lasting impact on the subsequent hiring process. Finally, this form of a first short exchange involves little effort on the part of employers, and the risk of turning down a real applicant due to the prospect of hiring one of the experimental testers seem, at least at this early stage, negligible.

Most research on the effects of foreign-accented speech in the workplace consists of laboratory experiments conducted in the United States (e.g., Carlson and McHenry 2006; Deprez-Sims and Morris 2013; Hosoda and Stone-Romero 2010; Hosoda et al. 2012; Huang et al. 2013; Segrest Purkiss et al. 2006; Timming 2017). Some laboratory studies have also addressed accent effects for immigrants in European destinations (e.g., Hansen et al. 2014; Rödin and Özcan 2011).

Evidence on accent-related labor market outcomes stemming from field experiments, however, is scarce. In contrast to the numerous correspondence tests on ethnic discrimination in hiring decisions, which signal an applicant's ethnicity by presenting distinct names (see e.g. the meta-analysis of Zschirnt and Ruedin 2016), the consequences of speaking with a foreign accent have hardly been examined in real-life workplace settings. In addition, the research designs of the few available field experiments do not allow separating speech influences from those of other attributes such as the person's name or appearance (e.g., Larja et al. 2012; United States General Accounting Office 1990).

Moreover, in laboratory experiments, accents can be implemented under ideal listening conditions without distracting noises or other kinds of interferences, while in everyday situations

these conditions are often less than ideal. Thus, also for reasons of external validity, it seems beneficial to complement existing laboratory studies with a field experimental design.

We contribute to the literature by studying the consequences of having a Turkish accent in the German labor market. Immigrants from Turkey constitute one of the largest migrant group in Western Europe. Today, they make up substantive proportions of the migrant populations in a number of European countries such as Austria, Belgium, Germany, the Netherlands, or Switzerland. About five million people of Turkish descent are living in Western Europe: of these, around 3.5 million are in Germany (Guveli et al. 2016). This group is struggling with difficulties in the labor market, mostly due to their low socioeconomic background (Kalter and Granato 2007), which resulted from the recruitment of low-skilled workers from rural areas with limited educational opportunities (Crul and Vermeulen 2003). Destination-language proficiency is central to immigrants' labor market success (e.g., Chiswick 2009; Dustmann and Fabbri 2003) and plays an important role in accounting for the disadvantages Turkish immigrants and their descendants experience in Germany (Kalter 2006). A foreign accent as one component of speaking skills may thus contribute to this migrant group's less favorable outcomes.

In addition, Turks have been considered to be "one of the toughest groups to integrate" (Crul and Vermeulen 2003, 970), and ongoing debates about the alleged unwillingness of the Turkish but also of other Muslim groups to adapt (Diehl, Fischer-Neumann and Mühlau 2016, 243) attest to bright boundaries that characterize the European context (Alba 2005; Foner and Alba 2008). In fact, social distances on behalf of the majority populations towards Turks as well as Muslims in general are quite pronounced throughout Western Europe (e.g. Strabac and Listhaug 2008), and individuals of Turkish origin seems to experience more discrimination than individuals from other migrant groups (Diehl et al. 2016; Zschirnt and Ruedin 2016). The few correspondence tests conducted in Germany further support this notion: they reveal that applicants with a Turkish name have a lower average probability of a callback or a response (Kaas and Manger 2011; Schneider, Yemane and Steinmann 2014).

We employ data from a field experiment, which covers the initial phone contact between applicants and employers. Trained individuals, so-called testers, inquire about a position that was just announced online or in a newspaper. They ask whether the job is still available. The response to this question is the key outcome of our study. In these short queries, we manipulated two signals of ethnic origin: the accent (Turkish accent/standard German) and the name (Turkish/German). As we employed a 'matched-guise technique', the same tester placed three phone calls to each ad – one with each treatment. In as far as we succeeded in selecting and training testers who differed exclusively in these attributes, variation in employers' responses can be traced back to differences in these stimuli (Blank, Dabady and Citro 2004; Pager 2007). In addition to varying applicant characteristics, information on job and firm characteristics was collected. This included information on the communicative demands or the extent of customer contact on the advertised job, as well as information on company orientation or the structure of the recruitment process. Moreover, from the Federal Employment agency and the German General Survey, context information was taken and merged to the experimental set. This included information on regional labor supply, job-specific vacancy rates or regional levels of anti-immigrant attitudes. The aim behind this procedure was to allow for tentative inferences on the reasons behind any accent effects.

Theoretical considerations

In order to address the question of why foreign-accented speech might reduce the likelihood of receiving a positive response in the early hiring process, we discuss two lines of arguments: The first centers on employers' productivity considerations, while the second refers to discrimination.

Foreign accents and productivity

Language skills can be viewed as a form of human capital (Chiswick and Miller 1995, p. 248): they are embodied in the person; their acquisition requires investments that involve costs; and they are productive in that they can be used to increase future benefits such as returns on the labor market. According to this reasoning, applicants who are fluent in the dominant language should be more productive than applicants who are not fully proficient.

Pronunciation enters the picture as one component of language proficiency. Individuals who speak with a foreign accent can be more difficult to understand than individuals who communicate with a standard accent (Deprez-Sims and Morris 2013; Munro and Derwing 1995). Empirical studies show that non-native accents affect the degree to which a listener understands a message (Floccia et al. 2009; Deprez-Sims and Morris 2013; Munro and Derwing 1995; van Wijngaarden 2001). Their presence can lead to a delay in word identification (Floccia et al. 2009), promote the misidentification of syllables, words, and sentences (van Wijngaarden 2001), and prolong the time it takes to process utterances (Adank et al. 2009; Munro and Derwing 1995). A foreign accent thus reflects a person's ability to communicate effectively (Deprez-Sims and Morris 2013, 356). Similar to the reasoning of economists, linguists in this context discuss the "costs" of speaking with a foreign accent (Adank et al. 2009; Munro and Derwing 1995). In the hiring process, employers may price in these perceived costs. As a result, they will reject applicants with a foreign accent more often than they will reject candidates who speak with a standard accent.

However, jobs typically differ in the linguistic skills they require. While for some occupations communicative demands are high (e.g., for teachers or call-center agents), other positions involve tasks, in which language proficiency is by far less relevant (e.g., for cleaners or IT-specialists). *In line with the productivity argument, we expect that in the early hiring process employers' response to applicants with a Turkish accent who call to inquire whether a vacancy is still open is conditional on the communicative demands of the job in question*. That is, employers might turn down foreign accented applicants more often, when the advertised job requires strong communicative skills, while being less prone to do so for jobs with lower communicative demands.

Accents and statistical discrimination

The productivity argument is based on the assumption that employers can reliably estimate the communication costs of (potential) applicants. However, especially in early hiring stages, such as the initial telephonic contact, insecurities and information deficits on any assessment should be particularly high. At this point, no standardized information, such as credentials or reference letters, is available. Models of statistical discrimination assume that in these situations, employers rely on easily accessible sources of information, such as an indicator of group membership (Arrow 1973; Aigner and Cain 1977; Phelps 1972). During the initial telephonic contact, this information should be particularly easy to obtain as the name is the first information an employer receives. For many immigrant groups, and particularly the Turkish group in the German context, a foreign name should be associated with lower proficiency in the host country language (Kalter 2006; vanTubergen and Kalmijn 2005) and the additional cue of accented-speech may confirm

this assumption. However, as majority employers should be less familiar with foreign accented speech, accents, for them, may not be reliable indicators for general linguistic abilities and result in an overestimation of communication costs. Employers may therefore reject Turkish accented applicants, even though understandability and language skills may be adequate for the position in question. The few available studies in the German context give evidence for statistical discrimination in the labor market (Kaas and Manger 2010) that may be accent-related (Horr et. al. 2018) and it has been argued that in competitive markets as such, this form of discrimination should be the most likely (Esser 2006). Additionally there is some evidence, that unfamiliarity with (foreign-) accented speech can result in less accurate judgements and decision-making (Adank et al. 2009), however, how this relates to applicant evaluations and the estimation of communication costs has not yet been investigated. Against this backdrop, the conditionality of accented speech on the communicative demands of an advertised job might at the initial hiring stage also reflect a form of statistical discrimination.

Foreign accents and discrimination

Tastes for discrimination

Foreign-accents, at the same time, saliently signal ethnic origin, which may trigger prejudice and in-group favoritism within majority members (Deprez-Sims and Morris, 2010). Models of taste-based discrimination consider prejudice and tastes for certain groups as the main motive for discriminatory behavior (Becker 1971). They build upon the idea that members of certain ethnic groups are preferred over members of other groups (Becker 1971). In order to act in accordance with their preferences, individuals are willing to pay or give up something of value to have something pleasant or avoid something unpleasant (ibid.). On the labor market, employers factor in the psychological costs of having to associate with someone or some group they dislike: they either pay members of this group less or avoid hiring them altogether (ibid.).

In the German context as well as in other Western European countries, distastes towards individuals of Turkish origin seem to be widespread and pronounced (Degner and Wentura 2010; Steinbach, 2004). *Therefore, we expect that, compared to applicants from the majority population, applicants with a Turkish accent will be told more often that an advertised position is no longer available.*

The reasoning on productivity, however, yields the same prediction. That is, individuals with a Turkish accent are more difficult to understand and therefore may have a higher chance of being turned down in an initial phone conversation. Accordingly, to distinguish between productivity considerations on the one hand and taste-based discrimination on the other hand, we build upon additional arguments. In the case of productivity, we suggested above that foreign accents are more important for positions with high communicative demands.

Regarding taste-based discrimination, in a similar manner, we now consider a series of moderating conditions, under which employers who have discriminatory preferences can be expected to be less likely or, vice versa, more likely to act in accordance with their tastes. The general reasoning underlying the subsequent discussion of these conditions is twofold: if the means for the rejection of candidates of certain origins are restricted or if the pursuit of discriminatory preferences incurs substantive costs, rational employers will adapt their behavior accordingly. Taste-based discrimination, consequently, should be less severe in situations, which provide fewer opportunities for discrimination and in which discrimination is an increasingly costly endeavor.

Opportunities and taste-based discrimination

Opportunities for discrimination are reduced in workplace settings, in which the recruitment process follows predefined rules*, because standardization leaves less scope for individual tastes to influence applicant selection* and thus limits the room for preferences to become effective (Kaas and Manger 2010). Typical measures of standardization include the implementation of tests, the appointment of specialized human resource personnel or the separation of the search for a suitable candidate from the final hiring decision. Usually, larger firms with more vacancies have established application processes that follow such predefined processes (Kaas and Manger 2010). Therefore, recruiters should here be less prone to taste-based discrimination.

A similar reasoning can be applied to external recruitment agencies whose business is the search for employees on behalf of other companies. These agencies benefit from providing their customers with candidates who perform well on the job. The supply of suitable employees thus plays an important role. Prior research has shown that the use of standardized procedures is an important and effective tool in personnel recruitment (Schmidt and Hunter 1998). Consequently, we expect external recruitment agencies to rely on standardized methods on a regular basis. As argued above, the implementation of such predefined proceedings restricts the opportunities for the pursuit of individual preferences for or against certain ethnic groups.

Moreover, companies, which promote diversity, may implement diversity-training programs, which address group-based attitudes and prejudices. Empirical evidence suggests that diversity training in organizational settings is effective in improving overall attitudes towards diversity, while the results are less consistent when addressing specific groups (Kulik and Roberson 2008). In general, diversity strategies could increase the sensitivity to equal treatment and provide measures, which reduce the scope for discrimination in hiring situations.

The costs of taste-based discrimination

In addition to opportunities, the costs discrimination incurs can vary across contexts and so should the extent of discriminatory behavior. Labor supply may serve as an important example. That is, if the supply of applicants for a certain occupation is limited in a country or region, the costs of hiring someone who, in addition to the required qualification, matches group-specific tastes are greater than they are in a less tight labor market (Baert et al. 2015; Birkelund, Heggebø and Rogstad 2017). In other words, in times of labor shortages, employers do not have much leeway for discrimination (Becker 1971). In these instances, the penalties for a foreign accent should be smaller than in contexts without recruitment difficulties.

Regional differences may also come into play. In environments, in which prejudice and negative attitudes towards certain immigrant groups are widespread, employers are more likely to hold such views themselves. More importantly, they have reason to assume that they will not be sanctioned when following their preferences, while the reactions of others might look different in environments with less prejudice. Evidence from a field experiment underlines this reasoning: in regions with more negative attitudes, there is more hiring discrimination (Carlsson and Eriksson 2017).

In Germany, anti-immigrants attitudes, in particular towards individuals of Turkish origin, are much more widespread in the East than in the West (Boehnke, Hagan and Hefler 1998; Wagner et al. 2003). At the same time, there is also substantive variation in these sentiments within the various regions in the East and in the West likewise. Irrespective of general location, the cited reasoning should apply. That is*, in regions with more pronounced anti-immigrant attitudes, we expect*

that a caller with a Turkish accent will be told more often that the position is no longer available than in a region, in which negative sentiments are less severe.

Customer preferences

In addition to acting upon their own preferences, employers might factor in their customers' tastes when hiring new employees (Becker 1971). If the current or prospective clientele is known to have preferences against certain groups, they may be inclined to avoid companies, which force them to engage with members of these groups. Employers might thus pay the workers in question less or not hire them at all.

A foreign accent is a very strong signal of group membership and, following Becker's (1971) considerations on customer preferences, may be a relevant factor in employers' considerations about the preferences of their clientele. Findings from numerous studies consistently reveal that listeners give higher ratings to products advertised by a standard-accented salesperson (Hill and Tombs 2011; Hendriks, van Meurs and van der Meij 2014; Morales, Scott and Yorkston 2012). Individuals with a foreign accent were also rated higher in non-costumer-facing jobs than in customer-facing jobs (Timming 2017). *We therefore expect that speaking with a Turkish accent will be perceived more negatively when the job in question involves customer contact.*

However, in addition to considering customers' tastes, productivity considerations may also become more relevant when filling jobs with customer contact. As pointed out before, it may be more challenging to understand speakers with a foreign accent. The difficulty can hamper message processing, which in turn may result in lower customer intentions to buy or recommend a product or a service (Mai and Hoffmann 2014). If employers consider how well their clientele can understand and process a message presented by a foreign-accented person, a negative response to such speakers is an increasingly likely outcome.

When filling positions with customer contact, employers' dismissive reactions to an accented applicant may thus represent both, perceived customer preferences or more productivity-related considerations of how well customers will understand the applicant.

Methods

Field experiment

We used a field experimental design to address the consequences of speaking with a Turkish accent in the early hiring process. In a short phone conversation testers inquired about a job that was just advertised. The key outcome of the present study is the response to the question of whether the position was still available. In addition to this central query, the phone call included two follow-up questions, which were used to disguise similarities between the conversations.

For each ad and within two weeks after the position was offered, the same tester placed three phone calls, in which the applicant's name (Turkish/German) and accent (Turkish accent/standard German) was manipulated. As the simultaneous presence of a German name with a Turkish accent is an unlikely combination (Lorenz 2017), calling with these attributes might create astonishment and produce an awkward situation. We therefore exclude this condition from the design. We used male and female testers and, in our analyses, control for this attribute.

To ensure that calling in response to an ad before sending an application is common in Germany, we conducted a small preliminary study. We used job announcements from different online job

sites and national newspapers, called the numbers provided on the ads and conducted 90 semi-structured interviews with the employers who had offered the positions. 72% of them indicated to have received at least one call in response to their job announcement. On average, they had received 11 calls per offer. We therefore concluded that this way of making contact is rather common and well established in the German context.

Moreover, these interviews revealed that the primary reason for calling, indeed, consists of the question of whether the job is still available. In addition, based on this preliminary inquiry, we were able to identify further questions and topics, which are typical for such conversations, such as salary, additional information on the hiring process or details about the advertised position. We used this information for our follow-up questions.

Manipulations

Following common linguistic definitions, a *foreign accent* primarily reflects in a difference in phonetic characteristics, such as articulation, intonation, prosody or stress (Lippi-Green 1994). In line with this notion, we restricted our manipulation to phonetic features, while other linguistic aspects such as grammar, syntax or vocabulary remained intact. To ensure that all elements of speech other than those that constitute the Turkish accent are comparable across the two experimental conditions, we used testers who were able to speak standard German as well as with a Turkish accent. This matched-guise technique (Lambert et al. 1960) allows holding constant other speech elements such as frequency or pitch, which also contribute to listeners' perceptions of a speaker.

We took several steps when selecting among candidates who were interested to work as a tester in our study. Almost all of the respective candidates were actors or acting students. Each applicant (n=30) was asked to read the same short text, one with a Turkish accent and one with a standard German accent. We recorded these audio samples and presented them to a sample of students (n=276). They assessed the accents as well as their vocal, personal and professional appeal. Based on these ratings, we preselected a set of candidates. In a second step, we presented the speech samples of these potential testers to linguistic experts who study foreign accents in the German context. They assessed the genuineness and credibility of both speech variants and additionally classified the accent strength. Based on their evaluations, we eventually chose four testers, even though initially we had planned for more. The reason for this reduced number of testers was that, in the course of our selection process, it became evident that suitable candidates who fulfilled the main requirement of being able to speak both accents in a credible manner are scarce. The testers we eventually chose were able to produce both accents at the same speed in a natural and credible way; they received comparable ratings for their vocal and personal attractiveness; and they were considered as professionally competent. All testers spoke with a medium to heavy accent. This, for instance, included a different pronunciation of the E-, Z-, and R sound, a general vocal reduction or the (over)pronunciation of the last syllable.

Our second signal of ethnic origin, the *name*, is the characteristic that has been most commonly used in field experiments on hiring discrimination (Zschirnt and Ruedin 2016). The name is an easily accessible cue, especially in written applications. On the phone, conversations usually start as well by providing the name. By randomly assigning different combinations of the two key attributes, the name and the accent, it is possible to separate between influences that are due to the presence of a Turkish accent and influences that result from the signal set by the use of a Turkish versus a German name.

In order to identify suitable names, we used the most common Turkish (Rodríguez 2010) and German (Digitales Familienwörterbuch Deutschlands 2018; Rudolph, Böhm and Lummer 2007) first and surnames present in the German population. From these lists and for each language, we took 6 first names for females, 6 first names for males and 8 last names. We then asked a small sample of students (n=71) to rate these names according to different attributes and, in addition, to write down any associations they had for each name. Based on these assessments, we selected names that allowed for an unequivocal identification of the two ethnic groups, that were perceived as timeless and which, within each group, ranged average on attributes such as intelligence or attractiveness. When not all criteria were met, clear indication of origin was prioritized[1].

Sampling and response rates

The sample is based on job advertisements from all over Germany, for any occupation and for any firm that offered a vacancy during that period. We covered two of the most common job advertising channels, namely, online job sites and newspapers. First, we included three prominent and frequently visited online websites[2]. In addition, we considered two national newspapers with high circulation and well-known sections of job offers[3]. Finally, we took a random sample of regional newspapers with high and low coverage from 15 different regions[4]. With this latter selection, we aimed at covering ads from smaller firms, such as craft enterprises, which frequently announce jobs in local media.

Taking into account that the number of ads in online media by far exceeds the number of ads in regional newspapers (von Stetten et al. 2013), which in turn exceeds the number of ads in national newspapers (Stepstone 2013), we sampled from these channels with a ratio of 6:3:1[5]. We divided the online ads into two strata: positions, which require high versus low qualifications and positions, which have high versus low communicative demands. From each strata, we sampled an equal number of ads. This step was necessary, as some website interfaces emphasize certain job categories and therefore present more ads for these than for other job types. From the newspapers, we drew random samples.

The field experiment was carried out between October 2014 and October 2015. During this period, in each week we identified all new ads that were published in any of the three channels. From these lists, we drew our weekly samples. In total, the sampling frame consisted of 8.690.791 online ads, 59.881 ads from regional newspapers, and 5.750 ads from the two national newspapers. We sampled 1.385 job announcements and were successful in contacting 719 firms (51.9%). As we planned three conversations for each offer, the number of phone calls that, in the ideal case, could have been reached, amounts to 2.157 calls. We realized 1.835 (85.1%) of them.

In 62.3% of all cases, the testers were able to make the intended number of calls for each of our three treatments: (1) Turkish name/Turkish accent; (2) Turkish name/standard German; (3) German name/standard German. For 35.6% of the ads, they managed to complete only one or two calls. The small remaining proportion (2.1%) consists of misplaced calls, for which testers by

---

1 We selected the following names: (1) Turkish first names: Ayşe, Hatice, Mehmet, Mustafa; (2) Turkish surnames: Yıldız, Yılmaz, Çelik, Özcan; (3) German first names: Claudia, Susanne, Christian, Matthias; (4) German surnames: Meyer, Müller, Schneider, Wagner.

2 http://jobboerse.arbeitsagentur.de; http://jobs.meinestadt.de; http://www.monster.de.

3 Süddeutsche Zeitung, Frankfurter Allgemeine Zeitung.

4 Flensburger Tageblatt, Ostseezeitung, Rheinische Post, Rheinpfalz, Berliner Zeitung, Stuttgarter Nachrichten, Augsburger Allgemeine, Fränkische Landeszeitung, Sächsische Zeitung, Leipziger Volkszeitung, Hildesheimer Allgemeine, Lippische Landeszeitung, Oberhessische Presse, Esslinger Zeitung, Oranienburger Generalanzeiger.

5 We also counted the number of ads within each of these channels prior to our experiment, to ensure the general gradation.

mistake called more than three times. As not fully completed sets and misplaced calls were randomly distributed across the different tester profiles, i.e. the different treatments, we included all calls in the analyses.

Information obtained from additional data sources

In addition to the field experiment, we used data from three additional sources: from (1) the job announcements and the homepage of the advertising firm to capture job and firm characteristics, and from (2) the Federal Employment Agency (FEA), and (3) the German General Social Survey (ALLBUS) to capture context characteristics.

Job and firm characteristics

To add information on relevant job characteristics, we used the job titles and the texts of the job announcements. They allowed us to specify job requirements as well as characteristics of the job. Moreover, we visited the companies' homepages to obtain relevant information on firm characteristics. In some cases, when we could not find the relevant information, we called and asked for certain attributes.

The *position's communicative demands* were captured on the one hand by the descriptions in the respective ads. Based on the texts of all ads that were included in our final data set, two coders compiled a list of 276 words, which indicate a communication requirement (e.g., negotiating or consulting). In a second step, these words were rated by a student sample (n=217) on an eleven point scale as to how strongly the expression signals the need for communication skills. We then weighted each word with the standardized median value of the student ratings so that words, which were perceived as powerful signals of communicative demands, obtain a higher value. In a final step, for each ad, we calculated the weighted share of these words. A higher value on this variable indicates that the position involves more communication tasks.

On the other hand, we considered the qualification that is usually necessary for the occupation in question. For this purpose, we assigned to each of the advertised job titles the required educational qualification using the International Standard Classification of Education. For the analyses, we distinguish between three levels, that is, low (ISCED 0, 1 and 2), medium (ISCED 3 and 4) and high (ISCED 5 and 6).

With three additional measures, we aimed at considering important aspects related to the *standardization of the recruitment process*. Firm size refers to the numbers of employees. Following common approaches, we distinguished between smaller (less than 50 employees) and larger (50 or more employees) firms. In addition, we took into account whether the job advertisement was launched by an external recruitment agency. Finally, we considered whether the firm had a diverse orientation. We assigned a value of 1, if the firm mentioned diversity strategies or programs on its homepage.

Customer contact was measured by applying a similar proceeding as for the communicative tasks. That is, two coders first compiled a list of 72 expressions from the ads, which indicate customer contact. Each word was then rated on an eleven point scale by a student sample (n=217). They indicated how strongly the expression signals that interactions with customers are part of the job. For each ad, we used these ratings to calculate the weighted share of words pointing to customer contact. A higher value on this measure implies that the position in question involves more contact with customers.

Context characteristics I - Federal Employment Agency

Furthermore, we merged data from the Federal Employment Agency (FEA)* to portray the region's labor supply. For each federal state, we calculated on a monthly basis the share of vacant positions in the employed population considering four levels of vocational qualification requirements, which capture the degree of complexity of an occupation, for each of 37 occupational main-groups (Paulus and Matthes 2013). We then assigned to each advertised position the relevant vacancy rate at the time of the experimental call. Higher vacancy rates point to a larger number of open positions and consequently to labor shortages.

High unemployment rates, in contrast, imply that there is more labor supply than the market can absorb. Information on unemployment is available on a monthly basis at the district level. Accordingly, we assigned the regional unemployment rate that was present at the time of the phone call to each of the job advertisements included in our study. As the FEA considers 401 districts in total, the unemployment measure is regionally more fine-grained than the vacancy rate, which is documented at the federal state level. However, in contrast to the vacancy rates, the unemployment rates are not specific to certain occupational categories or vocational requirements.

Context characteristics II - ALLBUS

Finally, to capture anti-immigrant attitudes in the regional context, we used data from the German General Social Survey (ALLBUS)** from 2010, 2012 and 2014. With four items in the years 2010 and 2012 and eight items in 2014, the surveys record sentiments towards ethnic minorities in Germany. For each respondent, we calculated the standardized mean over these items. We then aggregated the individual values for the three survey years across federal states. The resulting average captures how strong anti-immigrant attitudes are in the respective region. Higher values point to increasingly negative feelings towards ethnic minorities.

Table A.1 in the Appendix illustrates the distributions of our model variables. It also covers controls such as the tester's gender, the gender of the person who was reached on the phone, the ethnic origin of the person contacted (i.e., Turkish, German, other/unknown) and the number of words included in the ad.

Analyzing strategy

To examine the consequences of speaking with a Turkish accent in the initial stage of the hiring process, we estimate binary logistic regression models. Subsequent to analyzing how foreign accented applicants fare in early stages on the hiring process, we address the theoretical arguments via interaction effects. That is, we investigate, whether the extent to which a Turkish accent affects employers' responses in an initial phone conversation is conditional upon characteristics, which point to productivity considerations on the one hand, and which relate to taste-based discrimination on the other hand.

The design of our field experiment generates clustered data, as the testers placed up to three calls to each job advertisement. These ads, in addition, are nested in regional entities, which are relevant when considering local variation in labor supply and anti-immigrant attitudes. We consider the clustering by estimating multilevel-models (see Appendix, Table A2 for variance components of the unconditional ML model).

In the multivariate account, we deal with missing values by employing multiple imputation (Allison 2001). We run regression models for each of the 25 imputed datasets and then combine the estimates following Rubin's (1987) rule.

Results

Descriptive Overview

To address the first part of the research question of how foreign accented applicants fare in the initial stage of the hiring process, we investigate whether and to what extent applicants were told that the advertised position is still available. Table 1 tabulates the general distribution of employer replies; Table 2 indicates the probability of being able to partake in the selection process in a basic multilevel model. Both indicate the same pattern. Turkish applicants are told more often that an advertised position is already filled and are thus excluded from the selection process. This only applies however, when the applicant is identifiable by accent.

The general overview of employer replies (Table 1) suggests a hierarchy of positive availability replies. German applicants have the highest share of positive answers and are most often told that an advertised position is still available. For unaccented Turkish applicants, the distribution is almost comparable. The candidate calling with a Turkish accent lags behind both of the other applicants. The basic multivariate model (Table 2) confirms the overall pattern. For Turkish applicants, the probability of getting a positive availability reply is generally lower than for German applicants (Model 1, Model2) and this holds particularly true, if the applicant is identifiable by his accent (Model 3). Calling with a Turkish accent reduces the probability of being able to enter the hiring process by approx. 6 percentage points, irrespective of the applicant's name (Model 3). Having a Turkish name, does not seem to matter over and above accented speech (ibid.).

[Table 1]

[Table 2]

Multivariate Analyses

To examine why foreign accented speech affects the (initial stage) hiring prospects of immigrants, we estimate binary logistic multilevel models. In the basic model we regress the opportunity to partake in the hiring process on the manipulated applicant characteristics and on crucial job-, firm-, and context characteristics (Model 1, Table 3). Starting with Model 2 (in Table 3), we estimate interaction effects to test our Hypotheses.

Results indicate that the effect of foreign accented speech is not conditional on the communicative demands of the job in question (Model 2a, Model 2b), which suggests that an accent may not be considered in terms of productivity. If employers took accent related communication costs into account when deciding on whom to include in the hiring process, the influence of foreign accented speech should be conditional on the communicative demands of the advertised position. This reasoning also applies to statistical discrimination. If employers took accented speech as an indicator for language proficiency, the same conditionality should manifest.

There are certain indications however, that foreign accented speech may trigger taste discrimination in employers. Firstly, employers' *opportunities to act upon their preferences* seem to

matter. For instance, Model 3a indicates that a Turkish accent is particularly negative when applying for openings in smaller firms. Here the probability of receiving a positive reply for accented applicants is about eleven percentage points lower than the probability of unaccented callers. This negative effect is considerably reduced however, if applicants apply to larger firms instead (see also Fig.1). The same holds true for company focus (Model 3b, Fig.2) and firms that rely on external recruiters (Model 3c, Fig.3) As larger, diverse, and externally recruiting firms should employ more standardized recruitment processes, these findings might reflect employers' reduced opportunities to act upon their tastes. Secondly, *certain costs associated with acting upon personal preferences* seem to matter. Model 5 indicates that in regions with more negative attitudes towards immigrants, a Turkish accent considerably reduces the hiring prospects of Turkish accented applicants. If accented candidates apply for vacant positions in regions with more positive attitudes towards immigrants however, these disadvantages largely disappear (see also Fig.4). This could reflect the distribution of employer tastes per se as well as the anticipated social sanctions, when acting upon these preferences. However, other factors that should increase the costs of discrimination, such as the regional and sector specific labor supply (Model 4a, 4b) do not seem to matter.

Customer preferences or more productivity-related considerations of how well customers will understand the applicant also seem to be irrelevant for employers' decision making (Model 6).

[Table 3]

[Figures 1 – 4]

Summary & Discussion

We conducted a field experiment to investigate whether and to what extent the presence of a foreign accent can alter the early hiring prospects of immigrants and how accent induced disadvantages can be explained. We were primarily interested in whether foreign accents affect employers' perceptions in terms of productivity or if they are a potential source of bias triggering discrimination. Results indicate that whereas a Turkish name does not seem to affect the hiring prospects in a particularly negative way, there are significant differences in hiring success when an applicant is identifiable by accent. Independent of other applicant characteristics, a Turkish accent reduces the probability of partaking in the hiring process by about 6 percent. There are several indications that these disadvantages may be linked to employer tastes. First, the effect of foreign accented speech is not conditional on the communicative demands of the job in question, which suggests that an accent may not be considered in terms of productivity. Second, factors that shape employers' opportunities to act upon their tastes, such as a more or less standardized hiring process, moderate that the relationship between accented speech and hiring success; and third accented applicants face stronger disadvantages in regions, where ethnic aversions cluster and acting upon such distaste should be perceived as less costly. However, additional factors that might determine the costs of factoring in discrimination tastes, such as the labor supply, do not seem to matter. Customer preferences or more productivity-related considerations of how well customers will understand the applicant also seem to be irrelevant for employers' decision making.

Contrary to other studies in the German context, we do not find an independent effect of having a Turkish name. An explanation for this could be that all other studies employ correspondence

tests, where applicant characteristics are presented in formal writing (e.g. Kaas and Manger 2010). Certain evidence suggests that written information impacts cognitive processes, and particularly (short-term) memory functions, much stronger than orally delivered information (e.g. Daniel and Woody 2010). In our case, an employer may just not process the 'foreignness' of the name as much as he would when he sees it in writing. In addition to that, research demonstrates that verbal cues and particularly accented speech are dominant in impression formation (Rakic' et al. 2010; Kinzler et al. 2009). As in our case, applicants only identifiable by Turkish name, speak the German standard, the standard cue in speech may just overwrite the foreignness of the name (Rödin and Özcan 2011).

One *limitation* of our study is that we could not disentangle actual accent-related productivity considerations of employers from forms of statistical discrimination. To examine which of these processes may be of greater relevance, we merged additional data on the share of immigrants in the community of the hiring employer (INKAR data; BBSR Bonn [2018])**** to our set. The main argument here would be that everyday interactions between employers and immigrants should be more common in communities with a larger share of immigrants. Next to more positive attitudes, more frequent interactions also promote accent familiarity, which fosters a more accurate assessment of the communication costs that are associated with foreign accented speech. In other words, in regions with a higher share of immigrants, foreign accents might be a more reliable indicator for communication costs. Related forms of statistical discrimination should thus be less likely. To test such arguments, we estimated a three-way interaction between the applicants' accent, the communicative demands of the advertised job and the share of immigrants in the respective community. Results revealed no significant differences in the role of communicative demands for accented callers between regions with different shares of immigrants (supplementary materials, Model 1, Table I), which might point to a minor role of accent-based statistical discrimination. Yet, as the communicative demands of a job per se do not seem to matter for accented callers (insignificance of the two-way interaction in the main models), a more substantial interpretation of the above mentioned results might not be fruitful.

Similarly, we could not reliably untangle perceived customer preferences from more productivity-related considerations of how well customers will understand the applicant. To examine which of these arguments may be of greater relevance, we also estimated additional models. We assume that if employers primarily act upon customer preferences, accented applicants should be turned down more often for customer facing jobs in regions with higher levels of anti-immigrant sentiments. We again estimated a three-way interaction between the applicants' accent, the amount of customer contact within the advertised position and the regional level of anti-immigrant sentiments. Coefficients suggest that regional attitudes are less relevant, when employers evaluate job seekers for jobs with customer contact. While this, in tendency, suggests that productivity-related considerations of how well customers will understand the applicant may be of greater relevance, the three-way interaction effect fails to reach significance (supplementary materials, Model 2, Table I). However, as the customer contact on the job per se do not seem to matter for accented callers (insignificance of the two-way interaction in the main models), here, too a more substantial interpretation might not be fruitful.

Another limitation relates to our measure of the regional attitudes towards immigrants. This indicator is somewhat coarse as it aggregates the respective attitudes across federal states. To rule out ecological fallacies, we also estimated the model when aggregating on a lower level (planning regions, n=96)***. The general trends remained comparable, yet the interaction effect failed to reach significance. This could be due to the smaller number of participants within each individual

planning regions (< 20 respondents in 18 regions), or to the somewhat artificial distinction that is behind the respective regions. The planning regions represent a tool of city development and urban planning and aren't necessarily congruent with the actual living conditions and attitudes of the respective respondents.

Despite these limitations, our results suggest that foreign accented speech may trigger discrimination in employers and alter the hiring prospects of immigrants. This may be taken into account when interpreting the effect of language proficiency on labor market outcomes.

Data sources

* Federal Employment Agency (Bundesagentur für Arbeit). We therefore considered the following public access data: Arbeitsmarktdaten nach Zielberufen: Arbeitslose, Arbeitsuchende und gemeldete Arbeitsstellen, Oktober 2014 – November 2015 via https://statistik.arbeitsagentur.de/nn_31894/SiteGlobals/Forms/Rubrikensuche/Rubrikensuche_Form.html?view=processForm&resourceId=210368&input_=&pageLocale=de&topicId=287986&year_month=201504&year_month.GROUP=1&search=Suchen, as well as specific preparations by the FEA statistical office (Zentraler Statistik Service, BA, Arbeitsmarktberichterstattung, Nürnberg) on employees and registered vacancies by occupational group (monthly overview, oct 2015-nov 2015).

**Allgemeine Bevölkerungsumfrage der Sozialwissenschaften (ALLBUS):

- GESIS - Leibniz Institute for the Social Sciences (2011): ALLBUS/GGSS 2010 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2010). GESIS Data Archive, Cologne. ZA4610 Data file Version 1.1.0, https://doi.org/10.4232/1.10760
- GESIS - Leibniz Institute for the Social Sciences (2013): ALLBUS/GGSS 2012 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2012). GESIS Data Archive, Cologne. ZA4614 Data file Version 1.1.1, https://doi.org/10.4232/1.11753
- GESIS - Leibniz Institute for the Social Sciences (2018): ALLBUS/GGSS 2014 (Allgemeine Bevölkerungsumfrage der Sozialwissenschaften/German General Social Survey 2014). GESIS Data Archive, Cologne. ZA5240 Data file Version 2.2.0, https://doi.org/10.4232/1.13141

*** We thereby drew upon detailed regional information for the ALLBUS data, which is accessible at the Secure Data Center (www.gesis.org/en/sdc) of the GESIS Data Archive for Social Sciences in Cologne Germany. Researchers are required to sign a special usage agreement and to work within an individually tailored secure virtual workspace.

**** Indikatoren und Karten zur Raum und Stadtentwicklung (INKAR). BBSR Bonn [2018]. Datenlizenz Deutschland Version 2.0. (www.govdata.de/dl-de/by-2-0)

References

Alba, R. 2005. "Bright vs. Blurred Boundaries: Second-Generation Assimilation and Exclusion in France, Germany, and the United States." *Ethnic and Racial Studies* 28(1): 20-49.

Allison, P. D. 2001. *Missing Data*. Thousand Oaks, CA: Sage.

Adank, P., B. G. Evans, J. Stuart-Smith and S. K. Scott. 2009. "Comprehension of Familiar and Unfamiliar Native Accents Under Adverse Listening Conditions." *Journal of Experimental Psychology: Human Perception and Performance* 35(2): 520–529.

Aigner, D. & Cain, G. (1977). Statistical theories of discrimination in labor markets. *Industrial and Labor Relations Review* 30 (2): 175-187.

Arrow, K. (1973). The theory of discrimination. In: Ashenfelter, R. (ed.), *Discrimination in labor markets*. Princeton NJ: Princeton University Press, pp. 3-33.

Baert, S. B. Cockx, N. Gheyle and C. Vandamme. 2015. "Is there Less Discrimination in Occupations where Recruitment is Difficult?" ILR Review 68(3): 467-500.

Becker, G. S. 1971. *The Economics of Discrimination*. Chicago: University of Chicago Press.

Blank, R. M., M. Dabady and C. F. Citro. 2004. *Measuring Racial Discrimination: Panel on Methods for Assessing Discrimination*. Washington, D.C: The National Academies Press.

Birkelund, G. E., K. Heggebø and J. Rogstad. 2017. "Additive or Multiplicative Disadvantage? The Scarring Effects of Unemployment for Ethnic Minorities." *European Sociological Review 33*(1): 17-29.

Boehnke, K., J. Hagen and G. Hefner. 1998. "On the Development of Xenophobia in Germany: The Adolescent Years." *Journal of Social Issues* 54(3): 585-602.

Carlsson, M. and S. Eriksson. 2017. "Do Attitudes Expressed in Surveys Predict Ethnic Discrimination?" *Ethnic and Racial Studies 40*(10): 1739-1757.

Chiswick, B. R. 1995. The Economics of Discrimination: A Primer. In *The New Economics of Human Behaviour*, edited by M. Tommasi and K. Ierulli, pp. 15-26, Cambridge: Cambridge University Press.

Chiswick, B. R. 2009. "The Economics of Language for Immigrants: An Introduction and Overview." In *The Education of Language Minority Immigrants in the United States,* edited by T. G. Wiley, J. S. Lee and R. W. Rumberger, 72-91, Bristol: Multilingual Matters.

Chiswick, B. R. and P. W. Miller. 1995. "The Endogeneity between Language and Earnings: International Analyses". *Journal of Labor Economics* 13(2), 246-288.

Crul, M. and H. Vermeulen. 2003. „The Second Generation in Europe." *International Migration Review* 37(4): 965-986.

Daniel, D. B., & Woody, W. D. (2010). They hear, but do not listen: Retention for podcasted material in a classroom context. *Teaching of Psychology*, 37(3), 199-203.

Digitales Familienwörterbuch Deutschlands. 2018. *Liste der 50 häufigsten Familiennamen in Deutschland*. http://www.namenforschung.net/dfd/woerterbuch/faq, retrieved August 26, 2018.

Diehl, C., M. Fischer-Neumann, P. Mühlau. 2016. "Between Ethnic Options and Ethnic Boundaries: Recent Turkish and Polish Migrants' Identification with Germany." *Ethnicities* 16(2): 236-260.

Degner, J. and D. Wentura. 2010. "Automatic Prejudice in Childhood and Early Adolescence." *Journal of Personality and Social Psychology* 98(3): 356-374.

Deprez-Sims, A.-S. and S. B. Morris. 2013. "The Effect of Non-native Accents on the Evaluation of Applicants during an Employment Interview: The Development of a Path Model." *International Journal of Selection and Assessment* 21(4): 355-367.

Derwing, T. and M. Munro. 2009. "Putting Accent in its Place: Rethinking Obstacles to Communication. *Language Teaching* 42(4): 476-490.

Dustmann, C. and F. Fabbri. 2003. "Language Proficiency and Labour Market Performance of Immigrants in the UK." *Economic Journal* 113(489): 695-717.

Floccia, C., J. Butler, J., J. Goslin and L. Ellis. 2009. "Regional and Foreign Accent Processing in English: Can Listeners Adapt?" *Journal of Psycholinguistic Research* 38(4): 379-412.

Foner, N. and R. Alba. 2008. "Immigrant Religion in the US and Western Europe: Bridge or Barrier to Inclusion?" *International Migration Review* 42(2): 360-392.
Fuertes, J. N., W. H. Gottdiener, H. Martin, T. C. Gilbert and H. Giles. 2012. "A Meta-Analysis of the Effects of Speakers' Accents on Interpersonal Evaluations." *European Journal of Social Psychology*, 42(1):120-133.
Gluszek, A. and J. F. Dovidio 2010. "The Way *They* Speak: A Social Psychological Perspective on the Stigma of Nonnative Accents in Communication." *Personality and Social Psychology Review* 14(2): 214-237.
Guveli, Ayse, H. Ganzeboom, L. Platt, B. Nauck, H. Baykara-Krumme, Ş. Eroğlu, S. Bayrakdar, E. K. Sözeri and N. Spierings. 2016. *Intergenerational Consequences of Migration: Socio-economic, Family and Cultural Patterns of Stability and Change in Turkey and Europe*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
Hagendoorn, L. 1995. "Intergroup Biases in Multiple Group Systems: The Perception of Ethnic Hierarchies." *European Review of Social Psychology* 6(1): 199-228.
Hansen, K., T. Rakić and M. Steffens. 2014. "When Actions Speak Louder Than Words: Preventing Discrimination of Nonstandard Speakers." *Journal of Language and Social Psychology* 33(1): 68-77.
Hendriks, B., F. van Meurs and E. van der Meij. 2014. „Does a Foreign Accent Sell? The Effect of Foreign Accents in Radio Commercials for Congruent and Non-Congruent Products." *Multilingua* 34(1): 119-130.
Hill, S. R. and A. Tombs. 2011. "The Effect of Accent of Service Employee on Customer Service Evaluation." *Managing Service Quality* 21(6): 649-666.
Horr, A., Hunkler, C. & Kroneberg, C. (2010). Ethnic discrimination in the rental housing market - A field experiment on the underlying mechanisms. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (MZES); presented (Equalsoc final conference, 3-5 June 2010, Amsterdam).
Horr, A., C. Hunkler and C. Kroneberg. 2018. „Ethnic Discrimination in the German Housing Market." *Zeitschrift für Soziologie* 47(2): 134-146.
Hosoda, M. and E. F. Stone-Romero. 2010. "The Effects of Foreign Accents on Employment-Related Decisions." *Journal of Managerial Psychology* 25(2): 113-132.
Hosoda, M., L. T. Nguyen and E. F. Stone-Romero. 2012. "The Effect of Hispanic Accents on Employment Decisions." *Journal of Managerial Psychology* 27(4): 347-364.
Huang, L., M. Frideger and J. L. Pearce. 2013. "Political Skill: Explaining the Effects of Nonnative Accent on Managerial Hiring and Entrepreneurial Investment Decisions." *Journal of Applied Psychology 98*(6): 1005-1017.
Kaas, L. and C. Manger. 2011. "Ethnic Discrimination in Germany's Labour Market: A Field Experiment." *German Economic Review* 13(1): 1-20.
Kalter, F. 2006. „Auf der Suche nach einer Erklärung für die spezifischen Arbeitsmarktnachteile von Jugendlichen türkischer Herkunft." *Zeitschrift für Soziologie* 35(2): 144-160.
Kalter, F. and N. Granato. 2007. "Educational Hurdles on the Way to Structural Assimilation in Germany." In *Unequal Chances: Ethnic Minorities in Western Labour Market*s, edited by A. Heath and S. Y. Cheung, 271-319, Oxford: Oxford University Press.
Kulik, C. T. and L. Roberson. 2008. "Common Goals and Golden Opportunities: Evaluations of Diversity Education in Academic and Organizational Settings." *Academy of Management Learning and Education* 7(3): 309-331.
Lambert, W. E., R. C. Hodgson, R. C. Gardner and S. Fillenbaum. 1960. "Evaluational Reactions to Spoken Languages." *Journal of Abnormal and Social Psychology* 60(1): 44-51.
Larja, L., J. Warius, L. Sundbäck, K. Liebkind, I. Kandolin, I. Jasinskaja-Lahti. 2012. *Discrimination in the Finnish Labor Market: An Overview and a Field Experiment on Recruitment*. Ministry of Employment and the Economy.
Lippi-Green, R. 1994. "Accent, Standard Language Ideology, and Discriminatory Pretext in the Courts." *Language in Societies* 23(2): 163-198.

Lorenz, G. 2017. *Selbsterfüllende Prophezeiungen in der Schule: Leistungserwartungen von Lehrkräften und Kompetenzen von Kindern mit Zuwanderungshintergrund*. Wiesbaden: Springer VS.

Mai, R. and S. Hoffmann. 2014. „Accents in Business Communication: An Integrative Model and Propositions for Future Research.“ *Journal of Consumer Psychology* 24(1): 137-158.

Morales, A. C., M. L. Scott and E. A. Yorkston. 2012. "The Role of Accent Standardness in Message Preference and Recall." *Journal of Advertising* 41(1): 33-45.

Munro, M. J. and T. M. Derwing. 1995. "Processing Time, Accent, and Comprehensibility in the Perception of Native and Foreign-Accented Speech." *Language and Speech* 38(3): 289-306.

Paulus, W. and B. Matthes. 2013. "The German Classification of Occupations 2010: Structure, Coding and Conversion Table." *FDZ-Methodenreport 08/2013 (en)*, Nürnberg: Institute for Employment Research.

Pager, D. 2007. "The Use of Field Experiments for Studies of Employment Discrimination: Contributions, Critiques, and Directions for the Future." *The ANNALS of the American Academy of Political and Social Science* 609: 104-133.

Phelps, E. (1972). The statistical theory of racism and sexism. *The American economic* review 62 (4): 659-661.

Rakić, T., M. Steffens and A. Mummendey. 2011. "Blinded by the Accent! The Minor Role of Looks in Ethnic Categorization." *Journal of Personality and Social Psychology* 100(1): 16-29.

Rödin, M. and G. Özcan. 2011. "Is It How You Look or Speak That Matters? An Experimental Study Exploring the Mechanisms of Ethnic Discrimination." *Research Papers in Economics* 12, Stockholm University, Department of Economics.

Rodríguez, G. 2010. "Turksprachige Namen in Deutschland: Statistik und Tendenzen in der turksprachigen Vornamengebung." *Namenkundliche Informationen* 97: 95-107.

Rubin, D. B. 1987. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley and Sons.

Rudolph, U., R. Böhm and M. Lummer. 2007. "Ein Vorname sagt mehr als 1000 Worte: Zur sozialen Wahrnehmung von Vornamen." *Zeitschrift für Sozialpsychologie* 38(1): 17-31.

Schmidt, F. L. and J. E. Hunter. 1998. "The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings." *Psychological Bulletin* 124(2): 262-274.

Schneider, J., R. Yemane and M. Weinmann. 2014. *Diskriminierung am Ausbildungsmarkt: Ausmaß, Ursachen und Handlungsperspektiven*. Berlin: Sachverständigenrat deutscher Stiftungen für Integration und Migration.

Segrest Purkiss, S. L., P. L. Perrewé, P., T. L. Gillespie, B. T. Mayes and G. R. Ferris. 2006. "Implicit Sources of Bias in Employment Interview Judgments and Decisions." *Organizational Behavior and Human Decision Processes* 101(2): 152-167.

Steinbach, A. (2004). Soziale Distanz: Ethnische Grenzziehung und die Eingliederung von Zuwanderern in Deutschland. Wiesbaden: VS Verl. für Sozialwiss.

Stepstone. 2013. Jobsuche 2013: *Wie Recruiter und Bewerber vorgehen und was sie erwarten*. www.stepstone.de/content/de/de/fur_stellenanbieter/.../StepStone_jobsuche_2013.pdf, retrieved August 26, 2018.

Strabac, Z. and O. Listhaug. 2008. "Anti-Muslim Prejudice in Europe: A Multilevel Analysis of Survey Data from 30 Countries." *Social Science Research* 37: 268-286.

Timming, A. R. 2017. "The Effect of a Foreign Accent on Employability: A Study of the Aural Dimension of Aesthetic Labour and Customer-Facing and Non-Customer-Facing Jobs." *Work, Employment and Society* 31(3): 409-428.

van Tubergen, F. & Kalmijn, M. (2005), Destination-Language Proficiency in Cross-National Perspective: A Study of Immigrant Groups in Nine Western Countries. *American Journal of Sociology* 110 (5): 1412-1457.

Van Wijngaarden, S. J. 2001. "Intelligibility of Native and Non-Native Dutch Speech." *Speech Communication* 35 (1-2): 103-113.

Von Stetten, A., S. Laumer, C. Maier, T. Weitzel, A. Eckhardt and E. Guhl. 2013. *Recruiting Trends im Mittelstand 2013: Eine empirische Untersuchung mit 1.000 Unternehmen aus dem deutschen*

Mittelstand. Research report, Otto-Friedrich-Universität Bamberg und Goethe-Universität Frankfurt am Main.

Wagner, U., R. van Dick, R., T. F. Pettigrew, T. F., and O. Christ. 2003. "Ethnic Prejudice in East and West Germany: The Explanatory Power of Intergroup Contact." *Group Processes and Intergroup Relations* 6(1): 22-36.

United States General Accounting Office. 1990. *Immigration Reform: Employer Sanctions and the Question of Discrimination*. Washington, DC: Government Printing.

Zschirnt, E. and D. Ruedin. 2016. "Ethnic Discrimination in Hiring Decisions: A Meta-Analysis of Correspondence Tests 1990-2015." *Journal of Ethnic and Migration Studies* 42(7): 1115-11.

Table 1: Total distribution of employer replies (% and N)

Applicant characteristics	**Employer Reaction (Position Available)**			
	Yes	*No*	*Total*	
German name, Standard German	79.93	20.07	100.00	
	(482)	(121)	(603)	
Turkish Name, Standard German	79.77	20.23	100.00	
	(493)	(125)	(618)	
Turkish Name, Turkish accent	72.96	27.04	100.00	
	(448)	(166)	(614)	
Total	77.55	22.45	100.00	%
	(1423)	(412)	(1835)	(N)

Table 2: Logistic ML models. Employers' availability replies (yes/no) regressed on applicants' names and accents.

	Model (1) (AME)	Model (2) (AME)	Model (3) (AME)
Fixed part			
Applicant characteristics			
Tur. Name	-0.3*		0.00
	(0.01)		(0.01)
Tur. Accent		-0.06***	-0.06***
		(0.02)	(0.02)
Random part			
variance between ads (icc)	*24.15 (.86)*	*26.59 (.87)*	*26.89 (.87)*
variance between districts (icc)	*0.23 (.01)*	*0.15 (.01)*	*0.17 (.01)*
N	1835	1835	1835
DIC	973.62	930.98	*942.66*

*Note: Logistic ML models, AMEs shown; controls included; ***p<.001; **p<.01; *p<0.05; +p<.01.*

Table 3: Logistic ML models. Employers' availability replies (yes/no) regressed on applicant, job, firm, and context characteristics

DV: Position available (yes/no)	Model (1) AME (SE)	Model (2a) AME (SE)	Model (2b) AME (SE)	Model (3a) AME (SE)	Model (3b) AME (SE)	Model (3c) AME (SE)	Model (4a) AME (SE)	Model (4b) AME (SE)	Model (5) AME (SE)	Model (6) AME (SE)
Applicant characteristics										
Turkish Name	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Turkish Accent	**-0.06****	**-0.07****	**-0.07***	**-0.11****	**-0.07*****	**-0.09****	**-0.06*****	**-0.08**	**-0.21***	**-0.07****
	(0.02)	**(0.03)**	**(0.03)**	**(0.04)**	**(0.02)**	**(0.03)**	**(0.02)**	**(0.05)**	**(0.08)**	**(0.02)**
Job characteristics										
Communicative demands										
Communicative tasks	**0.03+**	**0.03+**	**0.03+**	**0.03+**	**0.04+**	**0.03+**	**0.04+**	**0.03+**	**0.03+**	**0.03+**
	(0.02)	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.02)**
Education (Ref. low)										
Medium	0.04	0.05	0.04	0.04	0.04	0.04	0.05	0.04	0.05	0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
High	0.00	0.00	-0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Customer contact	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Firm characteristics										
Standardization										
Size	**0.07***	**0.07***	**0.07***	**0.04+**	**0.07***	**0.07***	**0.07***	**0.07***	**0.07***	**0.07***
	(0.03)	**(0.03)**	**(0.03)**	**(0.02)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**
Diversity	0.04	0.05	0.04	0.04	0.02	0.05	0.05	0.04	0.05	0.05
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
External recruiter	**0.08****	**0.08****	**0.08****	**0.08****	**0.08****	**0.05****	**0.08****	**0.08****	**0.08****	**0.08****
	(0.03)	**(0.03)**	**(0.03)**	**(0.02)**	**(0.02)**	**(0.02)**	**(0.03)**	**(0.03)**	**(0.03)**	**(0.03)**
Context characteristics										
Labor supply										
Vacancyrate	-0.04	-0.04	-0.04	-0.04	-0.04	-0.05	-0.02	-0.05	-0.06	-0.04
	(0.28)	(0.28)	(0.28)	(0.28)	(0.28)	(0.28)	(0.25)	(0.28)	(0.28)	(0.28)
Unemployment rate	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)
Regional attitudes	0.02	0.03	0.02	0.02	0.03	0.03	0.03	0.03	-0.02	0.03
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.04)	(0.05)
Interactions with accent										
Tur. Accent x communicative tasks		0.01								
		(0.02)								
Tur. accent x education			0.00							
Medium			(0.04)							
High			0.02							
			(0.05)							
Tur. accent x firm size				**0.09***						
				(0.04)						
Tur. accent x diversity					**0.07***					
					(0.03)					
Tur. accent x external recruiter						**0.08****				
						(0.02)				
Tur. accent x vacancyrate							-0.03			
							(0.38)			
Tur. accent x unemploymentrate								0.00		
								(0.01)		
Tur. accent x regional attitudes									**0.15***	
									(0.06)	
Tur. accent x customer contact										0.01
										(0.02)
variance between ads (icc)	26.89(.88)	27.04 (.88)	27.06 (.88)	27.38 (.88)	27.49 (.88)	27.14 (.88)	27.02 (.88)	27.21 (.88)	28.48 (.88)	27.00 (.87)
variance between districts (icc)	0.17(.003)	0.19 (.006)	0.19 (.006)	0.22 (.007)	0.12 (.004)	0.15 (.005)	0.15 (.005)	0.20 (.006)	0.17 (.005)	0.22 (.007)
N	1835	1835	1835	1835	1835	1835	1835	1835	1835	1835
DIC	942.66	934.50	942.35	939.54	937.50	939.07	942.35	939.54	939.07	930.05

*Note: Logistic ML models, AMEs shown; controls included;***p<.001; **p<.01;*p<0.05; +p<.01.*

Figures 1- 4: Predicted probabilities of job vacancy by accent and job, firm and context characteristics (interaction effects, predictive margins)

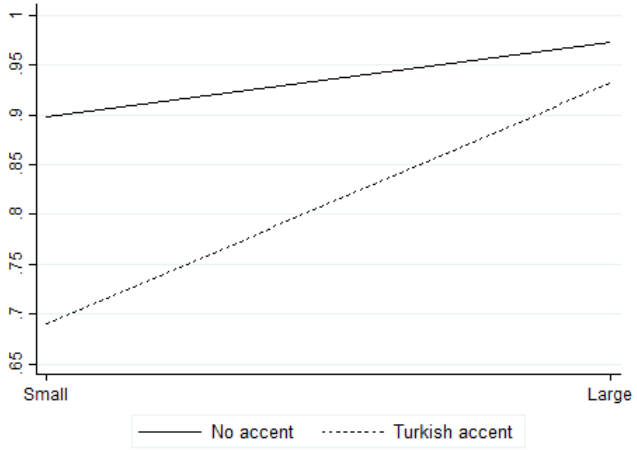


Figure 1: Predicted probability of job vacancy by Accent and firm size (interaction effects, predictive margins)

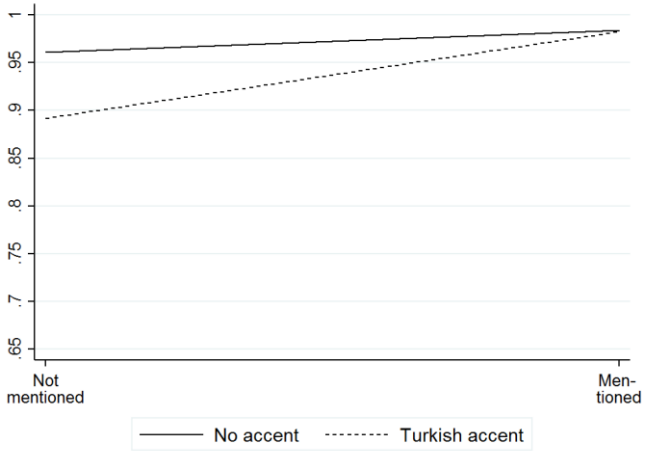


Figure 2: Predicted probability of job vacancy by Accent and company focus (interaction effects, predictive margins)

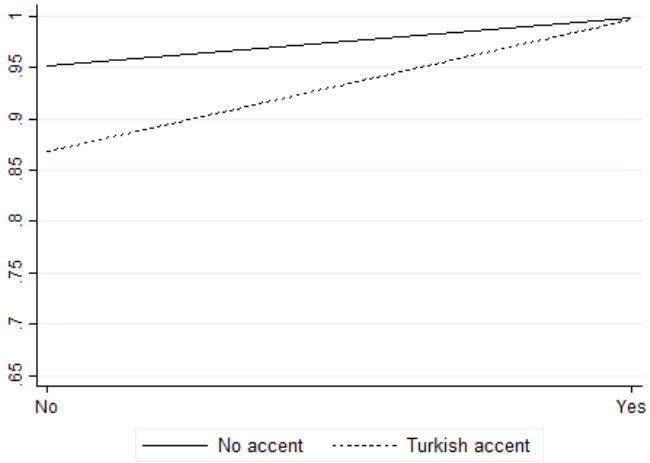


Figure 3: Predicted probability of job vacancy by Accent and external recruiters (interaction effects, predictive margins)

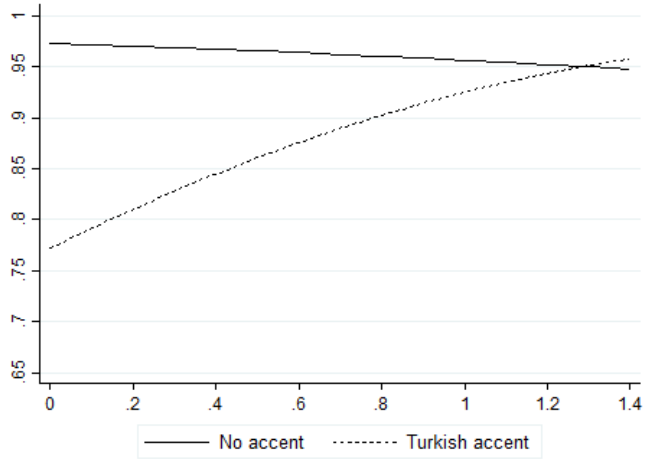


Figure 4: Predicted probability of job vacancy by Accent and regional attitudes (interaction effects, predictive margins)

Appendix, Table A.1: Distributions of model variables

	%	(N)	% Miss	Mean	(SE)	Min.	Max.	Data source
Position available		(1,835)						Field experiment
Yes	77,55	(1,423)	0,00					
No	22,45	(412)	0,00					
Applicant characteristics								
Name		(1,835)						Field experiment
German	32,86	(603)	0,00					
Turkish	67,14	(1,232)	0,00					
Accent		(1,835)						Field experiment
German	66,54	(1,221)	0,00					
Turkish	33,46	(614)	0,00					
Job Characteristics								
Communicative demands								
Communicative tasks		(1,835)	0,00	0,05	0,07	0	0,67	Firm / ad
Education*		(1,835)	0,38					Firm / ad
low	32,90	(605)	0,00					
medium	43,43	(797)	0,00					
high	23,22	(426)	0,00					
Customer contact		(1,835)	0,00	0,03	0,07	0	0,5	Firm / ad
Firm characteristics								
Standardization								
Firm size*		(1,835)	28,07					Firm / ad
<50	35,26	(647)						
>50	36,68	(673)						
External recruiter		(1,835)						Firm / ad
Yes	30,46	(559)	0,00					
No	69,54	(1,276)	0,00					
Diversity*		(1,835)						Firm / ad
No Strategies	68.61	(1,259)	23,98					
Strategies in place	7,41	(136)						
Context characteristics								
Labor supply								
Vacancy rate		(1,835)	0,00	0,02	0,05	0	0,5	FEA
Unemployment rate		(1,835)	0,00	6,79	2,58	1,7	15,9	FEA
Regional attitudes		(1,835)	0,00	0,35	0,02	0,29	0,43	ALLBUS
Control variables								
Gender applicant		(1,835)						Field experiment
Male	76,78	(1,409)	0,00					
Female	23,22	(426)	0,00					
Gender employer		(1,835)						Firm / ad
Male	37.55	(689)	0,00					
Female	28.99	(532)	0,00					
unknown	33.46	(614)						
Ethnicity employer		(1,835)						Firm / ad
German	51.83	(951)	0,00					
Turkish	10.57	(194)	0,00					
other/unknown	37.60	(690)	0,00					
N words per ad			0,00	57	55	2	499	Firm / ad

* imputed variables

Appendix, Table A.2: Variance components of the unconditional ML model

DV: Position available (yes/no)	Model (0)
Random part	
variance between ads (icc)	20.39 (.81)
variance between districts (icc)	1.01 (.04)
N	1835
DIC	1030.98

Supplementary materials, Table I: Three-way interaction models

DV: Position available (yes/no)	Model I ß (SE)	Model II ß (SE)
Applicant characteristics		
Turkish Name	0.01	0.02
	(0.27)	(0.27)
Turkish Accent		
	-1.26*	**-3.29*****
Job characteristics	**(0.56)**	**(0.79)**
Communicative demands		
Communicative tasks	0.60	**0.71+**
	(0.64)	**(0.39)**
Education (Ref. Primary, lower sec.)	0.97	1.08
Upper secondary	(0.66)	(0.67)
Teritary	-0.02	0.21
	(0.90)	(0.89)
Customer contact	-0.20	-0.80
	(0.47)	(1.28)
Firm characteristics		
Standardization		
Size	**1.73***	**1.68***
	(0.70)	**(0.70)**
Diversity	1.52	1.65
	(1.20)	(1.22)
External recruiter	**3.82*****	**3.84*****
	(0.94)	**(0.92)**
Context characteristics		
Labor supply		
Vacancyrate	-1.27	-0.82
	-(6.04)	(5.91)
Unemployment rate		
	0.05	0.05
	(0.11)	(0.11)
Regional attitudes	-0.19	-0.78
	(1.14)	(1.14)
Interactions with accent		
Tur. Accent x communicative tasks	-0.76	
	(0.51)	
Tur. Accent x share of foreigners	0.00	
	(0.04)	
Tur. Accent x communicative tasks x share of foreign	0.07	
	(0.05)	
Communicative tasks x share of foreigners	0.02	
	(0.01)	
Tur. Accent x customer contact		1.49
		(1.70)
Tur. Accent x regional attitudes		**2.61****
		(0.92)
Tur. accent x customer contact x regional attitudes		-1.64
		(2.1)
Customer contact x regional attitudes		1.47
		(1.70)
variance between ads (icc)		
variance between districts (icc)		
N	**1835**	**1835**
DIC	931.94	903.58

*Note: Logistic ML models, AMEs shown; controls included; ***p<.001; **p<.01;*p<0.05; +p<.01.*

Paper 3

Language minority students and the identification of special educational needs – findings from the English educational context

Planned (re)submission 2019: Journal of ethnic and migration studies
Author: Miriam Schmaus, Bamberg University

Language minority students and the identification of special educational needs – findings from the English educational context

Language minority status is commonly linked with ethnic educational inequalities. Next to academic achievement, debates also revolve around the identification of special educational needs (SEN). Language minority students can exhibit (linguistic) characteristics that resemble indications of SEN. Such ambiguities can entail diagnosing hesitancy as well as biased judgements, which foster inaccurate identification. The latter is associated with inadequate educational support, which might hamper educational achievement. This study examines the SEN identification of language minority children in England, about which there has been little research to date. Drawing upon risk-aversion and discrimination theories, I investigate whether the ambiguities of language minority students are associated with diagnosing hesitancy, or if they give room for discrimination. Building upon competence development models, I also assess how inaccurate SEN judgements affect students' educational achievement. Employing data from the Millennium-Cohort-Study, I examine discrepancies between children's performance-based predicted probability of being identified with SEN and teachers' actual SEN identification. Using logistic regression models I assess whether language minority children are more likely to experience diagnosing hesitancy or biased judgements upon school entry. Next, I employ linear regression models to assess how inaccurate SEN identification upon school entry affects children's Maths and English competences at the end of primary school. Results indicate group-specific effects. For black children, limited English proficiency was associated with over-identification, particularly of behavioural needs, which might reflect biased judgements. All other groups were less likely to be over-identified, which might reflect a greater (mis)diagnosing hesitancy. Inaccurate SEN identification was negatively related to educational achievement.

Introduction

Language minority students make up a considerable share of the student population in many countries. In the educational context, language minority status generally refers to all students whose native language differs from the official language in the respective country. These students commonly learn the official language as an additional language and can exhibit limited proficiency in the language of instruction upon school entry.

Limited proficiency in the language of instruction is widely accepted to affect educational outcomes. Next to academic achievement, debates also revolve around the identification of special educational needs (SEN) (Artiles et al. 2005; De Valenzuela et al. 2006; Samson and Lesaux 2009; Shifer et al.2011; Sullivan 2011). Language minority students with limited proficiency in the language of instruction can exhibit difficulties that resemble signs of certain special educational needs. On the one hand, linguistic features can overlap with indications of speech and language needs, as well as with signs of specific learning difficulties (Case and Taylor 2005; Klinger et al. 2006). On the other the hand, developing language proficiency might prevent children from fully following lessons, causing demeanours that can be interpreted as signs of behavioural difficulties, or general underachievement, which might be interpreted as being indicative of cognitive difficulties (Ortiz and Maldonado-Colón 1986). Effective methods for differentiating between these sources are scarce (Burr et al. 2015; Figueroa and Newsome 2006; Limbos and Geva 2001), which creates ambiguities and room for idiosyncratic decisions in the identification process. Such ambiguities could thus entail (mis)diagnosing hesitancy (Hibel and Jasper 2012; Samson and Lesaux 2009), as well as biased judgements (Harrison and Lakin 2018; Pettit 2011) and foster inaccurate identification.

Despite increasing awareness of these issues, research covering such processes in the English educational context is scarce. Large-scale quantitative studies tend to focus on the relationship between ethnic origin and SEN identification, irrespective of the child's language status (Keslair and McNally 2012; Strand and Lindsay 2009). The few that do consider language status employ somewhat coarse indicators that represent children's exposure to the language of instruction (e.g. at their home), rather than their actual proficiency (Lindsay and Strand 2016; Strand and Lindsay 2012,30-31), which might inadequately capture the population and processes of interest. Moreover, these studies tend not to consider information on children's academic performance and school-related behaviour, which is crucial when assessing whether SEN identification truly differs between children with different (language) backgrounds.

To bridge the gap, this study examines the SEN identification of language minority children in England. Drawing upon risk-aversion arguments and discrimination theories, I investigate whether the ambiguities that language minority students pose are associated with greater diagnosing hesitancy, or if they give room for discriminatory judgements. Drawing upon data from the Millennium cohort study (MCS), I examine discrepancies between children's performance-based predicted probability of being identified with different special educational needs and teachers' actual SEN identification. Predicted probabilities were obtained by regressing teachers' first SEN identification on children's school-related competences and behaviours upon school entry. Using logistic regression models I then assess whether language minority children are more likely to experience diagnosing hesitancy, or biased judgements upon school entry. I thereby employ two different indicators of language minority status that relate to exposure, as well as proficiency in the language of instruction.

Furthermore, I assess how inaccurate SEN judgements affect students' competence acquisition. Misidentification should be associated with inadequate educational support, as children are confronted with inappropriate learning materials, settings and expectations, which might be detrimental to their educational achievement. Using linear regression models, I assess how inaccurate SEN identification upon school entry affects competences in two core subjects (English and Maths) by the end of primary school. This allows tentative inferences on the consequences that are associated with inaccurate SEN identification and might shed new light on the processes behind the educational disadvantages of language minority children.

Language minority status and SEN identification

In the English educational context, (legal) definitions stipulate that a young person has special educational needs "if she/he has a learning difficulty or disability which calls for special educational provision to be made" (The Children and Families Act 2014,19). Literature distinguishes between "judgemental" and "non-judgemental" need categories (Strand and Lindsay 2009; Tomlinson 2016), where the former can be identified through subjective evaluations and "value judgements" whilst the latter gives less room for idiosyncratic decisions and sometimes allows for medical diagnoses. Judgemental categories for example comprise behavioural, emotional and social needs (BESD), communication and interaction needs (CI) and forms of cognition and learning needs (CL). Coincidentally these categories are also what are termed "high incidence" and cover the majority of children with special educational needs. Non-judgemental categories for example include sensory and/or physical needs (SP), which make up the "low incidence" categories.

While the difficulties associated with language minority status are explicitly excluded from all the above categories, distinguishing between developing English proficiency and learning difficulties

or disabilities can sometimes be somewhat difficult (Keller-Allen 2006; Klinger and Harry 2006; Limbos and Geva 2001; Sullivan 2011), especially when identifying educational needs in the judgemental categories. On the one hand, linguistic features of children who are learning English as a second language, such as problems with the (correct) production of sounds, articulation or understanding (and conveying) information through language, overlap with indications of communication and interaction needs (CI), or could be interpreted as forms of cognition and learning needs (CL), for instance as specific learning difficulties such as dyslexia (Case and Taylor 2005; Klinger and Harry 2006). On the other the hand, a still developing English proficiency might prevent children from fully following lessons, causing demeanours that can be interpreted as signs of behavioural difficulties (BESD), such as short attention spans, temper outbursts, low self-confidence, a general lack of interest or apathy (Ortiz and Maldonado-Colón 1986). It might also be associated with general underachievement, which could be interpreted as (moderate) learning- or cognitive difficulties (CL).

To date, there is no uncontested method for assessing the educational needs of (less proficient) language minority students unequivocally (Burr et al. 2015; Figueroa and Newsome 2006; Lane and Leventhal 2015) and testing those children has sometimes been called a case of "random chaos for psychometrics" (Figueroa 1989).

The role of early identification

In the English educational context, there is a distinct focus on early SEN identification to promote equal opportunities and high achievement for all children (DfES 2017). In general, SEN assessments start as children enter an early educational setting, such as nurseries, kindergartens or playgroups (DfES 2015). However, many children are initially identified during their first year(s) of primary school. At the beginning of compulsory education (age 5), children face the first comprehensive assessment of their development in crucial areas of learning (EYFS Profile). This then serves as a guide to teachers and practitioners, who place emphasis on early and accurate SEN identification.

In general, early SEN identification should be beneficial for students' academic achievement (e.g. Guralnick 2005). However, for language minority children it might be associated with inaccurate assessments. In the early years of schooling, many of these children still have limited proficiency in the language of instruction (e.g. Farina and Geva 2013; Strand and Malmberg 2015), which makes it particularly challenging to assess their SENs. When unequivocal assessment of these children is not possible, there might be different ways in which their ambiguities influence teachers' judgements.

The risk aversion argument

Teachers or other professionals might be more careful or even hesitant to identify language minority children, when their second language proficiency is (still) limited. As school personnel are often conscious of the ambiguities presented by these children, they might be very aware that early identification carries the risk of misdiagnosis. To secure the probable gains associated with correct early identification, such as improved learning trajectories (Guralnick 2005), teachers might be prone to act risk averse (Tversky and Kahneman 1992).

One strategy for avoiding the respective risk is to delay assessments until children's language proficiency has developed further, as this allows a more clear-cut evaluation of their difficulties. US studies indicate that teachers might postpone the SEN assessment of language minority

students until third or fourth grade (e.g. Hibel and Jasper 2012; Samson and Lesaux 2009). Some authors also report delays up to five years following school entry (Limbos and Geva 2001,136). Such delays should be particularly likely when assessing language related needs (CI, specific CL), as here ambiguities and consequently the risk of misdiagnosis are greatest. Building on these assumptions, one would expect that if risk-aversion and (mis)diagnosing hesitancy were core motives, *language minority children with lower language proficiency are more likely to be under-identified with judgement-based types of SEN in their early school years than (language) majority students, particularly if the needs are language-related (CI, CL) (H1a)*.

An alternative strategy to avoid the risk of misidentification is to be more meticulous and take additional actions when assessing less proficient language minority children at an early age (Burr et. al 2015). Teachers might try to ensure accurate assessments by including additional tests, observations or professionals and generally taking greater care when evaluating the respective information. Such meticulous assessments might increase identification accuracy over and above language related ambiguities and the respective language minority children should be more thoroughly assessed than (language) majority students. Consequently, if teachers paid additional attention one would assume that *language minority children with lower language proficiency are more likely to be accurately identified with judgement-based types of SEN in their early school years than (language) majority students (H1b)*.

Biased judgements and discrimination

SEN assessments of language minority students might also be linked to teacher bias and discrimination. On the one hand, limited language proficiency might foster ethnic discrimination and thus *solely affect members of specific ethnic origins*; on the other hand, it might elicit discrimination types that *affect all language minority children*.

Language minority status and ethnic discrimination

Language minority status could *firstly foster the application of ethnic stereotypes or negative beliefs* (England 1992). The emerging language proficiency represents an additional cue in impression formation, which when combined with other features of ethnic origin (such as name or looks) provides multiple consistent group stimuli, which is assumed to facilitate stereotype application (Fiske et. al 1999). If teachers and other professionals included such stereotypes in their decision-making, patterns of SEN misidentification should match group-specific beliefs about school-related attributes. Consequently groups associated with negative beliefs about intelligence, educability or deviance should be more likely to be identified with the respective SENs, whereas groups associated with positive beliefs should be less likely.

Negative assumptions about school-related aspects seem particularly to affect black (Caribbean) students (Burgess and Greaves 2013; Tomlinson 2016,523; Youdell 2003). Whereas negative associations about educability and intellect are sometimes discussed, stereotype content primarily revolves around deviant behavioural patterns, such as defiance, aggression or violence (Frosh and Pattman 2002; Gillborn 1997; Phillips 2011,181), which correspond to difficulties in the BESD spectrum. Thus, for these children behavioural anomalies associated with developing language proficiency, such as temper outbursts, apathy or attention deficits, might foster application of negative behavioural beliefs and be interpreted in a stereotype consistent manner.

Negative beliefs also relate to south Asian immigrants, for instance Pakistani and Bangladeshi groups. However, here unfavourable assumptions seem primarily to include beliefs about

religious fundamentalism and patriarchal gender roles (MacGhaill and Haywood 2015; Phillips 2011,181). Nevertheless, school-related assumptions might be more positive as, especially for current children, information about substantial improvements at every level of education (e.g. Strand and Hall 2015) has been reported in the press. Furthermore, qualitative research gives indications that teachers ascribe better abilities and behaviour to south Asian than to black (Caribbean) children, even if this includes the notion of suppression and passivity for girls (Gillborn 1997,385).

Particularly positive assumptions about school-related attributes are associated with immigrants from China and other parts of Eastern Asia. Here stereotypical attributes comprise assumptions about being docile, conformant or shy together with being high- or over-achieving (Zhang 2010). Consequently for these children, ambiguities associated with developing language proficiency, such as passivity, reluctance to speak or cognitive difficulties should foster the application of positive behaviour- and performance beliefs. Thus if language minority status were to foster the application of ethnic stereotypes, patterns of *misdiagnosis should be stereotype congruent, which would imply that with lower language proficiency, black students are more likely to be over identified with BESD, whereas Chinese, East Asian and to a lesser extent South Asian students are more likely to be under-identified with BESD, CI and CL than their majority peers (H2a)*.

Additionally, language minority status might enhance group-specific distastes (Becker 1976), as limited language proficiency can function as a particularly prominent reminder of the respective ethnic origin (Rakic' et al. 2011). SEN identification might then be seen as a strategy to decrease interactions with students of the respective groups and thus reduce the associated psychological costs (Becker 1976), as it makes them eligible for external help and more likely to spend time outside the regular classroom setting. In the English context, distaste and aversion are assumed to affect multiple ethnic minority groups, yet there might be a loose form of hierarchy in the extent to which these groups are non-preferred. Studies suggest that Muslim minorities, such as parts of the Pakistani and Bangladeshi population, face particularly strong distances and opposition (Storm et al. 2017), loosely followed by black Caribbean and Africans together with other Asian minorities (ibid.). These tendencies have been linked to increasing islamophobia (Storm et al. 2017) and prevailing "legacies of colonialism" (Tomlinson 2016). Consequently, if language minority status were related to group-specific distastes, patterns of SEN misidentification should be congruent with ethnic hierarchies and *with lower language proficiency, Pakistani and Bangladeshi students - and to a lesser extent those with black Caribbean and African backgrounds- are more likely to be over-identified with all judgement-based types of SEN than their majority peers (H2b)*.

Language minority-specific forms of discrimination

Alternatively, language minority status might also elicit forms of discrimination *that affect all of the respective students*. For all language minority children, lower second language proficiency might saliently signal their out-group membership. This might elicit negative attitudes (Tajfel and Turner 1986) and *increase the notion of competition over scarce resources* (Coser 1956). It has been argued that in such cases, SEN identification might be employed as an "early stratification measure" to keep out-group children from reaching valued resources, such as spots in distinguished colleges, or (good) employment positions (Sullivan and Artiles 2011,1531). Moreover, the salient out-group signal might foster cognitive bias, such as *"ultimate attribution errors"* (Pettigrew 1979). The associated model postulates that for in-group members unfavourable events, such as lower school performance, are ascribed to external circumstances, for instance an adverse home environment; however, for members of an out-group they are more likely to be

attributed to dispositions within out-group members. As limited language proficiency highlights out-group membership, majority teachers might be more prone to (mis)attribute school-related difficulties to unfavourable personality or ability dispositions within students themselves, rather than to the external factor of being a second language learner.

In addition to such processes on the individual level, language minority status has also been linked to a more far reaching, *structural form of disadvantage*. Initially discussions revolved around unreliable test instruments (e.g. Mercer and Brown 1973). *Commonly used IQ tests and other standardised measures can be less accurate in the case of language minority students*, as they often rely on (cultural) knowledge that is specific to the majority population and - when administered in the language of instruction - also function as a test of second language proficiency (Burr et al 2015; Duran 1988). Reliance on such measures might thus be associated with test-inherent over-identification of language minority children (e.g. MacSwan and Rolstad 2006).

Later contributions to this discussion also *address organisational processes within classrooms and schools*. It has been suggested that teachers might feel ill-prepared to deal with linguistically diverse students or to teach in multilingual classrooms (Ho 2004,32; Gandara et al. 2005). Establishing SENs ensures that the difficulties of language minority students will be tended to by somebody other than the classroom teacher, for instance by booster groups, one to one tuition, or learning mentors, which relieves some inter-lingual teaching burdens. It may thus be a more pragmatic, yet biased means of facilitating knowledge transfer in linguistically diverse classes and tackling the associated organisational difficulties. It has also been argued that schools might (informally) enforce such strategies at an administrative level, particularly when resources for specific second language learning programmes are scarce (Gomolla and Radtke 2003). If language minority status were related to such individual- or structural level forms of discrimination, one would *assume that all language minority students with lower language proficiency are more likely to be over-identified with all judgement-based types of SEN than (language) majority students (H2c)*

Consequences of inaccurate SEN assessments

Inaccurate SEN identification should alter multiple aspects of children's lives. However, first and foremost it might affect students' competence acquisition. To explain such processes, I draw upon a general model of competence development, which allows a systematic integration of all relevant arguments (Chiswick and Miller 1995). The main idea here is that competence acquisition per se is the result of investment in human capital and follows a function of incentives, exposure and efficiency. Incentives represent the gain that is associated with greater competence, such as higher certificates and ultimately higher earnings; exposure refers to learning opportunities, such as time spent on- and quality of learning materials, whilst efficiency is assumed to capture cognitive ability. Generally, high incentives, exposure and efficiency are assumed to promote competence development. Inaccurate SEN identification should negatively affect each of these parameters.

If children are *over-identified*, this might *firstly* have a *direct negative* effect on the parameters, as it reduces the *exposure* to their adequate level of learning materials and settings (time spent outside the regular classroom), and their learning *incentives*, as attainable outcomes might now seem lower. *Secondly,* over-identification might have an *indirect negative* effect on the parameters, as it changes external and internal perceptions of the child and its abilities. External perceptions might be affected as SEN designation typically comes with a set assumptions about capabilities (Ho 2004,88). These expectations are often lower than for non-designated children (Powell 2003). This might generate self-fulfilling prophecy effects (Rosenthal 1995) for children that are inaccurately

designated as SEN. Such effects should negatively affect *exposure*, as teachers might give them fewer opportunities to participate in the regular class setting. Internal perceptions of the child might be affected, as teachers might subtly convey these lower expectations or offer unnecessary help to children who are inaccurately designated as SEN, which might reduce their sense of self-efficiency and self-esteem (Powell 2003), or give them a fear of proving the lower expectations true (Steele and Aaronson 1995). This might reduce their learning *efficiency*, as these fearful thoughts and cognitive energy used to worry are assumed to detract from the cognitive energy available to process learning content or perform well in tests (ibid). This might also hold true for *exposure*, as children may feel incapable and consequently less willing to confront materials or situations that remind them of these feelings and thoughts (Powell 2003). Additionally, this might also decrease their motivation, or *incentives*, as the outcomes they (feel) they can reach are lower than those of other children.

If children are *under identified*, this might negatively affect *exposure*, as the available learning materials and settings might be less valuable to them than those specifically tailored to their needs; however, it might also affect *efficiency* as the inadequate learning environment might create feelings of being overwhelmed, scared or irritated, which again might detract from the cognitive energy available to process learning content.

Taken together, these arguments suggest that *an inaccurate SEN identification negatively affects students' competence acquisition (H3)*.

Data & methods

To test these assumptions empirically, I used data from the Millennium Cohort Study (MCS). The MCS represents one of Britain's larger birth cohort studies and follows approximately 19,000 children, born between 2000 and 2002, through their life course (Plewis 2007). Next to ample child and family characteristics, such as ethnic and language background, it includes teacher assessments, as well as externally administered competence tests. This provides unique opportunities to examine if teachers' SEN judgements match children's (objectively) tested abilities and how language minority status relates to potential discrepancies. Furthermore, it is possible to assess how inaccurate SEN judgements affect students' competence acquisition, as the study includes information on the development of school-relevant competencies.

As all arguments mentioned relate to processes within the primary school context, I draw upon data from the first five sweeps*, which include all information collected after birth (age 9 months) up to the end of primary school (age 11). As the SEN legislation and identification process differs slightly between the different UK countries, I restrict the sample to children who spent their whole (primary) school career in England and for whom teacher assessments of SEN are available[1]. As all analyses are (SEN) type-specific, final sample sizes range between n=4623 and n=4421.

Specification of variables I – Language minority status and inaccurate SEN identification

Dependent variable(s)

To assess the accuracy of teachers' judgements, scholars often adopt a residual approach (e.g. Madon et al. 1997). Here teacher evaluations are regressed on (objectively) tested or externally assessed information about children's skills and abilities. Residuals from the respective regression are then assumed to represent that part of teachers' judgements that is unrelated to students' abilities, namely the subjectively judged element that may potentially be inaccurate

(ibid.). I build upon such approaches and examine discrepancies between teachers SEN judgements and children's (externally) tested abilities. As SEN designation is a binary outcome, I draw upon predicted probabilities rather than residuals, as the latter are fixed in logistic models. Basic approach and interpretation are comparable to residual methods.

Accordingly, I regress the first teacher-reported SEN identification (age 7) on children's externally assessed abilities and predict students' probabilities of being identified with the respective SEN type. I use information on children's competences upon (or prior to) school entry (age 5) to rule out reverse causality. In concrete terms, I draw upon subsets of the British Ability Scales and the Bracken School Readiness test to predict cognitive (CL) and communication needs (CI)[2], on the Strength and Difficulties questionnaire to predict behavioural difficulties (BESD)[3] and on the International Classification of Diseases to predict physical needs (SP)[4]. Where necessary, the respective scores are age-adjusted. All models also included an early indicator of whether a child showed signs of developmental delay as a baby[5].

Children, whose externally assessed abilities suggest a high probability (more than 75%)[6] of being identified with a certain SEN type, but who are not recognised as such by their teacher, are classified as under-identified on that SEN type. Children with a low externally assessed probability (less than 25%), who are recognised by their teachers as having an SEN type, are classified as over-identified on that type. For each SEN type, over- and under identifications are thus indicated by separate dummy variables that take the value of 1 if students are classified as such.

Independent variables

In prior studies *language minority status* has commonly been indicated by information on children's minority language *exposure*, e.g. at their home or community (e.g. Samson and Lesaux 2008; Strand and Lindsey 2012,2016). While reflecting official (Census) definitions, such measures might inadequately capture the population and processes of interest. On the one hand measures of exposure are often congruent with ethnic minority status and might just "soak up" (additional) ethnic variation (Strand and Lindsey 2012,30). More importantly though, the link between language minority status and (inaccurate) SEN identification should primarily relate to limited second language proficiency. As children with the same minority language exposure can be heterogeneous in their second language proficiency (Strand et al. 2015), indicators of second language competence might allow for more direct tests of the respective arguments. I thus employed *two measures* that relate *to language exposure and proficiency*. This enables me to examine if second language proficiency is truly shaping inaccurate SEN assessments and if this holds true for all minority language exposed children, or if it affects only certain ethnic groups (H1-2).

To capture *second language exposure,* I employed a dummy variable that takes the value of 0 if parents stated that the language usually spoken at home was (mostly) English and the value of 1 if they spoke another language (half, most or all the time)[7] at the time when teachers were first asked about the child's SEN (age 7). To indicate *second language proficiency*, I drew upon children's scores on the BAS naming vocabulary scale. The respective test was externally (interviewer) administered close to school entry (age 5), which might reduce issues of reverse causality. It measures expressive language abilities and students' performance depends on their previous vocabulary development (Hansen et al. 2014,64). While test scores primarily reflect spoken vocabulary, it is also assumed to be indicative of general language development (ibid). Students' scores are age-adjusted and rescaled so that 0 represents the lowest observed value and a unit increase reflects one standard deviation. Higher scores are indicative of better abilities.

As the role of language minority status might be different for different *ethnic groups*, I drew upon parents' answers to the question to which ethnic group they regard their child as belonging (age 7). To ensure adequate case numbers, I distinguished between 6 main (census) categories and do not disaggregate the respective groups any further. I thus considered: White (majority, Ref.), Indian (1), Pakistani or Bangladeshi (2), Black or Black British (3), Chinese and Eastern Asians (4), Mixed and other heritage (5).

Control variables

In all models, information on *family- and individual level characteristics* is included. On the *family level*, I considered the mean number of times the *family lived in poverty* (sweep 0-4, 0-1 range), *parents' social class* (sweep 1-4, highest EGP) and *parents' educational background* (sweep 1-4, highest NVQ). Additionally, I considered *how involved the main caregiver is in school activities*, as indicated by the sum of all mentioned activities, such as parent-teacher committees or fundraising (age 7). On the *individual level,* a child's *age* and *gender* are considered. I also controlled for *school year*, as depending on birthday and parents' decision, children can be in year one, two or three at the time of teachers' first reported SENs (age 7).

Analytical strategy

To examine how language minority status relates to SEN misidentification, I estimated separate logistic regression models for each SEN type. I *firstly* assessed how minority language exposure and -proficiency per se affect SEN over (or under) identification (Model 1 & 2). *Secondly*, I estimated interaction models to examine if the effect of minority language exposure is conditional on second language proficiency (Model 3) and if second language proficiency moderates the identification accuracy for all origins, or if this affects only specific ethnic groups (Model 4). The main test of hypotheses thus relates to *the conditional main effects* of minority language exposure (M3) and ethnic origin (M4) and the associated *interactions*. Conditional main effects indicate how one interaction variable affects the outcome, when the second variable is zero, i.e. it indicates the likelihood of SEN over/under-identification for different language (M3) or ethnic groups (M4), when second language skills are low; interaction effects then signal how these group differences change if second language skills increase. Thus, if conditional main effects indicate that with low second language proficiency, all minority language (exposed) students are more likely to be under- (H1a) or accurately identified (H1b), than their (language) majority peers and that this difference reduces when language skills increase, this might reflect teachers' risk aversion strategies. Alternatively, if conditional main effects suggest that with low second language proficiency all minority language (exposed) students (H2c), or only specific ethnic groups (H2a-H2b) are more likely to be over-identified and that differences reduce with rising language skills, this might signal different forms of discrimination. In all models, the relevant survey weights were included.

To account for missing values, I used multiple imputations and estimated 25 datasets with complete information (e.g. Allison 2001). Regression analyses were performed on each dataset and afterwards combined following Rubin's (1978) approach. Furthermore, I followed an 'imputation then deletion' strategy (MID) (van Hippel 2007). Here all cases are used for imputation, but thereafter cases with imputed Ys values are deleted to increase precision. The distribution of model variables is illustrated in Table 4 (Appendix), imputed variables are marked.

As for all children SEN recognition is rare(r) than having no needs identified, all models were additionally estimated with a two-step correction technique (relogit), which is assumed to produce more accurate estimates in the presence of rare events (King and Zeng 2001)[8]. Moreover, as there

might be some kind of nesting structure in the data (students in teachers, in schools), I also estimated logistic multilevel models[9]. General trends remained comparable (supplementary materials, Table 7,8).

Specification of variables II – Inaccurate SEN identification and competence acquisition

Dependent variables

To assess how inaccurate SEN identification affects students' competence acquisition, I examined their *English and Maths* abilities at the end of primary school (age 11). On the one hand, at this stage enough time has passed since children's first SEN assessments to allow for potentially detrimental effects of inaccurate identification to unfold; on the other hand, with transition to secondary school, preceding achievements can shape school and subject choices (Werfhorst et al. 2003), which are of consequence for students' further academic careers.

To indicate the respective competences, I drew upon the MCS cognitive assessments[10]. For English skills I employed children's BAS naming vocabulary test scores (age 11). For Maths abilities I used the CANTAB Spatial Working Memory test (age 11). The latter indicates the extent of strategic thinking in a memory test, based on the ability to preserve and use spatial information (Atkinson 2015). While strategic thinking is related to the idea of problem solving in higher mathematics (Hyde et al 2008) and has been used as an indicator for similar constructs in the past (ibid), it is not a direct indicator of Maths competence. I supplemented this measure, therefore, with an indicator of children's arithmetic and mathematical reasoning abilities, based on a test that covers tasks on numbers, space, measures and data handling (NFER number skills), which was assessed at an earlier stage (age 7). While here the time between SEN identification and competence testing was considerably shorter, when taken together, both measures might provide a more comprehensive picture of mathematical competence and the role of (in)accurate SEN identification.

To ensure potential effects of SEN misidentification are truly due to the associated processes in the primary school context, information on children's competences in the respective subjects upon school entry was also included in these models (see robustness checks for alternative specifications). To indicate children's English proficiency, BAS naming vocabulary scores at age 5 were taken into account; to indicate their Maths abilities, BAS picture similarity scores, which reflect problem solving (age 5) and their Bracken number scores, which reflect basic counting and number recognition (age 3), were considered.

Independent variables

To indicate SEN misidentification, I drew upon the previously generated indicators that capture discrepancies between children's performance-based predicted probabilities and teachers' actual SEN judgements (age 7). I also replicated this procedure for the respective children at the end of primary school (age 11), to take potential posterior SEN adjustments into account. While early SEN misidentification might adversely affect competence acquisition irrespective of later adjustments, effects might be less pronounced, if children's needs are correctly reassessed at a later stage. I thus combined the two discrepancy indicators for each SEN type to distinguish between children *who during their whole primary school career were either (0) correctly identified as having no SENs, (1) correctly identified as having the respective SEN type, or (2) incorrectly identified as having the respective SEN type* and those who were *(3) inaccurately identified for only the first part of*

their primary school career (correct assessment at the later stage) and those who (4) had other changes of SEN status, e.g. between correct no SEN and SEN identification.

Control variables

In all models, information on *family- and individual level characteristics* was included. At family level, I considered the mean number of times the *family lived in poverty* (sweep 0-5, 0-1 range), *parents' social class* (sweep 1-5, highest EGP) and *parents' educational background* (sweep 1-5, highest NVQ). At individual level, I included information on *ethnic origin, minority language exposure, age and gender* – as described above. As competence acquisition within schools might be influenced by children's external learning opportunities, I included categorical variables that indicate whether parents arranged for extra lessons in reading, writing or Math (age 7), or in English, Maths or Science (age 11). Furthermore, as SEN children, whether inaccurately identified or not, might be more susceptible to *bullying* (Chatzitheochari et al 2016), which might hamper their competence acquisition, I also included information on how often children reported bullying (age 7), or being hurt or picked on by other children (age 11) [11].

Analytical strategy

To investigate how inaccurate SEN identification affects children's competence acquisition in English and Maths, I estimated linear regression models for all dependent variables. The test of hypotheses is linked to the main effect of the variables indicating SEN misidentification. If this is (substantially) negative, inaccurate SEN identification might be detrimental to children's educational attainment (H3).

As described above, I used multiple imputation (MID strategy) to account for missing values and estimated 25 datasets with complete information (e.g. Allison 2001). I additionally estimated all specifications using hierarchical linear models (HLM) to take potential nesting structures into account (supplementary materials, Table 9).

Results

Descriptives

In general, few children are inaccurately assessed as having SENs (Table 1). Across all SEN types, approximately 3% of all students are misidentified. Misidentification of CI is the most common, SP needs are least often inaccurately assessed, yet differences between SEN types are small. For all misidentified children, over-identification is more prevalent than under-identification. As the number of under-identified students is negligible, they are excluded from further analyses. Multivariate analyses thus examine the relationship between language minority status and SEN over-identification.

[Table 1]

Multivariate I – Language minority status and inaccurate SEN identification

For all judgement-based SEN types (BESD, CI, CL), minority language exposure affects SEN identification accuracy (Table 2,M1). When assessing behavioural difficulties (BESD), language minority exposure seems to promote over-identification (M1). Students that are exposed to a minority language at home are approximately 3 percentage points more likely to be falsely identified with BESD than their majority language peers. Second language proficiency seems to fully mediate this effect (M2). With other, more language-related difficulties (CI, CL) these tendencies reverse. Language minority exposure seems to foster more accurate assessments (M1) and independent effects remain over and above second language proficiency (M2).

Interaction models indicate that group-specific effects of second language proficiency might lie behind these differential findings. Conditional main effects reveal that lower second language proficiency does not lead to over-identification of behavioural difficulties (BESD) for all language minority children (M3), but that this solely affects Black British students (M4,Fig.3). Model 4 indicates that black British students with lower second language proficiency are around 40 percentage points more likely to be over-identified with BESD than their majority peers, whereas less proficient children from other origins do not differ considerably. However, when second language skills increase, the likelihood of being over-identified with behavioural difficulties also significantly reduces for black children (M4). *This might reflect the language-moderated application of adverse ethnic stereotypes for Black British children (H2a).*

For (almost) all other children that are exposed to a minority language at home, lower second language proficiency is associated with more accurate assessments, particularly of language related needs (CI). While here lower language proficiency does not fully explain the effect of language minority exposure (M1,M2), it moderates its impact (M3,Fig.1). Less proficient language minority children are approximately 14 percentage points less likely to be over-identified with CI than their majority peers, but this difference significantly reduces when their language skills increase (ibid). This applies to all ethnic groups, except black and mixed heritage children (M4,Fig.4). Results for the assessment of cognitive difficulties (CL) are comparable. Second language proficiency tends to shape the effect of language minority exposure (M3,Fig.2), which again holds true for all ethnic groups, except black and mixed heritage children (M4,Fig.5). This might reflect greater risk awareness when identifying less proficient language minority students, especially with language related needs. Contrary to expectations, risk awareness did not encourage

delayed assessments (H1a), but rather seems to foster *heightened attention during early identification (H1b), resulting in more accurate identifications for most language minority children.*

While for non-judgemental types (SP) identification accuracy varies between different ethnic groups, language minority status does not seem to make a difference. Minority language exposure does not notably affect misidentification on the respective type (M1,M2). Furthermore, second language proficiency does not differentially shape the likelihood of misidentification for children exposed to minority and majority languages at home (M3). Similarly, it does not differentially affect the likelihood for different ethnic groups (M4). However, Indian and Black British children with lower second language proficiency are considerably less likely to be over-identified with SP (ibid.).

Multivariate II - Consequences of inaccurate SEN identification

Inaccurate SEN identification negatively affects children's *English competence* acquisition (Table 3). At the end of primary school, consistently misidentified students have about 4.5 points lower naming vocabulary scores than their (correct) non-SEN peers. This corresponds to a disadvantage of approximately 0.5 standard deviations. Their lower performance could reflect adverse effects of exposure to inadequate learning materials and settings, reduced learning efficiency due to unfavourable self-perceptions and lower (perceived) incentives (H3). With exception to BESD, disadvantages also extend to students that were incorrectly assessed only for the first part of their primary school career. Here performance gaps are slightly smaller, but persistent at the end of primary school. In general, incorrectly assessed children perform slightly better than children that were (consistently) correctly identified with the respective SEN type. This might be because (correct) SEN children have been found to make less progress during primary school than their comparable peers (Parsons and Platt 2014) and despite receiving inadequate support, inaccurately identified children should be unaffected by the actual (cognitive, communicative or behavioural) difficulty. Despite small differences in significance and magnitude, general tendencies are comparable across all SEN types.

SEN misidentification also affects children's *Maths competence* acquisition. Inaccurate SEN judgements seem to influence strategic thinking abilities (SWM), as well as children's arithmetic and mathematical reasoning skills (NFER) - albeit to a different extent. Whereas misidentified students have only slightly lower strategic thinking scores than their (correctly identified) non-SEN peers (1.5 points, 0,2 SD), their arithmetic and mathematical reasoning skills are more strongly affected (10 points, 0,7 SD). Differences between these indicators could reflect the different aspects of Maths captured. Whereas strategic thinking is not directly taught in classes, arithmetic and mathematical reasoning are more closely related to school curricula. Inaccurate identification might thus affect the latter more strongly. However, differences between indicators could also reflect processes of self-adaption and resilience, as strategic scores are assessed at a later stage. Unlike with English competences, inaccurately identified students sometimes slightly underperform correctly identified SEN children on maths indicators. This could be due to different levels of test anxiety in Maths and English (Wolters and Pintrich 1998). In Maths, incorrectly identified children may have a greater fear of proving negative ability assumptions true (reduced efficiency). Yet differences are small and inconsistent, suggesting less systematic reasons.

[Table 2]

[Figures 1-5]

[Table 3]

Robustness of the findings

One objection towards these findings may be that the (dependent) variables indicating SEN misidentification are not sufficiently precise. When predicting the (pure) ability based probability of being identified with different SENs, it is imperative that all relevant competence aspects are taken into account. If the indicators employed do not fully represent all aspects of a child's (SEN type-specific) capabilities that a teacher can observe, children may be inadequately classified as over-, or under-identified.

To rule out such arguments, I first examined additional competence indicators. As prediction models take only type-specific abilities into account, e.g. cognitive indicators when looking at cognitive needs, I re-built all discrepancy indicators while including competence indicators from other domains, i.e. communicative, physical and behavioural indicators, when looking at the identification of cognitive needs. The main idea here is, that if all observable characteristics consistently suggest that the child's performance is higher (or lower) than its designated needs suggest, it might be unlikely that teachers observe traits that (a) differ from the child's measured competence level and (b) are not associated with any competences already included in the model. Whereas significance slightly differed in some of the respective models, the general trends remained similar, which might suggest that the respective measures capture some form of misidentification (Table 5, supplementary materials).

Moreover, if teachers did observe unmeasured competence aspects, this information should be missing for all children in the sample. Consequently group characteristics, such as ethnicity, or language status should not systematically affect the likelihood of misidentification, as this information was missing for children from all ethnic or language groups. Even if these 'unobservables' were correlated with language status, in that language minority children were more likely to lack skills that only teachers observe and this was driving SEN identification, there should be no systematic group-effects: the comparison category consists of children who were accurately assessed, which includes correct SEN and correct no SEN judgements, i.e. both low and high performing children. Consequently, unobserved competences should not produce systematic bias. As results (consistently) indicate group-specific effects of language minority status and ethnicity, this might be another hint that the indicators capture some form of misidentification.

A *second criticism* might pertain to how the role of inaccurate SEN identification for children's competence acquisition is examined. Here, the basic approach is to estimate the effect of inaccurate SEN identification on children's English and Maths competences, when relevant controls are taken into account. To ensure that potential effects are truly due to the assumed processes within the primary school context, children's English and Maths competences upon school entry are included in these models. However, it has been noted that if "baseline" competences, are not equal across groups, such specifications could produce misleading results (Lord 1967). As children with SENs might already lag behind non-SEN children in their

competence development upon school entry, meaning their "baseline" abilities might already be lower (Parsons and Platt 2014,16f), I re-conducted all analyses employing "value added" scores (ibid) that take baseline differences in children's competences into account (ibid)[12]. Results remained comparable (Table 6, supplementary materials).

Conclusion

This study investigated whether language minority status fosters inaccurate SEN identification and how such inaccuracies affect competence acquisition. Exploiting data from the Millennium Cohort Study, I examined whether the ambiguities that language minority children pose are associated with greater (mis)diagnosing hesitancy, or if they give room for discriminatory judgements. Additionally, I assessed how the respective misidentification affects children's competence acquisition for English and Maths. Results indicate group-specific effects. For black children, lower second language proficiency was associated with over-identification, particularly of behavioural needs, which might reflect biased judgements. However, for (almost) all other language minority children, lower second language proficiency was associated with more accurate identification, especially of language-related needs, which might reflect a greater misdiagnosing hesitancy that manifests in more thorough assessments. Inaccurate SEN identification was linked to lower educational achievement at the end of primary school.

With this, the study contributes to prior research in two main aspects. *Firstly*, it distinguishes between minority language exposure and –proficiency and systematically examines competing arguments as to how and why both aspects of language minority status affect SEN identification. This provides novel insights into SEN identification for language minority children in England, but might also enrich debates in other contexts and potentially unify contradictory findings that focus on singular aspects of language minority status. *Secondly*, this is the first study that examines how inaccurate SEN identification affects competence acquisition and can thus quantify the role of inaccurate assessments.

While giving novel insights into how language minority status affects SEN identification, there are limitations to this study. *The most crucial one* relates to case numbers. While sample size per se is sufficient, misidentification is comparatively rare. To ensure analyses are sound, I additionally estimated all models with a two-step correction technique (relogit), which is assumed to produce more accurate estimates in the presence of rare events (King and Zeng 2001). However, it might be informative to validate these findings further with a larger teacher-student sample.

The *second limitation relates* to the test of hypotheses. While results do suggest patterns that are consistent with different theoretical arguments, the proposed mechanisms could not be tested directly. It might thus be a potential route for future research to validate these mechanisms further, for instance with tailored teacher surveys, qualitative methods, such as in-depth interviews and videography, or stereotype indicating techniques, such as the IAT, or word-completion tests.

Despite these limitations this study provides new insights into the role of language minority status for ethnic educational inequalities and might offer strategies for early intervention, especially with regard to the SEN identification of Black British children.

Endnotes to chapter

1. When compared with the original English sample (sweep 1), final samples are slightly positively selected; while basic socio-demographic characteristics are comparable, children on average have slightly better skills and competences. This might limit generalisability as a whole, yet tests of hypotheses should remain unaffected and might be more conservative: If language minority status already affects the SEN assessments of more capable children, the same mechanisms should affect less advantaged children even more severely.
2. To predict CL needs, I draw upon BAS picture and pattern construction scores, indicative of children's problem-solving abilities and coordination (Hansen et al 2014,62); to predict CI needs, I draw upon BAS naming vocabulary- and the Bracken letters scores, indicative of expressive language skills, vocabulary knowledge and knowledge of letters (ibid).
3. To predict BESD, I draw upon the "Strength and Difficulties Questionnaire"- a behavioural screening test, which comprises 5 scales for 5 "problem" areas that parents rate to be: not, somewhat, or certainly true. For all five areas I include the individual aggregated score, along with an additional impact indicator that reflects how strongly these difficulties interfere with the child's life (Goodman 1997).
4. To predict SP needs, I draw upon parents' reports of their children's longstanding limiting illnesses (LSLI) and the International Classification of Diseases (ICD-10). Synthesising these two sources, I generate a variable that takes the value of 0 if parents report no longstanding physical illness, the value of 1 if parents report a longstanding physical illness that is not limiting and the value of 2 if parents report a longstanding limiting physical illness for their child. I also included a dummy variable that indicated whether parents explicitly mentioned problems with hearing or eyesight upon school entry.
5. This indicator is based on 13 questions taken from the Denver Development Screening Test and the MacArthur Communicative Development Inventories, indicating problems with commutative gestures or gross motor skills. Totalling the scores from all these items, a child is assigned the value of 1, if the total score was one standard deviation below the age-adjusted average, which is assumed to indicate mild developmental delays; with the value of 2 if scores were two standard deviations below the average score, which is assumed to indicate more severe delays and with the value of 0 otherwise (c.f. Parsons and Platt 2013,6).
6. Classifications were also obtained based on more and less conservative thresholds, such as probabilities higher than 85% (and lower than 15% respectively) or higher than 65% (and lower than 35% respectively). Results remained comparable.
7. Whereas this distinction roughly corresponds to the Department of Education's definition, all models were also estimated when contrasting only the extreme categories (only English vs. only other language at home). Results remained comparable.
8. While there are other correction techniques available (e.g. PMLE, Firth 1993), this method (relogit) is the only one that allows inclusion of the appropriate weights.
9. As it is not feasible to consider (final) survey weights and to simultaneously account for other nesting structures with unknown probabilities, I prioritised the inclusion of survey weights for the main analyses.
10. Students' abilities in English and Maths are also evaluated within national curriculum assessments. Here scores are determined by (school)teachers' judgements and students' performance on a Standardised Assessment Test (SAT). However, for children who are misidentified with SENs, both of these measures might inaccurately capture their "true" competences. If biased assumptions skew (school)teachers' SEN judgements, their competence ratings might be similarly affected. Moreover, research suggests that stereotype

threat effects are particularly pronounced when a good test performance is desirable to the individual; while misidentified SEN children might experience stereotype threat in any test situation, this might be particularly pronounced when taking the SATs, as here good results might be universally desirable. Students might thus be prone to underperforming on this specific test. Against this background MCS scores might (more) adequately capture the "true" abilities of misidentified children, as they are externally obtained within a "low-cost" setting.

11. Competence acquisition might also vary with the SEN support provided. However, for the aim of this paper, the respective information may bias results for two reasons: *First,* educational support is partially collinear with SEN status (correctly identified children receive no SEN support), which limits the interpretation of coefficient estimates and standard errors. *More importantly though,* this paper aims at assessing *baseline differences* between accurately and inaccurately identified children. SEN provisions reflect *mediating factors* that capture effects of inadequate materials and settings, ability perceptions and (reachable) incentives. Including such information might thus blur baseline discrepancies. Assessing the role of different SEN provisions is beyond the scope of this paper, but could be an avenue future research.

12. For both English and Maths, these scores were generated separately in a three step process. *First,* children were grouped based on their score on the first measure, i.e. at school entry (10 %-percentile groups). *Second*, the average score on the second measure (end of primary school) was calculated for each of these groups; *lastly,* this group-specific average was subtracted from children's individual scores on the second measure. The progress of individual children is thus compared with the progress of those children who started within the same 10^{th} of the ability distribution upon school entry (c.f Parsons and Platt 2014,16f).

Data sources

*University of London, Institute of Education, Centre for Longitudinal Studies. (2017). Millennium Cohort Study:

- Longitudinal Family File, 2001-2015. [data collection]. 2nd Edition. UK Data Service. SN: 8172, http://doi.org/10.5255/UKDA-SN-8172-2
- First Survey, 2001-2003. [data collection]. 12th Edition. UK Data Service. SN: 4683, http://doi.org/10.5255/UKDA-SN-4683-4
- Second Survey, 2003-2005. [data collection]. 9th Edition. UK Data Service. SN: 5350, http://doi.org/10.5255/UKDA-SN-5350-4
- Third Survey, Teacher Survey and Foundation Stage Profile, 2006. [data collection]. UK Data Service. SN: 6847, http://doi.org/10.5255/UKDA-SN-6847-1
- Fourth Survey, 2008. [data collection]. 7th Edition. UK Data Service. SN: 6411, http://doi.org/10.5255/UKDA-SN-6411-7
- Fifth Survey, 2012. [data collection]. 4th Edition. UK Data Service. SN: 7464, http://doi.org/10.5255/UKDA-SN-7464-4

I am grateful to the Centre for Longitudinal Studies (CLS), UCL Institute of Education, for the use of these data and to the UK Data Service for making them available. However, neither CLS nor the UK Data Service bear any responsibility for the analysis or interpretation of these data.

References

Allison, Paul. 2001. Missing data. Vol. 136. Sage.

Artiles, Alfredo, Robert Rueda, Jesús Salazar and Ignacio Higareda. 2002."English language learner representation in special education in California urban school districts." In Racial inequity in special education, edited by Daniel Losen & Gary Orfield (pp.117–135) Cambridge.

Artiles, Alfredo and Janette Klingner. 2006."Forging a knowledge base on English language learners with special needs: Theoretical, population, and technical issues." *Teachers College Record,108*(11):2187–2194.

Artiles, Alfredo, Robert Rueda, Jesús Salazar and Ignacio Higareda. 2005"Within-group diversity in minority disproportionate representation: English language learners in urban school districts" *Exceptional Children71,*(3):283–300

Atkinson, Mark.2015. Interpreting the CANTAB cognitive measures. *CLS:IOE*

Becker, Gary. 1971. The economics of discrimination. Chicago: UCP.

Burgess, Simon and Ellen Greaves.2013."Test scores, subjective assessment, and stereotyping of ethnic minorities." *Journal of Labor Economics,31(3):535-576.*

Burr, Elizabeth, Eric Haas and Karen Ferriere. 2015. Identifying and supporting English learner students with learning disabilities. Washington, DC: U.S. DoE,IES.

Case, Rod and Shannon Taylor. 2005."Language difference or learning disability? Answers from a linguistic perspective." *A Journal of Educational Strategies, Issues and Ideas,78(3):127-131.*

Chatzitheochari, Stella, Samantha Parsons and Lucinda Platt. 2016."Doubly disadvantaged? Bullying experiences among disabled children and young people in England." *Sociology,*50(4):695-713.

Chiswick, Barry and Paul Miller. 1995."The endogeneity between language and earnings: International analyses." *Journal of labor economics,*13(2):246-288.

Chu, Szu-Yin and Sobeida Flores. 2011."Assessment of English language learners with learning disabilities."*A Journal of Educational Strategies, Issues and Ideas,84*(6)

Coser, Lewis. 1956. The Function of Social Conflict. Glencoe, Illinois: Free Press.

De Valenzuela, Julia, Susan Copeland, Cathy Qi and Miwha Park. 2006."Examining educational equity: Revisiting the disproportionate representation of minority students in special education."*Exceptional Children,72(4):425-441.*

Duran, Richard. 1988."Validity and language skills assessment: Non-English background students." *Test validity:105-127.*

England, Paula 1992. Comparable worth: Theories and evidence. New York:Aldine.

Farnia, Fataneh and Ester Geva. 2013."Growth and predictors of change in English language learners' reading comprehension." *Journal of Research in Reading,36(4):389–421.*

Fiske, Susan, Monica Lin and Steven Neuberg. 1999."The continuum model: Ten years later." In Dual-process theories in social psychology, edited by Shelly Chaiken and Yaacov Trope: New York:GP.

Figueroa, Richard. 1989."Psychological testing of linguistic-minority students: Knowledge gaps and regulations." *Exceptional Children,56(2):145-152.*

Figueroa, Richard and Patricia Newsome. 2006."The diagnosis of LD in English learners: Is it nondiscriminatory?" *Journal of Learning Disabilities,(39):* 206–214.

Frosh, Stephen, Ann Phoenix and Rob Pattman. 2002. Young Masculinities. Palgrave:London.

Gandara, Patricia, Julie Maxwell-Jolly and Anne Driscoll. 2005."Listening to Teachers of English Language Learners."*Policy Analysis for California Education,PACE(NJ1)*

Gillborn, David. 1997."Ethnicity and educational performance in the United Kingdom: Racism, ethnicity, and variability in achievement." *Anthropology & Education Quarterly,28(3):375-393.*

Goodman, Robert. *1997.* "The Strengths and Difficulties Questionnaire: a research note."*Journal of child psychology and psychiatry,38(5):581-586.*

Gomolla, Mechtild and Frank-Olaf Radtke. 2003. Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule. Wiesbaden.

Guralnick, Michael 2005. "Early intervention for children with intellectual disabilities:Current knowledge and future prospects." *Journal of Applied Research in Intellectual Disabilities,18(4):313-324.*

Hansen, Kirstine, Jon Johnson and Heather Joshi.2012 Millennium Cohort Study first, second, third and fourth surveys: a guide to the datasets. London:CLS, IoE.

Harrison, Jamie and Joni Lakin. 2018."Mainstream Teachers' Implicit Beliefs about English Language Learners: An IAT Study of Teacher Beliefs". *Journal of Language, Identity & Education,17(2):85-102.*

Hibel, Jacob and Andrea Jasper. 2012."Delayed special education placement for learning disabilities among children of immigrants." *Social Forces,91*(2):503–530.

Huang, Jinyan, Kelly Clarke, Ericka Milczarski and Christie Raby. 2011. "The assessment of English language learners with learning disabilities." *Education 131*(4):732–739.

Ho, Anita 2004. "To be labelled, or not to be labelled: that is the question." *British Journal of Learning Disabilities,32(2):86-92.*

Hyde, Janet, Sara Lindberg, Marcia Linn, Amy Ellis and Caroline Williams. 2008."Gender similarities characterize math performance." *Science,321(5888):494-495.*

Keller-Allen, Chandra 2006. English language learners with disabilities: Identification and other state policies and issues. Alexandria:NASDSE.

Keslair, Francois, Eric Maurin Sandra McNally. 2012."Every child matters? An evaluation of Special Educational Needs programmes in England." *Economics of education review,31*(6):932-948.

King, Gary and Langche Zeng 2001. Logistic regression in rare events data. *Political analysis,9(2):137-163.*

Klingner, Janette, Alfredo Artiles and Laura Méndez-Barletta. 2006. "English language learners who struggle with reading: Language acquisition or LD?" *Journal of Learning Disabilities,39*(2):108–128.

Klingner, Janette and Beth Harry. 2006."The special education referral and decision-making process for English language learners: Child study team meetings and placement conferences." *Teachers College Record,108*(11):2247–2281.

Lane, Suzanne and Brian Leventhal. 2015."Psychometric challenges in assessing English language learners and students with disabilities." *Review of Research in Education,39(1):165-214.*

Limbos, Marjolaine and Ester Geva. 2001."Accuracy of teacher assessments of second-language students at risk for reading disability". *Journal of learning disabilities,34(2):136-151.*

Lindsay, Geoff and Steve Strand. 2016. "Children with language impairment: prevalence, associated difficulties and ethnic disproportionality in an English population." *Frontiers in Education,1(2).*

Lord, Frederic. 1967."A paradox in the interpretation of group comparisons." *Psychological bulletin,68(5).*

MacGhaill, Mairtin and Chris Haywood. 2015."British-born Pakistani and Bangladeshi young men: Exploring unstable concepts of Muslim, Islamophobia and racialization." *Critical Sociology,41(1):97-114.*

Madon, Stephanie, Lee Jussim and Jacquelynne Eccles. 1997."In search of the powerful self-fulfilling prophecy." *Journal of personality and social psychology,72(4):791.*

MacSwan, Jeff and Kellie Rolstad. 2006."How language proficiency tests mislead us about ability: Implications for English language learner placement in special education." *Teachers College Record,108*(11):2304–2328.

Mercer, Jane and Wayne Brown. 1973."Racial differences in IQ:Fact or artifact." *New York:TWP, pp.56-113.*

Ortiz, Alba. and Elba Maldonado-Colón. 1986."Recognizing learning disabilities in bilingual children" *Journal of Reading, Writing, and Learning Disabilities International,2(1):43-56.*

Parsons, Samantha and Lucinda Platt. 2014."Disabled children's cognitive development in the early years." *DoQSS WorkingPaper(14-15).*

Pettigrew, Thomas. 1979. "The ultimate attribution error: Extending Allport's cognitive analysis of prejudice." *Personality and social psychology bulletin,5(4):461-476.*

Pettit, Stacie. 2011. "Teachers' beliefs about English language learners in the mainstream classroom: A review of the literature." *International Multilingual Research Journal,5(2):123-147.*

Phillips, Coretta. 2011. "Institutional racism and ethnic inequalities: An expanded multilevel framework." *Journal of social policy,40(1):173-192.*

Plewis, Ian, Lisa Calderwood, Denise Hawkes, Gareth Hughes and Heather Joshi. 2007. "Millennium cohort study: technical report on sampling." London:CLS,IOE.

Powell, Justin. 2003."Constructing disability and social inequality early in the life course: The case of special education in Germany and the US." *Disability Studies Quarterly, 23(2).*

Rosenthal, Robert. 1974." On the social psychology of the self-fulfilling prophecy." *New York:MSS-Modular.*

Samson, Jennifer and Nonie Lesaux. 2009."Language-minority learners in special education: Rates and predictors of identification for services." *Journal of Learning Disabilities,42*(2):148–162.

Shifrer, Dara, Chandra Muller and Rebecca Callahan. 2011."Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language." *Journal of Learning Disab.,44(3):246-257.*

Steele, Claude and Joshua Aaronson. 1995."Stereotype Threat and the Intellectual Test Performance of African Americans." *69 J. of Personality & Soc.*

Strand, Steve, Lars Malmberg and James Hall. 2015."English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database." DfE.

Strand, Steve and Geoff Lindsay. 2009."Evidence of ethnic disproportionality in special education in an English population." *The Journal of Special Education, 43*(3):174-190

Strand, Steve and Geoff Lindsay. 2012."Ethnic Disproportionality in the Identification of Speech Language and Communication Needs and Autistic Spectrum Disorders: 2005-2011."DfE.

Storm, Ingrid, Maria Sobolewska and Robert Ford. 2017."Is ethnic prejudice declining in Britain? Change in social distance attitudes among ethnic majority and minority Britons." *The British journal of sociology, 68(3):410-434.*

Sullivan, Amanda 2011."Disproportionality in special education identification and placement of English language learners." *Exceptional Children 77*(3):317–334.

Sullivan, Amanda and Alfredo J. Artiles.2011."Theorizing racial inequity in special education." *Urban Education 46.6:1526-1552.*

Tajfel, Henry and John Turner. 1986: The social identity theory of intergroup behavior. Chicago: Nelson-Hall.

Tomlinson, Sally. 2016."Special education and minority ethnic young people in England." *Studies in the Cultural Politics of Education,37(4),513-528.*

Tversky, Amos and Daniel Kahneman. 1992."Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and uncertainty, 5(4):297-323.*

Von Hippel, Paul 2007."Regression with missing Ys: An improved strategy for analyzing multiply imputed data." *Sociological Methodology,37(1):83-117.*

Werfhorst, Herman, Alice Sullivan and Sin Yi Cheung. 2003."Social class, ability and choice of subject in secondary and tertiary education in Britain." *British Educational Research,29(1):41-62.*

Wolters, Christopher and Paul Pintrich 1998."Contextual differences in student motivation and self-regulated learning in mathematics, English, and social-studies classrooms." *Instructional science26: 27-47.*

Youdell, Deborah. 2003."Identity traps or how black students fail." *British Journal of Sociology of Education 24(1):3-20.*

Zhang, Qin. 2010.Asian Americans beyond the model minority stereotype. *Journal of international and intercultural communication 3(1):20-37.*

Table 1. Distribution of SEN (mis)identification by SEN type.

	SEN type			
	Judgemental			Non-Judgemental
	BESD	CI	CL	SP
Accurate identification	97,0 % (4,333)	96,5 % (4,464)	96,8 % (4,358)	97,1 % (4,298)
Over-identification	2,9 % (127)	3,4 % (159)	3,2 % (145)	2,8 % (123)
Under-identification	0,1 % (3)	0,1 % (2)	0,0 % (0)	0,1 % (3)
Σ	100 % (4,463)	100 % (4,625)	100 % (4503)	3,213 (100.00)

Note: percentages (%) and absolute numbers (N)

Table 2. Logistic models regressing SEN over-identification on different aspects of language minority status (average marginal effects, selected coefficients).

	Judgemental												*Non- Judgemental*			
	BESD				*CI*				*CL*				*SP*			
Ethnic group (Ref. White Maj.)	(M 1)	(M 2)	(M 3)	(M 4)	(M 1)	(M 2)	(M 3)	(M 4)	(M 1)	(M 2)	(M 3)	(M 4)	(M 1)	(M 2)	(M 3)	(M 4)
Indian	-.01	**-.02+**	-.02	.04	**-.04*** **	**-.04*** **	**-.04*** **	**-.14*** **	**-.04*** **	**-.04*** **	**-.04*** **	**-.17*** **	**-.02***	**-.03*** **	**-.03*** **	**-.22***
	(.01)	**(.01)**	(.01)	(.12)	**(.01)**	**(.01)**	**(.00)**	**(.04)**	(.01)	(.01)	(.01)	**(.04)**	**(.01)**	**(.01)**	**(.01)**	**(.10)**
Pakis. & Bangl.	**-.02*** **	**-.03*** **	**-.03*** **	-.07	**-.04*** **	**-.04*** **	**-.04*** **	**-.13*** **	**-.02*** **	**-.03*** **	**-.03*** **	**-.13** **	**-.02*** **	**-.03*** **	**-.03*** **	-.10
	(.01)	**(.01)**	**(.01)**	(.05)	**(.00)**	**(.00)**	**(.00)**	**(.04)**	**(.01)**	**(.01)**	**(.01)**	**(.05)**	**(.01)**	**(.01)**	**(.01)**	(.09)
Black British	.01	.00	.00	**.40+**	-.01	-.02	**-.02+**	-.02	-.02	**-.02***	**-.02***	.02	-.01	-.02	-.02	**-.20***
	(.01)	(.01)	(.01)	**(.24)**	(.01)	(.01)	**(.01)**	(.11)	(.01)	**(.01)**	**(.01)**	(.29)	(.02)	(.02)	(.02)	(.09)
Chinese & East Asian	**-.03*** **	**-.03*** **	**-.03*** **	-.06	**-.04*** **	**-.04*** **	**-.04*** **	**-.15*** **	**-.04*** **	**-.04*** **	**-.04*** **	**-.17*** **	**-.03*** **	**-.03*** **	**-.03*** **	-.15
	(.01)	**(.01)**	**(.01)**	(.08)	**(.00)**	**(.00)**	**(.00)**	**(.04)**	**(.00)**	**(.00)**	**(.00)**	**(.04)**	**(.00)**	**(.01)**	**(.01)**	(.12)
Mixed & Other	-.01	-.01	-.01	-.08	-.01	-.02	-.02	-.01	-.03	**-.03*** **	**-.03*** **	-.04	**-.03****	**-.03*** **	**-.02*** **	-.06
	(.01)	(.01)	(.01)	(.07)	(.02)	(.01)	(.01)	(.11)	(.01)	**(.01)**	**(.01)**	(.09)	**(.01)**	**(.01)**	**(.01)**	(.15)
Language minority status																
Language minority exposure (LE)	**.03+**	.02	.10		**-.02***	**-.03*** **	**-.14*** **		**-.02*** **	**-.03*** **	-.08		.02	-.00	.13	
	(.02)	(.02)	(.09)		**(.01)**	**(.01)**	**(.04)**		**(.01)**	**(.01)**	(.08)		(.01)	(.01)	(.19)	
Second language proficiency (LP)		**-.01*** **	**-.01*** **	**-.01*** **		**-.02*** **	**-.02*** **	**-.02*** **		**-.02*** **	**-.02*** **	**-.02*** **		**-.02*** **	**-.02*** **	**-.03*** **
		(.00)	**(.00)**	**(.01)**		**(.00)**	**(.00)**	**(.00)**		**(.00)**	**(.00)**	**(.00)**		**(.00)**	**(.00)**	**(.00)**
Interactions																
LE x LP			-.02				**.02*** **				**.01***				-.02	
			(.01)				**(.01)**				**(.01)**				(.02)	
Ethnic group x LP																
Indian x LP				-.01				**.02*** **				**.02*** **				.02
				(.01)				**(.00)**				**(.00)**				(.01)
Pakis. & Bangl. x LP				-.00				**.01*** **				.01				-.02
				(.01)				(.00)				(.01)				(.01)
Black British x LP				**-.07***				-.01				-.00				.01
				(.00)				(.02)				(.02)				(.02)
Chinese&East Asian x LP				.00				**.02*** **				**.02*** **				.01
				(.01)				**(.00)**				**(.00)**				(.01)
Mixed&Other x LP				.01				-.00				.00				.00
				(.01)				(.01)				(.01)				(.01)
MZ R2	.31	.34	.34	.34	.19	.23	.23	.24	.19	.24	.26	.25	.18	.28	.29	.33
N	4460				4623				4503				4421			

Note: dependent variable: overestimation on the respective type (yes/no); controls included; ***p<.001; p<.01;*p<0.05; +p<.01; significant coefficients are **bold**

Table 3. Linear models regressing English and Maths competence on SEN accuracy (selected coefficients).

	English competence				*Math competence*							
	BAS naming vocabulary (age 11)				*SWM test (age 11)*				*NFER number skills (age 7)**			
	BESD	CI	CL	SP	BESD	CI	CL	SP	BESD	CI	CL	SP
SEN misidentification (Ref. correct no SEN - whole primary)												
correct SEN identification - whole primary edu	**-5.3***	**-7.5*****	**-8.3*****	-7.1	-0.9	**-1.2+**	**-2.4*****	-1.3	**-9.8*****	**-13.2*****	**-16.7*****	**-9.2*****
	(2.0)	**(1.3)**	**(1.8)**	(4.7)	(1.1)	**(0.6)**	**(0.6)**	(1.2)	**(1.6)**	**(1.1)**	**(1.3)**	**(2.0)**
incorrect SEN identification - whole primary edu	**-4.3***	**-4.2***	**-4.6****	**-9.0***	**-1.4***	**-1.7***	**-1.8*****	-1.1	**-10.3*****	**-10.2*****	**-10.8*****	**-12.1*****
	(1.6)	**(1.7)**	**(1.7)**	**(3.9)**	**(0.6)**	**(0.8)**	**(0.5)**	(1.3)	**(1.4)**	**(1.6)**	**(1.3)**	**(1.4)**
incorrect SEN identification - part primary edu	-2.2	**-2.9***	**-3.5+**	**-3.8*****	-0.7	-0.6	-1.2	-1.1				
	(1.7)	**(1.2)**	**(2.1)**	**(1.4)**	(1.1)	(0.7)	(1.0)	(0.8)				
other	-1.4	**-2.8***	**-3.5*****	-2.2	-1.2	**-1.1+**	**-1.3*****	-0.3				
	(2.0)	**(1.3)**	**(1.0)**	(1.8)	(0.8)	**(0.6)**	**(0.5)**	(0.6)				
N	4460	4623	4530	4421	4460	4623	4503	4421	4460	4623	4503	4421
R2	.17	.21	.20	.18	.06	.06	.07	.06	.22	.26	.25	.22

Note: controls included;***p<.001; p<.01;*p<0.05; +p<.01; significant coefficients are **bold.** * NFER scores are obtained when there is only one report of children's SEN status available (age 7). Category changes are thus not considered in the respective analyses.

Figure 1. Interaction effects, margin plots (CI)

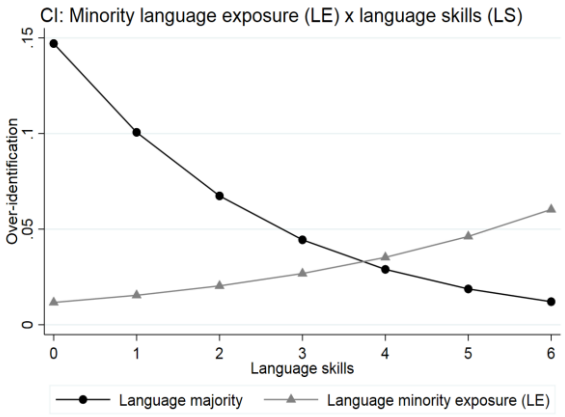


Figure 2. Interaction effects, margin plots (CL)

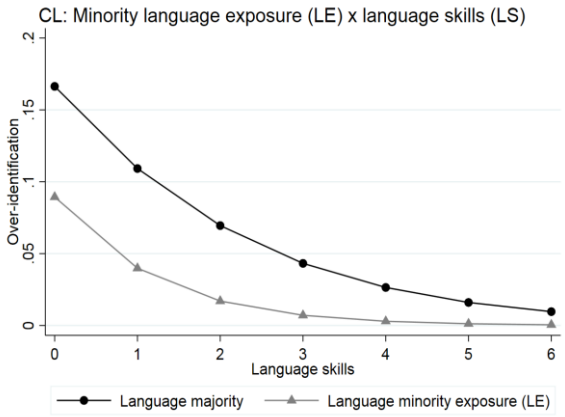
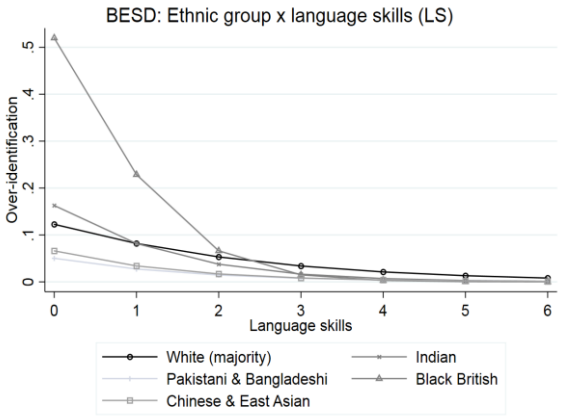
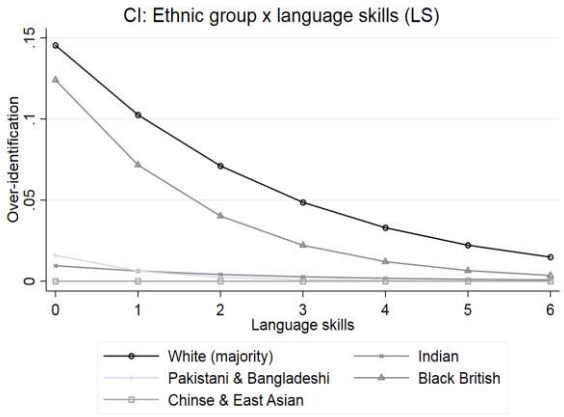


Figure 3. Interaction effects, margin plots (BESD)



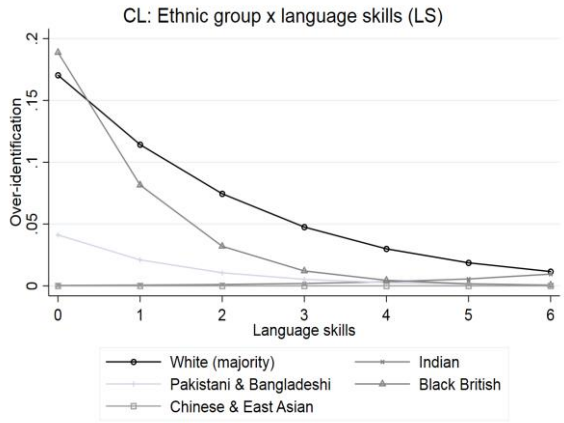
Note: Excluding mixed and other origins to facilitate ease of reading.

Figure 4. Interaction effects, margin plots (CI)



Note: Excluding mixed and other origins to facilitate ease of reading

Figure 5. Interaction effects, margin plots (CL)



Note: Excluding mixed and other origins to facilitate ease of reading

Appendix: Table 4. Sample characteristics.

Analyses I (Table 2)	BESD (n=4460) Distribution of model variables							CI (n=4623) Distribution of model variables							CL (n=4503) Distribution of model variables							SP (n=4421) Distribution of model variables						
	%	(N)	% MI	Mean	SE	Min.	Max.	%	(N)	% MI	Mean	SE	Min.	Max.	%	(N)	% MI	Mean	SE	Min.	Max.	%	(N)	% MI	Mean	SE	Min.	Max.
SEN identification accuracy	(100)	(4,460)	0.0					(100)	(4,623)	0.0					(100)	(4,503)	0.0					(100)	(4,421)	0				
Accurate identification	97.2	4333						96.56	4464						96.78	4358						97.22	4298					
Over identification	2.9	127						3.44	159						3.22	145						2.78	123					
Ethnic grop	(100)	(4,460)	0.0					(100)	(4,623)	0.0					(100)	(4,503)	0.0					(100)	(4,421)	0				
White	80.22	3578						80.25	3710						80.68	3633						80.32	3551					
Indian	3.54	158						3.48	161						3.51	158						3.62	160					
Pakistani & Bangladeshi	6.88	307						7.14	330						6.84	308						7.06	312					
Black British	3.97	177						3.85	178						3.69	166						3.66	162					
Chinese & East Asian	1.23	55						1.19	55						1.18	53						1.22	54					
Mixed & Other	4.15	185						4.09	189						4.11	185						4.12	182					
Language minority status																												
L2 Language exposure (LE)	(100)	(4,460)	0.0					(100)	(4,623)	0.0					(100)	(4,503)	0.0					(100)	(4,421)	0				
Language majority	91.39	4076						91.41	4226						91.67	4128						91.36	4039					
Language minority	8.61	384						8.59	397						8.33	375						8.64	382					
*L2 Language skills (LS)**		(4,259)	4.57	3.57	1.08	0	6		(4,046)	4.69	3.52	1.12	0	6		(4,303)	4.44	3.56	1.09	0	6		(4,233)	4.25	3.57	1.09	0	6
Controls																												
*Poverty (mean n., sweep 1-4)**		(4,458)	0.04	0.27	0.37	0	1		(4,621)	0.04	0.27	0.37	0	1		(4,501)	0.04	0.26	0.37	0	1		(4,419)	0.05	0.26	0.36	0	1
*Parents' social class (sweep 1-4, highest)**	97. 83	(4,363)	2.17					97.88	(4,525)	2.12					97.80	(4,404)	2.20					97.96	(4,331)	2.04				
Managers & Prof.	52.19	2277						51.73	2341						52.41	2308						52.41	2270					
Routine non-manual	14.12	616						14.25	645						14.06	619						14.13	612					
Self-employed	12.56	548						12.53	567						12.76	562						12.45	539					
Low grade tech.	10.04	438						10.10	457						9.90	436						10.18	441					
Semi-routine/routine workers	11.09	4363						11.38	515						10.88	479						10.83	469					
*Parental education (sweep 1-4, highest)**	99.98	(4,459)	0.02					99.98	(4,622)	0.02					99.98	(4,502)	0.02					99.98	(4,420)	0.02				
NVQ level 1	4.19	187						4.09	189						4.06	183						4.14	183					
NVQ level 2	21.57	962						21.79	1007						21.50	968						21.22	938					
NVQ level 3	14.87	663						14.67	678						15.02	676						15.02	664					
NVQ level 4	38.62	1722						38.45	1777						38.52	1734						38.78	1714					
NVQ level 5	13.66	609						13.57	627						13.86	624						13.85	612					
None of these	4.82	215						2.34	108						4.86	219						4.80	212					
Overseas qualification	2.27	101						5.1	236						2.18	98						2.19	97					
Parents school involvment (n)		(4,460)	0	1.56	0.96	1	6		(4,623)	0	1.56	0.96	1	6		(4,503)	0	1.56	0.97	1	6	(4,421)	0	1.57	0.97	1	6	
Students age (month, first int.)		(4,460)	0	15.10	3.42	10	21		(4,623)	0	15.14	3.41	10	21		(4,503)	0	15.11	3.42	10	21	(4,421)	0	15.09	3.42	10	21	
Students gender	(100)	(4,460)	0.0					(100)	(4,623)	0.0					(100)	(4,503)	0.0					(100)	(4,421)	0				
Female	51.77	2309						51.78	2394						52.52	2365						53.18	2351					
Male	48.23	2151						48.22	2229						47.48	2138						46.82	2070					
School year	(100)	(4,460)	0.0					(100)	(4,623)	0.0					(100)	(4,503)	0.0					(100)	(4,421)	0				
Year 1	0.27	12						0.32	15						0.27	12						0.25	11					
Year 2	94.28	4205						94.38	4363						94.38	4250						94.37	4172					
Year 3	5.36	239						5.15	238						5.24	236						5.27	233					
Other	0.09	4						0.15	7						0.11	5						0.11	5					

Analyses II (Table 3)	BESD (n=4460) Distribution of model variables							CI (n=4623) Distribution of model variables							CL (n=4503) Distribution of model variables							SP (n=4421) Distribution of model variables						
	%	(N)	% MI	Mean	SE	Min	Max	%	(N)	% MI	Mean	SE	Min	Max	%	(N)	% MI	Mean	SE	Min	Max	%	(N)	% MI	Mean	SE	Min	Max
English competence																												
BAS naming voc (age 11)		(3,940)	11.7	59.9	9.0	20	80		(4,074)	11.88	59.55	9.41	20	80		(3,977)	11.68	59.78	9.25	20	80		(3,914)	11.47	59,89	9.18	20	80
Math competence																												
SWM score (age 11)		(3,821)	14.33	33.54	5.89	0	48		(3,952)	14.51	33.61	5.94	0	48		(3,860)	14.28	33.59	5.87	0	48		(3,797)	14.11	33.54	5.87	0	48
NFER number skills (age 7)		(4,423)	0.83	31.39	14.94	0	67		(4,571)	1.12	30.82	15.17	0	67		(4,467)	0.80	31.11	15.04	0	67		(4,387)	0.77	31.41	14.88	0	67
SEN status primary school carrer*	63.45	(2,830)	36.55					62.97	(2,911)	37,03					64.76	(2,916)	35.24					63.13	(2,791)	36.87				
correct no SEN - whole primary edu	95.65	2,707						92,99	2707						92.87	2708						96.95	2706					
correct SEN - whole primary edu	0.78	22						2.13	62						1.34	39						0.36	10					
incorrect SEN - whole primary edu	0.74	21						0.72	21						1.30	38						0.43	12					
incorrect SEN - part primary edu	1.10	31						1.82	53						0.89	26						0.93	26					
other	1.73	49						2.34	68						3.60	105						1.33	37					
Ethnic grop	100.00	(4,460)	0					100.00	(4623)	0					(100)	(4,503)	0.0					(100)	(4,421)	0				
White	80.22	3578						80.25	3710						80.68	3633						80.32	3551					
Indian	3.54	158						3.48	161						3.51	158						3.62	160					
Pakistani & Bangladeshi	6.88	307						7.14	330						6.84	308						7.06	312					
Black British	3.97	177						3.85	178						3.69	166						3.66	162					
Chinese & East Asian	1.23	55						1.19	55						1.18	53						1.22	54					
Mixed & Other	4.15	185						4.09	189						4.11	185						4.12	182					
Language minority status																												
L2 Language exposure (LE)	(100)	(4,460)	0.0					(100)	(4,623)	0.0					(100)	(4,503)	0.0					(100)	(4,421)	0				
Language majority	91.39	4076						91.41	4226						91.67	4128						91.36	4039					
Language minority	8.61	384						8.59	397						8.33	375						8.64	382					
*L2 Language skills (LS)**		(4,259)	4.57	3.57	1.08	0	6		(4,046)	4.69	3.52	1.12	0	6		(4,303)	4.44	3.56	1.09	0	6		(4,233)	4.25	3.57	1.09	0	6
Controls																												
Poverty (mean n., sweep 1-5)		(4.460)	0.00	0.25	0.34	0	1		(4,623)	0.00	0.25	0.349	0	1		(4,503)	0.00	0.24	0.34	0	1		(4,421)	0.00	0.24	0.34	0	1
*Parents' social class (sweep 1-5, highest)**	98.12	(4,376)	1.88					98.16	(4,538)	1.84					98.09	(4,417)	1.91					98.24	(4,343)	1.76				
Managers & Prof.	52.49	2297						51.98	2359						52.71	2328						52.73	2290					
Routine non-manual	15.54	680						15.69	712						15.46	683						15.57	676					
Self-employed	12.50	547						12.56	567						12.70	561						12.43	540					
Low grade tech.	9.64	422						9.65	457						9.53	421						9.72	422					
Semi-routine/routine workers	9.83	430						10.11	515						9.60	424						9.56	415					
*Parental education (sweep 1-5, highest)**	96.35	(4,297)	3.65					96.52	(4,462)	3.48					96.51	(4,346)	3.49					96.58	(4,270)	3.42				
NVQ level 1	3.82	164						3.79	169						3.77	164						3.77	161					
NVQ level 2	19.60	842						19.70	879						19.42	844						19.32	825					
NVQ level 3	14.27	613						14.14	631						14.47	629						14.43	616					
NVQ level 4	39.17	1683						39.00	1740						38.96	1693						39.25	1676					
NVQ level 5	15.59	670						15.35	685						15.74	684						15.67	669					
None of these	5.24	225						5.56	248						5.32	231						5.29	226					

Overseas qualification	2.33	100						2.47	110						2.32	101						2.27	97					
Students age (month, first int.)		(4.460)	0	15.1	3.4	10	21		(4,623)	0	15.14	3.4	10	21		(4,503)	0	15.11	3.42	10	21		(4,421)	0	15.09	3.42	10	21
Students gender	100.00	(4.460)	0					100.00	(4,623)	0					(100)	(4,503)	0.0					(100)	(4,421)	0				
Female	51.77	2309						51.78	2394						52.52	2365						53.18	2351					
Male	48.23	2151						48.22	2229						47.48	2138						46.82	2070					
	99.98	(4.459)	0.02					100.00	(4,623)	0					(100)	(4,503)	0.0					(100)	(4,421)	0				
*Additional classes (age 7)**																												
No additional classes	94.68	4222						94.64	4375						94.71	4265						94.59	4182					
Math	1.55	69						1.58	73						1.58	71						1.70	75					
Reading	0.74	33						0.82	38						0.80	36						0.75	33					
Writing	3.03	135						2.96	137						2.91	131						2.96	131					
*Additional classes (age 11)**	89.17	(3,977)	10.83					89.21	(4,124)	10.79					89.16	(4,015)	10.84					89.46	(3955)	10.54				
No additional classes	77.14	3068						76.96	3174						76.84	3085						77.24	3055					
English	3.5	139						3.73	154						3.74	150						3.54	140					
Math	17.48	695						17.43	719						17.48	702						17.32	685					
Science	1.89	75						1.87	77						1.94	78						1.90	75					
*Bullying (n, age 7)**		(4,272)	4.22	2.45	0.63	1	3		(4,399)	4.85	2.45	0.63	1	3		(4,315)	4.17	2.45	0.62	1	3		(4,239)	4.12	2.45	0.62	1	3
*Hurt or picked on (n, age 11)**		(3,907)	12.40	4.70	1.53	1	6		(4,031)	12.81	4.70	1.54	1	6		(3,937)	12.57	4.71	1.53	1	6		(3,881)	12.21	4.71	1.52	1	6

**variables contain missing values, imputed for analyses*

Supplement: Table 5. Robustness of main models (Table 2), additional competence indicators used to predict SEN misidentification (overestimation shown).

	Judgemental												*Non- Judgemental*			
	BESD				*CI*				*CL*				*SP*			
Ethnic group (Ref. White Maj.)	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*
Indian	-.01	**-.02**	-.01	.08	**-.04*****	**-.04*****	**-.04*****	**-.14*****	**-.03*****	**-.03*****	**-.03*****	**-.16*****	-.01	-.01	-.01	-.05
	(.01)	**(.01)**	(.01)	(.11)	**(.01)**	**(.00)**	**(.00)**	**(.04)**	(.01)	(.01)	(.01)	**(.05)**	(.01)	(.01)	(.01)	(.10)
Pakis. & Bangl.	**-0.3*****	**-.03*****	**-0.3*****	-.04	**-.04*****	**-.04*****	**-.04*****	**-.14*****	**-.03*****	**-.03*****	**-.03*****	**-.13*****	**-.02*****	**-.03*****	**-.03*****	**-.10*****
	(0.1)	**(.01)**	**(0.1)**	(.04)	**(.00)**	**(.00)**	**(.00)**	**(.04)**	**(.00)**	**(.00)**	**(.00)**	**(.05)**	**(.00)**	**(.01)**	**(.00)**	**(.03)**
Black British	.01	.00	.00	**.49***	-.01	**-.02+**	**-.02+**	-.02	**-.02***	**-.02*****	**-.02*****	**-.14*****	-.00	-.01	-.01	-.04
	(.01)	(.01)	(.02)	**(.24)**	(.01)	**(.01)**	**(.01)**	(.11)	**(.01)**	**(.01)**	**(.01)**	**(.05)**	(.02)	(.02)	(.02)	(.08)
Chinese & East Asian	**-.02*****	**-.03*****	**-.03*****	-.01	**-.04*****	**-.04*****	**-.04*****	**-.15*****	**-.03*****	**-.03*****	**-.03*****	**-.16*****	**-.03*****	**-.03*****	**-.03*****	**-.10*****
	(.01)	**(.01)**	**(.01)**	(.01)	**(.00)**	**(.00)**	**(.00)**	**(.04)**	**(.00)**	**(.00)**	**(.00)**	**(.05)**	**(.00)**	**(.00)**	**(.00)**	(.03)
Mixed & Other	-.01	-.01	-.01	-.04	-.02	-.02	-.02	-.04	**-.03*****	**-.03*****	**-.03*****	**-.10+**	-.01	-.01	**-.02*****	-.05
	(.01)	(.01)	(.01)	(.06)	(.01)	(.01)	(.01)	(.10)	**(.01)**	**(.01)**	**(.01)**	**(.06)**	(.01)	(.01)	**(.01)**	(.06)
Language minority status																
Language minority exposure (LE)	.03	.02	.13		**-.02+**	**-.02*****	**-.13*****		**-.02***	**-.02*****	**-.10+**		.01	-.00	.00	
	(.02)	(.02)	(.09)		**(.01)**	**(.01)**	**(.04)**		**(.01)**	**(.01)**	**(.06)**		(.01)	(.01)	(.10)	
Second language proficiency (LP)		**-.01*****	**-.01*****	**-.01****		**-.02*****	**-.02*****	**-.02*****		**-.02*****	**-.01*****	**-.02*****		**-.01*****	**-.01*****	**-.01*****
		(.00)	**(.00)**	**(.00)**		**(.00)**	**(.00)**	**(.00)**		**(.00)**	**(.00)**	**(.00)**		**(.00)**	**(.00)**	**(.00)**
Interactions																
LE x LP			-.02				**.02*****				**.01*****				.00	
			(.01)				**(.01)**				**(.00)**				(.01)	
Ethnic group x LP																
Indian x LP				-.01				**.02*****				**.02*****				.00
				(.01)				**(.00)**				**(.00)**				(.01)
Pakis. & Bangl. x LP				-.00				**.01*****				**.01*****				**.01*****
				(.01)				(.00)				**(.00)**				(.00)
Black British x LP				**-.07*****				-.01				**.02+**				.00
				(.03)				(.02)				**(.01)**				(.02)
Chinese&East Asian x LP				-.00				**.02*****				**.02*****				**.01*****
				(.01)				**(.00)**				**(.00)**				**(.00)**
Mixed&Other x LP				.00				.00				**.01+**				.00
				(.01)				(.01)				**(.01)**				(.01)
MZ R2	.29	.31	.31	.31	.17	.21	.21	.21	.20	.24	.24	.25	.16	.19	.19	.19
N	4460				4624				4502				4421			

Note: logistic models (AMEs), DV: overestimation on type (yes/no); controls included; ***p<.001; p<.01;* p<0.05; +p<.01; significant coefficients are **bold;** Prediction of BESD additionally included BAS and BRACKEN subsets[2], as well as ICD-10 information[4;] Prediction of CI and CL additionally included SDQ scales[3] and ICD-10 information[4]; Prediction of SP additionally included BAS and BRACKEN subsets[2] , as well as SDQ scales[3]

Supplement: Table 6. Robustness of main models II (Table 3), "value added" scores[12].

	English competence				*Math competence*							
	BAS naming vocabulary (age 11)				*SWM test (age 11)*				*NFER number skills (age 7)**			
	BESD	CI	CL	SP	BESD	CI	CL	SP	BESD	CI	CL	SP
SEN misidentification (Ref. correct no SEN - whole primary)												
correct SEN identification - whole primary edu	**-5.1*****	**-7.6*****	**-7.8*****	-7.5	-1.0	**-1.2+**	**-2.5*****	-1,4	**-11.7*****	**-13.7*****	**-15.4*****	**-9.9*****
	(1.7)	**(1.3)**	**(1.8)**	(5.2)	(1.1)	**(0.7)**	**(0.7)**	(1.3)	**(1.8)**	**(1.8)**	**(1.4)**	**(2.1)**
incorrect SEN identification - whole primary edu	**-4.2*****	**-3.9***	**-4.3***	**-9.7***	**-1.6***	**-1.7+**	**-1.9*****	-1.7	**-10.7*****	**-10.4*****	**-11.6*****	**-13.4*****
	(1.5)	**(1.9)**	**(1.7)**	**(4.3)**	**(0.7)**	**(0.9)**	**(0.6)**	(1.4)	**(1.5)**	**(1.2)**	**(1.2)**	**(1.5)**
incorrect SEN identification - part primary edu	-2.5	**-3.1***	**-3.7+**	**-3.7***	-0.8	-0.8	-1.3	-1.0				
	(1.6)	**(1.2)**	**(2.2)**	**(1.5)**	(1.1)	(0.7)	(1.0)	(0.7)				
other	-0.7	**-3.1***	**-3.2*****	-2.1	-1.2	**-1.0+**	**-1.3***	-0.5				
	(2.0)	**(1.1)**	**(0.9)**	(1.6)	(0.9)	**(0.6)**	**(0.6)**	(0.7)				
N	4460	4623	4530	4421	4460	4623	4503	4421	4460	4623	4503	4421
(R2)	(.07)	(.08)	(.08)	(.07)	(.04)	(.04)	(.04)	(.04)	(.10)	(.13)	(.13)	(.10)

Note: linear regressions, select coefficients, controls included;***p<.001; p<.01;*p<0.05; +p<.01; significant coefficients are **bold.*** NFER scores are obtained when there is only one report of children's SEN status available (age 7). Category changes are thus not considered in the respective analyses.

Supplement: Table 7, additional analyses I, logistic regressions, rare events correction (relogit, selected coefficients).

	Judgemental												*Non- Judgemental*			
	BESD				*CI*				*CL*				*SP*			
Ethnic group (Ref. White Maj.)	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*
Indian	-.58	**-.67**	-.64	.36	**-2.5***	**-2.7****	**-2.8****	**-2.9***	**-2.4***	**-2.6***	**-2.6***	**-6.3*****	-.98	**-1.3+**	-1.3	-2.2
	(.05)	**(.43)**	(.43)	(1.0)	**(1.0)**	**(1.0)**	**(1.0)**	**(1.1)**	(1.0)	(1.0)	(1.0)	**(1.1)**	(.72)	**(.75)**	(.82)	(2.1)
Pakis. & Bangl.	**-1.5***	**-1.7****	**-1.8****	-1.1	**-2.4****	**-2.8*****	**-2.8*****	**-2.4****	**-1.0***	**-1.5****	**-1.5****	**-1.6***	**-1.1***	**-1.6*****	**-1.8*****	-.65
	(.60)	**(.60)**	**(.60)**	(.80)	**(.74)**	**(.73)**	**(.74)**	**(.82)**	**(.53)**	**(.53)**	**(.53)**	**(.77)**	**(.44)**	**(.42)**	**(.42)**	(.71)
Black British	.28	.07	.08	**2.5***	-.31	-.54	-.55	-.21	-.60	-.84	-.83	.12	-.31	-.67	-.62	-1.9
	(.43)	(.41)	(.41)	**(1.1)**	(.48)	(.46)	(.47)	(1.1)	(.60)	(.58)	(.58)	(1.8)	(.61)	(.63)	(.63)	(1.8)
Chinese & East Asian	**-1.5+**	**-1.7***	**-1.6***	-.80	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	**-1.8+**	**-2.2***	**-2.2***	-1.1
	(.83)	**(.83)**	**(.80)**	(1.4)	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	n.e.	**(1.1)**	**(1.1)**	**(1.1)**	(1.4)
Mixed & Other	-.44	-.50	-.50	-1.1	-.47	-.54	-.52	-.10	**-1.1+**	**-1.2+**	**-1.2***	-.41	**-1.3***	**-1.3***	**-1.4***	-.39
	(.61)	(.63)	(.63)	(1.4)	(.56)	(.55)	(.56)	(.92)	**(.62)**	**(.61)**	**(.61)**	(.84)	**(.62)**	**(.60)**	**(.55)**	(1.1)
Language minority status																
Language minority exposure (LE)	**.90***	.47	.82		-.81	**-1.2+**	**-2.7****		**-1.2***	**-1.4****	-.74		.43	-.10	.70	
	(.40)	(.40)	(.67)		(.60)	**(.60)**	**(.60)**		**(.60)**	**(.51)**	(.98)		(.40)	(.40)	(.81)	
Second language proficiency (LP)		**-.54*****	**-.52*****	**-.50*****		**-.44*****	**-.45*****	**-.41*****		**-.52*****	**-.52*****	**-.49*****		**-.77*****	**-.75*****	**-.75*****
		(.12)	**(.13)**	**(.13)**		**(.09)**	**(.09)**	**(.09)**		**(.10)**	**(.10)**	**(.10)**		**(.10)**	**(.11)**	**(.11)**
Interactions																
LE x LP			-.19				**.74+**				-.38				-.51	
			(.28)				**(.40)**				(.53)				(.35)	
Ethnic group x LP																
Indian x LP				-.37				-.01				**1.0*****				.41
				(.45)				(.15)				**(.19)**				(.70)
Pakis. & Bangl. x LP				-.16				**-.53***				-2.2				**-.94+**
				(.30)				**(.23)**				(.43)				**(.51)**
Black British x LP				**-1.1***				-.21				-.52				.51
				(.60)				(.44)				(.85)				(.66)
Chinese&East Asian x LP				-.23				n.e.				n.e.				**-.92+**
				(.64)				n.e.				n.e.				**(.50)**
Mixed&Other x LP				.23				-.21				-.34				-.44
				(.37)				(.23)				(.26)				(.47)
N	4460				4623				4503				4421			

Note: coefficients shown, dependent variable: overestimation on type (yes/no); controls included; ***p<.001; p<.01;*p<0.05; +p<.01; significant coefficients are **bold.**

Supplement: Table 8, additional analyses II, logistic multilevel models (selected coefficients).

	Judgemental												*Non- Judgemental*			
	BESD				*CI*				*CL*				*SP*			
Ethnic group (Ref. White Maj.)	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*	*(M 1)*	*(M 2)*	*(M 3)*	*(M 4)*
Indian	-.75	**-.83**	-.76	.68	**-2.6***	**-2.9****	**-3.1****	**-3.3***	**-2.8***	**-3.1***	**-3.1***	**-8.5*****	-1.2	**-1.7+**	-1.5	-3.6
	(.59)	**(.56)**	(1.0)	(1.6)	**(1.1)**	**(1.0)**	**(1.1)**	**(1.3)**	(1.1)	(1.2)	(1.2)	**(1.3)**	(.86)	**(1.0)**	(1.0)	(2.8)
Pakis. & Bangl.	**-1.9****	**-2.3****	**-2.3****	**-1.2**	**-2.7****	**-3.2*****	**-3.2*****	**-2.9****	**-1.2+**	**-1.8****	**-1.9****	**-1.9+**	**-1.5****	**-2.3*****	**-2.3*****	-.95
	(.70)	**(.69)**	**(.70)**	**(1.1)**	**(.85)**	**(.87)**	**(.89)**	**(1.0)**	**(.65)**	**(.70)**	**(.70)**	**(1.1)**	**(.57)**	**(.60)**	**(.61)**	(1.0)
Black British	.38	.04	.05	**3.3+**	-.13	-.48	-.49	-.31	-.52	-.94	-.92	.13	-.48	-.97	-.92	-2.9
	(.53)	(.51)	(.52)	**(1.7)**	(.60)	(.61)	(.61)	(1.4)	(.76)	(.76)	(.77)	(2.3)	(.80)	(.84)	(.85)	(2.2)
Chinese & East Asian	**-1.8+**	**-2.0+**	-2.0+	-.76	**-18.9*****	**-20.1*****	**-20.1*****	**-21.2*****	**-20.1*****	**-20.1*****	**-19.6*****	**-22.7*****	-2.2	**-2.9***	**-2.9***	-1.5
	(1.1)	**(1.0)**	(1.1)	(1.9)	**(.65)**	**(.80)**	**(.87)**	**(1.1)**	**(.94)**	**(1.1)**	**(1.3)**	**(1.4)**	(1.4)	**(1.4)**	**(1.5)**	(2.0)
Mixed & Other	-.60	-.67	-.69	-1.9	-.61	-.81	-.80	-.22	-1.1	-1.3	-1.3	-.44	**-1.5***	**-1.7***	**-1.8***	-.69
	(.69)	(.71)	(.72)	(1.4)	(.61)	(.60)	(.60)	(1.2)	(.82)	(.84)	(.85)	(1.2)	**(.74)**	**(.73)**	**(.74)**	(1.7)
Language minority status																
Language minority exposure (LE)	**1.3****	**.82+**	1.3		-.84	**-1.3+**	**-3.2***		**-1.2+**	**-1.6***	-.90		.67	-.07	.77	
	(.51)	**(.50)**	(.86)		(.73)	**(.75)**	**(1.3)**		**(.68)**	**(.68)**	(1.4)		(.49)	(.50)	(1.1)	
Second language proficiency (LP)		**-.64*****	**-.62*****	**-.60*****		**-.58*****	**-.59*****	**-.55*****		**-.74*****	**-.74*****	**-.71*****		**-1-1*****	**-1.0*****	**-1.0*****
		(.15)	**(.16)**	**(.17)**		**(.12)**	**(.12)**	**(.12)**		**(.10)**	**(.14)**	**(.15)**		**(.20)**	**(.17)**	**(.18)**
Interactions																
LE x LP			-.24				**.89+**				-.41				-.52	
			(.36)				**(.46)**				(.73)				(.68)	
Ethnic group x LP																
Indian x LP				-.46				.02				**1.5*****				.77
				(.61)				(.19)				**(.26)**				(.94)
Pakis. & Bangl. x LP				-.24				**-.55***				-.34				-1.1
				(.42)				**(.27)**				(.58)				(.70)
Black British x LP				**-1.4+**				-.15				-.54				.75
				(.81)				(.52)				(1.0)				(.78)
Chinese&East Asian x LP				-.29				**.53***				**.72***				-1.2
				(.76)				**(.24)**				**(.29)**				(.80)
Mixed&Other x LP				.44				-.25				-.36				-.46
				(.39)				(.29)				(.32)				(.70)
variance between teacher	3.7	3.7	3.7	3.5	2.2	2.7	2.8	2.8	2.0	2.4	2.4	2.5	4.3	4.8	4.7	4.8
variance between schools	.67	.50	.51	.52	1.0	.88	.89	.90	2.9	3.1	3.1	3.2	.20	.14	.13	.17
BIC	15318	15093	15193	15454	20226	19925	20015	20263	18413	18039	18141	18372	16400	15657	15745	156425
N	4460				4623				4503				4421			

Note: coefficients shown, dependent variable: overestimation on type (yes/no); controls included; ***p<.001; p<.01;*p<0.05; +p<.01; significant coefficients are **bold.**

Supplement: Table 9, additional analyses III, hierarchical linear models (selected coefficients).

	English competence								*Math competence*			
	BAS naming vocabulary (age 11)				*SWM test (age 11)*				*NFER number skills (age 7)**			
	BESD	CI	CL	SP	BESD	CI	CL	SP	BESD	CI	CL	SP
SEN misidentification (Ref. correct no SEN - whole primary)												
correct SEN identification - whole primary edu	**-5.2***	**-7.7*****	**-8.1*****	**-7.8+**	-1.0	**-1.3***	**-2.3*****	-1.5	**-10.4*****	**-13.6*****	**-15.4*****	**-9.4*****
	(2.0)	**(1.3)**	**(1.8)**	**(4.5)**	(1.1)	**(0.6)**	**(0.6)**	(1.2)	**(1.7)**	**(1.0)**	**(1.3)**	**(1.7)**
incorrect SEN identification - whole primary edu	**-4.4****	**-4.2***	**-4.5****	**-8.8***	**-1.4***	**-1.8***	**-1.8****	-1.2	**-10.4*****	**-10.5*****	**-11.7*****	**-12.1*****
	(1.6)	**(1.7)**	**(1.6)**	**(3.9)**	**(0.7)**	**(0.9)**	**(0.6)**	(1.3)	**(1.4)**	**(1.2)**	**(1.1)**	**(1.4)**
incorrect SEN identification - part primary edu	-1.9	**-2.9***	**-3.4+**	**-4.0****	-0.7	-0.7	-1.1	-1.0				
	(1.7)	**(1.2)**	**(2.0)**	**(1.4)**	(1.1)	(0.8)	(1.0)	(0.8)				
other	-1.7	**-3.2****	**-3.5*****	-2.6	-1.3	**-1.0+**	**-1.3***	-0.2				
	(1.7)	**(1.2)**	**(0.9)**	(1.6)	(0.7)	**(0.6)**	**(0.5)**	(0.6)				
variance between teachers	6.1	7.1	6.1	6.9	6.0	6.2	5.7	5.6	15.6	15.2	18.0	15.4
variance between schools	12.0	12.4	11.9	11.9	0.6	0.3	0.2	0.2	27.4	26.1	24.2	27.0
N	4460	4623	4530	4421	4460	4623	4530	4421	4460	4623	4530	4421

Note: controls included;***p<.001; p<.01;*p<0.05; +p<.01; significant coefficients are **bold.** * NFER scores are obtained when there is only one report of children's SEN status available (age 7). Category changes are thus not considered in the respective analyses.

Summary & Discussion

This dissertation aimed to assess the role of language-based discrimination for immigrants' structural integration into the educational system and the labor market. It was guided by two broad questions regarding a) the extent to which language-based discrimination affects immigrants' educational or labor market outcomes beyond the resource-related influences of L2 proficiency and b) the motives and mechanisms behind such forms of discrimination.

Taken together, the three research articles suggest that language-based discrimination could affect immigrants' educational and labor market outcomes. For instance, with regard to labor market outcomes, the results of paper 1 indicate that when L2 proficiency is low, immigrant groups associated with unfavorable attitudes are less likely to be employed, less likely to obtain a white collar position, or have lower job status scores (ISEI) than their native counterparts (Table I, Paper 1). When language skills increase however, these disadvantages reduce and the respective groups can catch up with the native population (Figure I-IV, Paper 1). In addition to that, the results of paper 2 indicated that applicants with deviant language cues are turned down more often at the beginning of the hiring process than comparable candidates without these cues, independent of other indications of ethnic origin, such as the name (Table 5, Paper 2).

With regard to educational outcomes, the results of paper 3 indicated that children of ethnic groups associated with unfavorable beliefs about school-related attributes were more likely to be over-identified with certain special educational needs. For instance, Black British students with lower L2 proficiency were found about 40 percentage points more likely to be misidentified with behavioral difficulties than their majority peers (Table 2, Paper 3). With increasing second language skills, this likelihood significantly reduced for the students in question (ibid.). For children of other ethnic groups, lower second language proficiency was associated with more accurate assessments of their special educational needs (ibid.).

For labor market outcomes, the cited patterns could be linked to employer tastes, as the extent of information that is available to employers did not seem to matter (Paper 1), and language-based disadvantages were found to be less pronounced in firms with a more standardized recruitment process, where employers might have less room to factor personal tastes into their decision-making (Paper 2). In addition to that, the negative effects of language cues were more pronounced in regions with higher levels of anti-immigrant attitudes, which might also hint at the potential role of employer tastes (Paper 2).
For the educational outcomes examined here, results could be related to ethnic beliefs, as patterns of misevaluations seemed to match the group-specific stereotype content (Paper 3).

However, there are several limitations to each of these studies. The first line of drawbacks concerns my assessment of language-based discrimination for immigrants' labor market outcomes with large-scale survey data (Paper 1). *Firstly,* I conducted cross-sectional analyses, as repeated test-based measures of L2 proficiency were not yet available within the employed data set (or elsewhere to the author's knowledge). Results may thus be affected by reverse causality. I conducted additional analyses, which did not hint at such relations, yet complete elimination of such arguments is impossible. A *second* limitation may relate to omitted variables. While I examined the role of certain alternative explanations, it is hard to establish that all relevant factors have been considered. Moreover, I could only address some of the alternative explanations for a subsample of respondents. *Lastly,* while being in line with certain theoretical arguments, I could not directly test the proposed assumptions.

Additional limitations may relate to the conducted field experiment (Paper 2). *Firstly,* it was not possible to univocally disentangle employers' accent-related productivity considerations from forms of statistical discrimination. Both arguments proposed the same empirical patterns within the respective design, and additional tests could only indirectly assess the individual relevance of each of these arguments. Similar difficulties relate to arguments on perceived customer preferences and more productivity-related considerations of how well customers will understand the applicant. *Secondly,* internal recruiting processes and recruitment via personal networks could not be considered within the design. As particularly recruitment via personal networks may place less emphasize on language cues, the results might not fully generalize to such forms of recruiting. *Thirdly,* while employing several measures to examine the motives behind language-based differential treatment, here too, underlying mechanisms could only be indirectly inferred.

The last line of drawbacks concerns the assessment of language-based discrimination for the generation of educational outcomes with large-scale survey data (Paper 3). Here*, the most crucial one* might concern case numbers. While sample size per se is sufficient, the outcome of interest, i.e. (mis)identification of special educational needs, is comparatively rare. Although I additionally employed a two-step correction technique (King and Zeng, 2001), which is assumed to produce more accurate estimates in the presence of rare events, it might be fruitful to validate these findings further with a larger teacher-student sample. The *second limitation* relates to the test of hypotheses. While results do suggest patterns that are consistent with different theoretical arguments, the proposed mechanisms, again, could not be tested directly. It might thus be a potential route for future research to validate these mechanisms further, for instance with tailored teacher surveys, qualitative methods such as in-depth interviews and videography, or techniques for revealing stereotypes such as the IAT or word-completion tests. *Moreover,* I examined only one very specific outcome – the identification of special educational needs. It is unclear whether the identified patterns generalize to other aspects of educational success. Prior research gives some indications that language-based discrimination also affects more general evaluations of students' academic abilities (e.g. Crowl and MacGinitie, 1974; Kanai et al., 1978) and that this relates to adverse beliefs (Lorenz, 2017); however, (current) research is scarce.

More general discussions might concern the role of the national context in which the theoretical assumptions were addressed. I examined how language-based discrimination affects the labor market outcomes of immigrants in Germany, and how such forms of discrimination shape the educational outcomes of students in England. While both contexts are fruitful to examine the respective processes, results might not fully generalize to other national contexts. Both, extent and motives behind language-based discrimination could vary between countries.

One the one hand, the prevalence of ethnic distaste and adverse beliefs should differ between contexts. For instance, unfavorable attitudes, perceived ethnic threats, and ethnic exclusionism seem to be more pronounced in contexts with larger shares of non-western immigrants or non-EU citizens (Scheepers et al., 2002; Schneider, 2008). Language cues might thus be associated with more or less unfavorable attitudes in different contexts to begin with. Moreover, the extent to which L2 proficiency shapes such attitudes might also vary between countries. In contexts where naturalization is linked to proficiency in the national language, as in both of the examined countries, lower L2 abilities might shape stronger aversions, as it is more closely linked to ideas about an 'unintegrated immigration status' and 'foreignness'. This may also increase insecurity about 'unobservable' individual traits and thus foster belief application.

On the other hand, the main reasons behind such language-related forms of discrimination may differ between countries. With regard to labor market outcomes, it has been noted that hiring processes in Germany, or all German-speaking counties, require comparatively comprehensive information on the prospective applicants and employers may thus "have less need to resort to mechanisms that can result in statistical discrimination" (Zschirnt and Ruedin 2016, 1119). Consequently, while in the German context examined here, lower L2 proficiency and deviant linguistic patterns were not associated with different forms of statistical discrimination, such processes may exist in other contexts where recruitment is based on less initial information.

Similarly, with regard to the examined educational outcome, certain forms of language-based discrimination may be less prevalent within the English educational context. While children with special educational needs receive additional measures and can spend time outside the regular classroom setting, legislation places emphasis on main stream school attendance (The Children and Families Act 2014). Language-based forms of institutional discrimination that build upon special school transferals may thus be more prevalent in contexts where this kind of school attendance is more common.

On an even more general level, one could also question whether language proficiency is truly 'special' in that it shapes ethnic discrimination. Similar arguments could be made with regard to other resources as well. For instance, immigrants' educational credentials might also shape distaste or the application of adverse beliefs.

Despite these limitations, this dissertation could contribute to the literature on ethnic inequalities, and might shed additional light on the mechanisms that link L2 language proficiency to immigrants' structural integration outcomes.

References

Aigner, D. & Cain, G. (1977). Statistical theories of discrimination in labor markets. Industrial and Labor Relations Review 30 (2): 175-187.

Allport, G. W. (1979). The nature of prejudice. 1954. Gergen, Kenneth J.,"The Significance of Skin Color in Human Relations.

Antidiskriminierungsstelle des deutschen Bundes (2012): Ausgewählte Entscheidungen deutscher Gerichte zum Antidiskriminierungsrecht.

Arrow, K. (1973). The theory of discrimination. In: Ashenfelter, R. (ed.), Discrimination in labor markets. Princeton NJ: Princeton University Press, pp. 3-33.

Barrett, A. N., Barile, J. P., Malm, E. K., & Weaver, S. R. (2012). English proficiency and peer interethnic relations as predictors of math achievement among Latino and Asian immigrant students. Journal of adolescence, 35(6), 1619-1628.

Becker, G. (1971). The economics of discrimination. Chicago: UCP. 2nd ed.

Blank, R., Dabady, M., & Citro, C. (2004). Measuring racial discrimination. Panel on Methods for Assessing Discrimination, National Research Council.

Blinder, A. (1973). Wage discrimination: Reduced-form and structural variables. Journal of Human Resources, 8 (3), 436–455.

Blommaert, L., Coenders, M., & Van Tubergen, F. (2013). Discrimination of Arabic-named applicants in the Netherlands: An internet-based field experiment examining different phases in online recruitment procedures. Social forces, 92(3), 957-982.

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y., & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. Language & Communication, 22(2), 171-185.

Brown, R. (1995). Prejudice: Its social psychology. John Wiley & Sons.

Burr, Elizabeth, Eric Haas and Karen Ferriere. 2015. Identifying and supporting English learner students with learning disabilities. Washington, DC: U.S. DoE,IES.

Campbell, Donald T. (1965): Ethnocentrism and other altruistic motives. In: David Levine (Hg.): Nebraska symposium on motivation. Lincoln: University of Nebraska Press, S. 283- 311.

Cediey, E., & Foroni, F. (2008). Discrimination in access to employment on grounds of foreign origin in France: A national survey of discrimination based on the testing methodology of the International Labour Office. ILO.

Chiswick, B. & Miller, P. (2001). Occupational language requirements and the value of English in the US labor market, *Journal of Population Economics*,23(1),353-372.

Chiswick, B. R. (2008). The Economics of Language: An Introduction and Overview.IZA. DP No. 3568

Chiswick, B. R. and P. W. Miller. 1995. "The Endogeneity between Language and Earnings: International Analyses". Journal of Labor Economics 13(2), 246-288.

Choy, S. J., & Dodd, D. H. (1976). Standard and nonstandard Hawaiian English-speaking children: Comprehension of both dialects and teacher's evaluations. Journal of Educational Psychology, 68(2), 184.

Crowl, T. K., & MacGinitie, W. H. (1974). The influence of students' speech characteristics on teachers' evaluations of oral answers. Journal of Educational Psychology, 66(3), 304.

Davila, A. (1993). Accent Penalties and the Earnings of Mexican. Social Science Quarterly, 74(4).

Diehl, C., & Fick, P. (2016). Ethnische Diskriminierung im deutschen Bildungssystem. In Ethnische Ungleichheiten im Bildungsverlauf (pp. 243-286). Springer VS, Wiesbaden.

Duran, Richard. 1988."Validity and language skills assessment: Non-English background students." Test validity:105-127.

Dustmann, C. & vanSoest, A. (2002). Language and the earnings of immigrants. *Industrial and Labor Relations Review* 55,(3),473-492.

Edele, A., & Stanat, P. (2016). The role of first-language listening comprehension in second-language reading comprehension. *Journal of Educational Psychology*, 108(2), 163.

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. Social Science Research, 52, 99-123.

Esser, Hartmut. Sprache und Integration: die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten. Campus, 2006.

Feagin, J. R., & Feagin, C. B. (1986). Discrimination American style (rev. ed.). Melbourne, FL: Krieger Publishing Co.

Fiske, Susan T.; Lin, Monica; Neuberg, Steven L. (1999): The continuum model: Ten years later. In: ShellyChaiken und Yaacov Trope (Hg.): Dual-process theories in social psychology. New York: GuilfordPress.

Gandara, Patricia, Julie Maxwell-Jolly and Anne Driscoll. 2005."Listening to Teachers of English Language Learners."Policy Analysis for California Education,PACE(NJ1)

Giles, H., & Johnson, P. (1987). Ethnolinguistic identity theory: A social psychological approach to language maintenance.

Gomolla, M. (2010). Institutionelle Diskriminierung. Neue Zugänge zu einem alten Problem. In Diskriminierung (pp. 61-93). VS Verlag für Sozialwissenschaften.

Gomolla, Mechtild and Frank-Olaf Radtke. 2003. Institutionelle Diskriminierung. Die Herstellung ethnischer Differenz in der Schule. Wiesbaden.

Gurr, T. R. (1970). Why men rebel, Princeton (NJ) Princeton Univ.

Holzer, H. J., & Ludwig, J. (2003). Measuring Discrimination in Education: Are Methodologies from Labor and Markets Useful?. Teachers College Record, 105(6), 1147-78.

Horr, A., Hunkler, C. & Kroneberg, C. (2010). Ethnic discrimination in the rental housing market - A field experiment on the underlying mechanisms. Mannheim: Mannheimer Zentrum für Europäische Sozialforschung (MZES); presented (Equalsoc final conference, 3-5 June 2010, Amsterdam).

Hosoda, M. and E. F. Stone-Romero. 2010. "The Effects of Foreign Accents on Employment-Related Decisions." Journal of Managerial Psychology 25(2): 113-132.

Jackson, M., & Cox, D. R. (2013). The principles of experimental design and their application in sociology. Annual Review of Sociology, 39, 27-49.

Kalter, F. (2006). Auf der Suche nach einer Erklärung für die spezifischen Arbeitsmarktnachteile von Jugendlichen türkischer Herkunft*, Zeitschrift für Soziologie*,35,144-160.

King, Gary and Langche Zeng 2001. Logistic regression in rare events data. Political analysis,9(2):137-163.

Kinzler, K. D., Shutts, K., DeJesus, J., & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. Social cognition, 27(4), 623-634.

Klingner, Janette and Beth Harry. 2006."The special education referral and decision-making process for English language learners: Child study team meetings and placement conferences." Teachers College Record,108(11):2247–2281.

Kossoudji, S. A. (1988). English language ability and the labor market opportunities of Hispanic and East Asian immigrant men. Journal of Labor Economics, 6(2), 205-228.

Kratzmann, J., & Schneider, T. (2009). Soziale Ungleichheiten beim Schulstart. KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 61(2), 211-234.

Kristen, C. (2006). Ethnische Diskriminierung im Schulsystem? Theoretische Überlegungen und empirische Ergebnisse. Discussion Paper Nr. SP IV 2006-601 der Arbeitsstelle Interkulturelle

Konflikte und gesellschaftliche Integration (AKI) am Wissenschaftszentrum Berlin. Berlin: AKI.

Kroneberg, C. (2005). Die Definition der Situation und die variable Rationalität der Akteure. ein allgemeines Modell des Handelns. Zeitschrift für Soziologie, 34(5), 344-363.

Lindemann, K., & Kogan, I. (2013). The role of language resources in labour market entry: Comparing Estonia and Ukraine. *Journal of Ethnic and Migration Studies*, 39(1), 105-123.

Lorenz, G. (2017). Selbsterfüllende Prophezeiungen in der Schule: Leistungserwartungen von Lehrkräften und Kompetenzen von Kindern mit Zuwanderungshintergrund. Springer-Verlag.

MacRae, Neil; Stangor, Charles; Hewstone, Miles (Hg.) (1996): Stereotypes and stereotyping. New York:Guilford Press

MacSwan, Jeff and Kellie Rolstad. 2006."How language proficiency tests mislead us about ability: Implications for English language learner placement in special education." Teachers College Record,108(11):2304–2328.

Mercer, Jane and Wayne Brown. 1973."Racial differences in IQ:Fact or artifact." New York:TWP, pp.56-113.

Oaxaca, R. (1973). Male-female wage differences in urban labor markets. International Economic Review, 14 (3), 693–709.

Ortiz, Alba. and Elba Maldonado-Colón. 1986."Recognizing learning disabilities in bilingual children" Journal of Reading, Writing, and Learning Disabilities International,2(1):43-56.

Pendakur, K., & Pendakur, R. (2002). Language as both human capital and ethnicity 1. International Migration Review, 36(1), 147-177.

Pettigrew, T. F. (1979). The ultimate attribution error: Extending Allport's cognitive analysis of prejudice. Personality and social psychology bulletin, 5(4), 461-476.

Phelps, E. (1972). The statistical theory of racism and sexism. The American economic review 62 (4): 659-661.

Phelps, E. (1972). The statistical theory of racism and sexism. The American economic review 62 (4): 659-661.

Radtke, Frank-Olaf (1996): Mechanismen ethnischer Diskriminierung in der Grundschule, in: Kersten, Ralph et al. (Hg.), Ausbilden statt Ausgrenzen. Jugendliche ausländischer Herkunft in Schule, Ausbildung und Beruf 22, Frankfurt: Haag + Herchen Verlag, S. 121-132.

Rakić, T., & Stössel, K. (2013). Die Wirkung fremder Akzente. Deutsch als Fremdsprache, 50(1), 11-18.

Rakić, T., M. Steffens and A. Mummendey. 2011. "Blinded by the Accent! The Minor Role of Looks in Ethnic Categorization." Journal of Personality and Social Psychology 100(1): 16-29.

Rödin, M. and G. Özcan. 2011. "Is It How You Look or Speak That Matters? An Experimental Study Exploring the Mechanisms of Ethnic Discrimination." Research Papers in Economics 12, Stockholm University, Department of Economics.

Runciman, W. G. (1966). Relative deprivation & social justice: Study attitudes social inequality in 20th century England.

Scheepers, P., Gijsberts, M. and Coenders, M. (2002).Ethnic exclusionism in European countries. Public opposition to civil rights for legal migrants as a response to perceived ethnic threat. European Sociological Review, 18, 17–34.

Schneider, S. L. (2008). Anti-immigrant attitudes in Europe: Outgroup size and perceived ethnic threat. European Sociological Review, 24(1), 53-67.

Sherif, M. (1966). In common predicament: Social psychology of intergroup conflict and cooperation. Houghton Mifflin comp.

Shifrer, D., Muller, C., & Callahan, R. (2011). Disproportionality and learning disabilities: Parsing apart race, socioeconomic status, and language. Journal of Learning Disabilities, 44(3), 246-257.

Sprietsma, M. (2013). Discrimination in grading: experimental evidence from primary school teachers. Empirical economics, 45(1), 523-538.

Stolzenberg, R. M., & Tienda, M. (1997). English proficiency, education, and the conditional economic assimilation of Hispanic and Asian origin men. Social Science Research, 26(1), 25-51.

Van Tubergen, F., Maas, I., & Flap, H. (2004). The economic incorporation of immigrants in 18 Western societies: Origin, destination, and community effects, *American Sociological Review,* 69(5),704-727.

Wagner, S. J., & Powell, J. J. (2003). Ethnisch-kulturelle Ungleichheit im deutschen Bildungssystem: zur Überrepräsentanz von Migrantenjugendlichen an Sonderschulen (pp. 183-208). Universitätsverlag Winter.

Werth, L., & Mayer, J. (2007). Sozialpsychologie. Spektrum Akad. Verlag.

Williams, F., Whitehead, J. L., & Miller, L. (1972). Relations between language attitudes and teacher expectancy. American Educational Research Journal, 9(2), 263-277.

Zschirnt, E., & Ruedin, D. (2016). Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. Journal of Ethnic and Migration Studies, 42(7), 1115-1134.

Appendix, Table A: List of considered studies, examining L2 effects for immigrants' labor market outcomes

Author (Name)	Date (year)	Country context	Research focus	Explanandum	Sample	Resource related, or other arguments	Discrimination mentioned	Discrimination adressed	SSCI, SCIE ranked	Impact factor	n citations	Link	Journal
Akresh	2007	US	labor market outcome	Earnings	NIS-P	1	0	0	YES	2802	63	https://w	Demography
Aldashev et al.	2009	GER	labor market outcome	Participation, Employ	SOEP (1996-2	1	1	0	YES	1036	67	https://d	Labour Economics
Berman et al.	2003	Israel	labor market outcome	Earnings	OS (1994)	1	0	0	YES	1036	172	https://d	Labour Economics
Bleakley & Chin	2004	US	labor market outcome	Earnings	Census 2000	1	1	0	YES	2926	516	https://w	Review of Econom
Boyd & Cao	2009	Canada	labor market outcome	Earnings	2001 Census	1	0	0	YES	0.633	62	https://j	Canadian Studies i
Budría & Swedberg	2012	Spain	labor market outcome	Earnings	2007 Spanish	1	0	0	YES	0.267	28	https://p	Revista de Econom
Carnevale et al.	2001	US	labor market outcome	Earnings	NALS (1992)	1	0	0	YES	4026	62	https://w	The American Eco
Chiswick	1991	US	labor market outcome	Earnings	Survey of ille	1	0	0	YES	2887	546	https://w	Journal of Labor Ec
Chiswick	1998	Israel	labor market outcome	Earnings	Census 1983	1	0	0	YES	2926	198	https://d	Journal of Populati
Chiswick & Miller	1995	Australia	labor market outcome	Earnings	Census data (	1	0	0	YES	2887	1023	http://ww	Journal of Labor Ec
Chiswick & Miller	2002	US	labor market outcome	Earnings	Census 1990	1	0	0	YES	2926	512	https://d	Journal of Populati
Chiswick & Miller	2010	US	labor market outcome	Earnings	Public Use M	1	0	0	YES	1136	121	https://li	Journal of Populati
Clark & Drinkwater	2008	UK	labor market outcome	Employment, Self-er	LFS	1	0	0	YES	1194	131	https://d	Oxford Review of E
Dávila & Mora	2004	US	labor market outcome	Earnings	Census (198(	1	1	0	YES	1647	25	https://d	Industrial Relation
Dávila, Bohara & Saenz	1993	US	labor market outcome	Earnings	Public Use M	1	1	1	YES	0.849	99	http://di	Social Science Qua
Di Paolo & Raymond	2012	Catalonia	labor market outcome	Earnings	Survey on Liv	1	0	0	YES	0.455	31	https://d	Journal of Applied
Dustmann	1994	GER	labor market outcome	Earnings	German Soci	1	1	0	YES	1136	424	https://d	Journal of Populati
Dustmann & Fabbri	2003	UK	labor market outcome	Employment, Earnin	Family and W	1	0	0	YES	2608	626	https://d	The economic jou
Dustmann & van Soest	2002	GER	labor market outcome	Earnings	German Soci	1	0	0	YES	2,137	370	https://d	ILR Review
Gonzales	2000	US	labor market outcome	Earnings	1992 Ntaional	1	0	0	YES	0.887	74	https://onlinelibrary.wiley.c	
Gonzales	2005	US	labor market outcome	Employment, Earnin	Public Use M	1	0	0	YES	2117	82	https://o	JOURNAL OF APl
Guven & Islam	2015	Australia	labor market outcome	socioeconomic outco	HILDA (2001	1	0	0	YES	2802	47	https://d	Demography
Hayfron	2001	Norway	labor market outcome	Earnings	survey 1993;	1	0	0	YES	0.648	82	https://w	Applied Economics
Hwang et al.	2010	US	labor market outcome	Earnings	Public Use M	1	0	0	YES	0.994	24	https://w	Ethnic & Racial Stu
Kalter	2006	GER	labor market outcome	Employment, Occup	German Soci	1	1	0	YES	0.447	267	https://w	Zeitschrift für Soz
Kossoudji	1988	US	labor market outcome	Occupation, Earnings	Survey of Inc	1	1	1	YES	2887	364	https://w	Journal of Labor Ec
Leslie & Lindley	2000	US	labor market outcome	Employment, Earnin	1994 Fourth	0	0	0	YES	0.920	118	https://o	Economica
Lindemann & Kogan	2013	nia and Uk	labor market outcome	Earnings	TIES survey (	1	0	0	YES	1362	20	https://w	Journal of Ethnic a
Miranda & Zhu	2013	UK	labor market outcome	Earnings	UK Househo	0	0	0	YES	0.558	31	https://d	Economics letters
Mora	1998	US	labor market outcome	Earnings	Public Use M	1	0	0	YES	0.849	28	https://w	Social Science Qua
Park	1999	US	labor market outcome	Earnings	1976 Survey (	1	1	0	YES	0.446	66	https://w	The American Jou
Pendakur & Pendakur	2002	Canada	labor market outcome	Earnings	Census 1991	1	1	1	YES	2195	135	https://w	The International
Rivera-Batiz	1990	US	labor market outcome	Earnings	NAEP Young	0	0	0	YES	0.558	120	https://d	Economics letters
Shields & Price	2002	UK	labor market outcome	occupational sucess	Fourth Natio	1	0	0	YES	1136	198	https://li	Journal of Populati
Stolzenberg & Tienda	1996	US	labor market outcome	Earnings	1980 Census	1	1	1	YES	1327	67	https://w	Social Science Res
Tainer	1988	US	labor market outcome	Earnings	from the Sur	1	0	0	YES	4047	197	https://w	The Journal of Hu
van Tubergen, Maas & Flap	2004	estern cou	labor market outcome	Labor force activity, e	European Lal	1	0	0	YES	4400	363	https://d	American Sociolog
Yao & van Ours	2015	Netherland	labor market outcome	Earnings	LISS Panel (2	1	0	0	YES	1036	33	https://d	Labour Economics

N Studies mentioning only resource-related, or additional arguments	28	0,74
N Studies mentioning language-based forms of discrimintion	10	0,26
N total studies	**38**	
N Studies pecifically examining language-based forms of discrimination	4 of 10	0,11 of 0,26

Considered studies are SSCI ranked, were conducted between 1990 and 2017 and examine the relationship between L2 proficiency and immigrants' educational or labor market outcomes with large-scale survey data. Tables include all studies that were located by the author and meet these criteria. However, completeness is not claimed.

Appendix, Table B: List of considered studies, examining L2 effects for immigrants' educational outcomes

Author (Name)	Date (year)	Country context	Research focus	Explanandum	Focus	Sample	Resource related, or other arguments	Discrimination mentioned	Discrimination adressed	SSCI, SCIE ranked	Impact factor	n citations	Link	Journal
Barrett et al.	2012	US	educational out	Competences	L2	Educatio	1	1	0+	YES	1.795	16	https://www.nc	Journal of adole
Beal, Adams & Cohen	2009	US	educational out	Competences	L2	*Sample	1	0	0	YES	1.194	96	http://journals.	Urban Educatio
Jasper & Hibel	2012	US	educational out	Special needs	LEP	ECLS-K	0	0	0	YES	2.113	35	https://academ	Social forces
Kratzmann & Schneid	2009	GER	educational out	Time of school	Social inec	SOEP	1	1	1	YES	0.476	69	https://link.spr	Kölner Zeitschı
OECD~	2006	OECD co	educational out	Competences	Immigran	PISA	1	1	0	NO			https://www.fa	Bildungsforsch
OECD~	2012	OECD co	educational out	Competences	Immigran	PISA	0	0	0	NO			http://www.oec	OECD Publishi
OECD~	2015	OECD co	educational out	Competences	Immigran	PISA	0	0	0	NO			http://www.oec	OECD Publishi
Perreira et al.	2006	US	educational out	Drop out rates	High scho	National	1	0	0	YES	2.802	349	https://link.spr	Demography
Samson & Lesaux	2006	US	educational out	Special needs	LEP	ECLS-K	0	0	0	YES	1.494	137	http://journals.	Journal of learn
Shifer et al	2011	US	educational out	Special needs	LM, SE, F	ELS	0	1	1	YES	2.296	123	http://journals.	Exceptional chil
Strobel	2016	GER	educational out	Competences	L1	National	1	0	0	YES	0.994	1	https://doi.org/	Ethnic and Raci

** does not actually meet the criteria of large-scale (representative) survey data*

+ examine how discrimination affects the level of motivation, which is assumed to be greater with higher L2 proficiency.

~ L2 proficiency itself is here seen as the school relevant outcome. Authors focus on the role of speaking the language of instruction at home.

N Studies mentioning only resource-related, or additional arguments	7	0,64
N Studies mentioning language-based forms of discrimintion	4	0,36
N total studies	**11**	
N Studies pecifically examining language-based forms of discrimination	2 of 4	0,18 of 0,36

Considered studies are SSCI ranked, were conducted between 1990 and 2017 and examine the relationship between L2 proficiency and immigrants' educational or labor market outcomes with large-scale survey data. Tables include all studies that were located by the author and meet these criteria (with a small number of marked exceptions). However, completeness is not claimed.

Appendix, Table C: Overview of experimental studies, examining L2 effects for immigrants' labor market outcomes

Author (Name)	Date (year)	Country context	Research focus	Explanandum	L2 aspect studied	Main finding	Type of experiment	Criticism	SSCI, SCIE ranked	Impact factor	n citations	Journal
Cargile	2000	US	labor market outcome	Employment su	L2 accent	L2 accent has minimal effects on employability	Laboratory	- Partially crossed manipulations (confounding nar - Student sample (systematically different routines) - Small n (192) (randomization wobbly) - Singular accent (postive group) - sparse theoretical arguments	YES	1.067	69	Journal of employm
Carlson & McHenry	2006	US	labor market outcome	Employment su	L2 accent	Pronounced L2 accents affect employability. Mini	Laboratory	- no control group (standard speech) - Student sample (systematically different routines) - Small n (60) (randomization wobbly)	YES	1.067	114	Journal of employm
Deprez-Sims & Morri	2010	US	labor market outcome	Hirability	L2 accent	L2 accents reduce hirability. This is due to lower	Laboratory	- accent origin not identifiable - Student sample (systematically different routines) - Small n (63) (randomization wobbly)	YES	1.778	65	
Hansen et al.	2017	GER	labor market outcome	Competemce ra	L2 accent	L2 accent affects competence ratings. A lack of e	Laboratory	- Generalizability (indirect link to labor market out - Student sample (systematically different routines) - Small n (226)	YES	1.829	3	Experimental psych
Hansen et al	2013	GER	labor market outcome	Competemce ra	L2 accent	L2 accent affects competence and hirability rating	Laboratory	- sparse theoretical arguments - Little individualizing information (favors certain t - Small n (46) (randomization wobbly) - Accent effect byproduct (intervention study)	YES	1.272	19	Journal of Languag
Hosoda & Stone-Ron	2010	US	labor market outcome	Job suitability, H	L2 accent	L2 accents reduce hiring propability, particularly	Laboratory	- Partially crossed manipulations (confounding nar - Employed accents differed in understandability - "verbal guise" technique (speakers differ in pitch, - Student sample (systematically different routines) - Generalizability (positive accents)	YES	1.195	104	Journal of Manageri
Hosoda & Stone-Ron	2012	US	labor market outcome	Job suitability, H	L2 accent	L2 accent reduces suitability ratings, hiring prob	Laboratory	- manipulations may evoke dissonance (underquali - Outcome judgements not always straightforward - Student sample (systematically different routines)	YES	1.195	104	Journal of Manageri
Kalin & Rayko	1978	Canada	labor market outcome	Applicant evalu	L2 accent	L2 accented speakers were rated less suitable for	Laboratory	- "verbal guise" technique (speakers differ in pitch, - aggregation of different accents - Student sample (systematically different routines) - Small n (203, 16 experimental groups)	YES	0.629	121	Psychological Repor
Rakić, Steffens & Mu	2011	GER	labor market outcome	Hirability, Com	L1 accent	L1 accents reduce perceived competence and hira	Laboratory	- Generalizability (L1 accents) - Student sample (systematically different routines) - Small n (98) (randomization wobbly)	YES	3.139	39	British Journal of Ps
Rödin & Özcan	2011	SWE	labor market outcome	Candidate selec	L2 accent	L2 accents override looks in impression formatio	Laboratory	- Generalizability (indirect simulation of labor mark - No individualizing information (favors certain typ - Student sample (systematically different routines) - Small n (103) (randomization wobbly)	No		14	Working Paper (SU
Segrest-Purkiss et al.	2006	US	labor market outcome	Applicant evalu	L2 accent	L2 accent amplifies negative effects of ethnic nam	Laboratory	- Potential effects of including interviewers in tapes - Student sample (systematically different routines)	YES	2.454	169	Organizational Beh
Singer & Eder	1989	New Zeal	labor market	Candidate Selec	L2 accent	L2 accent had negligible effects on selection deci	Laboratory	- Potential effects of including interviewers in tapes - Peer effects (group testing) - Student sample (systematically different routines) - Small n (210, 16 experimental groups)	YES	1778	32	International journa
Timming	2016	US	labor market outcome	Hiring decision	L2 accent	L2 accents reduce hiring propability, particularly	Online	- Sample selection (crowdfunding platform) - Small n (223) (randomization wobbly) - Potential spill-over effects (within design) - sparse theoretical arguments	YES	1.821	6	Work, employment a

Appendix, Table D: Overview of experimental studies, examining L2 effects for immigrants' educational outcomes

Author (Name)	Date (year)	Country context	Research focus	Explanandum	L2 aspect studied	Main finding	Type of experiment	Criticism	SSCI, SCIE ranked	Impact factor	n citations	
Williams et al.	1972	US	educational at	students' compete	L1 accent	L1 accent affects	Laboratory	- Priming effects (introduction of the experiment) - unstandardized behavior of children - "verbal guise" technique (speakers differ in pitch, rate, etc.) - Small n (175) (randomization wobbly)	YES	2.931	91	American Ec
Choy & Dodd	1976	Hawaii	educational at	students' compete	L2 accent	L2 accent affects	Laboratory	- unstandardized behavior of children - Priming effects (introduction of the experiment) - "verbal guise" technique (speakers differ in pitch, rate, etc.) - few and selective subjects	YES	3.459	45	Journal of E
DeMais & Turner	1978	US	educational at	students' compete	L1 accent	L1 accent affects	Laboratory	- Peer effects (group testing) - "verbal guise" technique (speakers differ in pitch, rate, etc.) - Small n (68) (randomization wobbly)	YES	2.877	93	Contempora
Seligman et al.	1972	Canada	educational at	students' compete	Different speec	Speech affects pe	Laboratory	- Combined effect of multiple speech cues - Small n (19) (randomization wobbly) - Priming effects (subsequent presentation of materials)	YES	1.409	180	Language in
Crowl & MacGinit	1974	US	educational at	students' compete	L1 speech vari	L1 speech varieti	Laboratory	- Peer effects (group testing) - Small n (62) (randomization wobbly)	YES	3.459	69	*Journal of Ec*
Ford	1984		educational at	students' compete	L2 inflictions ('	Foreign L2 inflic	Laboratory	- Peer effects (group testing) - Priming effects (subsequent presentation of materials) - Small n (62) (randomization wobbly)	YES	2.056	34	*TESOL Qua*
Granger et al.	1977	US	educational at	students' commu	L1 speech cues	L1 cues affect pe	Laboratory	- Small n (56) (randomization wobbly) -Priming effects (stressing language factors) - "verbal guise" technique (speakers differ in pitch, rate, etc.)	YES	3.459	23	Journal of E
Woodworth& Salz	1971	US	educational at	students' compete	L1 accent	L1 accent affects	Laboratory	- Small n (119) (randomization wobbly) - Peer effects (group testing) - "verbal guise" technique (speakers differ in pitch, rate, etc.)	YES	1.194	27	Urban Educ
Taylor & Janet B.	1983	US	educational at	students' compete	L1 accent	L1 accent affects	n.i.	- Small n (52) (gender effects?) - Potential order effects - "verbal guise" technique (speakers differ in pitch, rate, etc.)	YES	3.459	19	Journal of E
Hewett	1971	US	educational at	teachers perceptio	L1 accent	L1 accent affects	Laboratory	- very small n (24) (randomization questionable) - selective sample - selective stimulus materials (educated speakers) - "verbal guise" technique (speakers differ in pitch, rate, etc.)	YES	2.079	19	Language L