RESEARCH ARTICLE

Associations of students' linguistic distance to the language of instruction and classroom composition with English reading and listening skills

Nils Jaekel[1,2,3] (iD), Markus Ritter[4] and Julia Jaekel[1,3,5,6]

[1]University of Oulu, Finland; [2]University of Copenhagen, Denmark; [3]Finnish Institute of Public Health (THL), Finland; [4]Ruhr-Universität Bochum, Germany; [5]University of Warwick, UK; [6]University of Leicester, UK
Corresponding author: Nils Jaekel; Email: nils.jaekel@oulu.fi

Abstract

Globally classrooms are increasingly linguistically diverse. Research often oversimplifies lived linguistic heterogeneity as binary variables: native versus non-native. Linguistic distance (LD) measures allow a fine-grained operationalization of linguistic diversity in foreign language education. This study investigated associations of cognate LDs of students' home languages and classroom heterogeneity with English as a Foreign Language skills. Data were collected from a diverse sample of 5,130 Year 5 students in Germany. Mixed-effects linear models confirmed our hypotheses that higher individual LDs and a higher proportion of multilingual learners per classroom were both independently associated with lower English proficiency. Multilingual learners with higher cognate LDs to English and students in more linguistically heterogeneous classrooms had lower English proficiency. The results emphasize the need to assess LD in research to better differentiate between students. Foreign language classrooms seem not to address linguistic diversity adequately and need to readjust their focus to better meet multilingual learners' needs.

Background and rationale

In light of increasing human migration and mobility, linguistic and cultural diversity has been growing exponentially around the globe. Between 1990 and 2019, the number of immigrants—that is, persons living in countries different from their country of birth or citizenship, worldwide increased by 77% to 271 million (United Nations, 2019). According to the United Nations' International Migration Report (2019), out of 184 countries included, Europe (30%) and North America (21%) accounted for the largest increases in immigrant populations. The percentage of foreign-born residents varies greatly between countries, and many formerly monolingual societies have

become multilingual and multicultural. Among the 38 member countries of the Organization for Economic Co-operation and Development (OECD), the student population with an immigrant background has steadily increased to 12.8% (2019a). In the context of this study, a substantial 26.7% of the German population had an immigrant background in 2020 (Statistisches Bundesamt, 2021), and 38.2% of primary and secondary school students in its most populous state North-Rhine Westphalia were from immigrant families (Information und Technik Nordrhein-Westfalen, 2020).

Proficiency in a country's majority language is an important skill that correlates with individuals' income (Bousmah et al., 2021; OECD, 2012), educational attainment (OECD, 2018), and integration into society (Auer, 2018). In schools, non-native multilingual learners (MLs) face several challenges. International studies have consistently demonstrated that immigrant students may be at risk of showing lower academic achievement than their native-speaking peers (OECD, 2006, 2019b), particularly first-generation immigrants (Rodríguez et al., 2020). Students' non-native language status may negatively affect their educational trajectories, attainment (Flisi et al., 2016; Hippe & Jakubowski, 2018), and professional success across the life course (Tharmaseelan et al., 2010). However, research has also shown that both contextual and individual factors explain large proportions of the variation in immigrant students' educational, linguistic, and developmental outcomes (Genesee & Fortune, 2014; Jaekel & Leyendecker, 2016; Maluch et al., 2015), highlighting the need to control for such factors in research studies.

In addition to a country's majority language(s), English skills have gained critical importance for higher education attainment and professional success worldwide. In many educational contexts, English as a Foreign Language (EFL) plays an important role. In German secondary schools EFL, beyond its status as an international lingua franca, constitutes a core subject in the curriculum relevant for grade point average (GPA) calculations, which are important for university access and grade retention. Therefore, students' English proficiency directly affects their educational careers and opportunities for life-course success and well-being early on in life.

Emerging research in second and foreign language learning has provided evidence that learners benefit from language similarity or smaller linguistic distances between their native language(s) and the target language (Muñoz et al., 2018; van der Slik, 2010). Muñoz et al. (2018) demonstrated that Danish, as opposed to Spanish students, benefited in their learning of English from the close relationship their L1 and L2 share. Although these findings may sound trivial initially, they outline that language learners do not all start at the same language levels and that learners may be disadvantaged in learning more distant languages. This is particularly relevant if we consider that in schools, language learners, regardless of their linguistic background, are expected to progress at a similar or equal pace in learning a second or foreign language. Pedagogical approaches and educational research have not extensively focused on this individual difference, whereas linguistic research has shown that language learners may experience facilitative or nonfacilitative transfer, which may affect language-learning trajectories (Westergaard et al., 2017). Therefore, linguistic diversity poses considerable affordances for foreign language teachers as they support their students in bridging linguistic differences (Pulinx et al., 2017; Vigren et al., 2022).

Over and above individual students' linguistic characteristics, classroom composition—that is, the proportion of nonnative students in a learning group—has been negatively associated with academic achievement (Jensen, 2015; Stanat, 2006); however, others argue that all students benefit from immigrants in the classroom (Silveira et al., 2019). Although the true direction of this association deserves heightened

research attention due to its significance for educational policy planning, it is critical to note that studies have rarely controlled for relevant confounders or considered underlying contextual processes.

To shed light onto the processes affecting immigrant students' foreign language learning in education, this study assesses the role of different indicator/proxy variables in EFL. Specifically, we investigate the independent relationships of students' cognate linguistic distance (LD) to English and classroom linguistic composition with receptive English proficiency in Grade 5.

Linguistic distance

An increasingly recognized variable in the context of language learning and education is the linguistic distance (LD) between a student's mother tongue and the language of instruction. LD here refers to the lexical, phonological, or grammatical level of similarity between two languages. Despite growing evidence identifying LD as a primary individual-difference variable in language attainment, it has been starkly neglected in second language acquisition (SLA) and educational research (Munoz et al., 2018).

Outside of the field of linguistics, LD measures are used in economic research focused on migration (Bousmah et al., 2021; Chiswick & Miller, 2004; Isphording & Otten, 2013) and education (Borgonovi & Ferrara, 2020). For example, economic studies have shown that LD contributes to a wage gap between immigrant groups, with immigrants with a mother tongue (L1) that is closer to a country's official language (L2) earning higher wages (Bousmah et al., 2021; Strøm et al., 2018). Educational research has demonstrated that a greater LD is associated with lower mathematics and science scores in Programme for International Student Assessment (PISA) assessments due to the considerable reading and writing affordances to understand and solve problems (Borgonovi & Ferrara, 2020). Although the authors explain that the variance explained by LD is small, it may have a greater effect on outcomes if other protective factors are not in place—for example, arriving in a host country before age 12, high socioeconomic status (SES), or attending a school with a low non-L1 student body (Borgonovi & Ferrara, 2020).

Not surprisingly, research on linguistic outcomes has provided broad support for LD as a powerful predictor of L2 proficiency (Lindgren & Muñoz, 2013; van der Slik, 2010). Linguistic and educational research investigating immigration usually operationalizes language or country of origin with a binary variable—that is, "immigrant/non-L1" versus "native." However, doing so oversimplifies more complex and continuously distributed underlying process variables such as LD. From a linguistic perspective, the comparison "immigrant" versus "native" assumes that the former constitutes a homogeneous group with similar linguistic traits and distance to the target language(s). Depending on the research context, classrooms are likely much more diverse, and assessing their level of heterogeneity warrants more fine-grained approaches. Depending on the research focus, excluding linguistic diversity may overgeneralize findings, for example, because of set expectations that second language development should occur at a similar pace and with similar ease or difficulty for all students with an immigrant background. However, whereas some languages may, to varying degrees, be mutually intelligible within their language families, for example, Danish, Swedish, or Norwegian (Gooskens et al., 2018), speakers of other languages have to invest considerable effort in attaining a similar level of understanding if their language is more distinct from the target language.

LD measures offer an approximation of the complex relationships between two languages and have been shown to predict proficiency development in the majority language (Schepens, 2015; Schepens et al., 2020; van der Slik, 2010) as well as school-based foreign language learning (Edele et al., 2018; Muñoz et al., 2018). Language learners benefit from shorter LDs in both foreign and second language learning, as they can more easily draw from cross-linguistic transfers (Goriot et al., 2021; van der Slik, 2010), which facilitates the rate of L2/L3 acquisition (Paradis, 2011). For example, Munoz et al. (2018) showed that young (7–9 years of age) English learners from Denmark benefited from the close Danish–English relationship, whereas Spanish learners of English have to invest more effort. The authors argue that Danish students benefit from more frequent cognates in their L1 and L2 and a lower cognitive load on working memory, which may facilitate learning English faster. Similarly, Cenoz (2001) demonstrated how young learners of English tended to draw from Spanish, an Indo-European language, rather than from Basque, a non-Indo-European language. She suggests that when learners can draw from multiple languages, they tend to draw from linguistic resources closer to the target language and thus benefit from cross-linguistic transfer (Cenoz 2001). Research on cross-linguistic transfer has shown that language learners simultaneously activate all known languages and parallel process these to decode words in lexical decision tasks with cognates (de Groot et al., 2002; Lemhöfer et al., 2004; Westergaard et al., 2017). The linguistic proximity model builds on these and similar findings and proposes that both facilitative and nonfacilitative influence from all languages previously acquired by learners influence learning any additional languages (Westergaard et al., 2017). The linguistic proximity model suggests that linguistic properties, if they receive ample support through learners' L1 can facilitate their acquisition of the new language. Facilitative transfer from one language to another supports a faster integration of, for example, existing grammatical structures or vocabulary knowledge. On the other hand, nonfacilitative transfer will interfere with the acquisition, requiring more engagement with a particular construct (Murphy, 2003). Accordingly, learning a language that shares similar linguistic features and lexical items facilitates acquisition (Muñoz et al., 2018; Odlin, 1989). Drawing from one's L1 and recognizing similar principles in another language can free cognitive resources for language learners (Sweller, 2011), that may be used for other less familiar linguistic properties.

LD has been operationalized in different ways. Approximations of LD have, for example, been made based on language trees and families (Spolaore & Wacziarg, 2016), expert judgments of language characteristics as, for example, in the *World Atlas of Language Structure* (Dryer & Haspelmath, 2013), and automated procedures that calculate language similarity based on phonetic similarity (*Automatic Similarity Judgement Program* [ASJP]; Wichmann et al., 2020). Language trees assume cardinality and are limiting in comparing languages from distinct, isolated language families (Chiswick & Miller, 2008). Although expert judgments on LD have been a reliable resource for a small number of languages, they are less accurate for larger numbers (Schepens et al., 2013; Wichmann et al., 2010) and are limiting in larger samples with a multitude of languages. On the other hand, automated judgments are available for a majority of languages, offer a transparent, objective, and reliable LD based on cognate linguistics, and are continuously distributed. Accordingly, cognate LD, the measure used in this study, has been found to be an "impressive predictor" (Schepens et al., 2013, p. 224) and is more reliable than other LD measures in predicting language proficiency (van der Slik, 2010).

Multilingual learners and classroom linguistic composition

ML learners face manifold inequities in education due to their potential language gaps. They are at an initial linguistic disadvantage in content and language classes that may not adequately address their language-learning needs in the majority language to support age-appropriate education (De Backer et al., 2017). Depending on educational contexts, integration into mainstream classes may be deferred or delayed if majority language skills are lacking, resulting in postponed acceptance into peer groups and potentially falling behind in content classes.

In addition, primary and secondary education teachers are not always prepared to serve MLs well across the curriculum (Heikkola et al., 2022; Lucas & Villegas, 2013). In most countries, education is conducted monolingually in the majority language, whereas English is the language of instruction in EFL classes. Accordingly, ML students could be at a significant disadvantage depending on their proficiency in and LD to the language of instruction while developing their initial reading and writing skills.

However, results on ML learners' attainment of school-based L3s are not conclusive. Although studies in bilingual contexts such as Catalonia or the Basque region in Spain demonstrate advantages for bilingual students (Cenoz, 2003; De Angelis, 2015), evidence from monolingual immigrant contexts is mixed. Study outcomes of EFL classrooms range from small disadvantages in selected subskills (Goorhuis-Brouwer & de Bot, 2010; Nikolova & Ivanov, 2010), to significant advantages (Hopp et al., 2020; Steinlen & Piske, 2018). However, particularly in immigrant contexts, advantages are typically small or emerge only once analyses control for background variables such as parental SES, education, or students' cognitive abilities (Maluch et al., 2015). Interestingly, elementary bilingual or immersion programs provide a context in which ML learners have been found to be consistently on par with their native-speaking peers (Hirosh & Degani, 2018; Steinlen & Piske, 2018). Although such emerging evidence points to the benefits of bilingual or immersion programs, the role of students' LDs to English has never been assessed in the EFL context. Bridging the gap from their mother tongue (L1) to the foreign language (i.e., English L3) may place ML learners in double jeopardy in the classroom, which may also affect the learning progress of the whole class, depending on a classroom's linguistic composition. Importantly, integrating ML students into the classroom requires educators to attend to not only linguistic barriers and L1 literacy but also diverse cultural backgrounds, immigration histories, and various other individual differences due to their intersectionality. Addressing these needs may require additional resources from teachers, students, and parents, but unfortunately, deficit-based descriptions are most prevalent in the current scientific and political discourse.

Studies investigating the influence of classroom heterogeneity or multilingual composition—that is, the proportion of non-native, linguistically diverse learners per group—on attainment have not yet outlined a clear association. Although some studies have documented a negative association between heterogeneity and academic achievement (Jensen, 2015; Stanat, 2006), others have argued that effects are negligible once analyses are controlled for school effects, as affluent, native-speaking (NS) parents may selectively choose schools with a lower proportion of immigrant students (Figlio et al., 2021; Ohinata & van Ours, 2011). In general, MLs' achievement tends to be more severely affected by the classroom composition than NS learners' (Bredtmann et al., 2021; Jensen, 2015; Ohinata & van Ours, 2013; Schneeweis, 2015). Analyses of the 2015 PISA assessments from 41 countries found that immigrant students performed similarly to NS peers, and all students, ML and NS alike, benefitted academically from

classroom heterogeneity (Silveira et al., 2019). More recently, Bredtmann et al. (2021) investigated associations between the proportion of ML students in a classroom on reading and math skills in a large sample of fourth-grade students from Germany. They showed that a higher proportion of ML students per class was associated with lower attainment, pointing to linguistic barriers that need to be overcome. Nevertheless, some have argued that such effects may only apply to recently arrived, first-generation immigrants and that MLs who have resided in the country for a few years do not negatively affect NS attainment (Bossavie, 2018).

The association of the proportion of ML students on foreign language learning has, to our knowledge, not been investigated. There is however great value for policy makers and educators in understanding whether classroom composition has an influence on achievement. Beyond assessing the status quo, these results can provide valuable evidence-based information on a need for changes in pedagogical approaches, teacher training, and student-focused research with resource-based solutions in mind.

Exploring the association between classroom composition and EFL achievement is also warranted if we consider that students with ML backgrounds have different starting points in learning a foreign language—that is, language similarities due to language proximity (Muñoz et al., 2018; Westergaard et al., 2017). Even when the immigrant and foreign language are related, monolingual resources in the majority language may not facilitate a transfer between their L1 and L3. Therefore, it is both timely and important to investigate the role LD plays in ML students' language learning.

Accordingly, although the verdict on possible negative or positive associations between classroom composition and student performance is still open, all evidence points to complex contextual processes that heavily confound this relationship. Considering intersectionality, students in classrooms with a higher proportion of linguistic diversity may experience an accumulation of adverse factors such as a low SES, contextual resources, and cognitive abilities that place them at a disadvantage. Disentangling these confounding effects and their relationships with student achievement is difficult to operationalize (Stanat, 2006) but critically important. This is why the current study uses a mixed-effects design with a range of confounding variables included in all analysis models.

Research questions

In summary, variations in LD have been shown to explain differences in the language proficiency of immigrants, over and above the influence of immigrant status itself (Schepens et al., 2013). Considering the substantial effect of immigrant status on adult second language attainment and occupational success, it is surprising that few studies have assessed the role of LD as a potential underlying explanatory mechanism during primary and secondary education. Moreover, obtaining proficiency in EFL is critical to higher education and professional success as well as societal participation. Finally, studies have suggested that high proportions of ML students in classrooms may be negatively associated with learning outcomes, but findings have rarely been controlled for the various confounding factors. This study investigates the influence of LD on receptive English skills at the beginning of Year 5, the first year of secondary school in Germany. The objective is to test whether German NS and ML (immigrant) students' LD to English (LDE), the language of instruction in EFL is associated with their English reading and listening proficiency after controlling for student ML status (binary, German NS versus ML), country of birth (binary, Germany versus abroad), biological

sex, cognitive abilities (nonverbal figural analogy test), and cultural capital (books at home). First, we hypothesize that a higher LDE is associated with lower receptive English proficiency. Second, we will test whether variations in the proportion of ML students per classroom are associated with individual students' receptive English proficiency. We hypothesize that a higher proportion of ML students is associated with lower receptive English proficiency.

Methods

Context

This study was part of a multischool project in the state of North-Rhine Westphalia (NRW), Germany, called Ganz In—All-Day Schools for a Brighter Future [*Mit Ganztag mehr Zukunft*]. Students in Germany are streamed into different secondary school tiers after Year 4. In the state of North-Rhine Westphalia (NRW), the lower secondary (*Hauptschule*) and middle secondary (*Realschule*, *Sekundarschule*) finish with a middle years' degree. The grammar school (*Gymnasium*) and comprehensive school (*Gesamtschule*) offer the option to finish with a high school diploma (*Abitur*) that provides direct access to tertiary education. With the focus on grammar schools in this study, all participants had already shown better-than-average academic achievement in elementary school, as grammar schools generally attract higher-achieving students. About 41% of all elementary school students are streamed into grammar schools in NRW (of these, 31% were ML students in 2019/20; Information und Technik Nordrhein-Westfalen, 2018, 2020).

Participants

The data consisted of two Year 5 cohorts of students ($N = 5{,}130$) from 31 grammar schools in NRW, Germany. Both rural and urban schools participated in the study and were dispersed evenly across the state. Schools applied to participate and did so voluntarily.

NRW introduced EFL in elementary schools in 2003 for Year 3 and moved it to the second half of Year 1 in 2008. The two cohorts, therefore, differ concerning the initial introduction of EFL. Cohort 1 received two years (140 hr) starting at 8–9 years of age with EFL teaching in elementary school, whereas Cohort 2 was 6–7 years of age and had received 3.5 years (245 hr). The difference in exposure to English between the cohorts was 105 lessons (45 min), or a total of 78 ¾ hr (Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008). All analyses are controlled for this cohort difference, which is not a variable of interest here.

Instruments

English proficiency

English listening and reading proficiency were assessed using standardized tests at the beginning of Year 5 (Engel & Ehlers, 2013; Engel et al., 2009). The tests were designed and piloted for a statewide EFL assessment and aligned with the state curriculum. Test validity and reliability were validated and normed on >3,000 students (Engel et al., 2009). Specifically, English listening proficiency was assessed through 28 multiple-choice questions targeting picture recognition (17 items) and sentence completion (11 items). English reading comprehension used 20 multiple-choice and four open-answer items

(α= .71). Identical tests were used for both cohorts. A simple one-dimensional logistic item response model (Rasch, 1980) was calculated using weighted likelihood estimators (Warm, 1989), to calculate English reading and listening scores. Items were checked for conformity by assessing item characteristic curves, discrimination parameters, mean squared errors, and their respective *t* values. Acceptable item fit levels were based on guidelines used in large-scale assessment studies, and items that did not conform to thresholds were excluded (Adams & Wu, 2002; Wright & Linacre, 1994). Final scores were scaled to a mean of 500 points and a standard deviation of 100.

Cognitive abilities

The Figural Analogy subtest of the standardized *Kognitiver Fähigkeitstest* (KFT; Heller & Perleth, 2000) was used to estimate general nonverbal intelligence. The test items reached excellent reliability values in our sample (α = .91). Raw standard scores are used in analyses.

Demographic background information

In addition to students' biological sex, country of birth, and home language(s), demographic variables included self-reported cultural capital (number of books at home). Students were asked how many books their family owned. The five categories included: 0–10, 11–25, 26–100, 101–200, or more than 200 books. Five pictograms of bookshelves were used to provide a visual aid to estimate the number of books. For home languages, participants were asked which languages they spoke with their mothers and fathers, respectively. All variables were based on students' self-report, and parent questionnaires were used to fill in missing student responses. Scales originated from the PISA, Trends in International Mathematics and Science Study, and Progress in International Reading Literacy Study assessments and were specifically adapted for use in Germany (Bos et al., 2009).

Linguistic distance

Students' mothers' and fathers' individually reported home languages were used to calculate their average cognate LD to English (LDE) as one continuous index score (see Appendix 1 for list of languages and LDE scores). As established previously by Schepens et al. (2013), cognate-based LD best explains between-language variation across learners. LDs were operationalized using data from the *Automated Similarity Judgement Program* (ASJP; Wichmann et al., 2020). The LD was calculated using a software, *Programs for calculating ASJP distance matrices* 2.1 (Holman, 2011), based on Brown et al. (2008). The software first computes the normalized Levenshtein distance (Levenshtein, 1966)—that is, the number of changes, including deletions, insertions, or substitutions, necessary to change the phonetic representation of a specific target word from one language to another. For example, *house* (English) to *haus* (German) has a Levenshtein distance of 2, whereas house (English) to *ev* (Turkish) has a Levenshtein distance of 5. The ASJP LD score is based on a shorter 40-word Swadesh list (Swadesh, 1952, 1955). This list contains 40 everyday words that have been shown to be universally important across languages and culturally independent. The ASJP software normalizes LD (LDN) by correcting for differences in word length by dividing the LD by the number of symbols of the longer word (Wichmann et al., 2010). A final

normalization is run to adjust for chance similarities between all words, LDN divided (LDND; Wichmann et al., 2020). The continuously distributed LDND score is used to operationalize LD in this study.

In Germany, for instance, LD may indicate the level of similarities and differences between Turkish, German, and English. Turkish L1 students, for example, need to overcome a much larger LD to the language of instruction, EFL, than their German classmates (see Figure 1). Dutch-speaking students in our sample, on the other hand, benefit from a much closer distance to English with respect to their L1-L3.

LD constitutes an individual difference and is associated with learners' abilities to draw on the linguistic resources available to them in their L1. Learners with lower LD scores—that is, shorter LD—benefit from shared linguistic properties between their L1 and English, which can facilitate their foreign language learning, as proposed by the linguistic proximity model (Westergaard et al., 2017). LD also suggests that learners have different starting points in their learning of a foreign language. The learning process may be facilitated by a reduced cognitive load for those learners with a lower LD, allowing them to focus on other less familiar linguistic properties involved in the foreign language. More research is necessary to substantiate this hypothesis, which is outside of the scope of the current paper. Without differentiated and scaffolded language classes, learners with a higher LD could experience a considerably higher cognitive strain in contrast to their lower LD peers.

Here, each student's mother's and father's LDs to English were individually computed as continuous LDND scores, following the steps described above, then averaged. In single-parent families, the respective score was used as the average. Importantly, due to this study's design and focus on EFL, we excluded $n = 579$ students who reported English as one of their home languages from analyses.

Procedure

Participation in the study was voluntary, and written consent was obtained from parents before data collection. Students assented to participate in the study on the day of the assessments. Data were collected between Weeks 5 and 9 of the new school year. Data collection was conducted during regular school lessons.
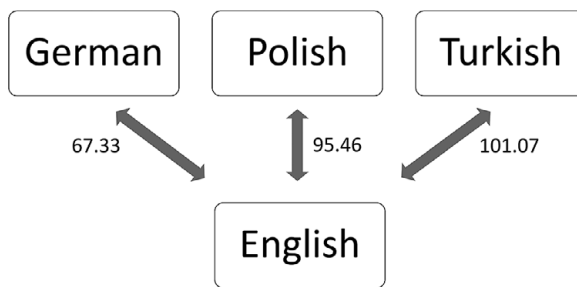
Figure 1. Examples of linguistic distance to English (LDE) scores for selected languages. Please note that a lower LD score indicates a higher similarity between two languages.

Analysis plan

Descriptive data were analyzed with IBM SPSS Statistics (version 27) and test data scaled with ConQuest 3.0.1 (Adams et al., 2012). For the multilevel main analyses, mixed-effects linear regressions were performed in STATA (version 16). As part of a stepwise model-building process, model fit was evaluated using log-likelihood goodness-of-fit tests and statistical parsimony criteria. We first ran unadjusted two-level models (i.e., Level 1: individuals, Level 2: classrooms), including fixed effects of students' LDEs on their English reading and listening proficiencies. Significant log-likelihood ratio chi-square tests for mixed-effects models versus simple linear models supported their fit; therefore, the models were retained (see detailed report below). We then tested whether adding another level (i.e., Level 3: schools) further improved model fit (rejected due to nonsignificant log-likelihood ratio chi-square tests). We then added a random effect for students' LDEs to the Level 2 models (nonsignificant effects, overall model fit not improved, rejected). Next, we tested fully adjusted Level 2 models by adding fixed effects of the confounding variables student ML status, country of birth, biological sex, cognitive abilities, cultural capital (books at home), and study cohort on English reading and listening proficiency. Significant log-likelihood ratio chi-square tests supported their fit over the unadjusted models. To test Hypothesis 2, a fixed and a random effect of the proportion of ML students per classroom on English reading and listening proficiency was added. Significant log-likelihood ratio chi-square tests supported their fit, and both fixed and random effects were retained. Finally, analyses were repeated separately for ML versus NS students to assess whether the associations of LD and the proportion of ML students per classroom on English reading and listening proficiency were different between the two subgroups.

Results

Comparative descriptive sample characteristics for the German native-speaking (NS) versus multilingual (ML) students are displayed in Table 1. By design, fewer students in the ML sample were born in Germany than in the German sample. However, 88.4% were second-generation immigrants who had been born in Germany. In addition, ML students' mothers' and fathers' average LDE scores were higher than their NS peers' parents' scores. On average, students in the ML sample were more often female, reported having fewer books at home, and had lower cognitive, English reading,

Table 1. Cognate list contrasting English with German (Germanic), Polish (Slavic), Spanish (Romance), and Turkish (Turkic)

English	German	Polish	Spanish	Turkish
you	**Du**	**ty**	**tú**	sen
woman	Frau	kobieta	mujer	kadın
house	**Haus**	dom	casa	ev
mountain	Berg	góra	**montaña**	dağ
blood	**Blut**	krew	sangre	kan
hair	**Haar**	włosy	cabello	saç
mouth	**Mund**	usta	boca	ağız
alphabet	**Alphabet**	**alfabet**	**alfabeto**	**alfabe**
lamp	**Lampe**	**lampa**	**lámpara**	**lamba**

Note: Cognates between English and other languages emphasized.

and listening test scores. Accordingly, students' biological sex, books at home, and cognitive abilities were included as control variables in all models.

We also calculated how many ML students were represented in each classroom and how they were distributed across schools. At the school level, the average proportion of ML students was 26.06% ($SD = 17.06$; range: 0%–89%), and there were, on average, 25.17% ($SD = 19.93$; range: 8%–74%) ML students per each participating classroom.

Linear mixed-effects regression models were run to test how students' individual LDEs were associated with their English reading and listening proficiency using the stepwise approach outlined above. Mixed-effects models indicated substantial heterogeneity between classrooms, supporting the use of multilevel models as opposed to linear regression models without random effects. Specifically, the log-likelihood ratio chi-square tests for mixed-effects models versus simple linear models were $\chi^2(1) = 10.23$, $p < .0001$ for reading, and $\chi^2(1) = 398.03$, $p < .0001$ for listening test scores. The inclusion of an additional third level for school (i.e., classrooms nested in schools) did however not result in improved model fit and was rejected for reasons of statistical parsimony. The initial, unadjusted models indicated negative associations of students' LDEs with their English reading ($R^2 = .16$) and listening proficiency ($R^2 = .22$). Next, we introduced the full set of control variables, including student ML status, country of birth, biological sex, cognitive abilities, cultural capital, and study cohort (Table 2), to test whether LDE had an independent added effect on English proficiency. The negative associations of students' LDEs with their English reading and listening proficiency remained significant with an improved overall model fit, indicated by a reduced log-likelihood value and a significant log-likelihood ratio chi-square test (assumption unadjusted model nested in fully adjusted model): for reading, $\chi^2(6) = 219.88$, $p < .0001$, $R^2 = .19$; for listening, $\chi^2(6) = 168.70$, $p < .0001$, $R^2 = .24$. Therefore, Hypothesis 1 was confirmed. Specifically, for instance, the effect of students' LDEs on English listening proficiency was -.66 (see Table 3). This corresponds to a decrease of 0.66 points in English listening proficiency scores with each 1-point increase in LD.

In other words, if all other factors in the model were held stable, a student's LDE of 101.07 points (for Turkish, see Figure 1) would result in a corresponding decrease in

Table 2. Descriptive sample characteristics

	ML sample (n = 1,086)	NS sample (n = 3,465)	χ^2 test / 95% confidence interval of the mean difference	p	d
Born in Germany (% yes)	88.4 %	99.2 %	$\chi^2 = 298.12$	<.001	.05
Biological sex (% female)	56.6 %	50.7 %	$\chi^2 = 11.74$	.001	1.26
Cultural capital (books)	2.86 (1.08)	3.65 (1.07)	−0.87 to −0.72	<.001	−.74
KFT test score (cognitive abilities)	16.16 (6.71)	18.07 (6.10)	−2.39 to −1.49	<.001	−.31
LDE mother	90.72 (12.94)	69.91 (8.18)	19.99 to 21.62	<.001	2.18
LDE father	88.93 (13.95)	69.75 (8.02)	18.31 to 20.05	<.001	1.96
LDE averaged	89.82 (12.39)	69.83 (7.49)	19.21 to 20.77	<.001	2.24
English reading proficiency	487.83 (97.41)	500.64 (97.65)	−19.47 to −6.16	.001	−.13
English listening proficiency	478.29 (96.20)	503.07 (95.95)	−31.32 to −18.23	<.001	−.26

Note: Data are reported as mean (*SD*) if not stated otherwise; ML = multilingual, NS = native speaking, and LD = linguistic distance.

Table 3. Multilevel linear mixed-effects models showing associations of students' individual LDEs with their English reading and listening proficiency test scores (N = 4,551)

Dependent variable	English reading proficiency	English listening proficiency
Fixed effects		
LDE	**−.63 (−.94, −.32)*****	**−.66 (−.95, −.36)*****
Biological sex	−18.28 (−23.67, −12.90)***	−10.57 (−15.75, −5.38)***
Country of birth	−19.55 (−34.95, −4.14)*	−15.23 (−30.06, −0.39)*
ML status	−8.67 (−17.73, 0.40)	−3.53 (−12.28, 5.22)
Books	4.63 (2.06, 7.20)***	3.80 (1.31, 6.28)**
Study cohort	27.66 (19.32, 36.00)***	33.85 (24.26, 43.45)***
Cognitive abilities	2.36 (1.92, 2.79)***	2.10 (1.67, 2.52)***
constant	526.23 (477.82, 574.64)***	497.20 (449.83, 544.57)***
Random effects by classroom		
SD (constant)	25.42 (21.83, 29.60)*	32.23 (28.33, 36.68)*
Log likelihood	−27,076.24	−27,013.58

Note: For fixed effects, *p < .05; **p < 0.01; ***p < 0.001. For random effects, * marks a 95% confidence interval not including 0. Significant effects of interest are marked in **bold**. LDE = linguistic distance to English; ML = multilingual.

English listening proficiency scores of 66.71 points, substantially more than half a standard deviation. Figure 2 shows unadjusted (i.e., not corrected for confounders) average English reading and listening proficiency scores by LDEs for selected language groups in our sample.

To test Hypothesis 2, we introduced a fixed and a random effect for the proportion of ML students per classroom into our models (Table 4). Again, model fit improved significantly: for reading, $\chi^2(2) = 13.30$, $p < .01$, $R^2 = .19$; for listening, $\chi^2(6) = 41.57$, $p < .0001$, $R^2 = .24$. Results showed that a higher proportion of ML students per classroom was associated with lower English reading and listening proficiency, thus confirming Hypothesis 2. In addition, the random effect of the proportion of ML students per classroom on English reading proficiency was estimated at $SD = 0.05$ 95% CI [0.00, 0.77], indicating variation between classrooms in the relationship between these two variables. The negative associations of students' LDEs with their English reading and listening proficiency remained significant.

Finally, to assess whether the associations between LDE and English performance were different among ML versus NS students, we repeated our analyses separately for these two subgroups. We found the same patterns as described above throughout the stepwise model-building process. Table 5 shows that, as could be expected, among ML students the negative association of LDE with their English reading and listening proficiency was more pronounced than for the whole sample analysis (total model $R^2 = .35$ and .40, respectively), whereas there was no effect of LDE on German NS students' English reading and listening proficiency (total model $R^2 = .19$ and .23, respectively). In both groups, a higher proportion of ML students per classroom was associated with lower English reading and listening proficiency. Additionally, significant random effects indicated substantial variation between classrooms with respect to the effect of linguistic composition on English proficiency. For instance, if all other predictors were held stable, the effect of the proportion of ML students per classroom on English listening proficiency was -.83 (ML) and -.75 (NS), respectively. This corresponds to a decrease of 0.83 (0.75) points in English listening proficiency scores
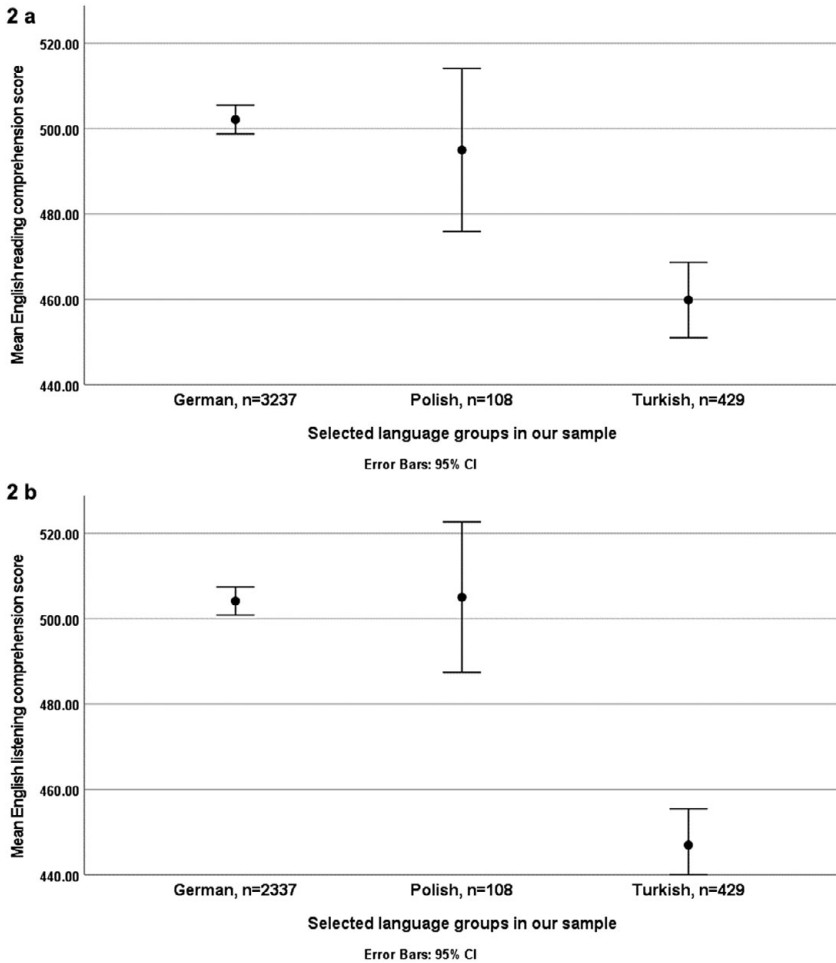
Figure 2. Unadjusted average English reading (2a) and listening (2b) proficiency scores by students' LDEs for selected language groups in our sample.

among ML (among NS) students with each percentage-point increase in the proportion of ML students per classroom.

Discussion

In this study, we investigated the association of students' cognate LDs to EFL with their receptive English proficiency. Mixed-effects analyses in a large and diverse sample demonstrated that both ML and NS students' LDEs predicted Grade 5 English listening and reading scores after controlling for student ML status, country of birth, biological sex, cognitive abilities, and cultural capital. Comparative separate analyses of NS and ML students uncovered a few distinct differences. Among ML students, LDE was strongly associated with English reading and listening proficiency. The separate parallel analyses among their NS peers did not uncover significant

Table 4. Multilevel linear mixed-effects models showing associations of students' individual LDEs and ML percentage per classroom with English reading and listening proficiency test scores (N = 4,551)

Dependent variable	English reading proficiency	English listening proficiency
Fixed effects		
LDE	**−.56 (−.87, −.25)*****	**−.56 (−.86, −.26)*****
%ML / classroom	**−.40 (−.62, −.19)*****	**−.78 (−1.01, −.55)*****
Biological sex	−18.17 (−23.55, −12.79)***	−10.40 (−15.58, −5.23)***
Country of birth	−18.90 (−34.30, −3.50)*	−14.29 (−29.10, 0.52)
ML status	−11.03 (−20.17, −1.88)*	−6.90 (−15.69, −1.89)
Books	4.13 (1.54, 6.71)**	3.13 (0.64, 5.62)*
Study cohort	27.29 (19.13, 35.46)***	33.15 (24.27, 42.04)***
Cognitive abilities	2.31 (1.87, 2.75)***	2.04 (1.62, 2.46)***
constant	537.27 (488.63, 585.91)***	518.42 (471.09, 565.75)***
Random effects by classroom		
SD (%ML / classroom)	**.05 (.00, .77)***	**.00 (.00, .00)***
SD (constant)	24.45 (20.92, 2.58)*	28.82 (25.15, 33.02)*
Log likelihood	−27,069.59	−26,908.45

Note: For fixed effects, *p* < .05; **p* < 0.01; ****p* < 0.001. For random effects, * marks a 95% confidence interval not including 0. Significant effects of interest are marked in **bold**. LDE = linguistic distance to English; ML = multilingual.

associations of LDE with receptive English skills, as was expected due to the largely homogeneous language group and resulting lack of statistical variance in LDE. Further, our findings showed that students in classes with a higher proportion of ML students tended to have lower English reading and listening skills and that this association was independent of individual students' LDs, ML status, country of birth, biological sex, cognitive abilities, and cultural capital. It is critically important not to interpret this finding as a student deficit or simply attribute the outcome to teachers; it rather represents a systemic issue. Content and language teachers are not well prepared through either teacher education programs or professional development to adequately address heterogeneous classrooms (Heikkola et al., 2022; Lucas & Villegas, 2013), and resources are often lacking.

Our data confirm that, in classroom contexts, students with a smaller LDE perform better, being able to draw on linguistic resources in their L1. The role of LDE in our analyses indicates that students who speak a language similar to English at home, such as German or Dutch, have an advantage in EFL over their ML peers speaking more distinct languages—for example, Turkish or Arabic. Therefore, our results support emerging evidence that a greater LD between L1 and EFL poses potential risks to successful language learning beyond immigrant background alone, particularly if it coincides with other individual difference risks such as male sex or low cognitive abilities (Blom et al., 2020; Borgonovi & Ferrara, 2020). The strong association between LDE and receptive English proficiency is not a surprising finding, as EFL classes in Germany rely almost exclusively on English as the language of instruction (Gogolin, 2021). Importantly, however, our data suggest that multilingualism, which is evident across our sample with an average of 25% MLs in each class and reaching up to 89% in some classes, is not addressed sufficiently during EFL instruction.

Linguistic heterogeneity in foreign languages and all classrooms deserves more attention from teachers, textbook publishers, and policymakers to address disparities based on LD. Teacher training in Germany involves coursework on pedagogical approaches to address heterogeneity; however, these may not specifically incorporate multilingualism in foreign language education. In-service teachers also require ongoing

Table 5. Multilevel linear mixed-effects models showing associations of students' individual LDEs and ML percentage per classroom with English reading and listening comprehension test scores separately for ML (n = 1.086) and NS students (n = 3,465)

ML students		
Dependent variable	English reading proficiency	English listening proficiency
Fixed effects		
LDE	**−.91 (−1.36, −.46)*****	**−.89 (−1.32, −.45)*****
%ML / classroom	**−.44 (−.76, −.13)****	**−.83 (−1.16, −.51)*****
Biological sex	−13.50 (−24.45, −2.56)*	−6.37 (−16.93, 4.11)
Country of birth	−20.64 (−37.74, −3.54)*	−15.96 (−32.44, .51)
Books	3.39 (−1.79, 8.57)	2.33 (−2.66, 7.32)
Study cohort	18.33 (4.74, 31.92)**	25.05 (11.30, 38.80)***
Cognitive abilities	2.34 (1.51, 3.17)***	2.10 (1.29, 2.90)***
constant	571.80 (509.40, 634.20)***	554.45 (493.83, 615.07)***
Random effects by classroom		
SD (%ML / classroom)	**.17 (.00, 6.99)***	**.21 (.01, 5.87)***
SD (constant)	26.51 (15.88, 44.24)*	29.03 (17.52, 48.10)*
Log likelihood	−6,452.30	−6,413.95
NS students		
Dependent variable	English reading proficiency	English listening proficiency
Fixed effects		
LDE	−.28 (−.72, .15)	−.27 (−.69, .14)
%ML / classroom	**−.34 (−.59, −.08)*****	**−.75 (−1.02, −.48)*****
Biological sex	−19.98 (−26.19, −13.77)***	−11.67 (−17.65, −5.69)***
Country of birth	−9.86 (−45.10, 25.37)	−3.25 (−37.19, 30.70)
Books	4.88 (1.88, 7.88)***	3.92 (1.02, 6.82)**
Study cohort	28.73 (20.10, 37.36)***	33.99 (24.63, 43.35)***
Cognitive abilities	2.37 (1.86, 2.89)***	2.03 (1.53, 2.52)***
Constant	473.29 (392.48, 554.10)***	459.65 (381.45, 537.85)***
Random effects by classroom		
SD (%ML / classroom)	**.00 (.00, .00)***	**.15 (.00, 47.18)***
SD (constant)	23.25 (19.30, 28.00)*	27.73 (22.85, 33.66)*
Log likelihood	−20,626.88	−20,510.55

Note: For fixed effects *p* < .05; **p* < 0.01; ****p* < 0.001. For random effects, * marks a 95% confidence interval not including 0. Significant effects of interest are marked in **bold**. LDE = linguistic distance to English; ML = multilingual.

professional development to learn about new approaches and adjust their teaching to better serve a growing population of ML students in most societies. Translanguaging, for example, is one of these new approaches that promise to support ML learners by allowing the purposeful use of other linguistic resources to resolve communication barriers in class (García & Li, 2014). However, translanguaging may not always show clear benefits for English proficiency (Hopp et al., 2020), partly because higher English exposure in classes has been linked to higher proficiency scores (Helmke et al., 2008). Our finding that the proportion of ML students affects all students' receptive English skills aligns with other research in the field (Bredtmann et al., 2021; Jensen, 2015; Stanat, 2006). Specifically, Bossavie's (2018) findings that the proportion of ML students does not affect achievement if MLs have lived in the country for a short while could not be replicated in our sample, which mainly consisted of second-generation, high-achieving ML students. The implications of these results are a cause for concern, as our sample only included students who had already demonstrated better-than-average academic promise in elementary school and therefore represents a selective

group streamed into grammar schools who are likely doing better than their peers at other secondary schools who were not assessed here. Considering that LDE was consistently negatively associated with English proficiency within such a selective sample suggests that adverse effects of higher LD on learning outcomes might be even more pronounced in other, more heterogeneous samples. Therefore, although our novel results may be considered seminal, replication in other more diverse and less selective samples and contexts is warranted.

Textbook resources, which EFL teachers in Germany heavily rely on, often encourage language contrasts between German and English. Considering students' ability to draw on all of their linguistic resources as outlined in the linguistic proximity model (Westergaard et al., 2017), it would be beneficial for ML students to have access to, for example, digital EFL resources that provide and encourage language contrasts between English and their L1 and that outline grammatical rules from a non-German perspective. Individual language teachers cannot easily provide this level of differentiation; however, school-book publishers should be required to better cater to a growing multilingual community. Providing digital or app-based English-L1 wordlists for ML students could also facilitate their vocabulary learning.

The success of ML students in immersive English programs in elementary schools may offer insights into possible solutions (Hirosh & Degani, 2018; Steinlen, 2017). Here, the increased exposure to English and meaningful use of the language in content learning could offer an equitable path for ML learners. The more heterogeneous a classroom, the more students are at risk of falling behind expected learning outcomes. Therefore, we need to raise awareness that higher linguistic diversity requires substantially higher resources from teachers, schools, and the education system in general in order to provide equitable education to all students. Future education and language studies need to better account for linguistic diversity in schools and classrooms.

With respect to ML parents, today's scientific and political discourse agrees that strengthening people's and families' native languages is critical to preserving cultural heritage and provides the foundation for learning any additional languages (Carreira & Kagan, 2011). Nevertheless, one recommendation to ML parents could be to offer their children age-appropriate extracurricular access to English language learning opportunities.

The fact that ML students lag behind their NS peers in Grade 5 also draws attention to the implicit learning approaches of foreign languages employed in elementary schools (Jaekel et al., 2021; Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen, 2008; Piske, 2017). Learning languages through playful means should not put ML students at a disadvantage due to their LD but rather benefit them as they have already managed to become bilingual. Here, teachers may feel uncomfortable discussing and contrasting other languages with English as they feel this may be outside their area of expertise. Consequently, educational policies, teacher education, and professional development need to address the changing linguistic landscape in classrooms (Suárez-Orozco et al., 2008). Beyond these measures, teachers' own backgrounds also must increasingly reflect societal changes—that is, more multilingual teachers need to be trained to represent the student body and lived experiences of growing up bilingually.

Finally, the current study emphasizes the important role of fine-grained linguistic differences between students, considering both individuals and whole classrooms. LD measures offer novel and differentiated pedagogical and research perspectives independent of learners' language proficiencies. What's more, LD proxy measures can be included in large-scale studies and surveys, offering valuable tools and possibly new

answers to long-standing research questions in multilingual education and cross-cultural psychology.

Strengths and limitations

This study used data from a large, well-documented sample that provided a good representation of ML and NS students in grammar school classrooms in Germany. The use of linear mixed models allowed us to account for the nestedness of students within classes while controlling for a broad range of possible confounders, thereby providing a robust assessment of the two hypotheses. Although the analyses controlled for cultural capital (books at home), SES was not included in the main analyses due to a large number of missing values for parents' occupation and income.

Our study focuses on EFL and English proficiency, and therefore it did not include assessments of students' German proficiency, nor did we assess productive English skills. Bilingualism alone does not facilitate L3 acquisition, but strong evidence points to biliteracy as a supportive factor (Cenoz, 2003; Rauch, 2014). Assessments of ML students' L1 literacy skills in L3 research focused on immigrants, particularly in larger samples, remain scarce. Recent studies on L3 attainment considered immigrant ML students' proficiency in their heritage language and its influence on L3 achievement, with mixed results. Edele et al. (2018) divided MLs into ability groups demonstrating that students' L1 proficiency predicted better English scores. In contrast, Lorenz et al. (2022) showed only a weak effect of students' heritage language skills on L3 skills. In contrast, our study operationalized LD via self-reported proxy scores without measuring heritage language proficiency, but we were able to show fairly strong and stable associations with English listening and reading, confirming previous evidence (Muñoz et al., 2018; Schepens et al., 2013). Future research on L3 learning of MLs should consider both L1 and L2 proficiency across both receptive and productive dimensions while controlling for LD.

Studies on L3 learning have consistently demonstrated that extralinguistic individual-difference variables, particularly student biological sex, SES, and cognitive skills, have a substantial influence on foreign language outcomes (Edele et al., 2018; Jaekel et al., 2017; Lorenz et al., 2022). In Hopp et al. (2019), ML students even outperformed NS students on some skills once analyses were controlled for background variables. Although our study confirms the importance of controlling for such confounding variables, it also shows that disentangling independent effects is statistically possible (see Tables 3–5). Nevertheless, across most classrooms and societies nonnative status of individual students and linguistic classroom heterogeneity both remain associated with accumulations of adverse contextual factors that place students at an academic disadvantage (Stanat, 2006), and statistically disentangling these variables does not improve the real lives of students facing cumulative challenges. One opportunity for future studies to identify such variables and shed more light on their complex intersectionality may be person-centered (e.g., latent class) analyses.

Conclusion

Societies across the globe are increasingly linguistically diverse, which is reflected in today's classrooms. Our results demonstrate that students' cognate LD to English as well as the proportion of MLs per classroom are robustly and independently associated

with receptive English skills. Consequently, EFL classes need to address linguistic diversity in order to bridge the continuous distance gap, which puts some learners at a disadvantage.

Competing interest. The author(s) declare none.

References

Adams, R., & Wu, M. (2002). *PISA 2000 technical report*. OECD.

Adams, R., Wu, M., & Wilson, M. R. (2012). *ACER Conquest* [Computer software]. ACER Press.

Auer, D. (2018). Language roulette: The effect of random placement on refugees' labour market integration. *Journal of Ethnic and Migration Studies*, *44*, 341–362. DOI: 10.1080/1369183X.2017.1304208

Blom, E., Boerma, T., Bosma, E., Cornips, L., van den Heuij, K., & Timmermeister, M. (2020). Cross-language distance influences receptive vocabulary outcomes of bilingual children. *First Language*, *40*, 151–171. DOI: 10.1177/0142723719892794

Borgonovi, F., & Ferrara, A. (2020). Academic achievement and sense of belonging among non-native-speaking immigrant students: The role of linguistic distance. *Learning and Individual Differences*, *81*, Article 101911. DOI: 10.1016/j.lindif.2020.101911

Bos, W., Bonsen, M., Gröhlich, C., Guill, K., & Scharenberg, K. (Eds). (2009). *KESS 7: Skalenhandbuch zur dokumentation der erhebungsinstrumente*. Waxmann.

Bossavie, L. (2018). *The effect of immigration on natives' school achievement: Does length of stay in the host country matter?* [Policy Research Working Paper No. 8492]. World Bank Group. https://openknowledge.worldbank.org/bitstream/handle/10986/29444/ WPS8492.pdf?sequence=4&isAllowed=y

Bousmah, I., Grenier, G., & Gray, D. M. (2021). Linguistic distance, languages of work and wages of immigrants in Montreal. *Journal of Labor Research*, *42*, 1–28. DOI:10.1007/s12122-020-09316-1

Bredtmann, J., Otten, S., & Vonnahme, C. (2021). Linguistic diversity in the classroom, student achievement, and social integration. *Education Economics*, *29*, 121–142. DOI:10.1080/09645292.2020.1866499

Brown, C. H., Holman, E. W., Wichmann, S., & Velupillai, V. (2008). Automated classification of the world's languages: A description of the method and preliminary results. *Language Typology and Universals*, *61*, 285–308. DOI:10.1524/stuf.2008.0026

Carreira, M., & Kagan, O. (2011). The results of the National Heritage Language Survey: Implications for teaching, curriculum design, and professional development. *Foreign Language Annals*, *44*, 40–64. DOI: 10.1111/j.1944-9720.2010.01118.x

Cenoz, J. (2001). The effect of linguistic distance, L2 status and age on cross-linguistic influence in third language acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic Influence in third language acquisition* (pp. 8–20). Multilingual Matters. DOI: 10.21832/9781853595509-002

Cenoz, J. (2003). The additive effect of bilingualism on third language acquisition: A review. *International Journal of Bilingualism*, *7*, 71–87. SAGE Publications.

Chiswick, B. R., & Miller, P. W. (2004). *Linguistic distance: A quantitative measure of the distance between English and other languages* [SSRN Scholarly Paper No. ID 575090]. Social Science Research Network. https://papers.ssrn.com/abstract= 575090.

Chiswick, B. R., & Miller, P. W. (2008). Linguistic distance: A quantitative measure of the distance between english and other languages. *Journal of Multilingual and Multicultural Development*, *26*, 1–11. DOI: 10.1080/14790710508668395

De Angelis, G. (2015). English L3 learning in a multilingual context: The role of parental education and L2 exposure within the living community. *International Journal of Multilingualism*, *12*, 435–452. DOI: 10.1080/14790718.2015.1071017

De Backer, F., Van Avermaet, P., & Slembrouck, S. (2017). Schools as laboratories for exploring multilingual assessment policies and practices. *Language and Education*, *31*, 217–230. DOI: 10.1080/09500782.2016.1261896

de Groot, A. M. B., Borgwaldt, S., Bos, M., & van den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, *47*, 91–124. DOI: 10.1006/jmla.2001.2840

Dryer, M., & Haspelmath. (2013). *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology. https://wals.info/

Edele, A., Kempert, S., & Schotte, K. (2018). Does competent bilingualism entail advantages for the third language learning of immigrant students? *Learning and Instruction*, *58*, 232–244. DOI: 10.1016/j.learninstruc.2018.07.002

Engel, G., & Ehlers, G. (2013). Hören–zuhören–verstehen. Möglichkeiten der analyse, diagnose und gezielten förderung des hörverstehens. In O. Börner, G. Engel, & B. Groot-Wilken. (Eds.), *Hörverstehen, leseverstehen, sprechen* (pp. 44–69). Waxmann.

Engel, G., Groot-Wilken, B., & Thürmann, E. (Eds). (2009). *Englisch in der primarstufe - chancen und herausforderungen: Evaluation und erfahrungen aus der praxis*. Cornelsen.

Figlio, D., Giuliano, P., Marchingiglio, R., Ozek, U., & Sapienza, P. (2021). *Diversity in schools: Immigrants and the educational performance of U.S. born students* [CEPR Discussion Paper No. DP15933]. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm? abstract_id=3816804

Flisi, S., Meroni, E. C., & Vera-Toscano, E. (2016). *Educational outcomes and immigrant background*. Publications Office of the European Union.

García, O., & Li, W. (2014). *Translanguaging: language, bilingualism and education*. Palgrave Macmillan.

Genesee, F., & Fortune, T. W. (2014). Bilingual education and at-risk students. *Journal of Immersion and Content-Based Language Education*, *2*, 196–209. DOI: 10.1075/jicb.2.2.03gen

Gogolin, I. (2021). Multilingualism: A threat to public education or a resource in public education? European histories and realities. *European Educational Research Journal*, *20*, 297–310. DOI: 10.1177/1474904 120981507

Goorhuis-Brouwer, S., & de Bot, K. (2010). Impact of early English language teaching on L1 and L2 development in children in Dutch schools. *International Journal of Bilingualism*, *14*, 289–302. DOI: 10.1177/1367006910367846

Gooskens, C., van Heuven, V. J., Golubović, J., Schüppert, A., Swarte, F., & Voigt, S. (2018). Mutual intelligibility between closely related languages in Europe. *International Journal of Multilingualism*, *15*, 169–193. DOI: 10.1080/14790718.2017.1350185

Goriot, C., van Hout, R., Broersma, M., Lobo, V., McQueen, J. M., & Unsworth, S. (2021). Using the Peabody Picture Vocabulary Test in L2 children and adolescents: Effects of L1. *International Journal of Bilingual Education and Bilingualism*, *24*, 546–568. DOI: 10.1080/13670050.2018.1494131

Heikkola, L. M., Alisaari, J., Vigren, H., & Commins, N. (2022). Requirements meet reality: Finnish teachers' practices in linguistically diverse classrooms. *Journal of Language, Identity & Education*. DOI: 10.1080/15348458.2021.1991801

Heller, K. A., & Perleth, C. (2000). *KFT 4-12+R - Kognitiver Fähigkeits-Test für 4. bis 12. Klassen: Revision*. Beltz.

Helmke, A., Helmke, T., Schrader, F.-W., Wagner, W., Klieme, E., & Nold, G. (2008). Wirksamkeit des Englischunterrichts. In E. Klieme (Ed.), *Unterricht und kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie* (pp. 382–397). Beltz.

Hippe, R., & Jakubowski, M. (2018). *Immigrant background and expected early school leaving in Europe: Evidence from PISA*. Publications Office of the European Union.

Hirosh, Z., & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin & Review*, *25*, 892–916.

Holman, E. W. (2011). *Programs for calculating ASJP distance matrices* (version 2.1) [Computer software]. https://asjp.clld.org/software

Hopp, H., Jakisch, J., Sturm, S., Becker, C., & Thoma, D. (2020). Integrating multilingualism into the early foreign language classroom: Empirical and teaching perspectives. *International Multilingual Research Journal*, *14*, 146–162. DOI: 10.1080/19313152.2019.1669519

Hopp, H., Vogelbacher, M., Kieseier, T., & Thoma, D. (2019). Bilingual advantages in early foreign language learning: Effects of the minority and the majority language. *Learning and Instruction*, *61*, 99–110. DOI: 10.1016/j.learninstruc.2019.02.001

Isphording, I., & Otten, S. (2013). The costs of Babylon—Linguistic distance in applied economics. *Review of International Economics*, *21*, 354–369. DOI: 10.1111/roie.12041

Information und Technik Nordrhein-Westfalen (2018). *Im sommer 2017 wechselten zwei von fünf kindern in NRW von einer grundschule zum gymnasium.* https://www.it.nrw/im-sommer-2017-wechselten-zwei-von-fuenf-kindern-nrw-von-einer-grundschule-zum-gymnasium-90833

Information und Technik Nordrhein-Westfalen. (2020). *Anteil der schülerinnen und schüler mit zuwanderungsgeschichte in NRW auf 38,2 prozent gestiegen.* https://www.it.nrw/anteil-der-schuelerinnen-und-schueler-mit-zuwanderungsgeschichte-nrw-auf-382-prozent-gestiegen

Jaekel, J., & Leyendecker, B. (2016). Inconsistent paternal behavior predicts Turkish immigrant and German children's and adolescents mental health. *Journal of Child and Adolescent Behaviour*, 4, Article 1000307. DOI: 10.4172/2375-4494.1000307

Jaekel, N., van Ackern, I., Schurig, M., & Markus, R. (2021). Investigating the transition from elementary to secondary school: student perspectives of the continuum and receptive language proficiency. *The Language Learning Journal*, 50, 1–16. DOI: 10.1080/09571736.2021.1879906

Jaekel, N., Schurig, M., Florian, M., & Ritter, M. (2017). From early starters to late finishers? A longitudinal study of early foreign language learning in school. *Language Learning*, 67, 631–664. DOI: 10.1111/lang.12242

Jensen, P. (2015, October). Immigrants in the classroom and effects on native children. *IZA World of Labor*, 2015, Article 194. https://wol.iza.org/uploads/articles/194/pdfs/immigrants-in-classroom-and-effects-on-native-children.pdf

Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and Cognitive Processes*, 19, 585–611. DOI: 10.1080/01690960444000007

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals | BibSonomy. *Cybernetics and Control Theory*, 10, 707–710.

Lindgren, E., & Muñoz, C. (2013). The influence of exposure, parents, and linguistic distance on young European learners' foreign language comprehension. *International Journal of Multilingualism*, 10, 105–129. DOI: 10.1080/14790718.2012.679275

Lorenz, E., Lorenz, E., Toprak-Yildiz, T. E., & Siemund, P. (2022). Why are they so similar? The interplay of linguistic and extra-linguistic variables in monolingual and bilingual learners of English. *Pedagogical Linguistics.* DOI: 10.1075/pl.21016.lor

Lucas, T., & Villegas, A. M. (2013). Preparing linguistically responsive teachers: Laying the foundation in preservice teacher education. *Theory Into Practice*, 52, 98–109. DOI: 10.1080/00405841.2013.770327

Maluch, J. T., Kempert, S., Neumann, M., & Stanat, P. (2015). The effect of speaking a minority language at home on foreign language learning. *Learning and Instruction*, 36, 76–85.

Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen. (2008). *Richtlinien und lehrpläne zur erprobung für die grundschule in Nordrhein-Westfalen.* https://www.schulentwicklung.nrw.de/lehrplaene/upload/klp_gs/LP_GS_2008.pdf

Muñoz, C., Cadierno, T., & Casas, I. (2018). Different starting points for English language learning: A comparative study of Danish and Spanish young learners. *Language Learning*, 68, 1076–1109. DOI: 10.1111/lang.12309

Murphy, S. (2003). Second language transfer during third language acquisition. *Studies in Applied Linguistics and TESOL*, 3, Article 1625. https://doi.org/10.7916/salt.v3i2.1625

Nikolova, R., & Ivanov, S. (2010). Englischleistungen. In W. Bos (Ed.), *KESS 8* (pp. 49–66). Waxmann.

Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning.* Cambridge University Press. DOI: 10.1017/CBO9781139524537

OECD. (2006). *Where immigrant students succeed: A comparative review of performance and engagement in PISA 2003.* DOI: 10.1787/9789264023611-en

OECD. (2012). *Equity and quality in education: supporting disadvantaged students and schools.* https://www.oecd.org/education/school/49603617.pdf

OECD. (2018). *Percentage of non-native speakers, by first language spoken at home and understood.* https://www.oecd-ilibrary.org/education/skills-on-the-move/percentage-of-non-native-speakers-by-first-language-spoken-at-home-and-understood_9789264307353-table45-en

OECD. (2019a). *PISA 2018 results: What students know and can do.* DOI: 10.1787/5f07c754-en

OECD. (2019b). *PISA 2018 results (volume II): Where all students can succeed.* DOI: 10.1787/b5fd1b8f-en

Ohinata, A., & van Ours, J. C. (2011). *How immigrant children affect the academic achievement of native Dutch children* [CEPR Discussion Paper. No. DP8718]. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3816804

Ohinata, A., & van Ours, J. C. (2013). How immigrant children affect the academic achievement of native Dutch children. *The Economic Journal*, *123*, F308–F331. DOI: 10.1111/ecoj.12052

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism*, *1*, 213–237. DOI: 10.1075/lab.1.3.01par

Piske, T. (2017). The earlier the better? Some critical remarks on current EFL teaching to young learners and their implications for foreign language teacher education. In E. Wilden & R. Porsch (Eds.), *The professional development of primary EFL teachers: National and International Research* (pp. 45–57). Waxmann.

Pulinx, R., Van Avermaet, P., & Agirdag, O. (2017). Silencing linguistic diversity: The extent, the determinants and consequences of the monolingual beliefs of Flemish teachers. *International Journal of Bilingual Education and Bilingualism*, *20*, 542–556. DOI: 10.1080/13670050.2015.1102860

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. University of Chicago Press.

Rauch, D. P. (2014). Effects of biliteracy on third language reading proficiency, the example of Turkish-German bilinguals. In P. Grommes & A. Hu (Eds.), *Plurilingual education: Policies - practices - language development* (pp. 199–218). John Benjamins.

Rodríguez, S., Valle, A., Gironelli, L. M., Guerrero, E., Regueiro, B., & Estévez, I. (2020). Performance and well-being of native and immigrant students. Comparative analysis based on PISA 2018. *Journal of Adolescence*, *85*, 96–105. DOI: 10.1016/j.adolescence.2020.10.001

Schepens, J. (2015). *Bridging linguistic gaps the effects of linguistic distance on the adult learnability of Dutch as an additional language*. LOT publications.

Schepens, J., van Hout, R., & Jaeger, T. F. (2020). Big data suggest strong constraints of linguistic similarity on adult language learning. *Cognition*, *194*, Article 104056. DOI: 10.1016/j.cognition.2019.104056

Schepens, J., van der Slik, F., & van Hout, R. (2013). Learning complex features: A morphological account of L2 learnability. *Language Dynamics and Change*, *3*, 218–244. Brill. DOI: 10.1163/22105832-13030203

Schneeweis, N. (2015). Immigrant concentration in schools: Consequences for native and migrant students. *Labour Economics*, *35*, 63–76. DOI: 10.1016/j.labeco.2015.03.004

Silveira, F., Dufur, M. J., Jarvis, J. A., & Rowley, K. J. (2019). The influence of foreign-born population on immigrant and native-born students' academic achievement. *Socius: Sociological Research for a Dynamic World*, *5*, Article 237802311984525. DOI: 10.1177/2378023119845252

Spolaore, E., & Wacziarg, R. (2016). Ancestry, language and culture. In V. Ginsburgh & S. Weber (Eds.), *The Palgrave handbook of economics and language* (pp. 174–211). Palgrave Macmillan. DOI: 10.1007/978-1-137-32505-1_7

Stanat, P. (2006). Schulleistungen von jugendlichen mit migrationshintergrund: Die rolle der zusammen-setzung der schülerschaft. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte disparitäten im bildungswesen: Differenzielle bildungsprozesse und probleme der verteilungsgerechtigkeit* (pp. 189–219). VS Verlag für Sozialwissenschaften. DOI: 10.1007/978-3-531-90082-7_5

Statistisches Bundesamt. (2021). *Migration und Integration*. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Migration-Integration/_inhalt.html

Steinlen, A. (2017). The development of English grammar and reading comprehension by majority and minority language children in a bilingual primary school. *Studies in Second Language Learning and Teaching*, *7*, 419–442. DOI: 10.14746/ssllt.2017.7.3.4

Steinlen, A., & Piske, T. (2018). Deutsch- und Englischleistungen von kindern mit und ohne migrationshinter-grund im bilingualen unterricht und im fremdsprachenunterricht: Ein vergleich. In A. Ballis & N. Hodaie (Eds.), *Perspektiven auf mehrsprachigkeit* (pp. 85–98). De Gruyter. DOI: 10.1515/9783110535334-006

Strøm, S., Piazzalunga, D., Venturini, A., & Villosio, C. (2018). Wage assimilation of immigrants and internal migrants: The role of linguistic distance. *Regional Studies*, *52*, 1423–1434. DOI: 10.1080/00343404.2017.1395003

Suárez-Orozco, C., Suárez-Orozco, M. M., & Todorova, I. (2008). *Learning a new land: Immigrant students in American society*. Belknap Press of Harvard University Press.

Swadesh, M. (1952). Lexico-statistic dating of prehistoric ethnic contacts. *Proceedings of the American Philosophical Society*, *96*, 452–463.

Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, *21*, 121–137.

Sweller, J. (2011). Cognitive load theory. In Jose P. Mestre & Brian H. Ross (Eds.), *Psychology of learning and motivation* (Vol. *55*, pp. 37–76). Elsevier. DOI: 10.1016/B978-0-12-387691-1.00002-8

Tharmaseelan, N., Inkson, K., & Carr, S. C. (2010). Migration and career success: Testing a time-sequenced model. *Career Development International*, *15*, 218–238. DOI: 10.1108/13620431011053712

UN General Assembly. (2015). *Transforming our world: The 2030 agenda for sustainable development* (Report No. A/RES/70/1).

United Nations. (2019). *International migrant stock 2019* (United Nations database, POP/DB/MIG/Stock/Rev.2019) [Data set]. https://www.un.org/development/desa/pd/content/international-migrant-stock

van der Slik, F. (2010). Acquisition of Dutch as a second language: The explanative power of cognate and genetic linguistic distance measures for 11 West European first languages. *Studies in Second Language Acquisition*, *32*, 401–432. DOI: 10.1017/S0272263110000021

Vigren, H., Alisaari, J., Heikkola, L. M., Acquah, E. O., & Commins, N. L. (2022). Teaching immigrant students: Finnish teachers' understandings and attitudes. *Teaching and Teacher Education*, *114*, Article 103689. DOI: 10.1016/j.tate.2022.103689

Warm, T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, *54*, 427–450 https://doi.org/10.1007/BF02294627

Westergaard, M., Mitrofanova, N., Mykhaylyk, R., & Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The linguistic proximity model. *The International Journal of Bilingualism*, *21*, 666–682. DOI: 10.1177/1367006916648859

Wichmann, S., Brown, C. H., & Holman, E. W. (Eds). (2020). *The ASJP Database* [Data set]. Max Planck Institute for the Science of Human History.

Wichmann, S., Holman, E. W., Bakker, D., & Brown, C. H. (2010). Evaluating linguistic distance measures. *Physica A: Statistical Mechanics and its Applications*, *389*, 3632–3639. DOI: 10.1016/j.physa.2010.05.011

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*, 370. https://www.rasch.org/rmt/rmt83b.htm

Appendix 1

Table A1. List of languages spoken by students in our sample with their parents including reported frequency (n) and linguistic distance) to English (LDE)

Languages	Language–mother *n*	Language–father *n*	LDE
Albanian	50	52	96,56
Arabic (various)	82	84	98,58
Armenian	3	0	97,09
Bosnian	43	39	88,94
Czech	4	3	90,78
Chinese	11	8	102,30
Croatian	35	29	37,09
Danish	2	1	68,05
Dutch	19	18	61,13
Finnish	0	1	104,04
Flemish	1	2	61,13
French	26	16	91,35
German	4,145	4,027	67,33
Greek	27	23	96,84
Hebrew	1	0	96,88
Hindi	5	3	96,06
Hungarian	7	3	96,19
Italian	48	47	89,88
Japanese	3	1	98,34
Korean	2	2	98,83
Kurdish	48	46	94,01
Latvian	1	1	95,53
Macedonian	6	5	91,08
Maltese	1	0	101,56
Malayalam	0	1	100,58
Norwegian	3	0	63,58
Pashto	8	10	96,89
Persian	15	21	95,34
Polish	148	108	95,46
Portuguese	17	16	96,16
Romanian	10	10	87,74
Russian	223	199	93,10
Serbian, Montenegrin	0	2	87,39
Slovenian	28	24	89,87
Slovak	2	0	91,38
Spanish	54	37	94,14
Swedish	4	4	64,26
Tamil	26	20	100,59
Thai	2	0	99,30
Turkish	467	441	101,07
Ukrainian	4	2	96,54
Urdu	2	1	95,03
Uzbek	1	0	97,67
Vietnamese	17	16	103,10
Wolof	1	1	99,25
	5,602	5,324	

Note: Listing more than two languages per student was possible; the language listed first was used for coding the LDE variable used for main analyses.