

Nine partner languages, one path: Minority language reading proficiency development among German two-way immersion students

Sandra Preusler^{a,*}, Steffen Zitzmann^b, Jürgen Baumert^c, Jens Möller^a

^a Institute for Psychology of Learning and Instruction at Kiel University, Olshausenstraße 75, 24118, Kiel, Germany

^b Medical School Hamburg, Am Kaiserkaai 1, 20457, Hamburg, Germany

^c Max Planck Institute for Human Development, Lentzeallee 94, 14195, Berlin, Germany

ARTICLE INFO

Keywords:

Minority language
Dual language
Two-way immersion
Bilingual education
Multilingualism
Partner language

ABSTRACT

Background: At the State Europe School of Berlin (SESb) students with different language backgrounds learn together in two languages of instruction: German and one of nine partner languages (English, French, Greek, Italian, Polish, Portuguese, Russian, Spanish, and Turkish).

Aims: This study investigates the reading proficiency trajectories in the minority or partner language among students in a two-way immersion (TWI) program.

Sample: Participants were 977 TWI students.

Methods: The study included longitudinal large-scale assessments in Grades 4 to 6. Latent growth curve models were utilized for analysis.

Results: The analyses revealed significant progress in partner language reading skills across all language combinations. However, notable differences in initial proficiency levels and developmental trajectories were observed between partner languages, even after controlling for background variables such as socioeconomic status and cognitive abilities. Furthermore, the results demonstrate that students' initial German reading proficiency positively influenced their reading proficiency in the partner language.

Conclusions: Overall, the study demonstrates the effectiveness of the TWI program in promoting partner language proficiency development in TWI programs, particularly within the unique context of a diverse linguistic environment. This research provides valuable insights into how students develop reading skills in a multilingual environment.

1. Introduction

Learning more than one language seems essential in an increasingly interconnected and multilingual world. One effective model promoting bi- or multilingualism among students is two-way immersion (TWI). In these dual-language education programs, instruction is offered in the majority language and a minority language, also known as the partner language (Howard et al., 2018). The TWI classes are composed of students whose first language (L1) is either the majority or the partner language, providing competent language models at all times (Baker & Wright, 2021; Christian et al., 2000). TWI aims to maintain and develop proficiency for all language groups in their L1 as they acquire proficiency in their second language (L2), and is therefore referred to as an additive bilingual program (Howard et al., 2003).

One theoretical explanation for the success of these bilingual programs can be found in Cummins' interdependence hypothesis (1979, 1984), which is based on the assumption of a common underlying proficiency (Cummins, 1981). Cummins (1981) assumes that, like an iceberg, linguistic abilities are connected below the surface by a common linguistic basis, and that language proficiency is, therefore, reciprocally dependent. Accordingly, positive development in one language has a positive effect on skills in the other language, as revealed by Gebauer, Zaunbauer, & Möller, (2013) in a one-way immersion program. Considering the growing number of TWI programs and the need for effective bilingual education, it is crucial to examine the specificities of the programs and their impact on language development.

Research on TWI has consistently demonstrated that TWI students perform at or above the level of their mainstream peers in the majority

* Corresponding author.

E-mail addresses: spreusler@ipl.uni-kiel.de (S. Preusler), steffen.zitzmann@medicalschooll-hamburg.de (S. Zitzmann), jmpobaumert@mpib-berlin.mpg.de (J. Baumert), jmoeller@ipl.uni-kiel.de (J. Möller).

<https://doi.org/10.1016/j.learninstruc.2024.102047>

Received 8 February 2024; Received in revised form 24 October 2024; Accepted 25 October 2024

Available online 6 November 2024

0959-4752/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

language, even with random assignment to the immersion group or when SES is taken into account in the analyses (Lindholm-Leary, 2001, 2011, Marian et al., 2013; Steele et al., 2017; for a comprehensive review, e.g., on pedagogical approaches and long-term outcomes, see Baker & Wright, 2021). While there is an abundance of research available on majority language proficiency, there is far less research on partner language proficiency (Gándara, 2021), mostly conducted in the United States with Spanish as the predominant partner language in TWI (Howard et al., 2018). Burkhauser et al. (2016) noted that standardized tests in U.S. TWI programs focus primarily on English, resulting in fewer studies assessing students' proficiency in the partner language, primarily due to time and cost constraints. As Watzinger-Tharp et al. (2021) noted, there is a need to conduct research on a wider range of partner languages. In particular, there is a lack of studies focusing on partner languages other than Spanish.

The State Europe School of Berlin (SESB) in Germany has developed a distinctive 50:50 TWI Program, which offers a wide range of partner languages at its various locations. In addition to German, the other half of the classes are taught in English, French, Greek, Italian, Polish, Portuguese, Russian, Spanish, and Turkish. In this study, we investigated the trajectories of the partner language reading proficiency among SESB students from Grades 4 to 6, considering their diverse backgrounds in terms of first language(s) and partner languages in school. We explore the influence of German reading proficiency and compare proficiency levels across the nine partner languages while controlling for several individual and familial background variables, such as socioeconomic status and cognitive abilities. By examining how TWI students with diverse backgrounds develop reading proficiency in partner languages, our research can help to understand language learning and thus support the best possible outcomes for these students. Additionally, it contributes valuable insights to our understanding of language acquisition in multilingual educational contexts.

1.1. Research on TWI

Several studies conducted in the United States showed comparable or better English (as L1 or L2) reading outcomes for English-Spanish TWI programs in comparison to mainstream schools (Lindholm-Leary, 2001; Marian et al., 2013). Similar results regarding the English proficiency were reported for programs with other partner languages such as Russian, Chinese, or Mandarin (Lindholm-Leary, 2011; Padilla et al., 2013; Steele et al., 2017). Many studies focused on the English performance of minority language students (i.e., students who had not learned the majority language (English) as an L1). These students seem to benefit particularly from TWI, where instruction is delivered in both the majority language and their L1 (for a review, see Genesee & Lindholm-Leary, 2013; Krashen, 2005).

In addition to developing proficiency in the majority language, TWI students also develop proficiency in the other language of instruction: the partner language. In their study of U.S. TWI programs involving the partner languages Spanish, Mandarin, and Japanese, Burkhauser et al. (2016) found that eighth-grade students performed at higher levels of proficiency in their partner language than students in traditional schools who had taken Spanish as an elective. Nevertheless, a more favorable result was anticipated given the higher amount of language instruction in TWI. Other studies (Lindholm-Leary, 2011; Lindholm-Leary & Block, 2010; Lindholm-Leary & Hernández, 2011; Padilla et al., 2013; Watzinger-Tharp et al., 2021) found that immersion students had remarkable linguistic proficiency, often above grade levels, in partner languages such as Spanish, French, Mandarin, and Cantonese, although their proficiency has not been directly compared to mainstream students.

Several studies showed a native language effect (Howard et al., 2003): Students who acquired the test language as their L1 got higher reading scores than students with the test language as their L2. This advantage for students with the test language as L1 could also be shown

in a study that controlled for gender and SES (Howard & Neugebauer, 2015).

In summary, the few studies on partner language proficiency show that students in TWI programs achieved skills in the partner language that were above those of students in mainstream schools.

Research gaps exist in understanding language development trajectories in the context of TWI programs. Here, investigating the effects of the majority language proficiency on the development of the partner language is crucial. Additionally, there is a need for comprehensive comparisons of language proficiency levels across various partner languages while controlling for individual and familial variables. These gaps set the stage for the present study, which delves into the trajectories of reading proficiency in several partner languages among SESB students from Grades 4 to 6.

1.2. The SESB instructional program

The SESB is an example of TWI in Germany, where students who have learned one or both languages of instruction as an L1 are taught together (Howard et al., 2018). In addition to German as the majority language, the sites each offer one of nine partner languages (English, French, Greek, Italian, Polish, Portuguese, Russian, Spanish, and Turkish).

Initially launched in the early 1990s with three language combinations (German alongside English, French, or Russian), both the languages offered and the locations were gradually expanded so that there are now over 30 school locations at primary and secondary schools. These schools offer bilingual lessons from the first grade through to graduation. Instruction is delivered in German and the partner language in equal amounts (50:50). Teachers instruct in their L1. The language of instruction differs according to the subject, with some subjects (e.g., mathematics) being taught in German and others (e.g., science) in the partner language. In the fifth grade, English (or French in the German/English site) is added as a regular foreign language class.

Registration is open to all students with sufficient language skills in at least one language of instruction. Students in one class are taught together in all subjects except German and the partner language from 1st to 8th grade. In these grades, students are divided into two groups in the language subjects depending on their (dominant) L1. The TWI model at SESB not only focuses on language acquisition but also promotes a deeper understanding and appreciation of different cultures and perspectives.

Previous studies on SESB showed comparable performance of TWI and mainstream students in German reading comprehension development from fourth to sixth grade (Preusler, Zitzmann, Baumert, & Möller, 2022). Despite reduced instruction in German, TWI students reached the same reading comprehension level as traditionally taught students, showing similar learning trajectories even when accounting for background variables such as L1, socioeconomic background, and cognitive abilities. Moreover, TWI students had a higher reading proficiency in German, the majority language, than in the partner language when comparing fourth-grade performance in both languages of instruction (Preusler, Zitzmann, Paulick, Baumert, & Möller, 2019). The findings also supported the native language hypothesis: Students whose L1 matched the test language demonstrated higher reading performance than students who acquired the test language as an L2 and those who grew up simultaneously bilingual in both languages.

At German-English SESB sites, Baumert, Köller, & Lehmann, (2012) demonstrated that students in the TWI program exhibit significantly higher English proficiency than those in mainstream programs who acquired English through traditional foreign language classes, even controlling for several covariates. However, the superiority of immersion students was expected due to the significantly greater amount of English instruction that 50:50 immersion students received. Also, there was an advantage in the English proficiency of immersion students who learned English as L1 compared to those whose L1 was German.

Preusler, Fleckenstein, Zitzmann, Baumert, & Möller, (2024) confirmed the superiority of TWI students from all other SESB sites (excluding the German/English sites) exhibiting higher English (as an L3) proficiency than mainstream students despite having received less English instruction at SESB. The results showed additional effects of German and partner language reading skills on English proficiency. Notably, the development in partner language reading proficiency has not been investigated to date, especially not separately for SESB's nine partner languages.

1.3. Research questions

This study investigates the trajectories of reading proficiency in the partner languages among students at SESB from Grades 4 to 6, aiming to fill existing gaps in the literature by examining the differences in reading proficiency among the partner languages, the development of reading skills over time, and the influence of German reading proficiency on partner language development. This study, therefore, aims to investigate the following three research questions.

RQ1 Does German reading proficiency impact partner language proficiency?

Due to the possible interdependence of language proficiency (Cummins, 1979, 1981, 1984), we expect that higher levels of German reading proficiency are positively associated with higher levels of partner language reading proficiency (Hypothesis 1).

RQ2 Do students with the partner language as L1 outperform students with the partner language as L2?

We also expect a native language effect (Howard et al., 2003): Students who have acquired the partner language as their L1 should, due to the higher exposure at home, demonstrate higher proficiency levels than students acquiring the partner language as an L2 (Hypothesis 2).

RQ3 How do the trajectories of partner language reading proficiency differ among the nine partner languages?

We anticipate significant differences in reading proficiency trajectories among the nine partner languages. This expectation is rooted in the heterogeneity among students concerning L1, socioeconomic status (SES), and cognitive abilities. In particular, earlier analyses of the SESB revealed differences between the partner languages in terms of first language composition, SES and cognitive abilities (Authors, 2017a). We aim to investigate whether background variables can explain the possible differences in reading proficiency among the nine partner languages, shedding light on the complex interplay of linguistic and non-linguistic factors influencing language proficiency trajectories in a multilingual educational setting. As there is no research on the trajectories of the partner language proficiency, we analyzed the antecedents in an exploratory manner.

2. Methods

2.1. Participants

The data for this study stems from the *Europe study*, a longitudinal evaluation of the SESB program conducted from 2014 to 2017 (Authors, 2017b). Approval for the Europe study was granted by the State of Berlin, represented by the Senator for Education, Science, and Research. The initial survey included all fourth-grade SESB students in the 2014 school year, as well as additional fourth-grade students from schools with lower enrollment included in the 2015 school year, making a total of 977 participants. Attrition was observed in subsequent assessments in Grades 5 and 6, with 138 participants not continuing after the

fourth-grade assessment and an additional 39 participants not participating in the study after the fifth-grade assessment due to reasons such as leaving school or temporary absence. The notable rate of attrition observed in our study is primarily attributed to the option provided to all Berlin elementary school students to transition into the academic track of secondary schooling at an earlier stage, i.e., before the usual start at 6th grade. Detailed information regarding attrition categorized by major sociodemographic characteristics is available in the appendix.

The total sample consisted of $N = 977$ participants, with 54.4% identifying as girls. The average age of the participants in the fourth-grade assessment was $M = 10.17$ years ($SD = 0.44$). Table 1 provides a breakdown of the number of participants across the different partner languages and the L1 distribution. Table 2 presents means and standard deviations of key variables, including performance scores and socioeconomic and cognitive background. The data is further broken down by partner languages at the school site.

2.2. Data collection

The assessments took place at the end of the fourth, fifth, and sixth grades. Students in these grades were required to participate in the study per Berlin School Law, except for an additional voluntary part. The voluntary part, which depended on parental consent, included information about the family, including details such as parents' occupations and language background. To ensure transparency and informed participation, parents received a comprehensive cover letter accompanying the voluntary part and the parent questionnaire. This letter outlined the study's objectives and measures to uphold data protection and confidentiality.

The data collection was conducted by the International Association for the Evaluation of Educational Achievement (IEA) Hamburg. This included training the test administrators, collecting the data, and coding the open questions. The questionnaires and tests were carried out with the whole school class.

2.3. Measures and covariates

2.3.1. Partner language reading proficiency

The partner language reading items used in this study were drawn from the Progress in International Reading Literacy Study (PIRLS) conducted by the IEA (Mullis & Martin, 2015). This allowed us to employ tests in all of the nine partner languages that adhered to established PIRLS standards, ensuring their international recognition and comparability. The PIRLS item review process includes examination for item-by-country interactions (Foy et al., 2012). This validation process enhances confidence in the measurement invariance of the tests across different languages. Each grade level featured two booklets, each

Table 1
Sociodemographic characteristics of participants at baseline in grade 4.

Variable	<i>N</i>	%
Gender (female)	531	54.3
L1 group		
L1 German	199	20.3
Simultaneous bilinguals (2L1)	398	40.8
L1 Partner Language	380	38.9
Partner Language at school site		
English	109	11.2
French	191	19.5
Greek	91	9.3
Italian	85	8.7
Polish	89	9.1
Portuguese	80	8.2
Russian	82	8.4
Spanish	131	13.4
Turkish	119	12.2
Total	977	

Table 2
Non-adjusted means (M) and standard deviations (SD) of the variables based on the imputed data.

	4th Grade Reading in PL	5th Grade Reading in PL	6th Grade Reading in PL	4th Grade Reading in German	Parents' SES	Cognitive abilities
	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Partner Language at school site						
English	-1.09 (1.12)	-0.11 (0.99)	0.22 (1.18)	0.81 (1.05)	69.27 (17.88)	28.43 (8.86)
French	-1.37 (1.22)	-0.76 (1.18)	-0.52 (1.4)	0.9 (0.88)	68.11 (16.36)	28.7 (8.75)
Greek	-1.58 (1.04)	-0.8 (0.91)	-0.07 (1.18)	0.22 (1.17)	51.71 (23.07)	25.02 (10.75)
Italian	-1.17 (1.2)	-0.78 (1.24)	-0.3 (1.54)	0.58 (0.95)	63.54 (21.05)	26.42 (9.68)
Polish	-1.19 (1.04)	-0.65 (1.08)	-0.76 (1.38)	-0.24 (1.15)	50.67 (21.28)	20.87 (9.38)
Portuguese	-2.4 (1.12)	-1.39 (1.14)	-1.55 (1.25)	0.4 (1.07)	57.89 (21.68)	25.81 (9.66)
Russian	-1.4 (1.03)	-0.86 (1.2)	-0.67 (1.15)	0.51 (0.96)	64.29 (18.25)	29.27 (8.17)
Spanish	-1.66 (1.03)	-0.78 (1.08)	-0.27 (1.26)	0.78 (1.03)	65.4 (19.01)	29.14 (9.08)
Turkish	-2.42 (1.07)	-1.58 (1)	-1 (1.08)	0.21 (1.18)	48.49 (20.25)	22.67 (9.43)
Total	-1.58 (1.19)	-0.85 (1.16)	-0.52 (1.36)	0.52 (1.09)	60.82 (20.96)	26.55 (9.65)

Note: PL = Partner Language. Cognitive abilities were measured in fourth grade (scoring range from 0 to 45).

containing a story (around 500–800 words) accompanied by 12–17 questions in open- and closed-ended response formats (for sample texts, questions and scoring guides, see Mullis et al., 2009). Using the PIRLS scoring guidelines, the open response-items were scored by the IEA Hamburg. The English, French, Italian, Polish, Portuguese, Russian, and Spanish booklets originated from PIRLS 2011. The Greek and Turkish booklets were taken from the 2001 PIRLS cycle, with one booklet corresponding to one used in Grades 4 and 5 for the other languages.

Since PIRLS booklets are primarily intended for fourth-grade students, we selected more challenging booklets for Grade 6. In Grade 5, we incorporated one booklet each from both Grade 4 and Grade 6, allowing us to include anchor items. This selection of reading booklets enabled us to explore reading proficiency development across different partner languages while facilitating comparisons between these languages.

The scaling and linking procedures were performed in R using the Test Analysis Modules (TAM) package (R Core Team, 2023; Robitzsch et al., 2021), applying Item Response Theory (IRT) with the 1-parameter logistic (1 PL) model. To assess measurement invariance across the three time points, a Differential Item Functioning (DIF) analysis was conducted using the Root Mean Square Deviation (RMSD) statistic, with a threshold of $RMSD \leq 0.12$, as used in the Programme for International Student Assessment (PISA) framework for acceptable item fit (OECD, 2024). In this analysis, only 2 out of 82 anchor items (2.44%) exceeded this threshold, indicating that the items are largely measurement invariant and that approximate measurement invariance holds. This suggests that the items relatively consistently measured the same construct across all three time points.

The fourth-grade data were scaled using item difficulties derived from the PIRLS datasets of countries where the partner language is a lingua franca. Specifically, we utilized data from England, France, Italy, Poland, Portugal, Russia, and Spain from the 2011 PIRLS dataset and data from Greece and Turkey from the 2001 PIRLS dataset.

In Grades 5 and 6, item difficulties were estimated freely. The linking of the three measurement points was accomplished using the Stocking-Lord method, employing the joint method within the TAM package (Kolen & Brennan, 2014; Robitzsch et al., 2021). Subsequently, Weighted Likelihood Estimates (WLEs) were calculated. Due to the scaling of the data using the PIRLS item difficulties and the subsequent linking, the mean WLE values of the sample were in the negative range between $M = -1.68$ (Grade 4) and $M = -0.61$ (Grade 6).

Reliabilities for the fourth, fifth, and sixth grade tests were satisfactory to good with $r_{WLE} = 0.86$, $r_{WLE} = 0.87$, and $r_{WLE} = 0.90$ for English, French, Italian, Polish, Portuguese, Russian, and Spanish, and $r_{WLE} = 0.78$, $r_{WLE} = 0.77$, and $r_{WLE} = 0.86$ for Greek and Turkish.

2.3.2. German reading proficiency

The German reading test originated from PIRLS in 2001 (Mullis et al., 2003). The measure comprised 13 items and achieved adequate reliability ($r_{WLE} = 0.76$).

2.3.3. First Language(s)

The participants' L1 was assessed through questionnaires administered to both parents and students. Parents were asked about the language(s) their child first acquired within the family, while students were asked about the language(s) they had learned from the beginning. In cases where parental responses were missing, student responses were used as supplementary information. These responses resulted in the identification of three L1 groups: L1 German speakers, L1 partner language speakers, and simultaneous bilingual children with German and a partner language as their first languages (2L1). Refer to Table 1 for a presentation of the distribution of these first language groups.

2.3.4. Background variables

To ensure comparability among students, we collected background variables, including socioeconomic status (SES), cognitive abilities, and gender. The mean values of the sample are presented in Table 2, and information on gender distribution is provided in Table 1.

The SES was measured using the International Socioeconomic Index of Occupational Status (ISEI; Ganzeboom & Treiman, 1996). ISEI scores were determined based on information provided by parents about their occupations, work-related activities, authority, and educational background. The ISEI has a possible range of 16–90, with higher values indicating a higher occupational status. In cases where parents had different ISEI scores, the higher score was used (HISEI = highest ISEI).

Cognitive abilities were evaluated using the 4th-grade Test of Cognitive Abilities (KFT; Heller & Perleth, 2000). The test consisted of a figural and verbal subtest, comprised a total of 45 items and had very good internal consistency (Cronbach's $\alpha = .91$).

Children's gender was obtained from school records and coded as 0 = male and 1 = female.

2.4. Missing data

Most of the variables used in our analyses had a low percentage of missing data, typically less than 10%. The only exception was the SES variable, which had a notably higher percentage of missing data, at 31.3%. In order to avoid bias and to obtain the largest possible sample, multiple imputation was used to address missing data. This approach utilizes available background variables to estimate complete data sets (Lüdtke et al., 2007; Schafer & Graham, 2002). In this study, information on school location and class, student age, L1, immigrant background, parental SES, cognitive abilities, grades, and performance on other test dimensions were used for the imputation.

In order to account for the multilevel structure of the data, the school ID was utilized as a clustering variable for predicting the reading test variables using the imputation method '2l.pan' from the 'pan' package as part of the imputation process with the 'mice' package (van Buuren & Groothuis-Oudshoorn, 2011; Zhao & Schafer, 2023). To ensure that an adequate number of imputed datasets were generated for the analyses,

the R package 'howManyImputations' was employed, which is based on an estimate of the fraction of missing information (FMI; von Hippel, 2018). As a result, 65 imputed datasets were generated, and analyses were run on each of these datasets. Subsequently, the results from these analyses were pooled using Rubin's (1987) rules.

2.5. Statistical analyses

We utilized latent growth curve models (LGCMS) to analyze the development of students' partner language reading proficiency over time. LGCMS involve two latent variables: the intercept and the slope. The intercept in the model represents the initial starting point or baseline level, while the slope captures the rate of change or growth over time (Bollen & Curran, 2006). In this study, we assumed a linear growth pattern. The factor loadings at each of the three measurement points were assigned values of zero for the initial measurement (T1), one for the second measurement (T2), and two for the final measurement (T3), reflecting the consistent annual intervals between measurements. The analyses were conducted using Mplus 8.4 (Muthén & Muthén, 1998) with the aid of the R package 'MplusAutomation' (Hallquist & Wiley, 2018). Considering the hierarchical structure of the data, with students nested within classes, the robust maximum likelihood estimator in Mplus was utilized to obtain robust standard errors.

The analyses were conducted using a stepwise construction of the LGCMS by sequentially incorporating predictor variables. First (Models A1-A4), the analyses were conducted across partner languages, and subsequently (Models B1-B3), the nine school partner languages were analyzed comparatively. Model A1 served as a baseline model without predictors, allowing for a general assessment of the changes in students' partner language reading proficiency over time. Model A2 explored the influence of German reading proficiency in the 4th grade. In Model A3, the influence of L1 was examined by including two dummy variables representing students' language group membership: one for L1 German speakers and the other one for simultaneous bilinguals. The group of students with the partner language as their L1 served as the reference group. Finally, Model A4 incorporated additional background variables, including socioeconomic status (SES), cognitive abilities, and gender, to consider potential confounding factors in the analysis of the development of students' partner language skills (see Fig. 1 for a graphical representation of Model A4).

In order to analyze the differences among the nine partner languages, eight dummy variables were created for French, Greek, Italian, Polish, Portuguese, Russian, Spanish, and Turkish. English was selected as the reference group because, based on the descriptive data (see Table 2), this group demonstrated the highest performance compared to the other languages at the third measurement point in Grade 6. Additionally, English has a special significance in Germany as the primary foreign language in most schools. It is also commonly used as a partner language in other bilingual school contexts. Using dummy variables allowed us to compare the multiple partner languages while controlling for potential confounding variables. To examine the differences among the nine partner languages, Model B1 was constructed without any additional predictor variables. In the last two models, additional predictors are included in the comparative analyses of students' partner language skill development between partner languages. Model B2 includes German language proficiency as a predictor, while Model B3 includes both German language proficiency and additional background variables (SES, gender, cognitive abilities).

Model fit was evaluated using four fit indices: Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) values above 0.90 indicate an acceptable fit of the model to the data, while values above 0.95 are considered a very good fit, as suggested by Marsh et al. (2004). The RMSEA and SRMR values should not exceed 0.08 for an acceptable fit (Hu & Bentler, 1999; Schreiber et al., 2006). As

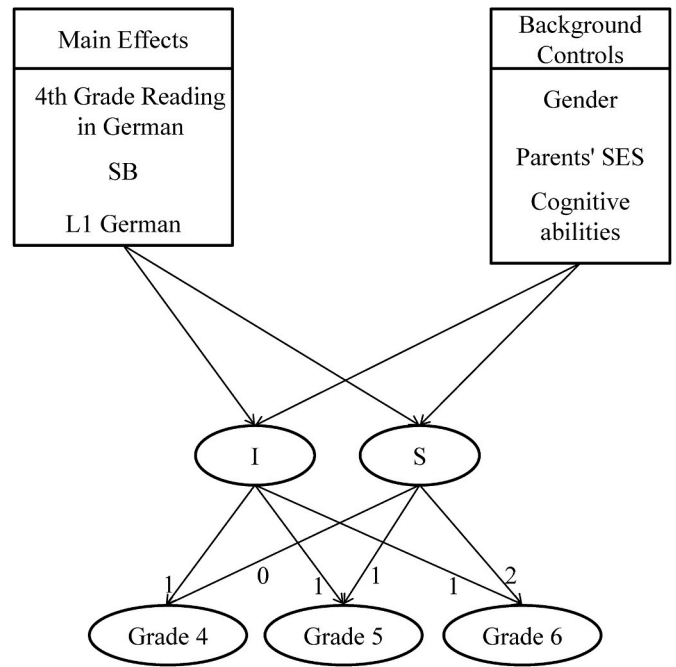


Fig. 1. Graphical representation of Model A4
 Note. (Residual) correlations are not shown to improve readability. SB = simultaneous bilinguals; Students with the partner language as L1 served as the reference group.

the data were multiply imputed, the mean values of the fit indices were utilized. No chi-square tests were conducted, as there is no theoretical basis for calculating chi-square tests with imputed data.

3. Results

The present study presents the results of two analyses examining language proficiency development. The first analysis investigates the influence of German reading proficiency and potential confounders on partner language proficiency development. Subsequently, the second analysis explores differences in reading proficiency development across the various partner languages.

3.1. Influence of German reading proficiency and confounders on partner language proficiency development

The model fit statistics for Models A0 to A4 are presented in Table 3. The initial model, Model A0, served as a baseline model that assumes no growth. As expected, this model showed a relatively poor fit to the data. Models A1 to A4 demonstrated a generally acceptable fit to the data.

Table 3
 Fit statistics for models A1-A4 on partner language reading proficiency development.

	df	CFI	TLI	RMSEA	SRMR
Model A0 (intercept-only)	5.00	0.81	0.86	0.27	0.34
Model A1	8.00	0.98	0.95	0.16	0.05
Model A2	10.00	0.97	0.92	0.14	0.04
Model A3	14.00	0.96	0.90	0.11	0.03
Model A4	20.00	0.97	0.90	0.08	0.03

Note. A1: Baseline (no predictors); A2: Included German proficiency (Grade 4); A3: Added language group (L1 German, simultaneous bilinguals; L1 partner language as reference); A4: Adds background variables (SES, cognitive abilities, gender). CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual.

Table 4
Latent growth curve models for partner language reading proficiency.

	Model A1			Model A2				Model A3				Model A4			
	Par.	SE	95% CI	Par.	SE	95% CI	d	Par.	SE	95% CI	d	Par.	SE	95% CI	d
Means															
Intercept	-1.49 ^c	0.09	[-1.66, -1.32]	-1.49 ^c	0.08	[-1.65, -1.34]	-	-1.20 ^c	0.10	[-1.39, -1.01]	-	-1.24 ^c	0.10	[-1.43, -1.04]	-
Slope	0.53 ^c	0.04	[0.45, 0.61]	0.53 ^c	0.04	[0.45, 0.61]	-	0.51 ^c	0.05	[0.40, 0.61]	-	0.40 ^c	0.06	[0.29, 0.51]	-
Residual Variances															
Intercept	0.92 ^c	0.10	[0.72, 1.11]	0.83 ^c	0.09	[0.65, 1.02]	-	0.76 ^c	0.09	[0.59, 0.94]	-	0.71 ^c	0.08	[0.55, 0.87]	-
Slope	0.06	0.05	[-0.04, 0.15]	0.06	0.05	[-0.03, 0.16]	-	0.07	0.05	[-0.02, 0.16]	-	0.06	0.04	[-0.03, 0.15]	-
Covariances															
Intercept x Slope	0.03	0.06	[-0.08, 0.14]	0.01	0.05	[-0.10, 0.11]	-	0.01	0.05	[-0.09, 0.11]	-	0.00	0.05	[-0.10, 0.10]	-
Effects															
4th Grade Reading in German^a															
Intercept	-	-	-	0.30 ^c	0.06	[0.19, 0.41]	0.31 ^c	0.38 ^c	0.06	[0.27, 0.49]	0.39 ^c	0.26 ^c	0.06	[0.15, 0.38]	0.27 ^c
Slope	-	-	-	0.06 ^b	0.03	[0.01, 0.11]	0.24	0.05	0.03	[-0.00, 0.10]	0.20	0.03	0.03	[-0.02, 0.09]	0.12
SB															
Intercept	-	-	-	-	-	-	-	-0.33 ^c	0.08	[-0.50, -0.17]	-0.36 ^c	-0.39 ^c	0.08	[-0.55, -0.23]	-0.42 ^c
Slope	-	-	-	-	-	-	-	0.02	0.06	[-0.08, 0.13]	0.12	0.02	0.05	[-0.08, 0.13]	0.10
L1GS															
Intercept	-	-	-	-	-	-	-	-0.78 ^c	0.12	[-1.03, -0.54]	-0.85 ^c	-0.84 ^c	0.12	[-1.08, -0.60]	-0.91 ^c
Slope	-	-	-	-	-	-	-	0.07	0.07	[-0.07, 0.21]	0.32	0.07	0.07	[-0.06, 0.21]	0.35
Gender (1 = female)															
Intercept	-	-	-	-	-	-	-	-	-	-	-	0.13	0.07	[-0.01, 0.27]	0.14
Slope	-	-	-	-	-	-	-	-	-	-	-	0.19 ^c	0.04	[0.11, 0.28]	0.92 ^c
Parents' SES^a															
Intercept	-	-	-	-	-	-	-	-	-	-	-	0.19 ^c	0.05	[0.10, 0.29]	0.20 ^c
Slope	-	-	-	-	-	-	-	-	-	-	-	-0.00	0.02	[-0.05, 0.04]	-0.01
Cognitive abilities^a															
Intercept	-	-	-	-	-	-	-	-	-	-	-	0.11 ^b	0.05	[0.01, 0.20]	0.11 ^b
Slope	-	-	-	-	-	-	-	-	-	-	-	0.02	0.03	[-0.04, 0.08]	0.08

Note. Unstandardized solution. Standard errors are in parentheses. CI = confidence interval; SB = simultaneous bilinguals; L1GS = L1 German speakers. Students with the partner language as L1 served as the reference group in Models A3 and A4. Cognitive abilities were measured in fourth grade. ***p* < .01.

^a standardized.

^b *p* < .05.

^c *p* < .001.

Table 4 shows the results of the LGCMs investigating the influence of German reading proficiency on the development of reading proficiency in the partner language. In Model A1, the positive slope factor ($\mu_{\text{slope}} = 0.53, p < .001$), significantly different from zero, indicated an overall positive growth in partner language proficiency over time.

Model A1 also showed a non-significant relationship between the initial starting point (intercept) and the developmental changes (slope) in partner language reading proficiency from Grades 4 to 6. This suggests that the proficiency level in Grade 4 did not significantly predict the subsequent growth rate in partner language reading proficiency, indicating that the growth curves may run parallel. Specifically, it suggests that, although starting at different levels in Grade 4 the growth curve is similar for initially strong readers as it is for weaker ones.

In Model A2, the German reading proficiency in Grade 4 was included as a predictor. The results showed a significant positive association between German reading proficiency and both the intercept and the slope in partner language reading comprehension. In support of Hypothesis 1, higher proficiency in German reading at Grade 4 was

associated with higher initial levels of partner language reading comprehension and more favorable developmental changes between Grades 4 and 6.

Model A3 examined the influence of students' linguistic background, specifically their L1. The results showed that students with the partner language as their L1 performed significantly better on the partner language reading test in fourth grade than students with German as their L1, thereby supporting the native language hypothesis (Hypothesis 2). Children with the partner language as their L1 were also superior to those who grew up speaking both languages simultaneously. No differences by L1 group were evident in the developments between the 4th and 6th grades. Fig. 2 shows the predicted trajectories for the three L1 groups. Model A4 extended the previous model by including the background variables gender, parental SES, and cognitive abilities. The findings of the previous models persisted: There was a significant positive association between German reading comprehension and partner language performance in Grade 4, even after controlling for relevant background variables. The advantages of children with the partner

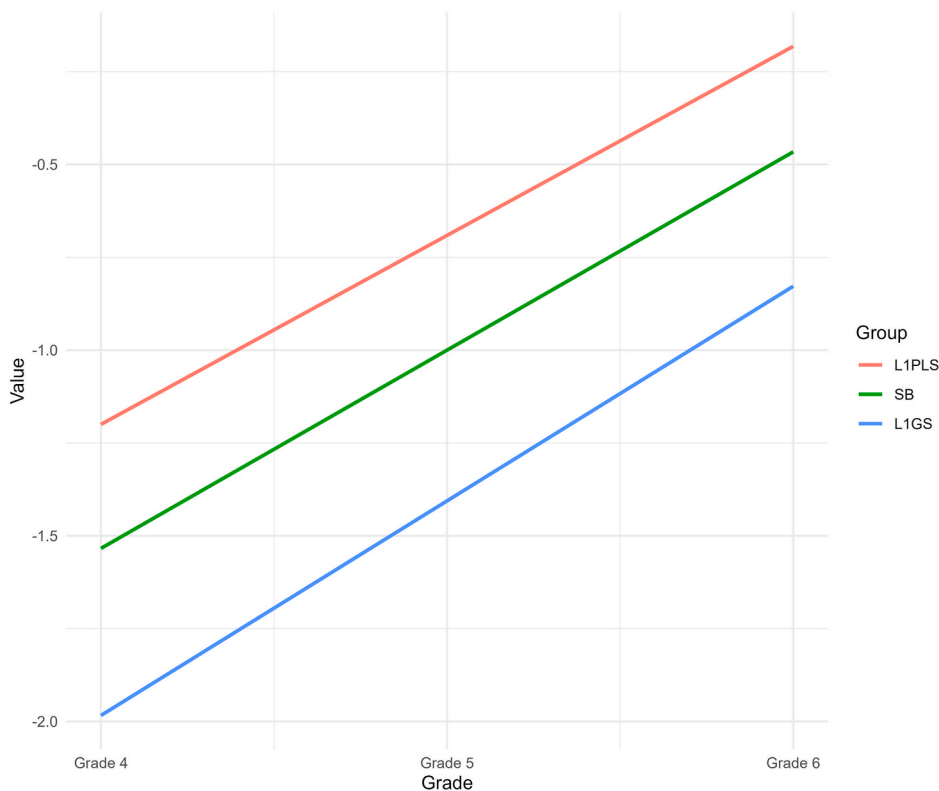


Fig. 2. Predicted trajectories for the three L1 Groups from Model A3
 Note. L1PLS = L1 Partner language speakers; SB = simultaneous bilinguals; L1GS = L1 German speakers.

language as L1 over the other two L1 groups also remained. Regarding the background variables, an influence of SES and cognitive ability on the partner’s language performance in K4 was shown. In addition, there was a significant slope effect of gender. This suggests a difference in development between girls and boys in partner language reading between Grades 4 and 6. In particular, girls showed faster growth in their partner language reading proficiency.

3.2. Differences in reading proficiency development across partner languages

The models examining the differences in reading proficiency development across the various partner languages showed an acceptable to good fit to the data, as indicated by the model fit statistics reported in Table 5. The results of the LGCMS comparing the partner languages are presented in Table 6. The reference group consisted of students in schools with English as a partner language. English was chosen as the reference group due to the highest performance compared to other languages at the third measurement point in Grade 6 and its significant role in German schools. Model B1 compared the partner languages without taking additional predictors into account (see Fig. 3 for the predicted trajectories for the nine partner languages). The results

Table 5
 Fit statistics for models B1-B3 comparing nine partner language groups.

	df	CFI	TLI	RMSEA	SRMR
Model B1	24.00	0.95	0.86	0.07	0.03
Model B2	26.00	0.96	0.87	0.07	0.03
Model B3	32.00	0.96	0.87	0.06	0.03

Note. B1: Baseline comparison of partner languages; B2: Added German proficiency; B3: Added background variables (SES, cognitive abilities, gender). CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root-mean-square error of approximation; SRMR = standardized root-mean-square residual.

showed that students from sites with Greek, Portuguese, Spanish, or Turkish as the partner language had lower reading proficiency levels in Grade 4 than students from the English sites. Conversely, the performance of students from French, Italian, Polish, or Russian sites was comparable to that of the English students. Furthermore Polish, Portuguese, and Russian students showed a less favorable developmental trajectory than the English students, while the trajectories of students in the other languages ran parallel to those of the English sites.

Subsequent models examined whether these differences between the partner languages could be explained by German reading achievement in the fourth grade (Model B2), gender, SES, and cognitive abilities (Model B3). Higher German reading achievement was related to higher partner language achievements. Girls had higher partner language achievement in Grade 4 and more favorable developmental trajectories. However, the differences between the partner languages remained mostly stable when the confounding predictors were included.

4. Discussion

In this study, we investigated the development of reading proficiency in the partner languages of students at the SESB across Grades 4 to 6. The study included analyses of LGCMS to describe the dynamics of growth in language proficiency and to examine associations with L1 affiliation and German reading proficiency as well as background variables. These analyses were divided into two sections: the impact of German reading proficiency on partner language development and differences in reading proficiency development across partner languages.

In line with Hypothesis 1, assuming an interdependence of language skills, a link between reading proficiency in German and the partner language was found in Grade 4. Thus, students’ proficiency in one language may positively impact their proficiency in the other language, which is consistent with previous research in other bilingual contexts (e. g., Gebauer, Zaunbauer, & Möller, 2013; Fleckenstein, Möller, & Baumert, 2018; Feinauer et al., 2013). However, no influence of Grade 4

Table 6
Latent growth curve models across partner languages.

	Model B1				Model B2				Model B3			
	Par.	SE	95% CI	d	Par.	SE	95% CI	d	Par.	SE	95% CI	d
Means												
Intercept	-0.95 ^d	0.18	[-1.31, -0.60]	-	-1.04 ^d	0.17	[-1.38, -0.69]	-	-1.16 ^d	0.16	[-1.48, -0.84]	-
Slope	0.66 ^d	0.08	[0.50, 0.82]	-	0.64 ^d	0.08	[0.48, 0.81]	-	0.55 ^d	0.09	[0.37, 0.72]	-
Residual Variances												
Intercept	0.73 ^d	0.08	[0.57, 0.89]	-	0.66 ^d	0.08	[0.50, 0.81]	-	0.63 ^d	0.08	[0.48, 0.79]	-
Slope	0.04	0.04	[-0.04, 0.13]	-	0.05	0.04	[-0.03, 0.13]	-	0.04	0.04	[-0.04, 0.13]	-
Covariances												
Intercept x Slope	0.05	0.05	[-0.04, 0.14]	-	0.03	0.04	[-0.06, 0.12]	-	0.02	0.04	[-0.07, 0.11]	-
Effects												
French												
Intercept	-0.34	0.23	[-0.78, 0.11]	-0.36	-0.36	0.22	[-0.79, 0.07]	-0.39	-0.36	0.21	[-0.78, 0.05]	-0.39
Slope	-0.23	0.13	[-0.49, 0.02]	-1.10	-0.24	0.13	[-0.49, 0.02]	-1.12	-0.25	0.13	[-0.51, 0.01]	-1.18
Greek												
Intercept	-0.61 ^c	0.21	[-1.03, -0.19]	-0.66 ^c	-0.45 ^b	0.20	[-0.84, -0.06]	-0.49 ^b	-0.39 ^b	0.19	[-0.76, -0.01]	-0.42 ^b
Slope	0.09	0.09	[-0.08, 0.27]	0.44	0.12	0.09	[-0.06, 0.31]	0.58	0.12	0.10	[-0.08, 0.31]	0.54
Italian												
Intercept	-0.23	0.33	[-0.89, 0.42]	-0.25	-0.17	0.28	[-0.73, 0.39]	-0.18	-0.12	0.27	[-0.66, 0.42]	-0.13
Slope	-0.23	0.12	[-0.47, 0.02]	-1.07	-0.21	0.12	[-0.46, 0.03]	-1.01	-0.19	0.13	[-0.44, 0.06]	-0.91
Polish												
Intercept	-0.10	0.22	[-0.53, 0.32]	-0.11	0.17	0.21	[-0.24, 0.59]	0.19	0.27	0.20	[-0.12, 0.66]	0.29
Slope	-0.44 ^d	0.10	[-0.64, -0.24]	-2.07 ^d	-0.39 ^d	0.10	[-0.59, -0.18]	-1.83 ^d	-0.38 ^c	0.11	[-0.60, -0.16]	-1.80 ^c
Portuguese												
Intercept	-1.20 ^d	0.19	[-1.56, -0.83]	-1.29 ^d	-1.09 ^d	0.19	[-1.46, -0.73]	-1.18 ^d	-1.04 ^d	0.18	[-1.39, -0.70]	-1.13 ^d
Slope	-0.23 ^b	0.10	[-0.43, -0.03]	-1.07 ^b	-0.21 ^b	0.10	[-0.41, -0.01]	-0.98 ^b	-0.21 ^b	0.10	[-0.41, -0.01]	-1.00 ^b
Russian												
Intercept	-0.37	0.22	[-0.80, 0.06]	-0.40	-0.29	0.21	[-0.71, 0.12]	-0.31	-0.31	0.20	[-0.71, 0.09]	-0.34
Slope	-0.29 ^c	0.10	[-0.49, -0.10]	-1.39 ^c	-0.28 ^c	0.10	[-0.47, -0.09]	-1.32 ^c	-0.30 ^c	0.10	[-0.50, -0.10]	-1.42 ^c
Spanish												
Intercept	-0.62 ^c	0.19	[-1.01, -0.24]	-0.68 ^c	-0.62 ^c	0.19	[-0.99, -0.25]	-0.66 ^c	-0.62 ^c	0.18	[-0.98, -0.27]	-0.67 ^c
Slope	0.03	0.10	[-0.16, 0.23]	0.16	0.04	0.10	[-0.16, 0.23]	0.17	0.02	0.10	[-0.19, 0.22]	0.08
Turkish												
Intercept	-1.41 ^d	0.24	[-1.89, -0.94]	-1.53 ^d	-1.25 ^d	0.23	[-1.70, -0.80]	-1.35 ^d	-1.14 ^d	0.22	[-1.57, -0.71]	-1.23 ^d
Slope	0.05	0.13	[-0.20, 0.30]	0.25	0.08	0.13	[-0.17, 0.33]	0.38	0.08	0.13	[-0.18, 0.34]	0.40
4th Grade Reading in German^a												
Intercept	-	-	-	-	0.29 ^d	0.04	[0.21, 0.38]	0.30 ^d	0.21 ^d	0.05	[0.12, 0.31]	0.22 ^d
Slope	-	-	-	-	0.05 ^b	0.02	[0.01, 0.10]	0.20	0.03	0.03	[-0.02, 0.08]	0.10
Gender (1 = female)												
Intercept	-	-	-	-	-	-	-	-	0.17 ^b	0.07	[0.02, 0.32]	0.18 ^b
Slope	-	-	-	-	-	-	-	-	0.19 ^d	0.04	[0.11, 0.27]	0.88 ^d
Parents' SES^a												
Intercept	-	-	-	-	-	-	-	-	0.11 ^b	0.05	[0.02, 0.20]	0.11 ^b
Slope	-	-	-	-	-	-	-	-	0.01	0.03	[-0.04, 0.07]	0.05
Cognitive abilities^a												
Intercept	-	-	-	-	-	-	-	-	0.10 ^b	0.04	[0.01, 0.19]	0.10 ^b
Slope	-	-	-	-	-	-	-	-	0.02	0.03	[-0.03, 0.08]	0.08

Note. Unstandardized solution. Standard errors are in parentheses. CI = confidence interval; Students with English as partner language served as the reference group. Cognitive abilities were measured in fourth grade.

^a standardized.

^b $p < .05$.

^c $p < .01$.

^d $p < .001$.

German reading on the development of reading comprehension in the partner language could be demonstrated between Grades 4 and 6, controlling for relevant background variables. The development of partner language reading skills is therefore parallel, with no evidence of scissors or compensation effects depending on Grade 4 German proficiency. Notably, this study did not analyze the growth in German proficiency alongside the proficiency in the partner language. Future studies should focus more closely on the reciprocal influences in the development of proficiency in both languages.

Furthermore, the results reconfirmed the native language hypothesis (Preusler, Zitzmann, Paulick, Baumert, & Möller, 2019; Lindholm-Leary & Howard, 2008): Students who have the partner language as their L1 and those who grew up bilingual outperformed those whose L1 is German, as expected by Hypothesis 2. No significant developmental differences were found between the language groups. As a result, the differences between language groups persisted until the end of primary school in 6th grade. The findings, therefore, suggest that L2 speakers

should receive more support at SESB to achieve language skills at the same level as L1 speakers.

The results also showed that the students progressed significantly in their partner language reading skills from fourth to sixth grade. These gains were observed in all language combinations offered at SESB. However, we found some differences in partner language reading proficiency between language combinations in the fourth grade and in developmental trajectories that persisted even when controlling for background variables such as socioeconomic status or cognitive abilities. The differences in SES between partner languages are reflected in initial performance in Grade 4 but not in the developmental trajectories, indicating parallel development regardless of SES. Thus, the differences cannot be fully explained by different socioeconomic or cognitive prerequisites among the students. This could indicate differences in the teaching context at the various locations, so that further investigations appear essential here. In fourth grade, the results also showed a link between the language proficiencies, indicating that higher German

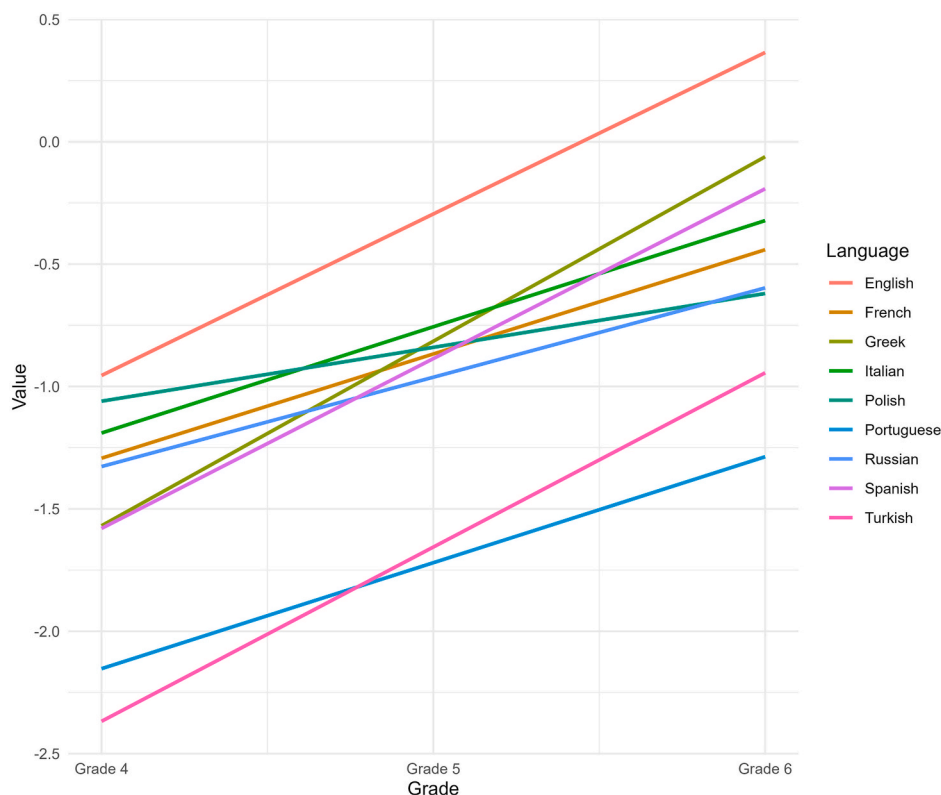


Fig. 3. Predicted trajectories for nine partner languages from Model B1.

reading performance was associated with higher performance in the partner language. This suggests a possible cross-linguistic influence, in which proficiency in one language has a positive effect on proficiency in another language. The gender-specific differences are also noteworthy, with girls exhibiting higher partner language achievement in fourth grade and following a more favorable developmental trajectory.

Given the exploratory nature of this study, further research is needed to identify potential factors and their contribution to reading comprehension in the different partner languages. On the student side, this could be due to different educational aspirations or different levels of language skills among parents. For students with an immigrant background, differences in the planned length of stay in Germany could also influence language development. On the teachers' side, there could be a variety of countries of origin and educational backgrounds of the partner language teachers or a suspected wide range of teaching materials and methods. Even if all school locations are based on the same school framework with the fixed allocation of languages, different approaches are possible on the school side, e.g., regarding linguistic diversity in the all-day school program. Thus, the instructional context in the language classroom can vary, which in other studies showed an influence on the connection between the L1 and the L2 (Kim et al., 2024). In summary, while our study identifies significant progress in partner language reading skills and highlights the influence of German reading proficiency, further research is necessary to understand the various factors that contribute to differences in language proficiency development. These factors include individual student characteristics, parental involvement, and the educational context provided by teachers and schools.

4.1. Strengths and limitations

The study offers some new insights into the area of partner languages and partner language development in TWI, revealing a positive link between German and partner language proficiency in Grade 4. However,

the development of the partner language proficiency between Grades 4 and 6 was unrelated to the German reading proficiency in Grade 4. Notably, differences in reading proficiency in the partner language between the language combinations were evident in fourth grade despite controlling for SES and cognitive abilities and persisted until the end of sixth grade. The strengths of the study lie mainly in its investigation of nine different partner languages and the strategic utilization of PIRLS reading items. This approach not only provides insights into diverse linguistic contexts but also allows comparisons between the proficiency levels in different languages. Moreover, the longitudinal design of the study enables the exploration of language proficiency trajectories over time.

However, there are also some limitations. First, this study only assessed receptive language skills in the partner language, while active oral and written language production was not assessed. Future research could attempt to capture both receptive and productive language skills to provide a more comprehensive understanding of language proficiency in TWI. Second, it should also be noted that the results of the study were obtained in the specific context of the SESB and, therefore, may only be partially transferable to other TWI programs. Third, the study did not include a mainstream comparison group, as it is not possible to find a sufficient number of pupils with adequate proficiency in all partner languages in the mainstream schools. Fourth, incorporating data on students' linguistic backgrounds (i.e., their L1) into the analyses of the various partner languages was not feasible, as the remaining samples within the different partner languages were too small. Fifth, the study did not assess the professional competencies or the instructional behavior of the partner language teachers, and potential differences in these aspects may contribute to the observed variations in student outcomes. Also, this study focused on individual-level predictors of language proficiency development. Future research could explore the potential impact of class-level factors.

4.2. Educational implications

The results of the study carry considerable educational implications. First, ensuring that all students develop substantial proficiency in the minority language is critical to overall academic success. School administrators and teachers need to be informed about the differences in performance in the various partner languages to provide additional support when needed. In addition, identifying instructional factors that contribute to success in specific partner languages is critical to transfer effective strategies to underperforming languages. The impact of the majority language on students' language development should be carefully considered, as it can influence the acquisition of partner languages. In this context, attention should be given to the native language effect, which indicates more support for L2 learners. The findings should improve language teaching practice, enable teachers to adapt their approaches, and better support students on their language learning journey.

4.3. Conclusion

Overall, the results of this study demonstrate that the TWI program used at SESB, with the distinctive feature of multiple partner languages, effectively promotes partner language proficiency among students. The study thus helps towards a better understanding of multilingualism in TWI. It provides insights into how students develop reading skills in a multilingual environment. Understanding these trajectories can inform language education practices and help educators tailor their approaches to better support students' language learning journeys.

CRedit authorship contribution statement

Sandra Preusler: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Steffen Zitzmann:** Writing – review & editing, Methodology. **Jürgen Baumert:** Writing – review & editing, Project administration, Funding acquisition. **Jens Möller:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT 3.5 in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

We have no conflicts of interest to disclose. This study is a part of the "Europe study", funded by the Senate of Berlin and the Stiftung Mercator.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.learninstruc.2024.102047>.

References

- Baumert, J., Hohenstein, F., Fleckenstein, J., & Möller, J. (2017a). Wer besucht die Staatliche Europa-Schule Berlin? Sprachlicher, ethnischer und sozioökonomischer Hintergrund sowie kognitive Grundfähigkeiten der Schülerinnen und Schüler [Who attends the State Europe School Berlin? Linguistic, ethnic and socioeconomic background as well as basic cognitive abilities of the pupils]. In J. Möller, F. Hohenstein, J. Fleckenstein, O. Köller, & J. Baumert (Eds.), *Erfolgreich integrieren - die Staatliche Europa-Schule Berlin* (pp. 75–94). Waxmann.
- Möller, J., Hohenstein, F., Fleckenstein, J., Köller, O., & Baumert, J. (Eds.). (2017b). *Erfolgreich integrieren - die Staatliche Europa-Schule Berlin* [Successful Integration – the State Europe School Berlin]. Waxmann.
- Baker, C., & Wright, W. E. (2021). Foundations of Bilingual Education and Bilingualism (7th ed.), 127. *Bilingual education & bilingualism*. Multilingual Matters <https://doi.org/10.21832/BAKER9899>.
- Baumert, J., Köller, O., & Lehmann, R. (2012). Leseverständnis im Englischen und Deutschen und Mathematikleistungen bilingual unterrichteter Schülerinnen und Schüler am Ende der Grundschulzeit. Ergebnisse eines Zwei-Wege-Immersionsprogramms [Reading comprehension in English and German and mathematics performance of bilingually taught students at the end of primary school. Results of a two-way immersion program]. *Unterrichtswissenschaft*, 40(4), 290–314.
- Bollen, K. A., & Curran, P. J. (2006). Latent curve models: A structural equation perspective. *Wiley series in probability and statistics*. Wiley-Interscience. <http://site.ebriary.com/lib/alltitles/docDetail.action?docID=10300097>.
- Burkhauser, S., Steele, J. L., Li, J., Slater, R. O., Bacon, M., & Miller, T. (2016). Partner-Language learning trajectories in dual-language immersion: Evidence from an urban district. *Foreign Language Annals*, 49(3), 415–433. <https://doi.org/10.1111/flan.12218>
- Christian, D., Howard, E. R., & Loeb, M. I. (2000). Bilingualism for all: Two-way immersion education in the United States. *Theory and Practice*, 39(4), 258–266. https://doi.org/10.1207/s15430421tip3904_9
- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222–251. <https://doi.org/10.3102/00346543049002222>
- Cummins, J. (1981). The role of primary language development in promoting educational success for Language Minority Students. In *California state department of education (ed.), schooling and language minority students: A theoretical framework* (pp. 3–49). California State University. <https://doi.org/10.13140/2.1.1334.9449>. Evaluation, Dissemination, and Assessment Center.
- Cummins, J. (1984). Bilingualism and special education: Issues in assessment and pedagogy. *Multilingual Matters*.
- Feinauer, E., Hall-Kenyon, K. M., & Davison, K. C. (2013). Cross-Language transfer of early literacy skills: An examination of young learners in a two-way bilingual immersion elementary school. *Reading Psychology*, 34(5), 436–460. <https://doi.org/10.1080/02702711.2012.658142>
- Fleckenstein, J., Möller, J., & Baumert, J. (2018). Mehrsprachigkeit als Ressource: Kompetenzen dual-immersiv unterrichteter Schülerinnen und Schüler in der Drittsprache Englisch [Multilingualism as a resource: dual-immersion students' competencies in English as a third language]. *Zeitschrift für Erziehungswissenschaft*, 21(1), 97–120. <https://doi.org/10.1007/s11618-017-0792-9>.
- Foy, P., Martin, M. O., Mullis, I. V. S., & Stanco, G. (2012). Reviewing the TIMSS and PIRLS 2011 achievement item statistics. In M. O. Martin, & I. V. S. Mullis (Eds.), *Methods and procedures in TIMSS and PIRLS 2011*. Boston College: TIMSS & PIRLS International Study Center. https://pirls.bc.edu/methods/pdf/tp11_reviewing_achievement.pdf.
- Gándara, P. (2021). The gentrification of two-way dual language programs: A commentary. *Language Policy*, 20(3), 525–530. <https://doi.org/10.1007/s10993-021-09595-z>
- Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally comparable measures of occupational status for the 1988 international standard classification of occupations. *Social Science Research*, 25(3), 201–239. <https://doi.org/10.1006/ssre.1996.0010>
- Gebauer, S. K., Zaunbauer, A. C. M., & Möller, J. (2013). Cross-language transfer in English immersion programs in Germany: Reading comprehension and reading fluency. *Contemporary Educational Psychology*, 38(1), 64–74. <https://doi.org/10.1016/j.cedpsych.2012.09.002>.
- Genesee, F., & Lindholm-Leary, K. (2013). Two case studies of content-based language education. *Journal of Immersion and Content-Based Language Education*, 1(1), 3–33. <https://doi.org/10.1075/jicb.1.1.02gen>
- Hallquist, M. N., & Wiley, J. F. (2018). Mplusautomation: An R package for facilitating large-scale latent variable analyses in Mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(4), 621–638. <https://doi.org/10.1080/10705511.2017.1402334>
- Heller, K. A., & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+R)* [Cognitive Abilities Test for grades 4 to 12, revision (KFT 4-12+R)]. Hogrefe.
- Howard, E. R., Lindholm-Leary, K., Rogers, D., Olague, N., Medina, J., Kennedy, B., ... Christian, D. (2018). Guiding principles for dual language education (3rd ed.). *Center for Applied Linguistics*.
- Howard, E. R., & Neugebauer, S. R. (2015). Moving towards biliteracy: Varying paths of bilingual writers in two-way immersion programs. *Revista Miriada Hispanica*, 10, 83–106.
- Howard, E. R., Sugarman, J., & Christian, D. (2003). *Trends in two-way immersion education. A review of the research (Report 63)*. Center for Research on the Education of Students Placed At Risk.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kim, Y.-S. G., Stern, J., Mohohlwane, N., & Taylor, S. (2024). Instruction influences cross-language transfer of reading skills: Evidence from a longitudinal randomized controlled trial. *Reading and writing*. Advance online publication. <https://doi.org/10.1007/s1145-023-10508-1>
- Kolen, M. J., & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practices. *Statistics for social and behavioral sciences* (3rd ed.). New York: Springer. <https://doi.org/10.1007/978-1-4939-0317-7>

- Krashen, S. (2005). The acquisition of academic English by children in two-way programs: What does the research say? In V. Gonzales, & J. Tinajero (Eds.), *Review of research and practice* (pp. 1–19). Lawrence Erlbaum Associates.
- Lindholm-Leary, K. (2001). Dual language education. *Bilingual education and bilingualism. Multilingual Matters*, 28.
- Lindholm-Leary, K. (2011). Student outcomes in Chinese two-way immersion programs: Language proficiency, academic achievement and student attitudes. In D. J. Tedick, D. Christian, & T. W. Fortune (Eds.), *Immersion education: Practices, policies, possibilities* (pp. 81–103). Multilingual Matters.
- Lindholm-Leary, K., & Block, N. (2010). Achievement in predominantly low SES/Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism*, 13(1), 43–60. <https://doi.org/10.1080/13670050902777546>
- Lindholm-Leary, K., & Hernández, A. (2011). Achievement and language proficiency of Latino students in dual language programmes: Native English speakers, fluent English/previous ELLs, and current ELLs. *Journal of Multilingual and Multicultural Development*, 32(6), 531–545. <https://doi.org/10.1080/01434632.2011.611596>
- Lindholm-Leary, K., & Howard, E. R. (2008). Language development and academic achievement in two-way immersion programs. In T. W. Fortune, & D. J. Tedick (Eds.), *Bilingual education and bilingualism: Vol. 66. Pathways to multilingualism: Evolving perspectives on immersion education* (pp. 177–200). Multilingual Matters.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung: Probleme und Lösungen [Handling of missing data in psychological research: Problems and solutions]. *Psychologische Rundschau*, 58(2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- Marian, V., Shook, A., & Schroeder, S. R. (2013). Bilingual two-way immersion programs benefit academic achievement. *Bilingual Research Journal*, 36(2). <https://doi.org/10.1080/15235882.2013.818075>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2015). *Pirls 2016 assessment framework* (2nd ed.). International Study Center <http://timssandpirls.bc.edu/pirls2016/framework.html>.
- Mullis, I. V. S., Martin, M. O., Gonzalez, E. J., & Kennedy, A. M. (2003). *Pirls 2001 international report: Iea's study of reading literacy achievement in primary school in 35 countries*. International Study Center.
- Mullis, I. V. S., Martin, M. O., Kennedy, A. M., Trong, K. L., & Sainsbury, M. (2009). *Pirls 2011 assessment framework*. TIMSS & PIRLS International Study Center. https://timssandpirls.bc.edu/pirls2011/downloads/PIRLS2011_Framework.pdf.
- Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- OECD. (2024). *Pisa 2022 technical report*. PISA. OECD Publishing. <https://doi.org/10.1787/01820d6d-en>
- Padilla, A. M., Fan, L., Xu, X., & Silva, D. (2013). A Mandarin/English two-way immersion program: Language proficiency and academic achievement. *Foreign Language Annals*, 46(4), 661–679. <https://doi.org/10.1111/flan.12060>
- Preusler, S., Fleckenstein, J., Zitzmann, S., Baumert, J., & Möller, J. (2024). Two-way immersion promotes additional language learning: performance of bilingual sixth-grade students in English as a third language. *International Journal of Bilingual Education and Bilingualism*, 27(7), 910–922. <https://doi.org/10.1080/13670050.2024.2307436>.
- Preusler, S., Zitzmann, S., Baumert, J., & Möller, J. (2022). Development of German reading comprehension in two-way immersive primary schools. *Learning and Instruction*, 79, 101598. <https://doi.org/10.1016/j.learninstruc.2022.101598>.
- Preusler, S., Zitzmann, S., Paulick, I., Baumert, J., & Möller, J. (2019). Ready to read in two languages? Testing the native language hypothesis and the majority language hypothesis in two-way immersion students. *Learning and Instruction*, 64, 101247. <https://doi.org/10.1016/j.learninstruc.2019.101247>.
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Robitzsch, A., Kiefer, T., & Wu, M. L. (2021). Tam: Test analysis Modules [R package Version 3.7-16] <https://CRAN.R-project.org/package=TAM>.
- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. Wiley. <https://doi.org/10.1002/9780470316696>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Steele, J. L., Slater, R. O., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2017). Effects of dual-language immersion programs on student achievement. *American Educational Research Journal*, 54(1S), 282S–306S. <https://doi.org/10.3102/0002831216634463>
- van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- von Hippel, P. T. (2018). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods & Research*, 49(3), 699–718. <https://doi.org/10.1177/0049124117747303>
- Watzinger-Tharp, J., Tharp, D. S., & Rubio, F. (2021). Sustaining dual language immersion: Partner language outcomes in a statewide program. *The Modern Language Journal*, 105(1), 194–217. <https://doi.org/10.1111/modl.12694>
- Zhao, J. H., & Schafer, J. L. (2023). pan: Multiple imputation for multivariate panel or clustered data [R package version 1.9] <https://CRAN.R-project.org/package=pan>.
- Sandra Preusler is a postdoctoral researcher at the Institute for Psychology of Learning and Instruction, Kiel University. Her main research interests include bilingual education, language proficiency, and assessing student achievement.
- Steffen Zitzmann is a full professor of quantitative methods at the Medical School Hamburg, Germany.
- Jens Möller is a full professor of educational psychology at the Institute for Psychology of Learning and Instruction, Kiel University.
- Jürgen Baumert is a professor emeritus at the Max Planck Institute for Human Development.