Multilingual Development - a Longitudinal Study

Desarrollo Multilingüe – un Estudio Longitudinal

Ingrid Gogolin, Thorsten Klinger, Birger Schnoor and Irina Usanova
Universität Hamburg

Abstract

Does multilingual development bolster or obstruct educational success? This was the starting question for the research project "Multilingual Development - a Longitudinal Perspective" that will be presented here. While multilingualism has often been cited as a disadvantage, there are also indications that it supports successful (language) learning. It was therefore a goal of the project to obtain primary information on the language development of multilingual secondary school pupils in Germany in order to contribute to the further clarification of this question. The project is based on data of receptive (reading) and productive (writing) language skills in German (the language of schooling), Russian and Turkish as immigrants' heritage languages, English (the first foreign language taught in primary school already) and for a part of the sample French or Russian (as second foreign languages taught in academic track schools). With this range, it was possible to examine the linguistic profiles of the participants as well as reciprocal influences between languages. The sample included roughly 2000 students, tested at four measurement points in the course of secondary schooling. This is the first interdisciplinary study worldwide which includes the complexity of multilingual development on the basis of a substantial sample of respondents, measured language data a wide range of background data for statistical control of influencing background factors. The project thereby creates new knowledge about the complexity of language development in the context of migration, as well as a new basis for the design of language education in linguistically heterogeneous schools. In this contribution, we focus on one of the premier outcomes of our project: the development of an innovative test for the measurement of multilingual writing skills.

Correspondencia: Ingrid Gogolin, ingrid.gogolin@uni-hamburg.de, Universität Hamburg, 20146 Hamburg, Alemania.

Resumen

El desarrollo multilingüe, ¿refuerza u obstruye el éxito educativo? Esta fue la pregunta inicial del proyecto de investigación "Desarrollo multilingüe - una perspectiva longitudinal" que se presenta en este artículo. Si bien el multilingüismo se ha citado a menudo como una desventaja, también hay indicios de que favorece el éxito del aprendizaje (de idiomas). Por consiguiente, uno de los objetivos del proyecto fue obtener información primaria sobre el desarrollo lingüístico de los alumnos de escuelas secundarias multilingües en Alemania, a fin de contribuir a aclarar más esta cuestión. El proyecto se basa en datos de competencias lingüísticas receptivas (lectura) y productivas (escritura) en alemán (el idioma de escolarización), ruso y turco como idiomas de herencia de los inmigrantes, inglés (el primer idioma extranjero que se enseña desde la escuela primaria) y, para una parte de la muestra, francés o ruso (como segundos idiomas extranjeros que se enseñan en las escuelas secundarias de carácter académico). Con esta variedad, fue posible examinar los perfiles lingüísticos de los participantes, así como las influencias recíprocas entre los idiomas. La muestra incluyó alrededor de 2.000 estudiantes, a los que se pasó una prueba en cuatro momentos a lo largo del curso escolar. Esta investigación constituye el primer estudio interdisciplinario en todo el mundo que incluye la complejidad del desarrollo multilingüe con una muestra sustancial de encuestados, que midió los datos lingüísticos una amplia gama de datos de contexto para el control estadístico de los factores contextuales que pudieran influir. De ese modo, el proyecto crea nuevos conocimientos sobre la complejidad del desarrollo lingüístico en el contexto de la migración, así como una nueva base para la concepción de la enseñanza de idiomas en escuelas lingüísticamente heterogéneas. En esta contribución nos centramos en uno de los principales resultados de nuestro proyecto: la elaboración de una prueba innovadora para la medición de las aptitudes de escritura multilingüe.

Palabras clave: diversidad lingüística; migración; lenguas de herencia; idiomas extranjeros; pruebas multilingües.

Introduction[1]

The number of international migrants worldwide rapidly increased since the turn of the century, and sources, motivations, types of migration vary considerably. Conflicts and crises are causes of forced migration, global mobility enables voluntary migrants to work or study in other countries – the world is 'on the move' (International Organization for Migration [IOM], 2019). Germany is a showplace and prototype of this development; by the following information, we aim to illustrate the actual situation briefly (detailed information in (Gogolin, McMonagle, & Salem, 2019). Since the early

---

1   This article is based on joint work and publications from our research group; see for example Gogolin, Klinger, Lagemann, and Schnoor (2017); Brandt, Lagemann, and Rahbari (2017); Klinger, Schnoor, and Usanova (2021). The project as a whole is a cooperative endeavour, carried out by a multidisciplinary team of researchers from Education Science, Linguistics and Educational Psychology, most of them based at Universität Hamburg (www.mez.uni-hamburg.de). An overview of all parts of the project is in print: Brandt, Krause, Usanova, and Rahbari (2021).

2000s, migrants from about 190 countries contribute to the economic, social, linguistic and cultural diversification of the country's population – that is to say: people from nearly all regions of the world live in Germany. The effects of migration are mainly visible in urban areas, for example in cities like Hamburg (where our research group is located). In all major German cities, one out of two school children come from a migrant family (Autorengruppe Bildungsberichterstattung, 2020). Many of these children live with more than one language at home. Unfortunately, no census data on languages in Germany is available; due to historical reasons, the respective information is officially considered to belong to the private sphere. Thus, neither the number of languages which are used on a daily basis in the country is reliably known nor the size of their groups of speakers. According to regional case studies however, the vitality of migrants' heritage languages is substantial, although different in the various migrant groups. According to population data and other estimates, the three most common family languages in Germany are German, Turkish and Russian.

Due to this constellation, language diversity is a day-to-day experience in the country. Not only immigrants' descendants face language diversity in their everyday world. Since the 1990s, learning one foreign language is obligatory at school; most children take part in English teaching. About half of the school population – i.e. all pupils who attend an academic track secondary school – are obliged to participate in second foreign language teaching. And last not least, also those children who grow up in monolingual families communicate with multilingual peers and friends, encounter different languages, language variations and dialects in their experience realm, when travelling and in the (social) media (Peukert & Gogolin, 2017). A number of studies show that migrant languages play a considerable role in juvenile communication, be it as common vernaculars of belonging or as variants acquired in everyday interaction with peers (Auer & Dirim, 2004; Rampton, 2018).

Linguistic diversity thus has become a key environmental condition of education. According to German basic law, the linguistic decent of a person shall not be a cause of disadvantage. With respect to education, the impartation of multilingual competences belongs to the general aims of schooling. The country adopted the EU policy agenda, according to which every person living in Europe should master the common language of their country or region of residence, and moreover at least two additional languages (see Acerca de la política de multilingüismo[2]). Given this context, multilingualism is a daily occurrence as well as subject of political and ethical precautions.

This contrasts with the recurring result of educational research, indicating that the linguistic origin of pupils is a source of disadvantage in educational careers. At any rate, this is a shared assumption in many national and international studies on performance of school systems (OECD, 2019a). The PISA studies for example, recurrently yield the result that pupils with migrant backgrounds achieve worse than their condisciples from autochthonous families. In Germany in particular, the achievement gap between pupils with and without a migrant background continues to be considerably large – irrespective of the country's longstanding tradition to deal with migration (OECD, 2019b).

---

2   https://ec.europa.eu/education/policies/multilingualism/about-multilingualism-policy_es; 1.9.2020.

The usage of a family language other than the common language of schooling is often understood as a highly operant source of disadvantage (Stanat & Christensen, 2006).

This, however, is only one side of the coin of multilingualism. Other relevant research – namely on the problem of language development in multilingual contexts – reveals that living in two or more languages is most likely a source of advantage for individual development (Woll & Wei, 2019). Being multilingual is, for example, repor-ted to come with cognitive vantages such as enhanced metalinguistic awareness and executive control, which are beneficial for language learning and learning in general (Blom, Küntay, Messer, Verhagen, & Leseman, 2014; Poarch & Bialystok, 2017). The project "Multilingual Development: A Longitudinal Perspective" (MEZ) was develo-ped against this inconsistent evidence. We intend to shed more light on the questions if, where applicable why and how linguistic diversity contributes to the continuing disadvantage of migrant pupils in schools, using the German constellation of language diversity as an exemplary case.

Method

Objectives

The MEZ study aims to shed light on processes and effects of multilingual develo-pment by examining the following research questions (see for the following (Brandt et al., 2017):

- Which individual, contextual and linguistic factors facilitate or constrain language development in a multilingual context?
- How do these factors interact?
- How is the development of multilingual competence affected by other influences on educational careers (e.g. effects of the types of school that were/ are attended)?
- The research questions asked in the MEZ study can be grouped as follows:
  a) Determinants of multilingual skills. We evaluate language skills in German (the language of schooling), in English (the first foreign language for all pupils), in heritage languages Turkish and Russian (the two largest migrants' languages in Germany) and in French (as second foreign languages taught in academic track schools).
  b) Language development over time (in secondary education). We investigate the development of competence in the languages over time in order to iden-tify differences in the language skills of pupils from mono- and multilingual families and explain variations in their skills by controlling for different individual, contextual and linguistic factors.
  c) Multilingualism, career aspirations and educational choices. Drawing on sociological rational-choice theories, we investigate the relationship between language skills, attitudes and pupils' self-concepts as influences of their career aspirations, expectations and eventual choices in the educational system, assuming that career aspirations and similar plans for the future become more relevant for investment in language learning towards the end of general education.

d) The MEZ study was carried out in a two-cohort sequence design. Data collection started in January 2016 with pupils in grade 7 (13+ years of age) resp. grade 9 (15+ years) and was carried out in four waves (figure1) over three years (until grade 9 and 11, respectively).
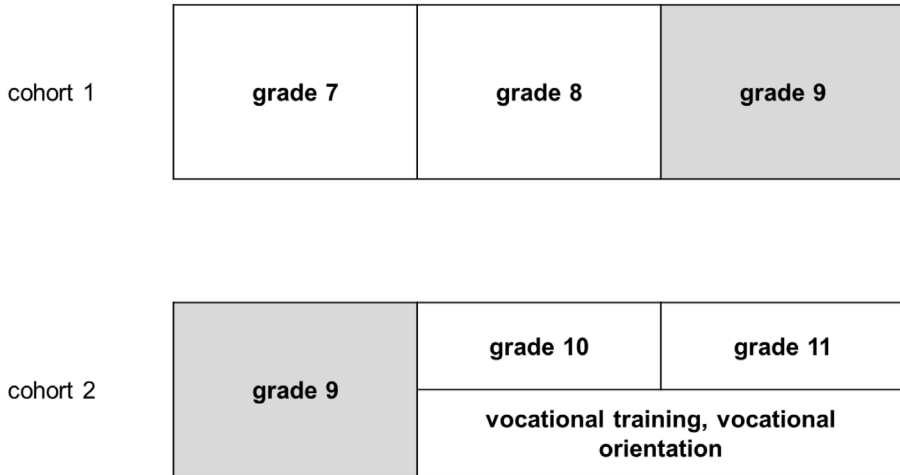

Figure 1. Design of the MEZ study

The repeated observation of the same individuals over time enables assessing the development of the language skills and allows for evidence-based assumptions of causal relationships between language development and influencing factors. The latter is essential to answer the question if and how multilingualism contributes to educational disadvantage.

Population and Sample Size

The MEZ sample consists of roughly 2000 pupils (at measurement point 1) with German-Russian, German-Turkish and monolingual German family backgrounds. All pupils attended a school in Germany since primary education[3]. The sampling of pupils took place via schools that were selected by the following criteria: they were attended by an average (or higher than average) percentage of pupils with Russian or Turkish family backgrounds (according to the Statistical Offices of the German Federal States), and their curricula had to include foreign language teaching in English, Russian and/or French (Heimler, 2019). The sampling procedure leads to a selective, non-representative sample. In order to control for specific features in comparison

---

3   We started designing our study before the recent new immigration of refugees and other persons seeking for shelter in 2015. Given the complexity of the research design and in order to control for relevant factors of influence on language development, we excluded pupils with potential educational careers in their countries of origin or elsewhere outside Germany in our sampling strategy.

with representative samples, we applied items on demographic and other relevant background information deriving from representative German studies, in particular the German National Educational Panel Study (NEPS; see https://www.neps-data.de) (Blossfeld & Roßbach, 2019).

Measurement Tools

A tool for measuring productive language abilities of secondary school students in multiple languages is essential for the MEZ project as a whole. Despite of the growing interest of education research in language related problems, a respective tool did not exist before we carried out our study. Thus, the development of an appropriate approach to testing is a first major, and innovative outcome of our project. It can build the basis for further research in the German as well as other international contexts. For this reason, we focus our further report on this aspect of our project: the development of a cross-lingual, comprehensive competence model of writing skills and its operationalisation through empirical indicators. Subsequent statistical modelling using a longitudinal latent measurement model reveals that the postulated theoretical construct has been measured in all MEZ-languages in a structurally comparable way. Thus, the tool is appropriate for the identification of inter-individual differences of writing skills in multiple languages.

Writing skills play a pivotal role in students' educational success and for their social participation in general. The vast majority of large-scale assessment studies however, focus on reading, whereas writing skills remain less explored Although research on writing is increasing in recent years, it is common practice – for example in international large-scale assessment studies such as PISA or PIRLS – to measure language-related skills only in the aspect of reading. Undeniably, reading skills belong to the basic cultural competences in complex societies. It is an open question, nevertheless, which functions can be ascribed to this partial linguistic capacity in educational contexts. In a linguistic perspective, language competence is a conglomerate of partial abilities (Rosebrock & Bertschi-Kaufmann, 2013). A single partial ability – such as reading – cannot represent 'language competence' pars pro toto. This also applies with regard to literacy skills. Language development which includes access to literacy via schooling, embraces the development of a formal register of language, and it is evident that both facets of this mode – the receptive (such as reading) as well as the productive (such as writing) – have to be acquired (Sürig, Şimşek, Schroeder, & Boness, 2016). Consequently, indicators of receptive as well as productive abilities must be considered if research aims at considering literacy in a generic sense. This assumption is supported by studies showing that both dimensions of literacy skills are interconnected, but neither the indicators for reading or nor those for writing represent literacy per se (Klinger, Usanova, & Gogolin, 2019). Their relation however, their respective 'shares' in a comprehensive construct of literacy skills is still underdetermined and not sufficiently explored in empirical studies.

The MEZ study contributes to closing this gap due to the study design which includes measures for both, reading and writing, in multiple languages. Whereas we could measure reading skills with a standardized test (Schneider, Schlagmüller, &

Ennemoser, 2007)[4], no such instruments were available for writing. The majority of studies on multilingual development is focused on early stages of language acquisition; studies on the development of writing skills emphasize emerging writers. The development in adolescents is scarcely explored (Morris Miller, Scott, & McTigue, 2018). Multilingual testing in particular, is a desideratum so far (Chalhoub-Deville, 2019). The development and psychometric evaluation of an appropriate test was thus one of our project's work packages.

The MEZ writing test is theoretically based on a functional pragmatic concept of communication which has proved particularly appropriate with regard to educationally relevant language skills (Halliday & Martin, 2004), and was supplemented and refined by research on writing (Wagner et al., 2011). Based on these approaches, the increase of learners' competences is indicated by growing complexity of their (oral or written) utterances in the dimensions of morpho-syntax, semantics, pragmatics, including the command of genre-specific characteristics. The expansion of educationally relevant language skills interacts with the enhancement of underlying cognitive skills (Cummins, 2002); in this understanding, respective tests represent compositions of linguistic and cognitive qualifications. For the purpose of measurement, context related factors must be taken into account, such as the time which is provided for the completion of tasks.

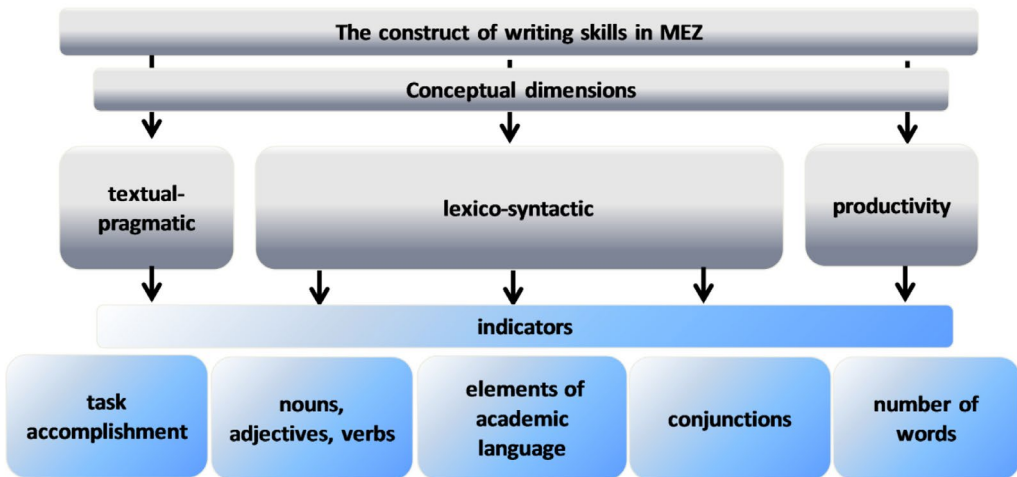These considerations have been incorporated in the MEZ construct of writing skills (see figure 2).

Figure 2. The MEZ construct of writing skills

4   This test was standardized for reading in German for pupils in secondary schooling. In the MEZ study, we developed functionally parallel versions for testing in Turkish and Russian; see Ennemoser, Usanova, and Gogolin (2021).

The choice of indicators takes account of the developmental nature of writing skills in adolescence (Nippold, 2016), with particular reference to traits specific to bilingual writing (Usanova, 2019). Furthermore, we deal with the specific constellation of biliteracy in different scripts. With regards to Turkish as a language of Latin script, only minor differences exist in comparison with the German or English. For Russian however, we have to deal with the difference of Latin and Cyrillic scripts. In general, scriptural difference is considered as an additional barrier for the development of biliteracy, whereas similarity of scripts is considered to be beneficial (Bialystok, Luk, & Kwan, 2005). The question however, how bilingual learners deal with the challenge of different scripts especially in contexts where students may not be systematically exposed to the different scriptural systems by instruction, has as yet not been sufficiently understood (Usanova 2019). In the MEZ study, we chose for an experimental approach to solve: We allow the test persons to choose the script when writing in Russian. At the level of indicators, the Russian language texts are treated identically, regardless of whether they are presented in Latin or Cyrillic script or whether mixtures of both systems are applied. This provides the opportunity to uncover concepts of script that are acquired in a range of language acquisition contexts – from exclusively oral development in one of the languages and literacy on one or more other languages to formal education.

The elicitation of written texts we applied in the MEZ study is based photo stories, each comprising of a series of nine pictures. Each story represents the production of an everyday object in individual steps. The task is to produce a text explaining this process, while also encouraging narrative expressions. The type of task is familiar to the participants because corresponding tasks also occur in schooling. In pilot studies we precluded the risk of the task being culturally or gender biased. Different prompts were used in waves 1 to 3 of the study. In wave four we used all prompts in a parallel rotated system in order to measure their comparability and to assess prompt-specific motivational effects.

In the following table, we give insight into the operationalization of the MEZ indicators for the analysis of writing skills:

Table 1.

Operationalisation of writing skills in the MEZ project

Indicators	Operationalisation
Task accomplishment	nine sub-categories, each characterizing a picture as part of the story line. Each of the nine sequences is assessed according to a four-step scale. A maximum of 27 points can be achieved.
Nouns, adjectives, verbs	Total numbers of the produced types of nouns, adjectives, and verbs.
Clause linkage	Total number of types of coordinating and subordinating conjunctions.

Indicators	Operationalisation
Elements of academic language	Total number of context appropriate lexical and morpho-syntactic elements of academic language
Productivity, text length	The total number of distinct words.

Data collection in the MEZ study was carried out by the International Association for the Evaluation of Educational Achievement (IEA) (https://www.iea.nl/)[5]. In order to avoid interaction between the language data, the tests were carried out on two test days with an intermediate waiting period. The resulting texts were coded by carefully trained research assistants. Training and coding were carried out on the basis of language-specific manuals. Regular quality controls per coder included continuous checking of the one-dimensionality as well as internal consistency of the total scales (see table 2):

Table 2.

Inter-rater reliability for writing scores in each of the investigated languages

Reliability by language and prompt: Intraclass correlation coefficients (ICC)

Language version	Prompt 1	Prompt 2	Prompt 3
German	.869	.935	.944
Russian	.988	.970	.966
Turkish	.971	.968	.865
English	.887	.983	.945
French	.969	.937	.957

Procedure and Statistical Analysis

Repeated measurements of the same construct over time requires invariance of the measurement model. Development in the measurements has to represent a change in the construct addressed by theory, not (only) in psychometric features of the tests that were applied. In order to evaluate our instruments, we used two-step analytic strategy: in step 1, we built a measurement model that replicates our theoretical model of

---

5   The procedures are explained in detail for each wave of data collection in reports per wave, see https://www.mez.uni-hamburg.de/1uebermez/working-papers.html.

writing ability into a statistical measurement model which accounts for psychometric features. In step 2, we tested for the level of measurement invariance of our model.

We applied a Longitudinal Confirmatory Factor Analysis (CFA) (Little, 2013). The coded categories serve as manifest indicators for a latent construct of writing ability which is comparable across languages. A persons' scores in a test are manifestations (manifest indicator) of the influence of one (or more) hypothetical constructs at a higher level of abstraction (latent factor). The CFA allows for separation of the constructs' true variance from indicator specific measurement errors; the explicit modelling of the measurement models' factorial structure also allows for testing its psychometric features by imposing restrictions to model parameters (Kline, 2016)[6].

Results

Four levels of factorial invariance are distinguished for measurement models (Kline, 2016): (I) Configural invariance (structurally identical models fit the data at all measurement occasions; the theoretical construct was measured structurally invariant over time; (II) metric invariance (the strength of the indicators' factor loadings is the same at each wave; the theoretical construct was measured on the same scale in all waves; (III) scalar invariance (the indicators' latent intercepts do not differ between the measurement occasions); (IV) residual invariance (the indicator-specific residual variances are invariant at each measurement occasion). [7] In the MEZ project, the level of measurement invariance was confirmed by a stepwise procedure, in which the most restrictive model was identified. The analyses reveal that we reached metric invariance for all our instruments, that is to say: It is possible to measure writing skills of adolescents across four waves in five languages with a uniform measurement model which is based on one comprehensive theoretical model.

For an illustration, we present the results we achieved with respect to German, the common language of schooling.

Table 3 presents an overview of means and standard deviations of the indicators of the "MEZ writing task for adolescents". For longitudinal analyses, the students' raw scores for the indicators were transformed into POMP scores[8] (Moeller, 2015).

The indicator means range for task accomplishment from 65.4 percent (wave 1) and 73.8 percent (wave 4), for verbs between 39.0 percent (wave 3) and 47.1 percent (wave 1), for conjunctions between 30.2 percent (wave 3), and for number of words between 42.5 percent (wave 1) and 32.2 percent (wave 3).

The high mean level for task accomplishment derives from the natural maximum of 27 points, which prevents outliers. This mechanism does not apply to the other

---

6 A detailed introduction to all methodological aspects of our evaluation is presented in Gogolin, Klinger, Schnoor, and Usanova (2021).

7 Level IV of measurement invariance is of little practical importance in applied research and, in most cases, not commended for investigation; see Cheung and Rensvold (2002).

8 POMP scores (percentage of maximum possible): The students' raw scores were standardized as the percentage of the maximum achievable [(raw score/maximum achievable score) * 100]. Since only task accomplishment has a natural maximum (27 points), the highest empirical score measured at all waves was used to set the maximum.

indicators. Task accomplishment is also the only indicator for which the means from wave 1 to wave 4 are continuously increasing. The other indicators show a trend of decreasing means from wave 1 to wave 3, followed by an increase in wave 4. This is a developmental pattern of the longitudinal mean structure of the indicators.

Table 3.

Means and standard deviation of the indicators for writing skills in German (percentage of maximum possible)

n = 2075[9]	wave 1 (n=1751)	wave 2 (n=1716)	wave 3 (n=1591)	wave 4[10] (n=1085)
Task accomplishment (points)	65,4 (14,0)	67.7 (10,6)	70.8 (11,5)	73.8 (11,0)
Verbs (types)	47.1 (13,6)	44.6 (11,8)	39.0 (12,0)	46.2 (15,2)
Conjunctions (types)	39.4 (15,8)	37.8 (15,4)	30.2 (17,0)	32.7 (15,5)
Number of words (words)	42.5 (12,4)	36.1 (9,2)	32.2 (10,3)	35.2 (12,6)

In table 4, we present the analysis of measurement invariance for German:

Table 4.

Model fit indices for the models corresponding to levels of factorial invariance (cf. Little, 2013)

Model (n = 2075)	χ^2	Df	P	RMSEA	RMSEA 90% CI	CFI	ΔCFI	Pass?
Null model	13592.3	144	<.000	-	-	-	-	-
Configural Invariance	488.7	73	<.000	.052	.048 / .057	.959	-	yes
Metric invariance	546.9	82	<.000	.052	.048 / .056	.954	.005	yes
Scalar invariance	1837.4	91	<.000	.096	.092 / .100	.828	.126	no
Partial scalar invariance	657.6	86	<.000	.057	.053 / .061	.944	.010	yes

The *null model* assumes complete independence of the indicators; it functions as a baseline model. It also provides information on the total amount of empirical information available for model specifications by using the variances, covariances, and mean

---

9   The total sample includes all cases for which information is available. By application of MPlus software package we can use all available information for estimation of models by conducting the FIML (full information maximum-likelihood) approach.

10   The increased drop-out of participants in wave 4 is due to pupils of the older cohort who had finished school or changed schools. It is therefore random.

values of the indicators for model parameter estimation (144 degrees of freedom). The model for *configural invariance* shows with a slight modification a good fit to the data.[11] Thus, the writing ability in German was measured structurally comparable over time. The model for *metric invariance* proves to be as good as the former model. The difference in ΔCFI = 0.005 indicates no poorer fit for the more restricted model, thus it can be assumed that the writing ability in German was measured in all waves on the same scale.

The model for *scalar invariance*, on the other hand, no longer fits the data. A review of the changes in the indicator means showed too much fluctuation in both task accomplishment and conjunctions. A model for partial scalar invariance that takes this into account shows a satisfactory fit with the data. For this purpose, the intercepts for task accomplishment were freely estimated for all four waves. Our further analyses show that metric invariance can be confirmed for all language versions tested in MEZ: German (the language of schooling), Turkish and Russian (heritage language of pupils from migrant families) as well as English and French (as first and second language taught at school).

Discussion and Conclusion

Our analyses confirm that the MEZ model of measurement for writing skills is in accord with the theoretical model which we presented above. With respects to indicators however, we had to reduce their number from seven (as originally assumed on the basis of theory and state-of-the-art research) to four. The reduction involves (I) the indicators for vocabulary. Here, the analyses brought to light that the three types of content words (nouns, adjectives, verbs) that were initially included are strongly correlated. The most relevant information is related to the type of verbs in a text. Furthermore, we had to delete "elements of academic language". These fit well into the overall construct when considering each prompt version separately, but less consistently across waves and versions. Thus, we can't consider them compatible with longitudinal measurement invariance. For cross sectional measurement however, they can well be applied.

Our analyses confirm that the latent construct was measured by our instruments in each language on the same scale over time. This is a remarkable result, as it confirms that it is possible to measure the development productive (written) competences of pupils in multiple languages on a unified theoretical basis. By this result, we can confirm that the application of our instruments allows for statements on the inter-individual multilingual development of the tested cohorts. At his moment, however, we not yet reached the possibility of scalar invariance, which would also allow for the analysis of intra-individual language development in multiple languages over time.

We consider our findings as preliminary and are about to further explore potential explanations for the absence of scalar measurement invariance over time. At the moments, we follow a number of assumptions deriving from the observations of our

---

11   For the first wave, a positive residual correlation of task accomplishment and text length was allowed, because - independent of the pupils' writing skills – the first-wave coders systematically rated longer texts higher in task accomplishment than shorter texts.

data. To these belongs the possibility that we induced a specific form of panel effect by the application of structurally very similar writing tasks in the four waves. On the one hand, this is one of the strengths of our design as it allows for the comparability of the latent construct across the different tested languages. On the other hand, it may cause unintended effects such as habituation to the type of task which might result in more parsimonious usage of different linguistic means or to a reduction of the motivation of participants. Explanations like the mentioned are now tested in our further investigation of the rich data we achieved.

In any case, the results of or study indicate that it will be possible in the future to apply instruments for the observation of multilingual language development in further large-scale research. This is innovative in at least two respects: (I) it is conceivable to measure not only the receptive but also the productive dimension of literacy, and thus gaining a more appropriate information on educationally relevant skills; and (II) the competences in multiple languages can be included in this information, and thus accommodating a comprehensive indication of the actual language competence of children and youth who are taught and have to learn in linguistically diverse constellations.

References

Autorengruppe Bildungsberichterstattung (2020). *Bildung in Deutschland 2020: Ein indikatorengestützter Bericht mit einer Analyse zu Bildung in einer digitalisierten Welt*. Bielefeld: wbv Publikation.

Bialystok, E., Luk, G., & Kwan, E. (2005). Bilingualism, Biliteracy and Learning to Read. Interactions among Languages and Writing Systems. *Scientific Studies of Reading.*, *9*(1), 43–61.

Blom, E., Küntay, A. C., Messer, M., Verhagen, J., & Leseman, P. P.M. (2014). The benefits of being bilingual: Working memory in bilingual Turkish-Dutch children. *Journal of Experimental Child Psychology*. (128), 105–199.

Blossfeld, H.-P., & Roßbach, H.-G. (Eds.) (2019). *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)*. Wiesbaden: Springer VS.

Brandt, H., Krause, M., Usanova, I., & Rahbari, S. (Eds.) (2021). *Language Development in Diverse Settings: Cross disciplinary findings from the Project "Multilingual Development: A Longitudinal Perspective (MEZ)"*. Westport: Springer.

Brandt, H., Lagemann, M., & Rahbari, S. (2017). Multilingual Development: A Longitudinal Perspective - Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ). *European Journal of Applied Linguistics*. (online first), 1–11. Retrieved from doi.org/10.1515/eujal-2017-0024

Chalhoub-Deville, M. B. (2019). Multilingual Testing Constructs: Theoretical Foundations. *Language Assessment Quarterly*, *16*(4-5), 472–480. Retrieved from DOI:10.1080/15434303.2019.1671391

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255.

Cummins, J. (2002). BICS and CALP. In M. Byram (Ed.), *Encyclopedia of Language and Teaching* (pp. 76–79). London: Routledge.

Ennemoser, M., Usanova, I., & Gogolin, I. (2021). Adapting the LGVT 5-12+ reading speed and comprehension test for the assessment of reading skills in the heritage languages Russian and Turkish. In H. Brandt, M. Krause, I. Usanova, & S. Rahbari (Eds.), *Language Development in Diverse Settings: Cross disciplinary findings from the Project "Multilingual Development: A Longitudinal Perspective (MEZ)"* (in print). Westport: Springer.

Gogolin, I., Klinger, T., Lagemann, M., & Schnoor, B. (2017). *Indikation, Konzeption und Untersuchungsdesign des Projekts Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ)* (MEZ Arbeitspapiere 1). Hamburg. Retrieved from Universität Hamburg website: https://www.mez.uni-hamburg.de/1uebermez/working-papers.html

Gogolin, I., Klinger, T., Schnoor, B., & Usanova, I. (2021). The Competence Model of Writing Skills in MEZ. In H. Brandt, M. Krause, I. Usanova, & S. Rahbari (Eds.), *Language Development in Diverse Settings: Cross disciplinary findings from the Project "Multilingual Development: A Longitudinal Perspective (MEZ)"* (in print). Westport: Springer.

Gogolin, I., McMonagle, S., & Salem, T. (2019). Germany: Systemic, sociocultural and linguistic perspectives on educational inequality. In P. A.J. Stevens & G. A. Dworkin (Eds.), *The Palgrave Handbook of Race and Ethnic Inequalities in Education* (2nd ed., pp. 557–602). Basingstoke: Palgrave Macmillan.

Halliday, M. A.K., & Martin, J.R. (2004). *Writing Science. Literacy and Discursive Power*. London: Routledge.

Heimler, J. (2019). *Stichprobenentwicklung im Projekt Mehrsprachigkeitsentwicklung im Zeitverlauf (MEZ)* (MEZ Arbeitspapier Nr. 3). Hamburg. Retrieved from Universität Hamburg website: https://www.mez.uni-hamburg.de/bilder/pdf/mezarbeitspapier03.pdf

International Organization for Migration (2019). *World Migration Report 2020*. Geneva. Retrieved from e-ISBN 978-92-9068-789-4

Kline, R. B. (2016). *Principles and practice of structural equation modeling: Methodology in the social sciences* (4th ed.). New York: The Guilford Press.

Klinger, T., Schnoor, B., & Usanova, I. (2021). Developing Multilingual Writing Skills: Bilingual Advantage in Foreign Language Learning? In H. Brandt, M. Krause, I. Usanova, & S. Rahbari (Eds.), *Language Development in Diverse Settings: Cross disciplinary findings from the Project "Multilingual Development: A Longitudinal Perspective (MEZ)"* (in print). Westport: Springer.

Klinger, T., Usanova, I., & Gogolin, I. (2019). Entwicklung rezeptiver und produktiver schriftsprachlicher Fähigkeiten im Deutschen. *Zeitschrift Für Erziehungswissenschaft*, *22*(1), 75–103.

Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. New York, NY: The Guiford Press.

Moeller, J. (2015). A word on standardization in longitudinal studies: Don't. *Frontiers in Psychology*. (6). Retrieved from DOI: 10.3389/fpsyg.2015.01389

Morris Miller, D., Scott, C., & McTigue, E. (2018). Writing in the Secondary-Level Disciplines: a Systematic Review of Context, Cognition, and Content. *Educational Psychology Review*, *30*(1), 83–120. Retrieved from DOI: 10.1007/s10648-016-9393-z

Nippold, M. A. (2016). *Later language development: School-age children, adolescents and young adults* (4th ed.). Austin, Texas: Pro-Ed.

OECD (2019a). *Education at a Glance 2019: OECD Indicators*. Paris: OECD Publishing. Retrieved from https://doi.org/10.1787/f8d7880d-en

OECD (2019b). *PISA 2018 Ergebnisse, Band I: Was Schülerinnen und Schüler können und wissen*. Bielefeld: wbv Media.

Peukert, H., & Gogolin, I. (Eds.) (2017). *Dynamics of Linguistic Diversity: Individuals and Multilingual Societies II*. Amsterdam: John Benjamins.

Poarch, G. J., & Bialystok, E. (2017). Assessing the implications of migrant multilingualism for language education. *Zeitschrift Für Erziehungswissenschaft (ZfE)*, *20*(2), 175–191.

Rampton, B. (2018). Stylisation and the dynamics of migration, ethnicity and class. In N. Braber & S. Jansen (Eds.), *Sociolinguistics in England* (pp. 97–125). London: Palgrave Macmillan.

Rosebrock, C., & Bertschi-Kaufmann, A. (Eds.) (2013). *Literalität erfassen: Bildungspolitisch, kulturell, individuell*. Weinheim u. München: Beltz Juventa.

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12*. Göttingen: Hogrefe.

Stanat, P., & Christensen, G. (2006). *Where Immigrant Students Succeed - A Comparative Review of Performance and Engagement in PISA 2003*. Paris: OECD.

Sürig, I., Şimşek, Y., Schroeder, C., & Boness, A. (Eds.) (2016). *Multilingualism. Acquisition in Schools in the Context of Migration and Multilingualism: A Binational Survey*. Amsterdam: John Benjamins.

Usanova, I. (2019). *Biscriptuality: Writing Skills among German-Russian Adolescents*. Amsterdam: John Benjamins.

Wagner, R. K., Puranik, C., Foorman, B., Foster, E., Gehron Wilson, L., Tschinkel, E., & Thatcher Kantor, P. (2011). Modeling the development of written language. *Reading and Writing*, *24*(2), 203–220. Retrieved from DOI: 10.1007/s11145-010-9266-7

Woll, B., & Wei, L. (2019). *Cognitive Benefits of Language Learning: Broadening our Perspectives*. London: The British Academy.