Classroom composition, classroom quality and German skills of very young dual language learners and German-only learners

Willard, Jessica A.; Agache, Alexandru; Leyendecker, Birgit

Veröffentlichungsversion / Published Version
Zeitschriftenartikel / journal article

Empfohlene Zitierung / Suggested Citation:
Willard, J. A., Agache, A., & Leyendecker, B. (2019). Classroom composition, classroom quality and German skills of very young dual language learners and German-only learners. *Early Childhood Research Quarterly*, 49, 269-281. https://doi.org/10.1016/j.ecresq.2019.07.004

Early Childhood Research Quarterly 49 (2019) 269–281

Contents lists available at ScienceDirect

Early Childhood Research Quarterly

Classroom composition, classroom quality and German skills of very young dual language learners and German-only learners

Jessica A. Willard*,[1], Alexandru Agache[1], Birgit Leyendecker

Ruhr-Universität Bochum, Germany

ARTICLE INFO

Article history:
Received 10 August 2017
Received in revised form 22 July 2019
Accepted 23 July 2019
Available online 3 September 2019

Keywords:
Early Childhood Education and Care (ECEC) quality
Childcare
Bilingual children
Immigrant children
Vocabulary

ABSTRACT

This study examined associations between the classroom percentage of dual language learners (DLLs), observed classroom quality, and children's German majority language skills. The cross-sectional sample of 2.5 years olds (n = 93 immigrant DLLs and n = 363 monolingual German-only learners) was clustered within n = 197 classrooms. Classrooms with higher percentages of DLLs demonstrated slightly lower levels of overall classroom process quality. DLLs scored about 1 *SD* below monolingual children on German language skills when adjusting for family and classroom covariates. Moderation analyses revealed that this difference did not depend on the percentage of DLLs in a classroom. In fact, the classroom percentage of DLLs was related to children's German skills only when omitting the child level language status (DLL vs. monolingual) from the analyses. However, classroom quality moderated the difference between DLLs' and monolingual children's German skills. This difference was estimated as about only 0.5 *SD* for DLLs and monolingual children experiencing higher classroom quality, but as about 1.5 *SD* for those experiencing lower quality. We conclude that high quality classrooms may promote the majority language skills of DLLs.

Harmonious dual language development – a "clearly positive experience with bilingualism" (De Houwer, 2015, p. 170) – is a desirable objective for the increasing numbers of children with immigrant roots that are likely to grow up with two languages (Organisation for Economic Co-operation and Development, 2018; United Nations Department of Economic and Social Affairs, 2017). This objective calls for an environment where dual language learners (DLLs) can develop both their heritage and majority languages to serve them in forming family and peer relationships, constructing cultural identities, navigating everyday life and succeeding academically. The present study contributes to understanding the role of early childhood education as a developmental context for children acquiring the majority language.

Our focus is on DLLs with an immigrant background and their German-only learning monolingual peers without an immigrant background. Around the beginning of primary school, these DLLs often score lower on tests of the majority language than monolingual children (Linberg, Schneider, Waldfogel, & Wang, 2019). During the later school years, students with an immigrant background are also outperformed on tests of academic achievement by non-immigrant peers (UNICEF Office of Research, 2018). While heritage language skills may also play a role in academic outcomes, the importance of majority language skills is undisputed (Kristen et al., 2011; Prevoo, Malda, Mesman, & van IJzendoorn, 2016). There is some evidence that early childhood education can support DLLs (Buysse, Peisner-Feinberg, Páez, Hammer, & Knowles, 2014). Thus, we examined how two features of early childhood education classrooms, namely their composition and overall quality, are connected to the majority language skills of young DLLs with an immigrant background and monolingual children without an immigrant background.[2]

We defined classroom quality as the so-called process quality consisting of teacher-child interactions and classroom features that foster children's safety, positive relationships and classroom climate, and provide a stimulating environment with many learning experiences (Harms, Cryer, & Clifford, 2005; Melhuish et al., 2008; Sylva et al., 2007). German language skills were assessed via early majority language vocabulary, an important indicator of overall skills in that language (Berendes, Weinert, Zimmermann, &

* Corresponding author at: Ruhr-Universität Bochum, Psychology Department, Child and Family Research, GAFO 04/608, 44797 Bochum, Germany.
E-mail address: Jessica.Willard@rub.de (J.A. Willard).

[1] First authorship is shared between the first two authors who made equal contributions to this publication.

[2] For readability's sake, we refer to the two groups as "immigrant DLLs" and "non-immigrant monolingual children".

https://doi.org/10.1016/j.ecresq.2019.07.004

Artelt, 2013), as well as a predictor of later literacy development and academic achievement (Bleses, Makransky, Dale, Højen, & Ari, 2016; Lee, 2011). We used the broad term "DLLs" to refer to young children who are exposed to two languages (Hammer et al., 2014). Our study provides findings from outside the United States using a nationally representative sample from the German National Study of Child Care in Early Childhood (NUBBEK study; Tietze et al., 2013). Moreover, our study focuses on 2.5 year olds, an age group for which research is sparse. Our review of the extant literature on effects of classroom composition and quality draws on publications available in English, usually from the United States, as well as on the few German studies.

1. Immigrant DLLs and non-immigrant monolingual children

In this study, we compared immigrant DLLs' majority language skills as assessed through vocabulary to those of non-immigrant monolingual children. The early differences between the majority language skills of immigrant DLLs and non-immigrant monolingual children are likely formed by multiple factors including differences in socioeconomic status (SES) and the home language and literacy environment (Hammer et al., 2014; Hoff, 2018). More specifically, the quantity and quality of exposure to each language have been found to be important predictors of DLLs' skills in each language (Unsworth, 2016). DLLs may on average receive less majority language exposure as their total language exposure is divided over two languages. Moreover, DLLs may also sometimes receive later exposure to the majority language (e.g., if the majority language is not spoken in the family). Thus, differences between DLLs' and monolingual children's skills are perhaps normative and may indeed still be evident into adulthood for outcomes such as vocabulary (Hoff, 2018). Importantly, when considering the heritage and the majority language together, DLLs' total vocabulary may be at least as large over even larger as that of monolingual children (Hoff, 2018). Thus, the goal may not be for DLLs to become indistinguishable from monolingual children, or "two monolinguals in one" (Grosjean, 1989, p. 3). Moreover, understanding how to support DLLs' majority language skills is only "half of the picture". From the United States, there is evidence that including the heritage language into early childhood education curricula can benefit DLLs (Buysse et al., 2014). However, the great wealth of different heritage languages represented in German early childhood education centers (Kohl, Willard, Agache, Bihler, & Leyendecker, 2019) does not lend itself to straightforward solutions on how to simultaneously incorporate all of them. Thus, we focus on understanding how early childhood education can support DLLs in building an early foundation of strong majority language skills.

2. Classroom composition and majority language skills

Attending a culturally diverse classroom is beneficial for children's intergroup attitudes (Pettigrew & Tropp, 2006). However, in the German public and media, classroom composition is primarily considered as a potential risk factor for language development (Frigelj, 2018; Wandt, 2018). The general expectation appears to be that attending a classroom with children who, for various reasons, have lower majority language skills, will slow individual children's majority language learning. One rationale for such expectations is that children may benefit from communicating with peers who have higher majority language skills (Justice, Petscher, Schatschneider, & Mashburn, 2011; Mashburn, Justice, Downer, & Pianta, 2009). Moreover, DLLs may also benefit from majority language exposure through peers (Palermo et al., 2014; Palermo & Mikulski, 2014).

In our study, we examined composition in terms of the percentage of DLLs in a classroom. Previous evidence on the effects of linguistic composition comes from preschool-age children. One study from the United States found a negative relation between a high percentage of DLLs and individual DLLs' English vocabulary (Garcia, 2018). Other studies from the United States and Germany produced null effects on DLLs and monolingual children (Hindman & Wasik, 2015; Kohl et al., 2019). Using a slightly different composition measure, German studies showed that only DLLs' but not monolingual children's majority language skills were negatively related to the classroom percentage of immigrant children (Ebert et al., 2013; Klein & Becker, 2017). One possible explanation is that children who, on average, have less exposure to the majority language at home, are more sensitive to variations in classroom majority language exposure. However, positive effects of a high classroom percentage of DLLs on majority language skills are also conceivable. For example, DLLs may feel more encouraged to actively use their emerging majority language when surrounded by many other beginning learners. We are not aware of studies on the connection between classroom composition and language outcomes of very young DLLs and monolingual children. Thus, we examined whether the classroom percentage of DLLs was related to children's German language skills, whether such an association was stronger for DLLs than for monolingual children, and whether the strength of such an association depended on classroom quality.

3. Classroom quality and DLLs' majority language skills

There is agreement that the overall quality of early childhood education is positively associated with children's language skills (often assessed via receptive vocabulary), albeit modestly (Burchinal, Kainz, & Cai, 2011). However, there are some caveats. Most of the studies stem from the United States, and there are only few studies of very young children. Further, several larger studies from the United States excluded families or children who were not fluent in English. Thus, they limited their analyses of how classroom quality can specifically support the development of DLLs (e.g., National Institute of Child Health and Human Development Early Child Care Research Network (NICHD ECCRN), 2002; Peisner-Feinberg et al., 2001). This is unfortunate as there is evidence that DLLs may especially benefit from higher quality. A German study revealed that classroom quality, as measured by the literacy subscale of the Environmental Rating Scales extension, was modestly connected to German skills at age three only for children whose parents were not native speakers of German (Ebert et al., 2013). Another German study suggested that quality was only related to the majority language skills of DLLs with relatively little exposure to German (Kohl et al., 2019). Similarly, a study from the United States showed that a specific measure of classroom language interactions was connected only to the English outcomes of DLLs with very low initial majority language skills (Sonnenschein, Thompson, Metzger, & Baker, 2013). Thus, we examined whether classroom quality was related to the size of the difference between DLLs' and monolingual children's German majority language skills.

4. German early childhood education

Publicly subsidized early childhood education centers ("Kita" in German) are a hallmark of the German education system. Centers are either public or run by church groups or other non-profit organizations. They are heavily regulated, and programs are thus very similar regardless of their provider. This type of "Kita" center is much more affordable than many options in the United States as fees are tied to parents' income (Cryer, Tietze, & Wessels, 2002). Twenty-eight percent of children under three years of age

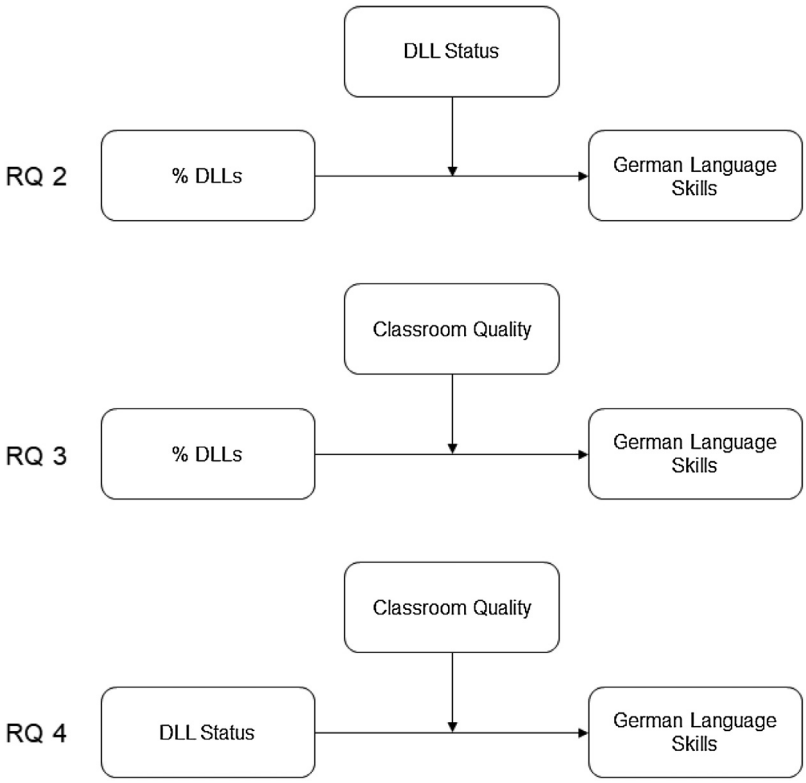

Fig. 1. Schematic depiction of the research questions 2–4.

attend such centers (Statistisches Bundesamt, 2018). The attendance rate for children from immigrant families under the age of three is relatively low even though it increases markedly later on (Alt, Berngruber, & Pötter, 2016; Schober & Spiess, 2013; Tietze et al., 2013). Overall, classroom quality in "Kitas" is only moderate (Suchodoletz, Fäsche, Gunzenhauser, & Hamre, 2014; Tietze et al., 2013). There is a shortage of placements, especially for children under three years of age (Bundesministerium für Familie, Senioren, Frauen und Jugend, 2018). Obtaining a spot in a center can be challenging and may entail multiple visits to sign up for waiting lists as well as follow-up calls.

Germany is highly diverse: every third child has an immigrant background (Statistisches Bundesamt, 2016). Turkish and Russian speakers form the two largest heritage language groups (Brehmer, 2007). Immigrant DLLs are likely to attend a center with many other peers from similar backgrounds (Gambaro, 2017). Kita classrooms with a higher proportion of children with an immigrant background may be of slightly lower quality (Becker & Biedinger, 2016; Kuger, Kluczniok, Kaplan, & Rossbach, 2016). Bilingual programs for "immigrant languages" such as Turkish and Russian are very rare (Frühe Mehrsprachigkeit an Kitas und Schulen e.V., 2019). Teachers in Kita centers almost exclusively use the German language (Bihler, Agache, Schneller, Willard, & Leyendecker, 2018).

5. This study

Our overarching aim was to uncover how early childhood education can support immigrant DLLs in developing their majority language skills. We had four research questions. First, we examined whether classroom composition in terms of the percentage of DLLs was related to classroom quality (RQ1). We expected to replicate the findings from Tietze et al. (2013) who used a larger sample of classrooms from the NUBBEK study and found a modest negative association while controlling for structural classroom characteristics. Second, we examined whether classroom composition was related to children's German skills (RQ2). Based on previous research, we expected a negative connection between the classroom percentage of DLLs and children's German skills, especially for DLLs. The third research question was whether classroom quality moderated a potential connection between classroom composition and children's German skills (RQ3). The fourth research question was whether the difference between DLLs' and monolingual children's German skills was smaller for children who attended classrooms of higher quality (RQ4). Based on the assumption that early childhood education classrooms provide a crucial portion of German exposure for DLLs, we expected that their German skills would be more strongly positively connected to classroom quality than the skills of their monolingual peers. Fig. 1 schematically depicts the last three research questions, all three involving the presence of interaction effects. For the last three research questions, we controlled for structural characteristics of the centers as well as for several socio-demographic, child and family characteristics that may be connected to either DLL status, self-selection into centers of differing composition or quality, or children's language learning (NICHD ECCRN & Duncan, 2003).

6. Method

6.1. Data and sample

Data stem from the subsample of 2.5 year olds attending Kita centers in the NUBBEK study (Tietze et al., 2013). For NUBBEK, a stratified random sample of classrooms was drawn from 32 geographic regions representative of Germany and located in eight federal states. A comparison between all mothers of 2.5-year-olds in the larger NUBBEK sample and German census data showed that mothers were representative in terms of their employment status and family status. Of NUBBEK immigrant mothers,

Table 1
Sample characteristics and descriptive data for dual language learners (DLLs) and monolingual children for all analyzed variables on the child level ($N = 456$).

	DLLs (total $n = 93$)			Monolinguals (total $n = 363$)			t or χ^2 – values
	M (SD) or %	Range	n	M (SD) or %	Range	n	
Focal outcome and predictor variables							
German language skills (PPVT-4)	18.2 (8.3)	0–35	83	29.3 (6.0)	0–40	361	11.5***
Classroom quality (ITERS-R)	3.6 (0.8)	1.9–5.3	91	4.0 (0.80)	1.9–6.1	360	3.8***
DLLs in classrooms (%)	42.5 (31.4)	0–100	93	10.8 (16.3)	0–100	363	−9.4***
Classroom control variables							
Mixed aged classroom	50.5		93	27.3		363	18.4***
Children–teacher ratio	6.1 (2.5)	1.4–14.0	91	5.7 (2.4)	1.5–15.0	359	−1.6
Teachers' education ≥upper secondary	34.4		93	34.2		363	0.002
Child and family control variables							
Child's age (months)	33.6 (2.1)	29.6–37.0	93	33.0 (2.0)	29.2–37.0	363	−2.3*
Girls	38.7		93	51.2		363	4.7*
Child's health[a]	4.31 (0.5)	2.5–5	93	4.5 (0.5)	3–5	350	2.80
No. of children in household <15 yrs.	1.8 (0.7)	1–7	93	1.7 (0.7)	1–4	363	−1.3
Age of enrollment into childcare (months)	22.5 (7.4)	2–34	91	16.0 (6.7)	0–34	360	−8.1***
Weekly hours in childcare	26.1 (7.0)	11.0–38.5	88	25.4 (7.4)	5.5–41.0	363	−0.7
Mother's age (years)	32.5 (4.8)	22.8–44.6	92	34.7 (5.2)	18.9–47.0	363	3.6***
Parents' education ≥ upper secondary[b]	37.6		93	67.2		363	27.3***
Parental low occupational status[c]	67.7		93	32.0		363	39.8***
Net equivalized household income (€)	1134 (548)	335–4048	82	1851 (834)	292–6667	328	9.4***
Difficulty finding daycare/preschool[d]	1.9 (0.5)	1.0–3.0	87	1.7 (0.5)	1.0–2.7	351	−4.0***
Lack of help with childcare in family[d]	2.3 (0.7)	1.0–3.0	92	2.0 (0.7)	1.0–3.0	362	−2.4*
Language development stimulation (z-scores)[e]	−0.3 (0.7)	−2.7–0.8	87	0.1 (0.6)	−2.4–0.8	352	4.3***
Organization and opportunities (z-scores)[e]	−0.4 (1.0)	−3.2–0.9	86	0.1 (0.8)	−6.6–0.9	345	4.4***

Notes: t-test for independent sample means (df = between 408 to 454), Pearson χ^2-test for counts ($df = 1$).
[a] See Section 6.
[b] ≥Level 4 of the International Standard Classification of Education 1997.
[c] Parental low occupational status according to Erikson–Goldthorpe–Portocarero ≥ III.
[d] Mothers' reports (1 = access to childcare/help within family, 2 = some access, 3 = difficulties finding daycare/lack of help within families).
[e] Indices of the Home Observation for Measurement of the Environment (see Section 6 on measurement details).
* $p < .05$.
** $p < .01$.
*** $p < .001$.

30% were employed (employment rate of immigrant mothers in Germany: 32%) and 9% were single mothers (percentage of immigrant mothers in Germany that are single: 7%). Of NUBBEK non-immigrant mothers, 62% were employed (non-immigrant mothers in Germany: 66%) and 13% were single mothers (non-immigrant mothers in Germany: 12%). However, the NUBBEK mothers' education was slightly higher than that of mothers in Germany as measured by the International Standard Classification of Educational Degrees 1997 (NUBBEK immigrant mothers: 3.2 vs. 2.6 for immigrant mothers in Germany, NUBBEK non-immigrant mothers: 4.1 vs. 3.6 for non-immigrant mothers in Germany) (Döge et al., 2013; Schneider, 2008). The NUBBEK study excluded children who had severe disabilities or were born very prematurely. The full data set is available at the GESIS Data Archive under the https://doi.org/10.4232/1.12297 (Leyendecker, Agache, & Madsen, 2014; Tietze et al., 2015).

Centers selected for participation handed out invitations to all potentially eligible families with children who spoke German, Turkish, or Russian at home. In order for children from Russian speaking families to take part in the NUBBEK study, the target child's mother had to report being born in the Former Soviet Union. For children from Turkish-speaking families, the selection criteria were somewhat different because there were also second generation mothers of Turkish origin in Germany (immigration from Turkey started decades earlier than immigration from the Former Soviet Union): the target child's mother or at least one maternal grandparent had to be born in Turkey. Mothers' or maternal grandparents' birthplaces were used as criteria as mothers frequently assume more responsibilities in caring for young children than do fathers (Tietze et al., 2013). While the NUBBEK study used children's immigrant background as a criterion, we were specifically interested in the children growing up as DLLs among those with an immigrant background. Thus, for our study, children with an immigrant background whose parents spoke to them solely in German (as indexed by mothers' reports) were excluded. All children received exposure to German in their classrooms. The remaining children were considered Russian–German or Turkish–German DLLs (referring to their two languages and not nationalities). From the non-immigrant subsample, we selected only the monolingual children and excluded children whose parent spoke to them in a language other than German. Lastly, only cases with complete information on their classroom's composition were retained.

The final sample for this study consisted of $n = 93$ DLLs with an immigrant background (63 Russian–German, and 30 Turkish–German), and $n = 363$ non-immigrant monolingual children. On average, the total of $n = 456$ children were 33 months old ($SD = 2$, range 29 –37). Of each the DLL and monolingual subsamples, 90% had been enrolled in no other center prior to the current one. All mothers of monolingual children were married or living with a partner, whereas this was the case for 88% of mothers of DLLs. Further descriptive information for the DLL and monolingual subsamples is presented in Table 1.

The children attended one of 197 classrooms in 188 centers. Of the classrooms, 55% were exclusively for children under three years of age, whereas the remaining 45% were "mixed-age" classrooms for children ranging in age from 0 to 6 years. An average classroom in had 22% DLLs, and 16% of the classrooms had a relatively high percentage of DLLs (over 35%). On average, there were six children per teacher in a classroom ($SD = 2.7$).

6.2. Procedure

NUBBEK was a cross-sectional study conducted from March 2010 to January 2011, with one time point of data collection for each target child. Data were collected when target children were approximately 2.5-years old, which means that some chil-

dren were assessed during the first and others during the second half of the year. Data collection in the centers was conducted by research assistants who were trained and certified to observe classroom process quality. Observations lasted about 3–4 h. Subsequently, research assistants administered standardized interviews and questionnaires to classroom teachers and center directors.

Families were visited at home by extensively trained research assistants. Research assistants visiting families of DLLs were bilingual. During a 2–3 h visit, the research assistants interviewed the mothers, administered the vocabulary test, and observed the quality of the home learning environment during the entire visit. The child received a small gift and parents 20 Euros as compensation.

6.3. Measures

German language skills. At the time of data collection, there was no standardized German Peabody Picture Vocabulary Test available for young children. Thus, a team of speech-language professionals adapted a German research version of the fourth edition of the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007; Glück, 2009). This test consisted of 40 items that were piloted and then ordered according to item difficulty. There were no basal or ceiling criteria and each child was given the same set of 40 items. Each item consisted of four color illustrations, and children were asked to select the one out of the four that matched a target word spoken by a research assistant. Split-half reliability (odd vs. even) was $r = .89$ for DLLs and .87 for monolingual children, and internal consistency was $\alpha = .90$ for DLLs and .83 for monolingual children. Correlations with teachers' ratings of communication skills that were also included in the NUBBEK study (shortened Vineland scales with $\alpha = .88$; Tietze et al., 2013) support the PPVT research version's validity. The PPVT score was moderately and significantly correlated with teachers' ratings of both DLLs' and monolingual children's communication skills ($r = .45$ and .43). Correlations with teachers' ratings of DLLs' and monolingual children's daily living skills ($r = .24$ and .11) and motor skills ($r = .24$ and .27), which can be viewed as measures of discriminant validity, were also significant but much smaller.

Classroom composition. Each classroom's lead teacher reported on the language(s) spoken at home for every child in the class. For our analyses, we considered all children for whom a language other than German was reported as DLLs (regardless of whether the other language was spoken solely or in addition to German). We then calculated the classroom shares of DLLs ranging from 0 to 1 (corresponding to 0%–100%).

Classroom quality. The German adaptation of the Infant/Toddler Environment Rating Scale-Revised (ITERS-R; Harms et al., 2005; German adaptation: Tietze, 2010) assesses classroom process quality. For this study, the first 37 observational items were mean averaged into the six subscales *Space and furnishings*, *Personal care routines*, *Language and cognitive stimulation, Activities, Interactions*, and *Program structure*. Items were rated on a seven point scale from 1 = *unacceptable quality* to 7 = *excellent quality*. A principal component analysis on the six subscale scores yielded a one factorial solution, supporting the use of a total quality score (standardized loadings ranging from .63 to .83). A recent review for the preschool version of the ITERS-R showed a much larger evidence base for predictions to vocabulary for the total score than for subscales (Brunsek et al., 2017). Thus, for our analyses we used the total score, which was computed by averaging the mean values of the six subscales ($\alpha = .83$).

6.4. Control variables

In all analyses, we controlled for structural characteristics of the classrooms which were mostly assessed through standardized questionnaires with the center directors. For the analyses on children's German skills, we included several socio-demographics and family level variables assessed through a standardized interview with the mother and an observation of the home learning environment.

Classroom covariates. We included the type of classroom by children's age span (0 = *toddler/infant classrooms for age 0–2* and 1 = *mixed-age classrooms for age 2–6*), classroom teacher's education levels as reported by center directors (recoded into 0 = *lower secondary education* and 1 = *upper secondary education with access to tertiary education*), and a measure of the ratio of children per teacher which was calculated from the average number of children and teachers present during the classroom observations.

Family SES covariates. To account for families' SES, we included three covariates. First, the highest level of education between both parents according to International Standard Classification of Educational Degrees 1997 (Schneider, 2008) was dichotomized into 0 = *lower secondary* and 1 = *upper secondary, equaling level 4 or higher*. Second, a measure of both parents' highest occupational status according to an adapted Erikson–Goldthorpe–Portocarero classification (1979) was dichotomized into 0 = *high occupational status: managerial/service occupations* and 1 = l*ow occupational status: small self-employed and non-service class occupations, equaling level III or higher*. Third, families' logarithmized net income was computed from the monthly household net income in Euros (equivalized according to family members' age using the Organisation for Economic Co-operation and Development's weighting scale; Förster, 1994).

Childcare covariates. Mothers reported on children's age in months at enrollment into childcare (*"At what age did you first enroll your child into daycare or a Kita?"*). They also reported a detailed weekly schedule from which we computed children's current weekly hours in childcare. To account for non-linear effects, we dichotomized age of enrollment into 0 = *under 18 months* and 1 = *18 months or older*, and weekly hours into 0 = *under 30 h* and 1 = *30 h or more*.

Access to childcare. To account for self-selection effects into centers, mothers reported on the potential availability of childcare for their child on a 3 point scale. We averaged three items (e.g., "*Can you currently utilize . . .a daycare/Kita?*" with 1 = *yes*, 2 = *only with difficulty*, 3 = *no*) into an index on *difficulty finding a daycare/Kita*. Another two items (e.g., "*Can you currently utilize care provided by grandparents?")* were averaged into an index *lack of help with childcare in family* from grandparents or other relatives. We dichotomized these scores (0 = *accessible childcare/help within family* and 1 = *difficulty finding daycare/lack of help within families*). These two indices are not homogeneous scales in which the items reflect an underlying construct. Instead, we considered these items as formative indicators that do not necessarily show high intercorrelations and internal consistency (Bollen & Bauldry, 2011). The same was the case for the index on children's health described below.

Quality of the home learning environment. The Home Observation for Measurement of the Environment assesses the quality of the home learning environment with six subscales. The subscales consist of 45 observational items and five interview questions answered by mothers, all in the form of yes/no statements (Caldwell & Bradley, 2001; Totsika & Sylva, 2004). A principal component analysis produced two components which we interpreted as capturing: (1) *Sensitive learning stimulation* (subscales: I, II, IV and V; higher scores indicate a more stimulating environment with sensitive mother-child interactions and play materials) and (2) *Organization and opportunities for learning stimulation* (subscales III and VI; higher scores indicate higher levels of family resources, more structured time or more opportunities for learning). From the subscales' *z*-scores, we computed a mean score for each component. Internal consistency was acceptable for sensitive learning stimula-

tion ($\alpha = .58$) and for organization and opportunities for learning stimulation ($\alpha = .57$).

Child's health. Children's health was rated by both mothers and teachers with one item ("*How would you normally describe the child's physical health and well-being?", with* 1 = *poor health* to 5 = *excellent health*). We calculated the mean score index of these two ratings.

6.5. Analytic approach

Data and missing values. We inspected the distribution and intercorrelations of all variables (available upon request). For replicability purposes, we report all the descriptives based on raw data. For calculating effect sizes as Cohen's *d* for differences between the DLL and monolingual groups and for estimating our multiple regression models, we report pooled results across multiple imputed datasets using Mplus 7 (Muthén & Muthén, 2015) and robust standard errors.

There were ten missing PPVT scores for the DLLs (for $n = 4$ Turkish-German and $n = 6$ Russian-German DLLs) and two missing scores for the monolingual children. There was no difference between the DLLs with missing and those with present PPVT scores for any of the focal predictors or covariates in Table 1 (results can be obtained upon request). Fifty datasets were imputed using the Bayesian Monte-Carlo-Markov-Chain imputation methods implemented in Mplus 7 and a two level estimation which accounts for the variance-covariance of the classroom variables measured only on the between classroom level (Asparouhov & Muthén, 2010). For the imputation model, we entered all variables listed in Table 1 but used the six ITERS-R subscale scores rather than the averaged total score.

Regression analyses. As a main analysis tool, we used multiple regressions with additive linear and interaction terms. We inspected the intercorrelations between all variables and found no high collinearity among predictors. For the first research question (RQ1) addressing the relationship between classroom composition and quality, the classroom was the unit of analysis ($n = 197$). Effects of group composition may only become evident at a certain threshold (Klein & Becker, 2017; Stanat, 2006). Thus, we inspected the association between classroom composition and classroom quality for non-linearity. For the remaining research questions on children's German skills, the child was the unit of analysis ($n = 456$). Children were nested in classrooms. We did not apply multilevel modeling as our data showed a sparse clustered pattern (in 62% of classrooms, there were only one or two target children), which may bias the variance parameters in multilevel modeling for the between and random slope effects (Clarke, 2008). Instead, for all analyses on the child level we used the standard error correction for clustering into classrooms with robust maximum likelihood estimation and the "type = complex" command as implemented in Mplus 7. There was no indication of non-linear associations between classroom composition and individual children's German skills.

For the analysis on children's German skills, we ran several regression models, adding focal predictors and covariates in a blockwise manner. Model 1 included only additive effects of the classroom variables. In model 2, we tested for additional explained variance by adding children's language status (0 = monolingual and 1 = *DLL*). In model 3, we added child and family level covariates. Theoretically, we expected that the proximal child and family level covariates would explain a major portion of variance in children's German skills. Yet, we entered the additive effects of classroom variables first because any change in the main effects from Model 1 to Model 3 would ease uncovering spurious patterns of association for the focal predictors (e.g., an effect of classroom composition might disappear after entering the child level covariates). Next, we included interaction terms that represented the effects of classroom composition differing by DLL status (RQ 2, model 4), and the effects of classroom composition (RQ3, model 5) and DLL status (RQ4, model 6) differing by classroom quality. These were the following interaction terms: Model 4 – *percentage of DLLs* × *DLL status*; model 5 – *percentage of DLLs* × *classroom quality*; model 6 – *DLL status* × *classroom quality*. All continuous predictors were grand-mean centered before calculating the interaction terms. To evaluate the models, we inspected the explained variance (R^2), the sample adjusted Bayesian information criterion (aBIC) and the Akaike information criterion (AIC), with smaller aBIC and AIC values indicating better fit. Effect sizes for the regression were evaluated as small, medium, and large according to Cohen (1992).

Table 2
Linear regression model for the association between classroom composition and classroom quality, controlling for teacher and classroom characteristics ($n = 197$ classrooms), $R^2 = .13$.

	b (SE)	β/*d*	*p*
% DLLs in classroom	−0.62 (.23)	−.20	.007
Children–teacher ratio	−0.07 (.03)	−.22	.014
Teachers' education ≥upper secondary[a]	.04 (.13)	.02	.773
Mixed-age classroom[a]	−.20 (.14)	−.24	.142

Notes. Pooled results based on 50 imputations. *SE* = Standard errors. *p* = two tailed *p*-values. β = standardized coefficients for continuous predictors.

[a] The standardized coefficients for dichotomous predictors were calculated by taking only the variance of the outcome variable into account and may be thus interpreted as *Cohen's d* coefficients.

7. Results

7.1. Descriptives

DLLs and monolingual children differed in several ways (Table 1). DLLs had much lower German skills ($d = 1.41$, 95% CI [1.21, 1.60]) and were enrolled in classrooms with a higher percentage of other DLLs ($d = 1.33$, 95% [1.10, 1.55]). The differences in classroom quality experienced by DLLs and monolingual children were small to moderate in size, with DLLs more likely to be enrolled in classes with lower classroom quality, $d = .43$, 95% CI [0.18, 0.68]. DLLs and monolingual children also differed on most other family level variables.

The average total classroom quality score was moderate ($M = 3.82$, $SD = 0.84$), with most classrooms ranging between 3 ("minimal") and 5 ("good") (Table 1). Classrooms with better quality were attended by fewer DLLs, had a better children-teacher ratio and no mixed age structure (table available upon request; classroom characteristics by language status in Table 1).

7.2. Multiple regression analyses

Classroom quality and classroom composition (RQ1). Fig. 2 illustrates both the linear and smoothed non-linear regression slopes for the association between classroom quality and shares of DLLs. The relationship appeared to be slightly non-linear with a drop in classroom quality around 35–40% DLLs. However, most classrooms in the sample were in the range of 0–20% DLLs. The classrooms with over 40% of DLLs had lower classroom quality scores, but there were only very few classrooms with very high percentages. Overall, both a linear and non-linear slope fit the data equally well. Therefore, we ran several multiple regression analyses comparing linear, quadratic, and cubic effects. Fig. 2 includes information on the children-teacher ratio, as it was significantly negatively related to classroom quality in the analyses. The model including only the linear term fit the data slightly better (Table 2, other models available upon request). The linear model estimated that two otherwise equal classrooms that were 10% apart with regard to enrolled DLLs

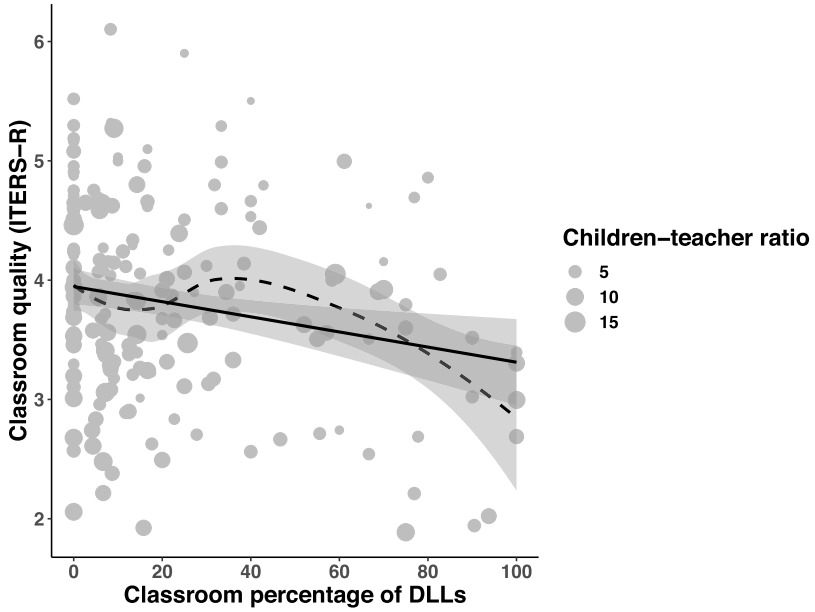

Fig. 2. Bivariate linear and non-linear associations between classroom composition and classroom quality. Data points are sized according to children-teacher-ratios (from 1.4 to 15 children per teacher). The dashed line represents the locally estimated scatterplot smoothing regression fitted slope and the straight line represents the linear fitted slope (with 95% confidence intervals as gray shaded areas). Based on non-imputed data on the classroom level.

(corresponding to about 2–3 DLLs), differed only by 0.06 points on the ITERS-R scale – a very modest difference in observed quality. A somewhat more sizeable difference in quality would be expected only for classrooms that were very far apart with regard to the percentage of DLLs, such as a classroom made up of DLLs entirely and another made up of only monolingual children. Holding everything else constant, these two classrooms would be estimated to be a little over 0.5 points apart on the ITERS-R scale.

Predicting children's German skills (RQ2-4). A series of multiple regressions was conducted to address all further research questions. While not the central focus of our investigation, these regression models provided information on the size of the difference between DLLs' and monolingual children's German skills. Our first model included only the percentage of DLLs, the classroom quality score, and the classroom level covariates described in the method section (Table 3, model 1). This model explained only a small amount of variance (10%). In the second model, we added individual DLL status. This model explained a third of the variance in German skills, and the model fit was strongly improved compared to model 1. The estimated difference between DLLs' and monolingual children's German skills from model 2, adjusting for the classroom variables, was large, equaling 11 words on the 40 point vocabulary test (Table 3, model 2). After entering the child and family covariates, this difference decreased but was still at nearly 9 words (95% CI [6.87, 19.95]), which corresponds to more than one standard deviation ($d = -1.13$, Model 3). Adding the child and family covariates in model 3 explained an additional 12% of variance and further improved fit.

Classroom composition × DLL status (RQ 2). Our second research question concerned whether there was an association between classroom composition and German skills that was stronger for DLLs than for monolingual children. When only accounting for other classroom covariates, a higher percentage of DLLs significantly predicted lower German skills (model 1). However, in model 2, where individual children's DLL status was entered as a predictor, the association between the percentage of DLLs and German skills vanished. Further investigation showed why this was the case: DLLs were much more likely to be enrolled in classrooms with many other DLLs. Neither within the DLL nor the monolingual group was there a connection between composition and German skills. This finding was thus an example of a Simpson's paradox, where associations evident in the total group disappear when examining subgroups separately (or the other way around). Accordingly, there was also no significant interaction between classroom composition and DLL status (RQ 2, model 4), and adding this interaction term did not increase explained variance or improve fit. Thus, the term was not retained in further models. In sum, the connection between classroom composition and German skills was not stronger for DLLs than for monolingual children – as there was no connection at all within either group.

Classroom composition × classroom quality (RQ 3). Our next research question concerned whether the strength of associations between classroom composition and German skills were moderated by classroom quality. Adding an interaction between classroom composition and classroom quality (RQ 3, model 5) did not improve the amount of explained variance or model fit compared to model 3, and the interaction was not a significant predictor. Thus, the interaction term was not retained for the next model. Together with the non-significant main effect of classroom composition, this means that neither in low nor in higher quality classrooms was composition connected to German skills.

Classroom quality × DLL status (RQ 4). Our final research question concerned whether classroom quality was more strongly connected to DLLs' German skills than to those of their monolingual peers. There was a significant interaction between DLL status and classroom quality in the final model 6, which had the best fit and explained an additional 11% of variance when compared to model 3. The significant main effect of DLL status in model 6 showed that at mean classroom quality values, the difference between DLLs and monolingual children was around 9 words. The positive sign of the interaction term indicated that the difference between DLLs and monolingual children declined with increasing classroom quality scores. The interaction effect corresponded to an overall standardized effect of classroom quality of $\beta = .21$; 95% CI [.04, .38] for DLLs and no significant effect for monolingual children.

We further probed the interaction effect with a simple slopes analysis (Preacher, Curran, & Bauer, 2006) and calculated differences in German skills at mean as well as lower and higher

Table 3
Results of the regression models predicting children's German language skills ($n = 456$).

	Model 1		Model 2		Model 3		Model 4 (RQ2)		Model 5 (RQ3)		Model 6 (RQ4)	
	b (SE)	β/*d*	*b (SE)*	β/*d*	*b (SE)*	β/*d*	*b (SE)*	β/*d*	*B (SE)*	β/*d*	*b (SE)*	β/*d*
Classroom variables												
% DLLs in classroom	−9.53 (1.95)	−.29***	.08 (1.65)	.00	.08 (1.50)	.00	.54 (1.71)	.02	.37 (1.59)	.01	.77 (1.45)	.02
Classroom quality (ITERS-R)	.76 (.49)	.08	.55 (.45)	.06	.53 (.40)	.06	.52 (.39)	.06	.39 (.46)	.04	.06 (.43)	.01
Children–teacher ratio	−.04 (.18)	−.01	−.00 (.15)	−.00	.00 (.13)	.00	.00 (.13)	.00	.01 (.13)	.00	.00 (.13)	.00
Mixed age classroom†	−.05 (.96)	−.01	.68 (.71)	.09	.21 (.69)	.03	.20 (.68)	.03	.20 (.68)	.03	.08 (.66)	.01
Teachers' education ≥upper secondary	.87 (.76)	.11	.44 (.66)	.06	.14 (.60)	.02	.15 (.60)	.02	.16 (.60)	.02	.19 (.59)	.02
Child and family variables												
DLL			−11.10 (.98)	−1.41***	−8.91 (1.04)	−1.13***	−8.82 (1.08)	−1.10***	−8.92 (1.04)	−1.13***	−8.64 (1.00)	−1.10***
Age of enrollment ≥18 months†					−1.63 (.66)	−.21	−1.61 (.65)	−.20	−1.59 (.64)	−.20	−1.51 (.65)	−.19
≥30 h/week in childcare†					−.95 (.66)	−.12	−.94 (.66)	−.12	−.94 (.66)	−.12	−1.10 (.66)	−.14
Girl†					.86 (.54)	.11	.86 (.54)	.11	.88 (.54)	.11	.90 (54)	.11
Child's age (months)					.94 (.15)	.24***	.94 (.15)	.24***	.94 (.15)	.24***	.95 (.15)	.24***
No. of children in household <15 years					−.18 (.37)	−.02	−.18 (.37)	−.02	−.19 (.37)	−.02	−.20 (.37)	−.02
Child's health					.38 (.59)	.02	.38 (.59)	.02	.38 (.59)	.02	.38 (.59)	.02
Mother's age (years)					−.05 (.06)	−.03	−.05 (.06)	−.03	−.05 (.06)	−.03	−.03 (.06)	−.02
Parents' education ≥upper secondary†					1.32 (.60)	.17*	1.33 (.60)	.17*	1.30 (.61)	.17*	1.38 (.60)	.18*
Low occupational status†					−.14 (.79)	−.01	−.13 (.79)	−.01	−.13 (.79)	−.01	−.11 (.78)	−.01
Net equivalized household income					1.17 (.88)	.07	1.18 (.89)	.08	1.16 (.88)	.07	.99 (.87)	.06
Difficulty finding daycare/preschool†					.29 (.60)	.02	.30 (.61)	.02	.29 (.60)	.02	.27 (.59)	.02
Lack of help with childcare in family†					−.57 (.69)	−.07	−.59 (.69)	−.07	−.58 (.69)	−.07	−.62 (.69)	−.08
Sensitive learning stimulation					2.17 (.46)	.18***	2.17 (.46)	.18***	2.19 (.46)	.19***	2.29 (.46)	.19***
Organization and opportunities for learning					−.04 (.47)	−.00	−.04 (.47)	−.01	−.03 (.47)	−.00	−.05 (.47)	−.01
Interaction terms												
% DLLs in classroom × DLL							−.96 (3.18)	−.03				
% DLLs in classroom × classroom quality									.69 (1.50)	.03.		
DLL × classroom quality											2.32 (1.07)	.11*
AIC	3142.76		3012.87		2949.71		2951.53		2951.34		2944.37	
aBIC	3149.40		3020.46		2970.58		2973.35		2973.16		2966.19	
R^2	.10		.33		.45		.45		45		.54	

Note. All continuous variables were grandmean centered. Pooled results based on 50 imputations. The robust standard errors were adjusted for clustering within classrooms. Two-tailed *p*-values. Standardized coefficients for predictors marked with a dagger were calculated by taking only the variance of the outcome variable into account and can be interpreted as *Cohen's d* coefficients. There were no differences between the *p*-values calculated for the unstandardized and standardized coefficients.

* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

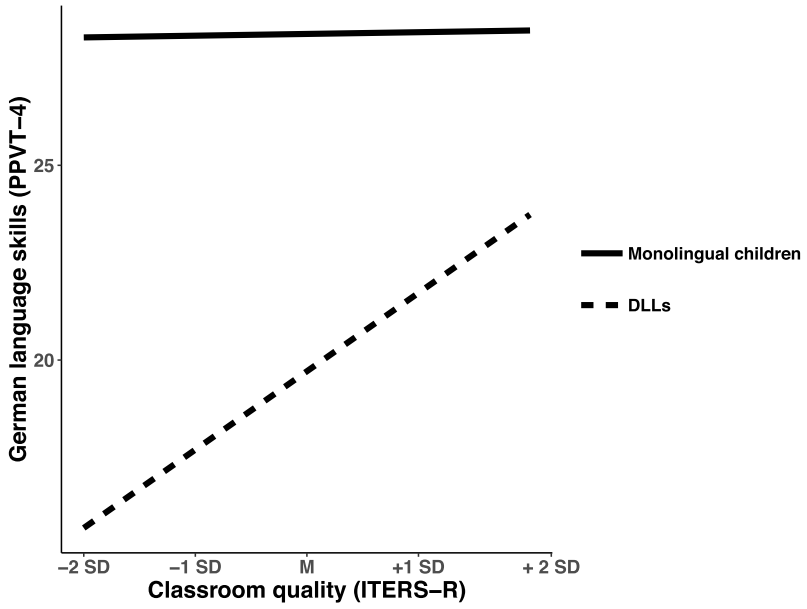

Fig. 3. Simple slope analysis of the interaction between dual language learner (DLL) status and classroom quality on children's German language skills.

classroom quality (mean ± 1 *SD* and ± 2 *SD*). Fig. 3 shows that the estimated difference between DLLs and monolingual children was 10.5 points ($p < .001$) for children attending classrooms with "minimal" quality (at mean −1*SD*), and 12.4 points ($p < .001$) at "inadequate to minimal" quality (mean −2 *SD*), the latter equaling around 1.5 *SD* in German skills. For children in classrooms with "minimal to good" quality (mean + 1 *SD*), the difference decreased to only 6.8 points ($p < .001$). Finally, for children in classrooms with "good" levels of quality (at mean + 2 *SD*), the difference decreased to 4.9 points ($p = .017$), corresponding to around 0.5 *SD*.

8. Discussion

Understanding how the classroom context is connected to majority language vocabulary is essential for supporting DLLs. To summarize our findings, children in classrooms with a higher percentage of DLLs experienced lower levels of overall quality (RQ1). Turning to children's language, when adjusting for a host of child, family, and classroom covariates, the very young DLLs had much lower German skills than their monolingual peers. This difference was not dependent on classroom composition in terms of the percentage of DLLs (RQ2). In fact, classroom composition was connected to children's German skills only when all child and family level covariates were disregarded. Accordingly, classroom quality did not moderate any association between classroom composition and German skills (RQ3). However, the difference between DLLs' and monolingual children's German skills was smaller for children who experienced higher classroom quality (RQ4).

8.1. Classroom composition and classroom quality (RQ1)

Our findings from a nationally representative German sample of classrooms indicated that classrooms with higher percentages of DLLs were of slightly lower overall quality than classrooms with fewer DLLs. This is in line with a recent German study showing that children from immigrant families are also more often enrolled in classrooms with lower structural quality (Stahl, Schober, & Spiess, 2018). Interestingly, there appeared to be a "drop" in quality for classrooms with more than 40% of DLLs. This is consistent with previous findings from Germany suggesting a threshold for negative effects of composition in this percentage range (Klein & Becker, 2017; Stanat, 2006). Our result contrasts with those studies from the United States that reveal similar levels of quality experienced by DLLs and monolingual children. However, comparisons are difficult because the studies pertain to different age groups and use inconsistent measures of DLL status (Espinosa et al., 2017; Karoly, Dastidar, Zellman, Perlman, & Fernyhough, 2008; Peisner-Feinberg et al., 2014).

The extant literature points to several processes that may contribute to the negative association between composition and classroom quality. For example, immigrant parents of DLLs may be selecting lower quality centers because they have less information about the early childhood education system than other parents (Becker, 2010; Becker & Biedinger, 2016). Moreover, centers with a higher percentage of DLLs may attract less effective teachers (Reid, Kagan, Hilton, & Potter, 2015). Another possible explanation is that there are bidirectional effects between composition and quality. Future studies may weigh different explanations against each other.

8.2. Classroom composition and majority language skills (RQ2 and RQ3)

When only adjusting for classroom covariates, composition in terms of a higher percentage of DLLs was related to lower German skills. We interpreted this as a methodological artifact (i.e., a case of the so-called Simpson's Paradox) which arose because DLLs were likely to attend classrooms together with many other DLLs. When children's individual language status was accounted for, classroom composition was no longer associated with children's German skills. As this result held true for both DLLs and monolingual children, it indicated that DLLs were not more sensitive to classroom composition. Moreover, it gave us no reason to believe that children's majority language learning was more effective with monolingual than with DLL peers. There was also no indication that children experiencing different levels of quality were differentially sensitive to classroom composition. In sum, our investigation did not suggest that classroom composition was related to children's German skills, neither in low nor in high quality classrooms and neither for monolingual children nor for DLLs.

The non-association between classroom composition and German skills contrasts findings from a study in the United States

(Garcia, 2018). Our sample was very young at not yet three years old. Verbal communication with peers had perhaps just begun within the last year of life and might have still been infrequent (Eckerman, Davis, & Didow, 1989). Moreover, composition effects may develop only after spending an extended period of time in a classroom. On average, the DLLs in our sample had been exposed to early childhood education for less than a year and the monolingual children for a little over a year. However, two previous studies also failed to uncover any effects of classroom composition in terms of the percentage of DLLs on somewhat older children (Hindman & Wasik, 2015), including children who had attended for several years (Kohl et al., 2019).

The lack of association between composition and German skills may also be connected to methodological issues. There were only few classrooms with over 70–80% of DLLs in our sample. Further studies with larger samples of classrooms will provide a wider range of percentages and allow simultaneously considering composition in terms of DLL status, immigrant status, ethnicity, and SES. Finally, further studies should also gather information about the actual use of the majority language (Gámez, Griskell, Sobrevilla, & Vazquez, 2019) in classrooms and in individual children's friendship networks. Parents and policy-makers will often have to base their decisions on very general classifications (such as DLL vs. monolingual child). However, more proximal measures of children's actual interactions are presumably much more powerful predictors of their language skills.

8.3. Classroom quality and majority language skills (RQ4)

Classroom quality was connected only to DLLs' but not to monolingual children's German skills. DLLs in classrooms with higher quality demonstrated German skills that were much more similar to those of monolingual children. This suggests that high classroom quality can indeed help support DLLs' early majority language skills. This conclusion is in accord with several previous findings for DLLs and monolingual children suggesting that high classroom quality is related to smaller differences in DLLs' majority language skills (Ebert et al., 2013; Kohl et al., 2019; Sonnenschein et al., 2013). These studies also found that monolingual children in higher quality classrooms did not outperform their peers in lower quality classrooms. Monolingual children, who on average have more experience with the majority language, may only start benefiting from higher levels of quality than those observed in the three studies. There is previous evidence for such non-linear effects of classroom quality (Burchinal et al., 2011)

That classroom quality supports DLLs' majority language skills is only one possible interpretation of our study. An alternative interpretation is that parents of DLLs with stronger German language skills selected higher quality centers for their children. We controlled for eleven family level covariates, among them three SES related measures as well as observations of the home environment, to rule out that they were responsible for any connection between classroom quality and German language skills. NICHD ECCRN and Duncan (2003) suggest that for quality effects on cognitive outcomes, even a small set of important selection factors including parental education can remove a major portion of the bias. Moreover, a strong bias due to self-selection into centers of varying quality appears implausible for two reasons: Firstly, at the time of data collection, there was a serious shortage of extraparental care options for toddlers (Tietze et al., 2013) and parents are likely to have been compelled to accept any opening they could find. Secondly, parents' judgments of quality diverge from observed levels of quality (Cryer et al., 2002; Cryer & Burchinal, 1997). Parents of DLLs with higher German skills may still have been attracted to centers of higher quality without intentionally selecting such centers. In that case, it is not immediately clear why such forces might have differentially affected parents of DLLs and parents of monolinguals. In sum, we would conclude that we cannot rule out that the effect size connecting classroom quality and DLLs' German skills is biased to some degree. Still, we would argue that our extensive covariates together with the parents' limited options in intentionally selecting high quality care make it unlikely that the effect is completely spurious.

If there was indeed an effect of classroom quality on DLLs, we can only speculate on the exact pathways of the connection between classroom quality and majority language skills. This highlights the need for studies more closely tapping classroom quality in terms of the classroom language environment. For example, teachers' use of complex language, teachers' facilitation of conversations, and teachers' language use during shared book reading appear to be related to preschoolers' majority language skills (Gámez, Neugebauer, Coyne, McCoach, & Ware, 2017; Gómez, Vasilyeva, & Dulaney, 2017; Justice, Jiang, & Strasser, 2018). In-service professional development in the form of coaching may enable teachers to provide higher quality classroom language environments (Egert, Fukkink, & Eckhardt, 2018). However, for DLLs, research is still very much in the process of discovering what exactly constitutes high classroom quality (Peisner-Feinberg et al., 2014). Castro, Páez, Dickinson, and Frede (2011) propose that DLLs may benefit from very specific supports: bilingual teachers who value the heritage language, who use the heritage language to teach majority language vocabulary, who use gestures and visual aids, and who encourage frequent small group or one-to-one interactions.

8.4. Limitations

First, we analyzed cross-sectional data and causal interpretations should be made very cautiously. The statistical effects associated with DLL status or classroom quality may have been affected by so-called endogeneity issues (Duncan, Magnuson, & Ludwig, 2004). The true direction of causality is unclear, and spuriousness cannot be fully ruled out. Longitudinal studies with larger samples may better control for children's selection into centers of varying quality. Moreover, considering that selection factors are difficult to assess in practice, these measurement shortcomings could be addressed with longitudinal data by employing fixed effects models which can account for unmeasured time invariant heterogeneity (Allison, 2009).

Second, we did not include DLLs' heritage language skills as a predictor of their German skills. Cross-linguistic transfer is thought to aid children's oral language development, but there is evidence that such effects are relatively small and that transfer is only one of many influential factors (Melby-Lervåg & Lervåg, 2011). The internal validity of our findings on classroom quality might be threatened if, for example, parents of children with higher heritage language skills were more likely to enroll their children in higher quality classrooms. While this seems to be in the realm of the possible, in Germany, children of less educated parents have been found to have stronger heritage language skills (Willard, Agache, Jäkel, Glück, & Leyendecker, 2015).

Third, while our sample is representative on the regional level, the unbalanced sample size between the DLL and monolingual group may lead to an overestimation of moderation effects. Future larger studies sampling more children per classroom will also provide better possibilities for multilevel analyses. Such analyses would allow a much clearer estimation of possible main and cross-level effects for contextual classroom and quality factors, e.g., by accommodating for varying differences between DLLs and monolingual children within each analyzed classroom.

Lastly, before generalizing our results, one should reconsider the specific characteristics of both DLLs in Germany and the German early childhood education system. The non-effect of classroom

composition may not be generalizable to contexts in which many DLLs share the same heritage language, such as the United States where Spanish is very dominant (Pew Research Center, 2017). Moreover, the connection between classroom quality and DLLs' majority language skills may not be generalizable to countries such as Canada where immigrants are likely to have a relatively high SES (Statistics Canada, 2017).

8.5. Implications

For policy-makers, one pressing issue is how to best support DLLs' majority language development in order to smooth their path to academic success. The current early childhood education system may actually be inadvertently widening differences between DLLs' and monolingual children's majority language skills instead of narrowing them. Our findings suggest that DLLs from immigrant families are more likely to profit from higher classroom quality than other children are. Yet, these DLLs were likely to attend classrooms of lower quality. In addition, other studies show that immigrant DLLs tend to enter early childhood education at an older age (e.g., Tietze et al., 2013), which further limits their chances of being exposed to high quality education early on.

Our study suggests that targeting classroom composition is not a principal starting point for supporting DLLs. Instead, investing into classroom quality, for example by increasing funds and improving teacher education, is likely to have positive effects. To create effects that are also long-lasting, it does not suffice to provide high quality to very young children, but throughout children's educational trajectories (Brooks-Gunn, 2003; Li, Farkas, Duncan, Burchinal, & Vandell, 2013). Parents have great difficulty judging classroom quality (Cryer et al., 2002; Cryer & Burchinal, 1997). In Germany, even if they strive for the best care, many parents must accept any opening they can find. If parents have difficulties in recognizing high quality and they have few options, there is no apparent reason for child care providers to invest in quality above minimal standards (Camehl, Schober, & Spiess, 2018). This suggests that it is necessary for policy makers to push for the setting of very high quality standards and to provide the means for early childhood education centers to attain these standards.

Finally, to truly support the harmonious development of two languages, research is urgently needed to understand how early childhood education can effectively support the many different heritage languages represented in linguistically diverse classrooms. High classroom quality may stimulate not only the majority but also the heritage language (Hindman & Wasik, 2015). Nurturing the heritage language may be especially vital for children who enter the early education system, and thus begin experiencing the powerful draw of the majority language, at a very young age.

Acknowledgement

The NUBBEK project was supported by the German Ministry of Family Affairs, the Jacobs Foundation, the Bosch Foundation and the Family Ministries of several German Federal States. We would like to thank all NUBBEK partners and the participating families. Research for this paper was also supported by NORFACE grant #292.

References

Allison, P. D. (2009). *Fixed effects regression models*. Los Angeles: Sage Publications.

Alt, C., Berngruber, A., & Pötter, U. (2016). Wer bemüht sich um einen Kitaplatz und wer nimmt ihn in Anspruch? Ein Vergleich zwischen Migranten- und autochthonen Familien mit Kindern unter drei Jahren Who seeks a Kita-spot and who enrolls? A comparison between immigrant and autochthonous families with children under three. *Zeitschrift für Pädagogik*, *5*, 690–706.

Asparouhov, T., & Muthén, B. O. (2010). *Bayesian analysis using Mplus: Technical implementation (Mplus Technical Report)*.. Retrieved from http://statmodel.com/download/Bayes3.pdf

Becker, B. (2010). Ethnische Unterschiede bei der Kindergartenselektion: Die Wahl von unterschiedlich stark segregierten Kindergärten in deutschen und türkischen Familien, [Ethnic differences in preschool choice – Turkish and German families choose preschools with differing levels of segregation]. In B. Becker, & D. Reimer (Eds.), *Vom Kindergarten bis zur Hochschule: Die Generierung von ethnischen und sozialen Disparitäten in der Bildungsbiographie [From preschool to college - The generation of ethnic and social disparities in educational trajectories]* (pp. 17–47). Wiesbaden: VS Verlag für Sozialwissenschaften.

Becker, B., & Biedinger, N. (2016). Ethnische Ungleichheiten in der vorschulischen Bildung [Ethnic disparities in preschool education]. In C. Diehl, C. Hunkler, & C. Kristen (Eds.), *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten* (pp. 433–474). Wiesbaden: Springer.

Berendes, K., Weinert, S., Zimmermann, S., & Artelt, C. (2013). Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online*, *5*(2), 15–49.

Bihler, L.-M., Agache, A., Schneller, K., Willard, J. A., & Leyendecker, B. (2018). Expressive morphological skills of dual language learning and monolingual German children: Exploring links to duration of preschool attendance, classroom quality, and classroom composition. *Frontiers in Psychology*, *9*, 888. http://dx.doi.org/10.3389/fpsyg.2018.00888

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476.

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: Causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–284. http://dx.doi.org/10.1037/a0024448

Brehmer, B. (2007). Sprechen Sie Qwelja? Formen und Folgen russisch-deutscher Zweisprachigkeit in Deutschland [Do you speak Qwelja? Russian-German bilingualism in Germany]. In T. Anstatt (Ed.), *Mehrsprachigkeit bei Kindern und Erwachsenen: Erwerb, Formen, Förderung* (pp. 163–185). Tübingen: Gunter Narr.

Brooks-Gunn, J. (2003). Do you believe in magic? What can we expect from early childhood intervention programs. *SRCD Social Policy Report*, *17*(1).

Brunsek, A., Perlman, M., Falenchuk, O., McMullen, E., Fletcher, B., & Shah, P. S. (2017). The relationship between the Early Childhood Environment Rating Scale and its revised form and child outcomes: A systematic review and meta-analysis. *PLos ONE*, *12*(6), 1–29. http://dx.doi.org/10.1371/journal.pone.0178512

Bundesministerium für Familie, Senioren, Frauen und Jugend. (2018). *Kindertagesbetreuung Kompakt: Ausbaustand und Bedarf 2017*.. Retrieved from http://www.fruehe-chancen.de/fileadmin/PDF/Fruehe_Chancen/Betreuungszahlen/Kita_Kompakt_Dritte_Ausgabe-BF.PDF

Burchinal, M., Kainz, K., & Cai, Y. (2011). How well do our measures of quality predict child outcomes? A meta-analysis and coordinated analysis of data from large-scale studies of early childhood settings. In M. Zaslow, I. Martinez-Beck, K. Tout, & T. Halle (Eds.), *Quality measurement in early childhood settings* (pp. 11–31). Baltimore: Brookes.

Buysse, V., Peisner-Feinberg, E., Páez, M., Hammer, C. S., & Knowles, M. (2014). Effects of early education programs and practices on the development and learning of dual language learners: A review of the literature. *Early Childhood Research Quarterly*, *29*(4), 765–785. http://dx.doi.org/10.1016/j.ecresq.2013.08.004

Caldwell, B. M., & Bradley, R. H. (2001). *HOME inventory and administration manual* (3rd ed.). Little Rock: University of Arkansas for Medical Sciences and University of Arkansas at Little Rock.

Camehl, G. F., Schober, P. S., & Spiess, C. K. (2018). Information asymmetries between parents and educators in German childcare institutions. *Education Economics*, *26*(6), 624–646. http://dx.doi.org/10.1080/09645292.2018.1463358

Castro, D. C., Páez, M. M., Dickinson, D. K., & Frede, E. (2011). Promoting language and literacy in young dual language learners: Research, practice, and policy. *Child Development Perspectives*, *5*(1), 15–21.

Clarke, P. (2008). When can group level clustering be ignored? Multilevel models versus single-level models with sparse data. *Journal of Epidemiology and Community Health*, *62*(8), 752–758.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Cryer, D., & Burchinal, M. (1997). Parents as child care consumers. *Early Childhood Research Quarterly*, *12*(1), 35–58. http://dx.doi.org/10.1016/S0885-2006(97)90042-9

Cryer, D., Tietze, W., & Wessels, H. (2002). Parents' perceptions of their children's child care: A cross-national comparison. *Early Childhood Research Quarterly*, *17*(2), 259–277.

De Houwer, A. (2015). Harmonious bilingual development: Young families' well-being in language contact situations. *International Journal of Bilingualism*, *19*(2), 169–184. http://dx.doi.org/10.1177/1367006913489202

Döge, P., Weyer, E., Resa, E., Eckhardt, A. G., Lee, H.-J., Agache, A., . . . & Spieß, C. K. (2013). Untersuchungsanlage [Design]. In W. Tietze, F. Becker-Stoll, J. Bensel, A. G. Eckhardt, G. Haug-Schnabel, B. Kalicki, & B. Leyendecker (Eds.), *Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit (NUBBEK)*. Weimar: verlag das netz.

Duncan, G. J., Magnuson, K. A., & Ludwig, J. (2004). The endogeneity problem in developmental studies. *Research in Human Development*, *1*(1–2), 59–80.

Dunn, L. M., & Dunn, D. M. (2007). *PPVT4: Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: NCS Pearson.

Ebert, S., Lockl, K., Weinert, S., Anders, Y., Kluczniok, K., & Rossbach, H.-G. (2013). Internal and external influences on vocabulary development in preschool children. *School Effectiveness and School Improvement*, *24*(2), 138–154.

Eckerman, C. O., Davis, C. C., & Didow, S. M. (1989). Toddlers' emerging ways of achieving social coordinations with a peer. *Child Development*, *60*(2), 440–453.

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Review of Educational Research*, *88*(3), 401–433. http://dx.doi.org/10.3102/0034654317751918

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *The British Journal of Sociology*, *30*(4), 415–441.

Espinosa, L. M., LaForett, D. R., Burchinal, M., Winsler, A., Tien, H.-C., Peisner-Feinberg, E. S., & Castro, D. C. (2017). Child care experiences among dual language learners in the United States. *AERA Open*, *3*(2), 1–15.

Förster, M. F. (1994). *Measurement of low incomes and poverty in a perspective of international comparisons (OECD labour market and social policy occasional paper, No. 14). Paris.*. Retrieved from http://www.oecd.org/dataoecd/45/58/1895548.pdf

Frigelj, K. (2018, August). (*Islamization of preschools? "Would put it differently today"*) Islamisierung in der Kita? "Würde es heute anders formulieren". In *Welt*. Retrieved from https://www.welt.de/vermischtes/article181309098/Kita-Beschwerde-wegen-vieler-Migranten-Ich-wuerde-es-heute-noch-einmal-anders-formulieren.html

Frühe Mehrsprachigkeit an Kitas und Schulen e.V. (2019). *Bilinguale Kindergärten*. Retrieved from http://www.fmks-online.de/bilikitas.html

Gambaro, L. (2017). Children from migrant backgrounds: Who are their Kita peers? *DIW Economic Bulletin*, *51+52*, 559–566.

Gámez, P. B., Griskell, H. L., Sobrevilla, Y. N., & Vazquez, M. (2019). Dual language and English-only learners' expressive and receptive language skills and exposure to peers' language. *Child Development*, *90*, 471–479. http://dx.doi.org/10.1111/cdev.13197

Gámez, P. B., Neugebauer, S. R., Coyne, M. D., McCoach, D. B., & Ware, S. (2017). Linguistic and social cues for vocabulary learning in Dual Language Learners and their English-only peers. *Early Childhood Research Quarterly*, *40*, 25–37. http://dx.doi.org/10.1016/j.ecresq.2017.01.003

Garcia, E. B. (2018). The classroom language context and English and Spanish vocabulary development among dual language learners attending Head Start. *Early Childhood Research Quarterly*, *42*, 148–157. http://dx.doi.org/10.1016/j.ecresq.2017.09.005

Glück, C. W. (2009). *Receptive Vocabulary Test Research Version Modeled on PPVT-4th Edition for NUBBEK*. Heidelberg, Germany: Unpublished Research Version.

Gómez, L. E., Vasilyeva, M., & Dulaney, A. (2017). Preschool teachersó read-aloud practices in Chile as predictors of children's vocabulary. *Journal of Applied Developmental Psychology*, *52*, 149–158. http://dx.doi.org/10.1016/j.appdev.2017.07.005

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. http://dx.doi.org/10.1016/0093-934X(89)90048-5

Hammer, C. S., Hoff, E., Uchikoshi, Y., Gillanders, C., Castro, D., & Sandilos, L. E. (2014). The language and literacy development of young dual language learners: A critical review. *Early Childhood Research Quarterly*, *29*(4), 715–733.

Harms, T., Cryer, D., & Clifford, R. M. (2005). *Infant/Toddler Environment Rating Scale – Revised Edition (ITERS-R)*. New York: Teachers College Press.

Hindman, A. H., & Wasik, B. A. (2015). Building vocabulary in two languages: An examination of Spanish-speaking dual language learners in Head Start. *Early Childhood Research Quarterly*, *31*, 19–33. http://dx.doi.org/10.1016/j.ecresq.2014.12.006

Hoff, E. (2018). Bilingual development in children of immigrant families. *Child Development Perspectives*, *12*(2), 80–86. http://dx.doi.org/10.1111/cdep.12262

Justice, L. M., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children's language growth? *Early Childhood Research Quarterly*, *42*, 79–92. http://dx.doi.org/10.1016/j.ecresq.2017.09.003

Justice, L. M., Petscher, Y., Schatschneider, C., & Mashburn, A. (2011). Peer effects in preschool classrooms: Is children's language growth associated with their classmates' skills? *Child Development*, *82*(6), 1768–1777. http://dx.doi.org/10.1111/j.1467-8624.2011.01665.x

Karoly, L. A., Dastidar, M. G., Zellman, G. L., Perlman, M., & Fernyhough, L. (2008). *Prepared to learn: The nature and quality of early care and education for preschool-age children in California*. Santa Monica, CA: RAND Corporation.

Klein, O., & Becker, B. (2017). Preschools as language learning environments for children of immigrants. Differential effects by familial language use across different preschool contexts. *Research in Social Stratification and Mobility*, *48*, 20–31.

Kohl, K., Willard, J. A., Agache, A., Bihler, L.-M., & Leyendecker, B. (2019). Child care quality, classroom composition, and age at entry: Child care experiences and single and dual language learners' German skills. *AERA Open*, *5*(1), 1–16. http://dx.doi.org/10.1177/2332858419832513

Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., & Will, G. (2011). The education of migrants and their children across the life course. *Zeitschrift für Erziehungswissenschaft*, *14*(2), 121–137. http://dx.doi.org/10.1007/s11618-011-0194-3

Kuger, S., Klucziniok, K., Kaplan, D., & Rossbach, H.-G. (2016). Stability and patterns of classroom quality in German early childhood education and care. *School Effectiveness and School Improvement*, *27*(3), 418–440.

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics*, *32*(01), 69–92.

Leyendecker, B., Agache, A., & Madsen, S. (2014). NUBBEK – a national German study on early childhood education and care: Design, methods, overview, data access, and the potential for multilevel analyses. *Zeitschrift für Familienforschung*, *26*(2), 244–258.

Li, W., Farkas, G., Duncan, G. J., Burchinal, M. R., & Vandell, D. L. (2013). Timing of high-quality child care and cognitive, language, and preacademic development. *Developmental Psychology*, *49*(8), 1440.

Linberg, T., Schneider, T., Waldfogel, J., & Wang, Y. (2019). Socioeconomic status gaps in child cognitive development in Germany and the United States. *Social Science Research*, *79*, 1–31. http://dx.doi.org/10.1016/j.ssresearch.2018.11.002

Mashburn, A. J., Justice, L. M., Downer, J. T., & Pianta, R. C. (2009). Peer effects on children's language achievement during pre-kindergarten. *Child Development*, *80*(3), 686–702. http://dx.doi.org/10.1111/j.1467-8624.2009.01291.x

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114–135.

Melhuish, E. C., Phan, M. B., Sylva, K., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2008). Effects of the home learning environment and preschool center experience upon literacy and numeracy development in early primary school. *Journal of Social Issues*, *64*(1), 95–114.

Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.

NICHD ECCRN. (2002). Early child care and children's development prior to school entry: Results from the NICHD study of early child care. *American Educational Research Journal*, *39*(1), 133–164.

NICHD ECCRN, & Duncan, G. J. (2003). Modeling the impacts of child care quality on children's preschool cognitive development. *Child Development*, *74*(5), 1454–1475.

Organisation for Economic Co-operation and Development. (2018). *The resilience of students with an immigrant background: Factors that shape well-being*. Paris: OECD Publishing.

Palermo, F., & Mikulski, A. M. (2014). The role of positive peer interactions and English exposure in Spanish-speaking preschoolers' English vocabulary and letter-word skills. *Early Childhood Research Quarterly*, *29*(4), 625–635.

Palermo, F., Mikulski, A. M., Fabes, R. A., Hanish, L. D., Martin, C. L., & Stargel, L. E. (2014). English exposure in the home and classroom: Predictions to Spanish-speaking preschoolers' English vocabulary skills. *Applied Psycholinguistics*, *35*(6), 1163–1187.

Peisner-Feinberg, E., Buysse, V., Fuligni, A., Burchinal, M., Espinosa, L., Halle, T., & Castro, D. C. (2014). Using early care and education quality measures with dual language learners: A review of the research. *Early Childhood Research Quarterly*, *29*(4), 786–803.

Peisner-Feinberg, E. S., Burchinal, M. R., Clifford, R. M., Culkin, M. L., Howes, C., Kagan, S. L., & Yazejian, N. (2001). The relation of preschool child-care quality to children's cognitive and social developmental trajectories through second grade. *Child Development*, *72*(5), 1534–1553.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783. http://dx.doi.org/10.1037/0022-3514.90.5.751

Pew Research Center. (2017). *Facts on U.S. Latinos, 2015: Statistical portrait of Hispanics in the United States.*. Retrieved from http://www.pewhispanic.org/2017/09/18/facts-on-u-s-latinos-current-data/

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Education and Behavioral Statistics*, *31*(4), 437–448.

Prevoo, M. J. L., Malda, M., Mesman, J., & van IJzendoorn, M. H. (2016). Within- and cross-language relations between oral language proficiency and school outcomes in bilingual children with an immigrant background: A meta-analytical study. *Review of Educational Research*, *86*(1), 237–276.

Reid, J. L., Kagan, S. L., Hilton, M., & Potter, H. (2015). *A better start: Why classroom diversity matters in early education.*. Retrieved from http://www.prrac.org/pdf/A_Better_Start.pdf

Schneider, S. L. (Ed.). (2008). *The international standard classification of education: An evaluation of content and criterion validity for 15 European countries*. Mannheim: MZES.

Schober, P. S., & Spiess, C. K. (2013). Early childhood education activities and care arrangements of disadvantaged children in Germany. *Child Indicators Research*, *6*(4), 709–735.

Sonnenschein, S., Thompson, J. A., Metzger, S. R., & Baker, L. (2013). Relations between preschool teachers' language and gains in low income English language learners' and English speakers' vocabulary, early literacy and math skills. *NHSA Dialog*, *16*(4), 64–87.

Stahl, J. F., Schober, P. S., & Spiess, C. K. (2018). Parental socio-economic status and childcare quality: Early inequalities in educational opportunity? *Early Childhood Research Quarterly*, *44*, 304–317.

Stanat, P. (2006). Schulleistungen von Jugendlichen mit Migrationshintergrund: Die Rolle der Zusammensetzung der Schülerschaft [Academic achievement of youths with an immigrant background – The role of composition]. In J. Baumert, P. Stanat, & R. Watermann (Eds.), *Herkunftsbedingte Disparitäten im Bildungswesen: Differenzielle Bildungsprozesse und Probleme der Verteilungsgerechtigkeit [Educational disparities related to background]* (pp. 189–219). Wiesbaden: VS Verlag für Sozialwissenschaften.

Statistics Canada. (2017). *Education in Canada: Key results from the 2016 census.*. Retrieved from https://www150.statcan.gc.ca/n1/daily-quotidien/171129/dq171129a-eng.htm

Statistisches Bundesamt. (2016). *Bevölkerung und Erwerbstätigkeit: Bevölkerung mit Migrationshintergrund. Ergebnisse des Mikrozensus 2015.*. Retrieved from https://www.destatis.de/DE/Publikationen/Thematisch/Bevoelkerung/MigrationIntegration/Migrationshintergrund2010220157004.pdf?__blob=publicationFile

Statistisches Bundesamt. (2018). *Statistiken der Kinder- und Jugendhilfe: Kinder und tätige Personen in Tageseinrichtungen und in öffentlich geförderter Kindertagespflege am 01.03.2018.*. Retrieved from https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Soziales/Kindertagesbetreuung/Publikationen/Downloads-Kindertagesbetreuung/tageseinrichtungen-kindertagespflege-5225402187004.pdf?__blob=publicationFile&v=6

Suchodoletz, A., Fäsche, A., Gunzenhauser, C., & Hamre, B. K. (2014). A typical morning in preschool: Observations of teacher–child interactions in German preschools. *Early Childhood Research Quarterly*, *29*(4), 509–519. http://dx.doi.org/10.1016/j.ecresq.2014.05.010

Sylva, K., Taggart, B., Siraj-Blatchford, I., Totsika, V., Ereky-Stevens, K., Gilden, R., & Bell, D. (2007). Curricular quality and day-to-day learning activities in pre-school. *International Journal of Early Years Education*, *15*(1), 49–65.

Tietze, W. (2010). *Krippen-Skala (KRIPS-R)*. Berlin: Forschungsversion.

Tietze, W., Becker-Stoll, F., Bensel, J., Eckhardt, A. G., Haug-Schnabel, G., Kalicki, B., . . . & Leyendecker, B. (2013). *Nationale Untersuchung zur Bildung, Betreuung und Erziehung in der frühen Kindheit (NUBBEK)*. Weimar: verlag das netz.

Tietze, W., Becker-Stoll, F., Bensel, J., Haug-Schnabel, G., Kalicki, B., Keller, H., & Leyendecker, B. (2015). *NUBBEK – National Survey on Education, Care, and Development in Early Childhood*. Köln: GESIS Datenarchiv. Retrieved from https://doi.org/10.4232/1.12297

Totsika, V., & Sylva, K. (2004). The home observation for measurement of the environment revisited. *Child and Adolescent Mental Health*, *9*(1), 25–35.

UNICEF Office of Research. (2018). *An unfair start: Inequality in children's education in rich countries (Innocenti Report Card no. 15). Florence.*

United Nations Department of Economic and Social Affairs. (2017). *International migrant stock: The 2017 revision. Table 3. International migrant stock as percentage of the total population by age and sex and by major area, region, country or area, 1990–2017.*. Retrieved from http://www.un.org/en/development/desa/population/migration/data/estimates2/estimates17.shtml

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis, & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 103–121). Washington, DC: De Gruyter.

Wandt, C. (2018). *Jedes dritte Kita-Kind in Essen spricht kaum Deutsch [Every third preschooler in Essen barely speaks German]*. WAZ. January 31, Retrieved from https://www.waz.de/staedte/essen/jedes-dritte-kita-kind-in-essen-spricht-kaum-deutsch-id213275233.html

Willard, J. A., Agache, A., Jäkel, J., Glück, C. W., & Leyendecker, B. (2015). Family factors predicting vocabulary in Turkish as a heritage language. *Applied Psycholinguistics*, *36*(04), 875–898. http://dx.doi.org/10.1017/S0142716413000544