

BILEX: A new tool measuring bilingual children's lexicons and translational equivalents

First Language

2018, Vol. 38(3) 263–283

© The Author(s) 2017

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0142723717736450

journals.sagepub.com/home/fla

**Anja Gampe, Ira Kurthen
and Moritz M. Daum**

University of Zurich, Switzerland

Abstract

The current study describes the development and validation of a novel scale (BILEX) designed to assess young bilingual children's receptive vocabulary in both languages, their conceptual vocabulary, and translational equivalents. BILEX was developed to facilitate the assessment of vocabulary size for both of the children's languages within one session without any transfer from one language to the other. One-hundred-and-eighty-two 3-year-old children participated in the studies of reliability and validity. Psychometric properties have very good consistency and reliability, along with good concurrent, construct, and criteria validity.

Keywords

Assessment, bilinguals, language acquisition, lexicon, translational equivalents

Learning two or more languages simultaneously from birth influences how children perceive, learn about, and interact within the world in comparison to monolingual children. Advantages for bilinguals have been reported in the domains of language and communication (Bosch & Sebastián-Gallés, 1997; Davidson, Jergovic, Imami, & Theodos, 1997; Fennell, Byers-Heinlein, & Werker, 2007; Ferjan Ramírez, Ramírez, Clarke, Taulu, & Kuhl, 2017; Wermelinger, Gampe, & Daum, 2017), as well as in social competence and cognition (Adesope, Lavin, Thompson, & Ungerleider, 2010; Bialystok, 2001; Kovács, 2009; Yow & Markman, 2011). The mechanisms behind these overt differences in behavior are currently being investigated, with some studies pointing towards associated

Corresponding author:

Anja Gampe, Department of Psychology, University of Zurich, Binzmühlestr 14, Box 21, 8050 Zurich, Switzerland.

Email: a.gampe@psychologie.uzh.ch

differences in brain structure (see Bialystok, 2017 for an overview; Burgaleta, Sanjuán, Ventura-Campos, Sebastian-Galles, & Ávila, 2016). But research comparing performance between monolinguals and bilinguals is still challenging because many potentially influencing covariates and moderators of overt behavior are still unknown or under-investigated. The target groups of monolingual and bilingual children might, for example, differ in terms of socio-economic status, language use, general cognitive abilities, or other variables which have yet to be identified (Chiat & Polišenská, 2016; Mueller Gathercole et al., 2016; Wood, Diehm, & Callender, 2016). But even beyond this, groups of bilingual children in different studies might differ in terms of their cultural background, language similarity, language exposure, and balanced language status (Marton, 2016; Thomas-Sunesson, Hakuta, & Bialystok, 2018). All these factors contribute to the mechanisms of bilingualism and its apparent advantages.

One important example of these possible covariates between monolingual and bilingual children is language skills. Recent research shows converging evidence for a monolingual advantage in lexicon acquisition (Bialystok & Luk, 2012; Bialystok, Luk, Peets, & Yang, 2010; Umbel, Pearson, Fernandez, & Oller, 1992). However, this advantage is observed only when bilinguals and monolinguals are compared regarding lexical skills in the monolinguals' language exclusively. When combining the lexicons of both of the bilinguals' languages, bilinguals and monolinguals have at least similar vocabularies (Bedore, Peña, García, & Cortez, 2005; De Houwer, Bornstein, & Putnick, 2014; Hoff et al., 2012; Pearson, Fernández, & Oller, 1993). These findings nicely illustrate the opposition between lexical advantages depending on which vocabulary is been analyzed. Furthermore, they underline the importance of assessing lexical skills in a way that meets the preconditions of the bilingual group, namely assessing lexical skills in both of the child's languages. Assessing both languages enables not only the child's total vocabulary (the number of words in both languages combined) to be measured, but also their conceptual vocabulary (the number of concepts for which the child has a word in either language) and the number of translational equivalents (concepts for which the child has a word in both languages). Furthermore, the language in which the child can express more concepts (sometimes used as a definition for language dominance) can be identified. Despite the need to assess the lexicon in both languages, currently only a few instruments exist that are able to do so.

Most instruments for assessing vocabulary are constructed for one given language only; in some rare cases, adaptations or translations are available. In very young children aged between 12 and 30 months who are at the onset of language acquisition, parents are usually asked to indicate which words (e.g., of the corresponding MacArthur–Bates inventories) children produce in the two languages (Bosch & Ramon-Casas, 2014; Marchman & Martínez-Sussmann, 2002; Perez-Pereira, 2008; Thordardottir, Rothenberg, Rivard, & Naves, 2006). However, recent research has demonstrated that parents tend to overestimate their children's knowledge in the two languages, resulting in a higher number of translational equivalents reported by parents compared to the translational equivalents actually measured directly (Legacy, Reider, et al., 2016). For slightly older children, starting from 3 years of age, some vocabulary tests like the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007) can be used in different languages (Umbel et al., 1992). However, these instruments have mainly been constructed for monolinguals, and the

items and the number of items differ between languages (e.g., English: $n = 228$, French: $n = 170$, Spanish: $n = 125$). Consequently, vocabulary size can be measured for each language separately and standards for each language allow an estimation of skills. However, the individual vocabulary sizes in the two languages cannot be combined to estimate the conceptual vocabulary or the number of translational equivalents because each version assesses different items.

Beyond these monolingual scales, a few multilingual tools do exist (Blumenthal, Giannuzzi, Holstvoogd, De Ruiter, & Bos, n.d.; Haman, Łuniewska, & Pomiechowska, 2015; Hovsepian, 2018; Munoz-Sandoval, 1998; Peña, Gutierrez-Clellen, Iglesias, Goldstein, & Bedore, 2014). But again, these tools include different numbers of items for each language or require the researcher to be proficient in the test language. One example for children aged 5 years and older is the Bilingual Verbal Ability Test (BVAT; Munoz-Sandoval, 1998), which is available in 18 languages. Every item that is not correctly answered in English is repeated in the child's second language. In this procedure, examiners need to be fluent in both languages themselves and they assess the second language for a small number of items only – those the child does not know in the first language. As such, it does not assess translational equivalents. It is possible to test children on all items in both languages. But this results in methodological problems: Receptive vocabulary tests often use pictures showing three distractors and one target (Dunn & Dunn, 1997, 2007). When these pictures are presented for both languages consecutively, carry-over effects are very likely to occur because children might remember which of the four objects they pointed at when the picture was presented the last time. Administering scales like BVAT in both languages would therefore not yield usable data.

Accordingly, the goal of the present study was to develop an instrument that (a) assesses lexicons and translational equivalents in 3-year-old bilinguals, (b) is easy to administer without the need for the administrator to be fluent in the bilingual child's two languages, (c) has the same number of items for each group of bilinguals, (d) keeps transfer effects caused by memorizing targets from one language to another to a minimum, and (e) provides the possibility of assessing vocabulary skills in bilinguals with different language combinations with a primary focus on scientific studies. We report construction details and psychometric properties of this instrument, called 'BILEX', assessing bilinguals' lexicons in the eight most frequently acquired languages in Switzerland: Swiss German, High German, Latin American Spanish, European Spanish, British English, American English, Italian, and French. In the first section, we describe the item validity and test construction of BILEX, followed by Study 1 in which reliability and internal and external validity were investigated. In Study 2 we analyzed convergent validity of BILEX with PPVT-4 (Peabody Picture Vocabulary Test, Fourth Edition). Finally in Study 3, we report on prognostic validity with H-S-E-T (Heidelberger Sprachentwicklungstest).

Item validity and test construction

In this section, we describe item selection and validity together with considerations and solutions for test construction for the BILEX instrument.

Selection of word category

In the course of language acquisition, children acquire words for the concepts in their languages, learn to pronounce them and use them in the correct context, and they learn to combine words into sentences following the morphological and syntactic rules of their language. In the chosen target age range of 3 years, that is the age for which parental reports like M-CDI are no longer available or are less accurate, children have already acquired a great number of words and have begun to combine words into sentences with a simple structure. We chose to focus on vocabulary development because it is a cross-linguistically important marker of language acquisition (Dale & Goodman, 2005). Within vocabulary, we decided to test nouns because children's early words in various languages fall into a small number of the same semantic categories like food, body parts, clothing, animals, vehicles, toys, and household objects (Clark, 2007). We did not include verbs for the following reasons: The acquisition of verbs depends on the specific verb morphology of the language (Slobin, 1973; Stoll et al., 2012), which means that the more complex and the richer the morphology in a specific language, the slower the acquisition of verbs. Furthermore, cross-linguistic influences of verb morphology between the two languages children acquire have been found in a number of studies (Conboy & Thal, 2006; Döpke, 1998). These influences can be ascribed, for example, to differences in saliency, frequency of verbs or pragmatics of verbs across languages (Bornstein et al., 2004; Gopnik & Choi, 1995; Kim, McGregor, & Thompson, 2000). Taken together, several factors make it difficult to assess verbs in bilinguals with different language combinations.

For it to be feasible for researchers to assess bilinguals' linguistic skills, we decided to assess receptive knowledge of words. This procedure does not require researchers to have knowledge in all the languages spoken by the bilingual children tested. We used pictures and assessed whether children select the target picture from a number of related items, similar to the procedures of PPVT-4 (Dunn & Dunn, 2007) and BPVS (British Picture Vocabulary Scale, Dunn & Dunn, 1997).

Selection of items

For BILEX, we chose to present realistic, ecologically valid representations of real-life objects. Accordingly, we decided to use items from the Moreno-Martínez and Montoro (2012) database. Those authors created 360 high-quality color images belonging to 23 semantic subcategories, and identified the ages of acquisition (AoAs) of the respective terms. We determined the target age range for BILEX items from the Moreno-Martínez and Montoro (2012) database by matching the ages of acquisition of the items prompted in the German PPVT-4 (Lenhard, Lenhard, Segerer, & Suggate, 2015) for 3-year-olds to those of Moreno-Martínez and Montoro. The target age range was between 1.5 and 2.9 on the scale of Moreno-Martínez and Montoro. This makes it possible to have easy-to-answer items to keep children motivated and more difficult items to distinguish between children. We selected 48 items in 12 subcategories used early on: animals ($n = 16$), body parts ($n = 4$), buildings ($n = 2$), clothing ($n = 4$), fruits ($n = 4$), furniture ($n = 4$), kitchen utensils ($n = 2$), musical instruments ($n = 2$), tools ($n = 2$), vegetables ($n = 2$), vehicles ($n = 4$), and weapons ($n = 2$).

All pictures of the target items were shown to native female speakers of the respective eight target languages (French, Italian, Latin American Spanish, European Spanish,

British English, American English, Swiss German, and High German) to generate the labels in the different languages, which were recorded. These audio recordings were presented to the children during the test. This way, the examiner was not required to pronounce any of the target items.

Validity of item selection

We aimed to determine content validity in terms of representative cross-linguistic phonological similarity between the target items as well as validity in terms of cross-linguistically valid AoAs.

Cross-linguistic phonological similarity of items. Languages share phonological similarities to a different degree. Languages with a common ancestor have more similarities than languages from different language families. When constructing a cross-linguistic vocabulary scale, it is important to reproduce the similarity features across languages. We therefore investigated whether the items chosen for BILEX are representative of the similarity found between languages.

We determined the number of cognates between the Swiss German labels of the target items and the other languages' labels in BILEX. Swiss German was used as the reference category because this is the shared language between all monolinguals and bilinguals. Cognates are regarded as words with high cross-linguistic meaning and orthographic or phonetic similarity, e.g., German *boot* and English *boat* (Ciobanu & Dinu, 2014). Cognate words between the two languages were calculated using the LexStat algorithm for cognate detection (SCA method with threshold of 0.3, see List, 2014 for details) as implemented in the LingPy software package (<http://lingpy.org>, List & Forkerl, 2015; for details see List & Moran, 2013). The thresholds and the method for the LexStat algorithm were chosen in such a way that the cognates identified correspond to the language-learning definition of cognacy, meaning they are words of similar meaning and pronunciation which are usually also recognized as such by second language learners and bilinguals.

The overall phonological similarity score between the languages was taken from the ASJP database (Automated Similarity Judgment Program, Holman et al., 2011; Wichmann et al., 2013). This measure is based on Levenshtein distance between the translational equivalents of the phonological forms of the 40 most stable lexical items of Swadesh's 100-item list (Swadesh, 1955; for details see Holman et al., 2011).

Finally, we analyzed whether the number of cognates between the languages of BILEX was associated to the phonological distance of the ASJP score, which is a prerequisite for the items to be used within a cross-linguistic scale. The number of cognates shared between the Swiss German and the other languages' versions of BILEX was significantly correlated with the phonological similarity of ASJP, $r(6) = -.87, p = .01$. This suggests that the cross-linguistic phonological properties of the items were representative of the phonological similarity between the languages used.

Cross-linguistic age of acquisition of items. Cross-linguistic collections of AoAs showed that reported ages for Spanish correlated highly with those of German, Italian, and English for example (Łuniewska et al., 2015). This suggests that the words in these

languages are acquired at a similar time in development. We investigated whether the items chosen for BILEX are also acquired at similar age ranges in the BILEX languages. This is important in order to show that children speaking a certain BILEX language are not prone to lower scores because the items are acquired later in their language.

We extracted the AoAs for English, High German, Spanish, and Italian (no French data were collected) for the items used in BILEX from the database of Łuniewska et al. (2015). For each language pair, we recalculated the correlation strength for BILEX items. Because not all items were available for cross-check in all languages, sample sizes differed between the languages. We calculated the correlations between the AoAs of High German and the English, Italian, and Spanish, because no Swiss German AoAs were collected by Łuniewska et al. (2015). All correlations were significant (German–English: $r(26) = .805$, $p < .001$, German–Italian: $r(26) = .751$, $p < .001$, German–Spanish: $r(25) = .851$, $p < .001$). Correlation coefficients are in the range found by Łuniewska et al. (2015). We furthermore tested whether the strengths of the correlations are different between the languages using Fisher r -to- z transformation and found no differences (English–Italian: $p = .681$, English–Spanish: $p = .686$, Italian–Spanish: $p = .171$). This suggests that the items have similar AoAs and are similarly early acquired in all of the languages used.

Test construction

Considerations. Similar to other instruments assessing children's vocabulary, we intended to present children with test cards displaying a picture of the target item together with other distractor pictures. This procedure reduces the likelihood that children choose the target randomly. For the purpose of measuring vocabulary in both of the children's languages, a further requirement for BILEX was to reduce the likelihood of children memorizing the target position from the administration of the first language in the administration of the second language. To fulfill this requirement, we adhered to the following design features:

1. We grouped items on test cards in a way that each test card was shown multiple times. With this procedure, each item was shown to the child several times, either as the target or as a distractor. In the typical scenario of PPVT-4 or BPVS, in which each item is shown only once with a fixed set of distractors, children might memorize the test card and choose the same picture again in the second administration, even though this would not necessarily mean that the child knows the target in the second language as well. By grouping target items of the current trial with target items of later trials, transfer effects are less likely to occur.
2. To reduce the selection of targets using exclusion criteria, we added distractor items to the test cards that were acquired later in life. If children have to respond to a label whose concept they have not yet acquired in the respective language, they might exclude the other distractors as targets because they know the labels of these (early-acquired) pictures. By adding late-acquired distractors, the chance of selecting the target through exclusion is kept much lower than if only distractors from the same age range are chosen.
3. Each time the test card was shown, we changed the positions of the pictures. This procedure meant that the location of a given item changed during testing, further

minimizing memorization effects. Finally, the probability of choosing the target picture randomly was kept at the same level.

4. We created two orders of trials, one for the Swiss German vocabulary and one for the vocabulary in the child's other language. Children were therefore less likely to remember the order of targets in the two versions. Again, transfer effects were minimized between the administrations of the two BILEX versions.

Implementation of test cards and trial order. For arrangement (1) we chose to group each item together with three other items to build a test card. We built groups on the level of semantic categories. We grouped two semantic categories on one test card: Each test card comprised items from the same semantic category (i.e., a tomato and an onion) together with two items from another semantic category (i.e., a chair and a lamp).

Furthermore, to fulfill arrangement (2) each test card included two late-acquired distractors. Late-acquired distractors had an AoA of between 3.8 and 6.5. The two distractors were split by semantic category: one late-acquired distractor was added for each semantic category (i.e., leek and a cabinet). The grouping resulted in 12 test cards (test card 'a' to test card 'm'), each with four items and two distractors. Table 1 provides an overview of how objects were grouped together on cards and which objects served as target items and as distractors.

For arrangement (3) we created six positions of items and distractors on the test cards: Pictures were positioned in two rows with three columns. During administration, each test card was shown four times in each language (because four items were grouped together on one card). The location of the pictures was different every time the card was shown, meaning every item had its own arrangement of picture positions and four different compositions were created. Picture locations were counterbalanced so that all six positions were served equally often as target locations across all card sets, items, and versions.

To fulfill arrangement (4) we created two trial orders, one for the Swiss German BILEX and one for the Non-Swiss German BILEX. The Non-Swiss German BILEX refers to the children's second L1 and was available in the seven other language variants (High German, Latin American Spanish, European Spanish, British English, American English, Italian, and French). The trial orders were designed in such a way that a fixed order of test cards was used: The trials were run four times from test card 'a' to test card 'm' for the Swiss German version, and backwards from card 'm' to card 'a' in the Non-Swiss German version. Each card was therefore shown every 12 trials. In each version, the order of target items was also fixed. For both versions, all items of semantic group1 target1 were run first, followed by all targets of set semantic group1 target1, followed by semantic group2 target1 and semantic group2 target2; see Table 1 and Figure 1. This procedure was preferred in comparison to a random presentation of trials in order to minimize transfer effects between both language versions.

Study I

In this section, we describe the testing procedure of BILEX together with a first set of psychometric properties. In Study 1, we investigated reliability as well as internal and external validity. Reliability describes the overall consistency of a measure: We analyzed

Table 1. Test cards with items and late-acquired distractors.

Test card	Semantic group1 target1	Semantic group1 target2	Semantic group 2 target1	Semantic group2 target2	Semantic group1 late-acquired	Semantic group2 late-acquired
a	cat	snake	church	castle	platypus	pagoda
b	tomato	onion	chair	lamp	leek	cabinet
c	bus	boat	spider	butterfly	paragliding	termite
d	lemon	strawberry	eye	leg	currant	vertebra
e	table	sofa	horse	tiger	lectern	tapir
f	ant	ladybird	Sword	arrow	praying mantis	boomerang
g	arm	mouth	car	plane	pelvis	skateboard
h	elephant	turtle	sock	vest	lynx	clog
i	fork	cup	rooster	duck	churrera	kiwi
k	shoe	trousers	grapes	pear	beret	custard apple
l	fly	wasp	shovel	hammer	moth	chisel
m	guitar	tambourine	cow	giraffe	saxophone	armadillo

internal consistency (consistency of results across items within BILEX) and split-half reliability (consistency of results across two halves of the test). We furthermore investigated validity: whether BILEX measures vocabulary size effectively in bilingual children. We analyzed internal validity (validity of the internal structure) and external validity (relations to other variables). For internal validity, we investigated the degree to which the different measures of BILEX (Swiss German and Non-Swiss German vocabulary, conceptual vocabulary, and translational equivalents) are related. Given previous findings on cross-language associations, we expected a moderate relationship between performance in the two language versions (Hoff, Quinn, & Giguere, 2018) as well as between the single language performances and conceptual vocabulary and translational equivalents (Legacy, Zesiger, Friend, & Poulin-Dubois, 2016). For external validity, we aimed to replicate previous contributing factors to children’s vocabulary size, namely an influence of sex (Eriksson et al., 2012) and input frequency (Place & Hoff, 2011, 2016; Unsworth, 2015), as well as the difference between monolingual and bilingual dissociation of lexicon size depending on how many languages are assessed as explained above. Furthermore, since BILEX is a tool for use in developmental psychology, we analyzed whether performance in BILEX increases with age. As a novel research question, we investigated the role of language similarity between the two languages. Recent findings in simultaneous and sequential bilingualism point to the fact that similarity between the two languages to be acquired eases word learning (Bialystok & Hakuta, 1994; Bosch & Ramon-Casas, 2014; Cysouw, 2013; Schepens, Van Der Slik, & Van Hout, 2013; Snow & Kang, 2006).

Methods

Participants. Participants were 144 3-year-old children (77 female, 67 male, $M = 40$ months, $SD = 3$ months, range = 36–47 months). Children were either monolingual ($n = 57$; 27 female; 30 male; both parents only speak Swiss German), simultaneous bilingual

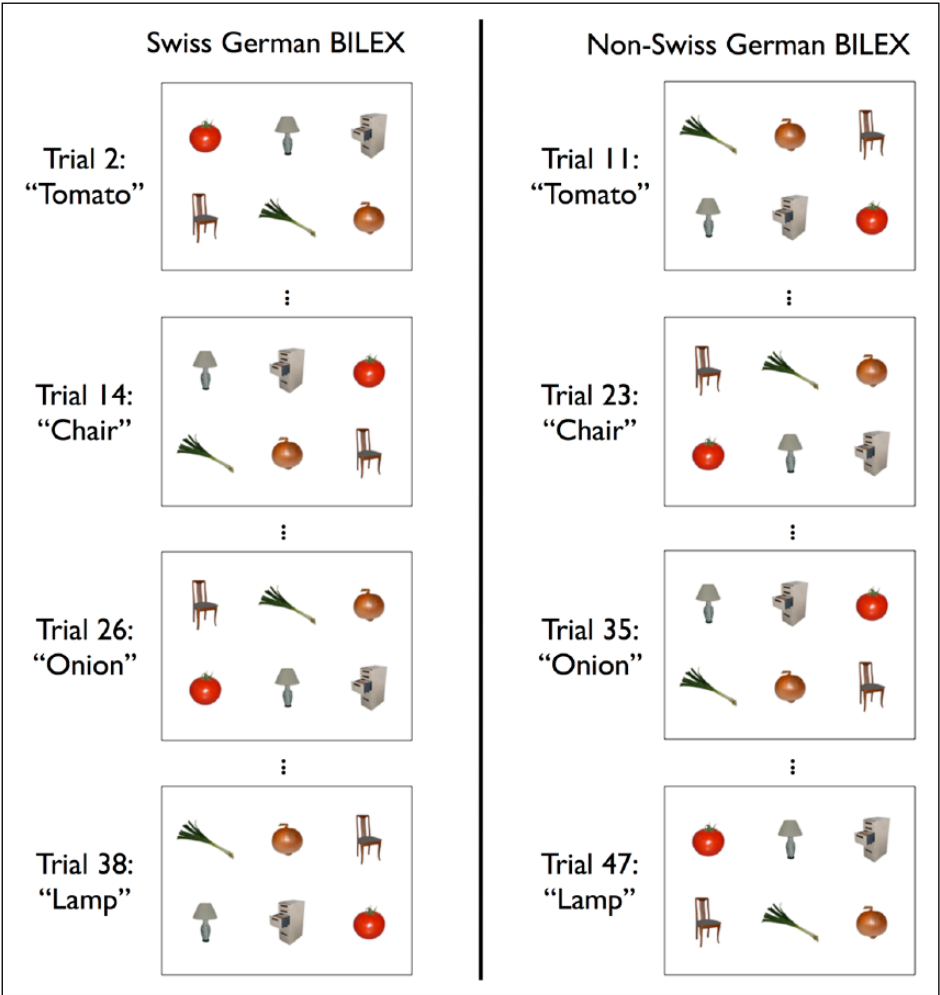


Figure 1. Trial order and targets shown for the example of test card ‘b’ in the Swiss German BILEX and in the Non-Swiss German BILEX. Picture locations change for each administration of the same test card. Furthermore, trial order changes between both versions of BILEX.

($n = 75$; 45 female; 30 male; one parent speaks Swiss German, the other parent speaks another language, both from birth on) or sequential bilingual ($n = 12$; 5 female; 7 male; both parents speak the same language that is not Swiss German, but children have visited Swiss German day-cares for more than one year). See Table 2 for an overview of the language combinations of the two bilingual groups. Children were recruited from a database of parents who consented to volunteer in developmental research and received a small gift at the end of the study. The procedure was approved by the ethics committee of the University of Zurich and in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments.

Table 2. Number of participants in the different language groups, Studies 1, 2, and 3.

Group	Number of participants
<i>Study 1</i>	
Simultaneous bilingual, Swiss German +	75
British English	6
American English	8
French	15
High German	22
Italian	15
European Spanish	7
Latin American Spanish	2
Sequential bilingual, Swiss German +	12
American English	1
French	4
High German	5
Italian	1
Spanish	1
Monolingual Swiss German	57
<i>Study 2</i>	
Simultaneous bilingual, Swiss German +	16
Swedish	1
Czech	3
Greek	1
Russian	1
Polish	1
Georgian	1
Dutch	2
Moroccan	1
Bosnian	1
Serbian	1
Turkish	1
Kurdish	1
Armenian	1
Monolingual Swiss German	22
<i>Study 3</i>	
Simultaneous bilingual, Swiss German +	12
British English	1
American English	2
French	2
High German	1
Italian	1
Latin American Spanish	1
European Spanish	3
Greek	1
Sequential bilingual, Swiss German +	4
French	2
High German	1
Italian	1
Monolingual Swiss German	5

We assessed parents' highest education as an approximation of socio-economic status on a scale ranging from 1 to 10, with 10 being equivalent to a university degree. We calculated the mean of both parents' education. The three language groups did not differ in terms of parental education, $F(2,140) = 2.116$, $p = .124$, nor did monolinguals differ from bilinguals (simultaneous and sequential together), $F(1,141) = 2.691$, $p = .103$.

All bilingual children had at least 20% input in each of their two languages. This criterion was adopted from Pearson, Fernandez, Lewedeg, and Oller (1997) and is commonly used in bilingual research (Byers-Heinlein, 2013; Fiestas, Bedore, Peña, & Nagy, 2005; Gillam, Ho, Bedore, & Pen, 2010; Gutiérrez-Clellen & Kreiter, 2003). The children's exposure to these languages was estimated through a parental questionnaire similar to those used in previous research (Bosch & Sebastián-Gallés, 1997; Fennell et al., 2007; Poulin-Dubois, Bialystok, Blaye, Polonia, & Yott, 2012). It included questions on parents' working hours, children's waking hours, the amount of time spent in childcare and the caregiver's language backgrounds. We calculated how much time children spend with each language and converted this to a percentage based on their waking hours. Monolinguals and bilinguals had no more than 10% of other languages in their input. Children's mean exposure to Swiss German was 57% ($SD = 17\%$) for the simultaneous bilinguals and 33% ($SD = 16\%$) for the sequential bilinguals; the mean exposure to non-Swiss German was 40% ($SD = 16\%$) for the simultaneous bilinguals and 67% ($SD = 16\%$) for the sequential bilinguals.

Procedure. Testing took place in the rooms of the university's lab. Children were familiarized with the experimenter prior to testing and were accompanied to the testing room. Once in the room, children sat opposite the experimenter at a table. Stimulus material was presented using the software Presentation® (Neurobehavioral Systems) on a touchscreen laptop. The Swiss German version of BILEX was always administered first, to have monolinguals and bilinguals perform the test at the same point in time during testing.

Children were told that they were going to play a computer game where a woman's voice would say labels of objects and children were asked to touch the object that corresponded to the label they had heard. Two warm-up trials served to make sure that children understood these instructions. In the warm-up trials (asking for apple and dog), children were presented with three prerecorded sentences: 'Apple. Where is the apple? Point to the apple.' If children failed to touch the correct picture, the experimenter repeated the instructions. The two warm-up trials were untimed to ensure that the children understood the task. Piloting showed that two warm-up trials were enough to ensure children's comprehension of the task. These instructions were given in Swiss German.

During the 48 test trials, only the labels were used as a prompt ('Snake'), since piloting showed a decrease in attention when children were presented with whole sentences ('Where is the snake?') for each item. In each trial, children were presented with one of the 12 card sets consisting of one target object and five distractors. The test trials were timed; if children did not respond within 7 seconds, the next trial was presented without later repetition (this time window was also determined through piloting). The experimenter gave no feedback during the test trials. Nevertheless, if necessary, the administration of items could be paused and restarted. The two versions (one for each of the child's languages) were administered with a break of 30 minutes between. In the break, children

took part in other social-cognitive tasks with the experimenter. Instructions were repeated before the administration of the Non-Swiss German version.

Coding. For each trial, the program script logged which of the areas on the screen were touched. For each trial, children received 1 point if the target position and touch position matched, and 0 points if the positions did not match or children did not choose any object. The sum of these trials was calculated for Swiss German BILEX and Non-Swiss German BILEX. Furthermore, the conceptual vocabulary of each child was calculated from the number of all items that were identified correctly in one or both languages. The number of translational equivalents was calculated from all targets that were identified correctly in both language versions.

Results

The average accuracy levels for all language groups and all measures of BILEX are depicted in Figure 2.

Reliability. For internal consistency, we analyzed data from the 48 test trials for all 144 children, which yielded a Cronbach's alpha for conceptual vocabulary of $\alpha = .87$. For the internal consistency of translational equivalents, we analyzed all 87 bilingual children, which yielded a Cronbach's alpha of $\alpha = .90$. For split-half reliability, we divided the whole set of items into the first and the second half (24 items each), which yielded a Spearman-Brown coefficient of $r = .83$ for conceptual vocabulary and $r = .86$ for translational equivalents.

Internal construct validity. From all bilingual children, internal construct validity was calculated via correlations between test scores in the different BILEX measures. All correlations were significant: Swiss German BILEX and Non-Swiss German $r(85) = .487, p < .001$, Swiss German BILEX and conceptual vocabulary $r(85) = .804, p < .001$, Swiss German BILEX and translational equivalents $r(85) = .741, p < .001$, Non-Swiss German BILEX and conceptual vocabulary $r(85) = .786, p < .001$, Non-Swiss German BILEX and translational equivalents $r(85) = .924, p < .001$, conceptual vocabulary and translational equivalents $r(85) = .816, p < .001$.

External construct validity. We analyzed whether measuring one or both languages influences the scores in monolinguals compared to bilinguals. A 2×2 (Measure [Swiss German BILEX, conceptual vocabulary] $\times$ Language Group [monolingual, bilingual]) ANCOVA with all children speaking monolingual Swiss German or bilinguals speaking two of BILEX languages, controlling for age in days, sex, and educational level of the parents, revealed a significant interaction between measure and language group, $F(1,142) = 219.297, p < .001$, showing that bilinguals performed better than monolinguals when conceptual vocabulary was assessed via BILEX, $p = .009$, but bilinguals performed worse than monolinguals when only Swiss German was assessed, $p = .006$.

Furthermore, we investigated the influence of several demographic factors (sex, age, and educational level of the parents) on Swiss German BILEX scores. We ran a linear

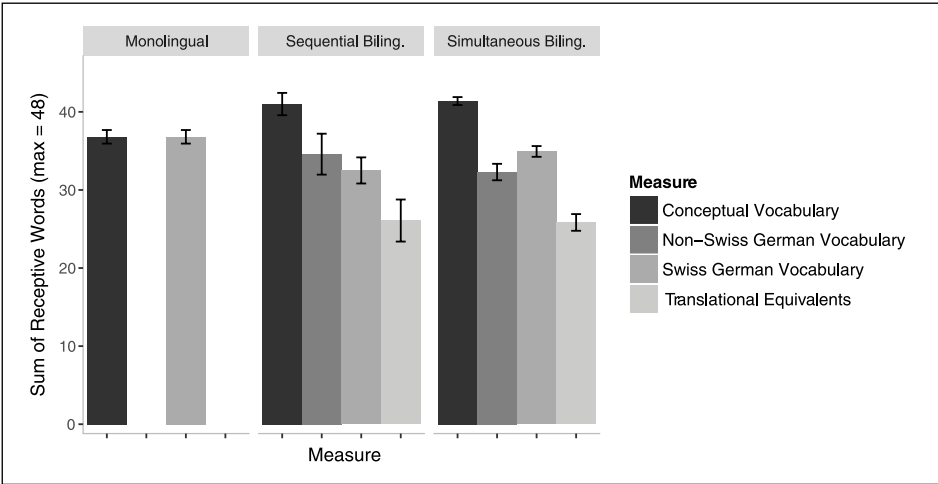


Figure 2. Children’s average accuracy level on BILEX for monolinguals, sequential bilinguals, and simultaneous bilinguals; error bars indicate ± 1 SE.

regression on Swiss German vocabulary for all bilingual children with all of the demographic factors as predictors. We found that all predictors but age significantly influenced the Swiss German vocabulary, see Table 3.

Finally, we analyzed whether the bilingual children with language combinations sharing more cognates have higher scores in conceptual vocabulary or translational equivalents. This result would support the hypothesis that language similarity positively influences bilingual language acquisition. We again ran a linear regression with all of the demographic predictors as well as the number of cognates and found that all predictors but input infrequency significantly influenced the conceptual vocabulary, see Table 4.

Discussion

Study 1 described the procedure of using BILEX as a research instrument and assessed reliability and validity measures. The analyses showed very good reliability along with replications of the impact of external factors such as education of the parents, sex, and age. Although the age of the children in this study only spanned a few months, it nevertheless revealed age-related changes in the acquisition of conceptual vocabulary, which points to a very precise scaling. BILEX measures did not replicate previous findings concerning input frequency. Previous findings showed that the more input children receive in one of their languages, the poorer their performance in their second language (Cattani et al., 2014; David & Wei, 2008; Hoff et al., 2012; Pearson et al., 1997). However, the best way of testing the bilingual experience is still a matter of debate. New measures like the Bilingualism Profile Index look more promising to capture overall input than the current input as used in our study (De Cat, Gusnanto, & Serratrice, 2018). Furthermore, bilinguals tested with BILEX performed better when both of their languages were assessed (Bedore et al., 2005; Pearson et al., 1993). These replications of

Table 3. Results of the regression on Swiss German vocabulary for monolinguals and bilinguals.

Predictors	B	SE B	p
Sex (female, male)	3.11	0.99	.002
Parental education (1 = formal schooling, 10 = university degree)	0.78	0.25	.003
Age in days (min = 1087, max = 1407)	0.01	0.01	.148
Language group (monolinguals, bilinguals)	3.79	1.17	.001

Table 4. Results of the regression on conceptual vocabulary for the bilingual group.

Predictors	B	SE B	p
Sex (female, male)	1.02	0.87	.245
Parental education (1 = formal schooling, 10 = university degree)	0.84	0.21	< .001
Age in days (min = 1087, max = 1407)	0.02	0.01	.001
Input frequency in Swiss German (min = 21%, max = 80%)	-0.01	0.02	.440
Number of cognates	0.10	0.03	.003

known factors driving language acquisition point to good psychometric properties for internal and external validity.

BILEX was furthermore used to investigate the effect of language similarity between the two languages on lexicon size. Previous findings in adult L2 acquisition have demonstrated rather strong effects of the learners' first language (L1) on L2 language proficiency (Cysouw, 2013; Schepens et al., 2013; Van der Slik, 2010). For example, the phonological similarity between the first languages of immigrants (35 different languages) and the L2 of the country they migrated to influences the immigrants' L2 proficiencies (Schepens et al., 2013). In the case of early bilingual children, vocabularies comprise a relatively higher proportion of translational equivalents that are cognate words (e.g., *boot* and *boat*) as compared to translational equivalents that are non-cognates (e.g., *keks* and *biscuit*; Bosch & Ramon-Casas, 2014; Schelleter, 2002). This suggests that the similarity of the phonological forms of words across languages eases word learning. We were able to provide further evidence for this hypothesis with our findings from Study 1.

Study 2

In Study 2 we investigated the convergent validity of BILEX. A strong correlation between preschoolers' lexicon size assessed with BILEX and a second measure would suggest that the two tests measure the same construct.

Methods

Thirty-eight children (20 female, 18 male, $M = 43$ months, $SD = 2.2$ months, range = 35–46 months, 22 monolingual children and 16 bilingual children) took part, see Table

2. All bilingual children spoke Swiss German and a second language that was not part of BILEX. They were recruited in the same way as in Study 1 and received a small gift at the end of the study.

We used the Swiss German version of BILEX and PPVT-4 (Dunn & Dunn, 2007) to assess children's receptive vocabulary. The PPVT-4 is available in a standardized High German version (Lenhard et al., 2015). It contains nouns, verbs, and adjectives and is designed as a power test with a block design: The age of the participants determines which block is presented first. Every block contains 12 items, and if children have more than seven incorrect targets, the test ends. We created a Presentation® (Neurobehavioral Systems) version of the PPVT-4 similar to the one used during standardization of the High German version. We again prerecorded the Swiss German labels of the targets to make the administration of the PPVT-4 comparable to the administration of BILEX.

The procedure for BILEX was the same as in Study 1 and BILEX was administered at the beginning of the study. Half an hour later, the PPVT-4 was administered according to the instructions of the PPVT-4. We calculated the number of correctly touched words for BILEX and PPVT-4.

Results and discussion

The scores of the two instruments BILEX and PPVT-4 were highly correlated, $r(38) = .605$, $p < .001$. The two vocabulary scales shared about 36% of variance. The receptive word knowledge assessed by BILEX items was associated with the receptive word knowledge assessed by PPVT-4. This suggests that BILEX shows good convergent validity with an already established scale on children's lexicon size.

Study 3

A further requirement to evaluate psychometric properties of a new scale is prognostic validity. This means that the results of BILEX should be able to predict language skills at a later point in time. We investigated this in Study 3.

Methods

We tested 21 of the children who took part in Study 1 or 2 once again one year later to assess prognostic validity. Children were selected randomly on the basis of parents' consent to take part in a continuation study of the original study. These were 12 simultaneous bilinguals, 4 sequential bilinguals, and 5 monolinguals (9 female, second measurement point: $M = 52$ months, $SD = 3$ months, time between first and second measurement: $M = 11.5$ months, $SD = 1.3$ months), see Table 2.

For prognostic validity, we decided to measure grammatical skills, to ensure that vocabulary skills were representative of overall linguistic skills. The only test for grammatical skills available in Swiss German is the Swiss German version of the Heidelberger Sprachentwicklungstest (H-S-E-T; Grisseman, Baumberger, & Hollenweger, 1991). We conducted the subtests on syntax (SB, IS), morphology (PS, AM), and semantics (WF) according to the instructions of the scale. We followed the scoring procedure of each

subtest and calculated the mean of standardized T-scores for grammatical performance in the subtests of H-S-E-T.

Results and discussion

We correlated vocabulary size in the Swiss German BILEX with performance in the H-S-E-T, controlling for the delay between the two tests, sex, and parental education, and found a significant relationship, $r(21) = .525, p = .014$. The performance in BILEX was associated with a later assessment using a standardized test. This is indicative of prognostic qualities for later language skills.

General discussion

In the present study, we introduced BILEX, an instrument for the assessment of bilingual children's receptive vocabulary. In particular, we described the development and the validation of BILEX. It was developed due to a shortage of scales assessing bilinguals' lexicons in both of their languages. Previous research has shown that bilinguals' linguistic skills are underestimated if only one language is investigated (Bedore et al., 2005; De Houwer et al., 2014; Hoff et al., 2012; Pearson et al., 1993). Existing instruments for bilinguals (Blumenthal et al., n.d.; Haman et al., 2015; Peña et al., 2014) have been constructed for each language separately, or assess some of the items in the second language (Munoz-Sandoval, 1998), which hinders the assessment of conceptual vocabulary or the number of translational equivalents. For this reason, the vocabulary of both languages needs to be measured at the same time, using a comparable measure.

BILEX was constructed to measure receptive word knowledge in both bilingual children's languages within one session. It can be applied by examiners who do not speak both of the bilingual children's languages. Furthermore, the construction of the test minimizes transfer effects between the administrations of the bilingual child's languages. A novel scale needs to prove its applicability in research by showing adequate psychometric properties. We investigated various types of reliability as well as internal and external validity, and found that the psychometric properties of BILEX are excellent. This is the basis for further utilization of BILEX in research on bilingual children's language and other cognitive skills. Reliability was .90 for internal consistency and .86 for split-half reliability. The correlation between BILEX and the PPVT-4 that determined convergent validity was .61, while the prognostic validity between BILEX and H-S-E-T about one year later was .53. These psychometric properties are comparable with those of BVAT (Munoz-Sandoval, 1998) for older children and adults. Further external validity measures were high and in line with previous research, and support the fact that assessing both languages improves the correct measurement of bilinguals' lexicons (Bedore et al., 2005; De Houwer et al., 2014; Hoff et al., 2012; Pearson et al., 1993).

BILEX was constructed to be used as an estimate of linguistic skills for bilinguals with different language combinations and monolinguals. It is not designed to be used in a clinical setting and therefore does not identify children whose vocabulary knowledge is below average. However, its applications in developmental bilingualism

research are numerous: It can be employed in research addressing potential advantages of bilingualism in cognitive domains such as theory of mind and inhibition (Bialystok, 2011; Carlson & Meltzoff, 2008; Poulin-Dubois, Blaye, Coutya, & Bialystok, 2011). To match groups of monolingual and bilingual children properly, researchers depend on the assessment of bilinguals' linguistic skills in both languages (Adesope et al., 2010). Beyond measuring vocabulary skills in itself, such an instrument has further utilizations in research on bilinguals. In the area of bilingual language and its acquisition, researchers are interested in the organization of the bilingual lexicon (Serratrice, 2013) and investigate the acquisition and processing of cognate and non-cognate translational equivalents (Costa, Miozzo, & Caramazza, 1999; Pearson, Fernández, & Oller, 1995). BILEX also offers the possibility to determine the dominant language of a bilingual child (Gathercole & Thomas, 2009; Hoff et al., 2012). The use of the current version of BILEX is limited because it is applicable only for 3-year-olds and a small number of languages. Future versions of BILEX are planned to extend both aspects and will include additional language variants and will be applicable to a broader age range. However, BILEX overcomes the shortcomings of previous scales for receptive vocabulary for total vocabulary and conceptual vocabulary, as well as the number of translational equivalents, accessible to developmental researchers concerned with bilingual language acquisition.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Adesope, O. O., Lavin, T., Thompson, T., & Ungerleider, C. (2010). A systematic review and meta-analysis of the cognitive correlates of bilingualism. *Review of Educational Research, 80*, 207–245.
- Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: When does it make a difference? *Language Speech and Hearing Services in Schools, 36*, 188–200.
- Bialystok, E. (2001). *Bilingualism in development : Language, literacy, & cognition*. Cambridge, UK: Cambridge University Press.
- Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology [Revue Canadienne de Psychologie Expérimentale], 65*, 229–235.
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin, 143*, 233–262.
- Bialystok, E., & Hakuta, K. (1994). *In other words : The science and psychology of second-language acquisition*. New York, NY: Basic Books.
- Bialystok, E., & Luk, G. (2012). Receptive vocabulary differences in monolingual and bilingual adults. *Bilingualism: Language and Cognition, 15*, 397–401.

- Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition*, 13, 525–531.
- Blumenthal, M., Giannuzzi, B., Holstvoogd, L., De Ruiter, E., & Bos, P. (n.d.). *Development of a speech production screening tool in two Arabic dialects for multilingual children in the Netherlands*. Retrieved from <https://static1.squarespace.com/static/57ab4d27725e25c7daf8f1d6/t/5800cc2a59cc685e14182152/1476447373437/DevelopmentOfASpeechProductionScreeningTool.pdf>
- Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S.-Y., Pascual, L., . . . Vyt, A. (2004). Cross-linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, 75, 1115–1139.
- Bosch, L., & Ramon-Casas, M. (2014). First translation equivalents in bilingual toddlers' expressive vocabulary: Does form similarity matter? *International Journal of Behavioral Development*, 38, 317–322.
- Bosch, L., & Sebastián-Gallés, N. (1997). Native-language recognition abilities in 4-month-old infants from monolingual and bilingual environments. *Cognition*, 65, 33–69.
- Burgaleta, M., Sanjuán, A., Ventura-Campos, N., Sebastian-Galles, N., & Ávila, C. (2016). Bilingualism at the core of the brain: Structural differences between bilinguals and monolinguals revealed by subcortical shape analysis. *NeuroImage*, 125, 437–445.
- Byers-Heinlein, K. (2013). Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary. *Bilingualism: Language and Cognition*, 16, 32–48.
- Carlson, S. M., & Meltzoff, A. N. (2008). Bilingual experience and executive functioning in young children. *Developmental Science*, 11, 282–298.
- Cattani, A., Abbot-Smith, K., Farag, R., Krott, A., Arreckx, F., Dennis, I., & Floccia, C. (2014). How much exposure to English is necessary for a bilingual toddler to perform like a monolingual peer in language tests? *International Journal of Language & Communication Disorders*, 49, 649–671.
- Chiat, S., & Polišenská, K. (2016). A framework for crosslinguistic nonword repetition tests: Effects of bilingualism and socioeconomic status on children's performance. *Journal of Speech, Language, and Hearing Research*, 59, 1179–1189.
- Ciobanu, A. M., & Dinu, L. P. (2014, June 22–27). Automatic detection of cognates using orthographic alignment. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Short Papers)* (pp. 99–105). Baltimore, MD.
- Clark, E. V. (2007). *First language acquisition* (6th ed.). Cambridge, UK: Cambridge University Press.
- Conboy, B. T., & Thal, D. J. (2006). Ties between the lexicon and grammar: Cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*, 77, 712–735.
- Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, 41, 365–397.
- Cysouw, M. (2013). Predicting language-learning difficulty. In L. Borin & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 57–82). Boston, MA: De Gruyter.
- Dale, P. S., & Goodman, J. C. (2005). Commonality and individual differences in vocabulary growth. In M. Tomasello, & D. I. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 41–78). Mahwah, NJ: Lawrence Erlbaum.
- David, D. A., & Wei, L. (2008). Individual differences in the lexical development of French–English bilingual children. *International Journal of Bilingual Education and Bilingualism*, 11, 598–618.
- Davidson, D., Jergovic, D., Imami, Z., & Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, 24(1), 3–24.
- De Cat, C., Gusnanto, A., & Serratrice, L. (2018). Identifying a threshold for the executive function advantage in bilingual children. *Studies in Second Language Acquisition*, 40, 119–151.

- De Houwer, A., Bornstein, M. H., & Putnick, D. L. (2014). A bilingual-monolingual comparison of young children's vocabulary size: Evidence from comprehension and production. *Applied Psycholinguistics*, 35, 1189–1211.
- Döpke, S. (1998). Competing language structures: The acquisition of verb placement by bilingual German–English children. *Journal of Child Language*, 25, 555–584.
- Dunn, L. M., & Dunn, D. M. (1997). *The British Picture Vocabulary Scale* (2nd ed.). Slough, UK: NFER-Nelson.
- Dunn, L. M., & Dunn, D. M. (2007). *PPVT-4: Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: Pearson.
- Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez Pereira, M., Wehberg, S., . . . Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *British Journal of Developmental Psychology*, 30, 326–343.
- Fennell, C. T., Byers-Heinlein, K., & Werker, J. F. (2007). Using speech sounds to guide word learning: The case of bilingual infants. *Child Development*, 78, 1510–1525.
- Ferjan Ramírez, N., Ramírez, R. R., Clarke, M., Taulu, S., & Kuhl, P. K. (2017). Speech discrimination in 11-month-old bilingual and monolingual infants: A magnetoencephalography study. *Developmental Science*, 20(1), e12427.
- Fiestas, C. E., Bedore, L. M., Peña, E. D., & Nagy, V. J. (2005). Use of mazes in the narrative language samples of bilingual and monolingual 4- to 7-year old children. In J. Cohen, K. T. McAlister, K. Rolstad, & J. MacSwan (Eds.), *Fourth international symposium on bilingualism* (pp. 730–740). Somerville, MA: Cascadilla Press.
- Gathercole, V. C. M., & Thomas, E. M. (2009). Bilingual first-language development: Dominant language takeover, threatened minority language take-up. *Bilingualism: Language and Cognition*, 12, 213–237.
- Gillam, R. B., Ho, T.-H., Bedore, L. M., & Pen, E. D. (2010). Language sample measures and language ability in Spanish–English bilingual kindergarteners. *Journal of Communication Disorders*, 43, 498–510.
- Gopnik, A., & Choi, S. (1995). Names, relational words, and cognitive development in English and Korean speakers: Nouns are not always learned before verbs. In M. Tomasello & W. E. Merriman (Eds.), *Beyond names for things: Young children's acquisition of verbs* (pp. 63–80). Hillsdale, NJ: Lawrence Erlbaum.
- Grissemann, H., Baumberger, W., & Hollenweger, J. (1991). *H-S-E-T: Heidelberger Sprachentwicklungstest: Schweizer Version*. Bern, Switzerland: Hans Huber.
- Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, 24, 267–288.
- Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. De Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–240). Bristol, UK: Multilingual Matters.
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M., & Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39, 1–27.
- Hoff, E., Quinn, J. M., & Giguere, D. (2018). What explains the correlation between growth in vocabulary and grammar? New evidence from latent change score analyses of simultaneous bilingual development. *Developmental Science*, 21, e12536.
- Holman, E. W., Brown, C. H., Wichmann, S., Mueller, A., Velupillai, V., Hammarstrom, H., . . . Egorov, D. (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52, 841–875.
- Hovsepian, A. (2018). Vocabulary growth in Armenian-English bilingual kindergarteners. *First Language*, 38, 47–71.

- Kim, M., McGregor, K. K., & Thompson, C. K. (2000). Early lexical development in English- and Korean-speaking children: Language-general and language-specific patterns. *Journal of Child Language*, 27, 225–254.
- Kovács, Á. M. (2009). Early bilingualism enhances mechanisms of false-belief reasoning. *Developmental Science*, 12, 48–54.
- Legacy, J., Reider, J., Crivello, C., Kuzyk, O., Friend, M., Zesiger, P., & Poulin-Dubois, D. (2016). Dog or chien? Translation equivalents in the receptive and expressive vocabularies of young French–English bilinguals. *Journal of Child Language*, 44, 881–904.
- Legacy, J., Zesiger, P., Friend, M., & Poulin-Dubois, D. (2016). Vocabulary size, translation equivalents, and efficiency in word recognition in very young bilinguals. *Journal of Child Language*, 43, 760–783.
- Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *PPVT 4* (Deutsche Fassung). Frankfurt am Main, Germany: Pearson.
- List, J.-M. (2014). *Sequence comparison in historical linguistics*. Düsseldorf, Germany: Düsseldorf University Press.
- List, J.-M., & Forkerl, R. (2015). *LingPy. A Python library for quantitative tasks in historical linguistics*. With contributions by S. Moran, P. Bouda, J. Dellert, T. Rama, F. Nagel, & S. Greenhill. Version 2.3. Available at <http://lingpy.org>
- List, J.-M., & Moran, S. (2013, August 4–9). An open source toolkit for quantitative historical linguistics. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations* (pp. 13–18). Sofia, Bulgaria.
- Łuniewska, M., Haman, E., Armon-Lotem, S., Etenkowski, B., Southwood, F., Andelković, D., . . . Ūnal-Logacev, Ö. (2015). Ratings of age of acquisition of 299 words across 25 languages: Is there a cross-linguistic order of words? *Behavior Research Methods*, 48, 1–24.
- Marchman, V. A., & Martínez-Sussmann, C. (2002). Concurrent validity of caregiver/parent report measures of language for children who are learning both English and Spanish. *Journal of Speech, Language, and Hearing Research*, 45, 983–997.
- Marton, K. (2016). Executive control in bilingual children: Factors that influence the outcomes. *Linguistic Approaches to Bilingualism*, 6, 575–589.
- Moreno-Martínez, F. J., & Montoro, P. R. (2012). An ecological alternative to Snodgrass & Vanderwart: 360 high quality colour images with norms for seven psycholinguistic variables. *PLoS ONE*, 7(5), e37527.
- Mueller Gathercole, V. C., Thomas, E. M., Viñas Guasch, N., Kennedy, I., Prys, C., Young, N., . . . Jones, L. (2016). Teasing apart factors influencing executive function performance in bilinguals and monolinguals at different ages. *Linguistic Approaches to Bilingualism*, 6, 605–647.
- Munoz-Sandoval, A. (1998). *Bilingual verbal ability tests*. Itasca, IL: Riverside Publishing.
- Pearson, B. Z., Fernandez, S. C., Lewedeg, V., & Oller, D. K. (1997). The relation of input factors to lexical learning by bilingual infants. *Applied Psycholinguistics*, 18, 41–58.
- Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43, 93–120.
- Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1995). Cross-language synonyms in the lexicons of bilingual infants: One language or two? *Journal of Child Language*, 22, 345–368.
- Peña, E., Gutierrez-Clellen, V., Iglesias, A., Goldstein, B., & Bedore, L. M. (2014). *BESA : Bilingual English-Spanish assessment manual*. San Rafael, CA: AR-Clinical Publications.
- Perez-Pereira, M. (2008). Early Galician/Spanish bilingualism: Contrasts with monolingualism. In C. Pérez Vidal (Ed.), *A portrait of the young in the new multilingual Spain* (pp. 39–62). Clevedon, UK: Multilingual Matters.
- Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development*, 82, 1834–1849.

- Place, S., & Hoff, E. (2016). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, 19, 1023–1041.
- Poulin-Dubois, D., Bialystok, E., Blaye, A., Polonia, A., & Yott, J. (2012). Lexical access and vocabulary development in very young bilinguals. *International Journal of Bilingualism*, 17, 57–70.
- Poulin-Dubois, D., Blaye, A., Coutya, J., & Bialystok, E. (2011). The effects of bilingualism on toddlers' executive functioning. *Journal of Experimental Child Psychology*, 108, 567–579.
- Schelleter, C. (2002). The effect of form similarity on bilingual children's lexical development. *Bilingualism: Language and Cognition*, 5, 93–107.
- Schepens, J., Van Der Slik, F., & Van Hout, R. (2013). The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In L. Borin & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 199–230). Boston, MA: De Gruyter.
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism*, 3, 3–25.
- Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 175–208). New York, NY: Holt, Rinehart and Winston.
- Snow, C. E., & Kang, J. Y. (2006). Becoming bilingual, biliterate, and bicultural. In W. Damon & R. M. Lerner (Eds.), *Handbook of child psychology* (pp. 1–3). Hoboken, NJ: John Wiley.
- Stoll, S., Bickel, B., Lieven, E., Paudyal, N. P., Banjade, G., Bhatta, T. N., . . . Rai, N. K. (2012). Nouns and verbs in Chintang: Children's usage and surrounding adult speech. *Journal of Child Language*, 39, 284–321.
- Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 21, 121–137.
- Thomas-Sunesson, D., Hakuta, K., & Bialystok, E. (2018). Degree of bilingualism modifies executive control in Hispanic children in the USA. *International Journal of Bilingual Education and Bilingualism*, 21, 197–206.
- Thordardottir, E., Rothenberg, A., Rivard, M.-E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders*, 4, 1–21.
- Umbel, V. M., Pearson, B. Z., Fernandez, M. C., & Oller, D. K. (1992). Measuring bilingual children's receptive vocabularies. *Child Development*, 63, 1012–1020.
- Unsworth, S. (2015). Amount of exposure as a proxy for dominance in bilingual language acquisition. In C. Silva-Corvalán & Treffers-Daller (Eds.), *Language dominance in bilinguals: Issues of measurement and operationalization* (pp. 156–173). Cambridge, UK: Cambridge University Press.
- Van Der Slik, F. W. P. (2010). Acquisition of Dutch as a second language. *Studies in Second Language Acquisition*, 32, 401–432.
- Wermelinger, S., Gampe, A., & Daum, M. M. (2017). Bilingual toddlers have advanced abilities to repair communication failure. *Journal of Experimental Child Psychology*, 155, 84–94.
- Wichmann, S., Müller, A., Wett, A., Velupillai, V., Bischoffberger, J., Brown, C. H., . . . Valenzuela, P. (2013). *The ASJP database* (version 16).
- Wood, C., Diehm, E. A., & Callender, M. F. (2016). An investigation of language environment analysis measures for Spanish–English bilingual preschoolers from migrant low-socioeconomic-status backgrounds. *Language, Speech, and Hearing Services in Schools*, 47, 123–134.
- Yow, W. Q., & Markman, E. M. (2011). Young bilingual children's heightened sensitivity to referential cues. *Journal of Cognition and Development*, 12, 12–31.