Bilingualism: Language and Cognition

cambridge.org/bil

A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs?

Draško Kašćelan[1], Philippe Prévost[2], Ludovica Serratrice[3,7], Laurie Tuller[4], Sharon Unsworth[5] and Cécile De Cat[6,7]

[1]University of Leeds, Leeds, United Kingdom; [2]Université de Tours, Tours, France; [3]University of Reading, Reading, United Kingdom; [4]Université de Tours, Tours, France; [5]Radboud University, Nijmegen, Netherlands; [6]University of Leeds, Leeds, United Kingdom and [7]UiT The Arctic University of Norway, Norway

Review Article

Cite this article: Kašćelan D, Prévost P, Serratrice L, Tuller L, Unsworth S, De Cat C (2022). A review of questionnaires quantifying bilingual experience in children: Do they document the same constructs? *Bilingualism: Language and Cognition* **25**, 29–41. https://doi.org/10.1017/S1366728921000390

Received: 16 October 2020
Revised: 9 July 2021
Accepted: 11 July 2021
First published online: 8 November 2021

Keywords:
bilingualism; questionnaires; language exposure and use; language skills; activities

Address for correspondence:
Draško Kašćelan, University of Leeds, Michael Sadler Building, Room B.20, Woodhouse, Leeds LS2 9JT, United Kingdom; d.kascelan@leeds.ac.uk, kascelandrasko@gmail.com

Abstract

Questionnaires documenting children's bilingual experience have been used frequently in research on language and cognitive development. However, there has been little investigation of the comparability between these tools. In this review, we (i) provide a list of available questionnaires used to quantify bilingual experience in children; (ii) identify the components of bilingual experience documented across questionnaires; and (iii) discuss the comparability of the measures used to operationalise these components. In doing so, we review 48 questionnaires and identify 32 overarching constructs, manifested as 194 components, and we calculate the frequency with which they are documented. Finally, by focusing on a subset of overarching constructs (language exposure and use, activities, and current language skills), we observe high variability in how they are operationalised across tools. These findings highlight the need for greater transparency in how we document bilingualism and for more comparable measures.

1. Introduction

Bilingualism is intrinsically a multifaceted and heterogeneous construct, relevant to many scientific and applied fields, including linguistics, psychology, neuroscience, education, and speech-language pathology, among others. While these fields might focus on different aspects of bilingualism, their mutual relevance requires a common understanding of how bilingualism is operationalised. Interdisciplinary research requires some objective common ground. Furthermore, within each field, and as an essential condition for robust science, research findings need to be replicable. An example of how the lack of such common ground may potentially complicate our understanding of research findings is the prolific debate on whether bilingualism confers cognitive benefits (e.g., Adescope, Lavin, Thompson & Ungerleider, 2010; Lehtonen, Soveri, Laine, Järvenpää, de Briun & Antfolk, 2018). Some recent opinion papers have tried to provide reasons for the conflicting results and suggested that diverging approaches to the multidimensional nature of bilingualism might be partly responsible for the lack of replicability in findings (Valian, 2015; Bak, 2016). Marian and Hayakawa (2020) and Kremin and Byers-Heinlein (2020) have convincingly argued that bilingualism research would benefit from greater transparency regarding the measures used to operationalise the variables of interest.

Many studies rely on questionnaires to identify the bilingualism profiles of their participants or to estimate how bilingual they are. In the case of children, the information tends to be obtained from parents/caregivers, teachers, and to a lesser extent from the children themselves. While it is now widely acknowledged that several components of bilingualism vary along a continuum (e.g., language proficiency, age of first exposure), data obtained from questionnaires are regularly used to characterise bilinguals through discrete categories, such as simultaneous vs. sequential bilinguals, (child/adult) second language learners, speakers of additional languages, heritage speakers. In some cases, attribution to a category depends on a relatively arbitrary cut-off point. It is however unclear to what extent different questionnaires tap into the same constructs, even when they use identical labels (e.g., language proficiency, language exposure, language mixing, etc.).

The variability in documenting bilingualism has been observed in a number of recent reviews. For instance, Surrian and Luk (2017) reviewed 186 studies from 165 empirical articles published between 2005 and 2015. They noted that specific components of bilingual experience were measured to different degrees and they offered several explanations as to why this might be. For instance, whether a specific variable inquired about differed depending on the geographic region where the study was conducted. In particular, the language of schooling was reported in all of the studies from Asia, Australia, and Africa, while only in 46% of studies from the US, most likely as it was assumed that the language of schooling was English.

© The Author(s), 2021. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Similarly, the inclusion of information on language history depended on the age of participants (children vs. adults). More specifically, language history was reported in 56% of child studies and in 83% of adult studies, which the authors assumed was probably due to the fact that adults have a longer history to report.

Variability in the documentation of bilingual experience was also observed by Li, Sepanski, and Zhao (2006). In a review of 41 studies, Li et al. (2006) identified the dimensions that most researchers measured through language background questionnaires. Although their review included mostly studies conducted with adults, it informed their creation of a freely available questionnaire (the Language History Questionnaire) which can be adapted for use with children. This questionnaire has since been adopted in many studies, enabling the use of common scales to measure various components of bilingualism, and it has been updated twice since (see Li, Zhang, Tsai & Puls, 2014; Li, Zhang, Yu & Zhao, 2019). While Li and colleagues observed that some components of bilingualism were documented less often than others (e.g., writing ability in 36.6% of studies, frequency of speaking L1 at home in 7.3% of studies, etc.), they did not review or discuss differences in how these dimensions were operationalised across studies. For instance, reading or speaking ability were both documented in 41.5% of the studies surveyed, but no further information was provided about the comparability of scales used across the studies.

De Bruin (2019) offers some discussion of the importance of describing measures used to quantify bilingualism in relation to the following components: age of acquisition, proficiency, language use, language switching and language context. Her review focuses on a number of questionnaires designed for use with adults (Language Experience and Proficiency Questionnaire, LEAP-Q, Marian, Blumfield & Kaushanskaya, 2007; Language and Social Background Questionnaire, LSBQ, Anderson, Mak, Keyvani Chahi & Bialystok, 2018; Bilingual Switching Questionnaire, BSWQ, Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman & Münte, 2012) and on specific studies (e.g., in case of age of acquisition: Luk, De Sa & Bialystok, 2011; Pelham & Abrams, 2014; Paap, Johnson & Sawi, 2014; Tao, Marzecová, Taft, Asanowicz & Wodniecka, 2011). While focusing on how the documentation of these components could be improved, de Bruin (2019) illustrates the variability in their operationalisation, especially regarding age of acquisition and proficiency. For instance, age of acquisition can be documented differently across different studies: as the start of language acquisition/learning, as the age of arrival in the new country (in case of immigrants), as the age of fluency in the second language, as the age at which bilinguals started using languages actively on a daily basis, as the age of first exposure to a language, or as the age of formal classroom instruction. Similarly, when proficiency is estimated based on questionnaires, the resulting scales can vary widely (e.g., 1-7-point scales, 1-10-point scales).

In sum, previous reviews have considered how frequently specific components of bilingual experience are operationalised or how much variability there is in the operationalisation of particular components of bilingual experience. What is currently missing is a comprehensive review of the components of bilingual experience documented across questionnaires, considering the commonality of the components included as well as how they are operationalised. We aim to fill that gap.

Our study focuses on questionnaires used to document bilingualism in children (0–18 years) as estimating their bilingualism poses a specific set of challenges. Some of these include practicality issues, such as generally having to rely on carers' reports because children are often unable to complete reports themselves. Nevertheless, while we might rely on different informants and while different components of bilingual experience might be documented for children as opposed to adults (see Surrian & Luk, 2017), we expect there to be substantial overlap. As such, our findings will also be relevant to research on adults. The use of questionnaires with adult participants faces similar issues, such as determining the most felicitous ways of grouping bilinguals in specific categories or using questionnaire data to create composite scores that characterise bilinguals.

Our aim is not to argue for or against the use of composite measures - such as language entropy (Gullifer & Titone, 2020) or the continuous factor score method advocated by Anderson et al. (2018). Furthermore, while acknowledging the multidimensional nature of bilingualism, our purpose is not to discuss the relative importance of each dimension - such as language proficiency (Hulstijn, 2012) or child-level and context-level variables (Byers-Heinlein, Esposito, Winsler, Marian, Castro & Luk, 2019a), nor to argue about the relative merits or the comparability of categorical vs. continuous approaches to bilingualism (Kremin & Byers-Heinlein, 2020). Rather, we will focus on the raw measures used to operationalise bilingualism in order to:

1. provide a list of available questionnaires used to quantify bilingual experience in children;
2. identify the components of bilingual experience that are documented across questionnaires;
3. discuss the comparability of the measures used to operationalise these components.

2. Methodology

2.1 A systematic review

Our initial aim was to conduct a systematic review. To this end, in October 2019, we searched PsycINFO, Embase Classic + Embase, ERIC, Web of Science, and Scopus by using a list of key terms. These were: Key terms 1 [bilingual* OR multilingual*] AND Key terms 2 [questionnaire* OR assess* OR tool* OR measur* OR report* OR estimat* OR rating* OR instrument* OR quantif* OR survey*] AND Key terms 3 [dominance OR proficiency OR 'length of exposure' OR 'length of residency' OR impairment* OR codeswitching OR code-switching OR code-mixing OR ses OR 'parental education' OR 'input quality' OR 'input quantity' OR 'diversity of linguistic environment'] AND Key terms 4 [child* OR pupil* OR kid*]. This yielded a large number of papers across databases: PsycINFO (1,098), Embase Classic + Embase (418), ERIC (1,442), Web of Science (1,026), Scopus (655).

Several reasons prevented us from further pursuing the systematic review, however. First, questionnaires quantifying bilingual experience were often not the focus or the main topic of research articles. The use of questionnaires was therefore rarely mentioned in the abstract, requiring scanning the entire paper. Second, the very large range of research topics associated with the operationalisation of bilingualism made it extremely challenging (if not impossible) to identify them all with a single search string. Third, even among studies reporting the use of questionnaires to quantify bilingual experience, the questionnaires themselves tended not to be included in the papers or even available. As an illustration, in the review by Li et al. (2006), out of 41 studies which used questionnaires, only 7 included them in the

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

publication. These challenges led us to adopt an alternative approach to the identification of relevant questionnaires.

2.2 An alternative approach

Initially, 13 questionnaires were already available to us at the beginning of the review process (these questionnaires are marked with an asterisk in section A in the supplementary material). We created a Google Form survey open to all bilingualism researchers, asking them whether they had used any of these 13 questionnaires, modifications of these questionnaires, or any other questionnaires designed by themselves or by anyone else. Respondents were also asked to email us any questionnaires or questionnaire modifications that they were familiar with and that were not on our original list. This survey was advertised on social media (Facebook, Twitter, LinkedIn), via specialist websites, groups and mailing lists (The LINGUIST List, Info-CHILDES, Bi-SLI COST Action) and at the Dutch national research meeting on language development in children (Amsterdam, 20 November 2019). At the same time, we did a non-systematic search for available questionnaires across the bilingualism literature, including a Google search exploiting some of the key terms listed in section 2.1, and through the resources available at the UCLA National Heritage Language Resource Center. Finally, several researchers familiar with our review project emailed us their own questionnaires. The collection process lasted between October 2019 and June 2020. The number of relevant questionnaires identified, as well as the exclusion criteria, are provided in the results section.

2.3 Analytic strategy

To identify the components of bilingual experience that the questionnaires were designed to document, we adopted an inductive method, allowing categories to emerge from the list of raw measures documented by the questionnaires. Using this method, we identified (a) overarching constructs, (b) specific components of these overarching constructs, and (c) how each component was operationalised. An illustration is provided in Table 1. The coding procedure was as follows: the first author examined each questionnaire and classified the questions into the categories which were emerging. Subsequently, the first author repeated the process to check for any inconsistencies in coding. Questions which were difficult to categorise were flagged and discussed between the first and the last author until consensus was reached. In the list of overarching constructs and their components (see supplements, section B), the first author noted any peculiarities in classification or explanations for components that might not be straightforward to understand. All authors looked at these notes and commented if further clarifications were necessary, which were then addressed by the first author.

In total, 32 overarching constructs were identified, as well as 194 components, each operationalised in a variety of ways. A detailed list of overarching constructs and their components can be found in section B in the supplementary materials. Where necessary, an explanation of what each component embodies is provided.

In order to be able to compare questionnaires, we had to decide on a set of terms to use which do not necessarily reflect the circumstances of all of the communities the questionnaires were intended for. We chose to use the labels home language (HL) and societal language (SL) to refer to the languages of bilinguals. This enabled us to capture a variety of labels in the questionnaires (e.g., mother tongue, L1-L4, language A/B/C, other language, target language, additional language, country language, language X/Y, etc.). Whenever the questionnaire included specific language names (e.g., English and Spanish), we assigned them the label HL or SL depending on the context. For instance, if a questionnaire was created by a US team of researchers and it was aimed at Spanish–English bilinguals in the US, Spanish was labelled as the HL, and English as the SL. While the HL/SL labels are not universally adequate, they provide a practical solution to identify the child's languages for the purpose of this review. For the questionnaires documenting more than two languages, we mapped the two main ones onto HL and SL, and noted that information on one or more other languages was also collected.

Finally, in order to address our third research aim (i.e., discuss the comparability of the measures used to operationalise identified components), we looked into the ways in which questionnaires documented components of the overarching constructs. Due to space limitations, and by way of illustration, we focus on a few overarching constructs which are common across questionnaires: exposure and use, activities, and current skills in HL and SL.

3. Results

3.1 Questionnaires

The first research aim was to provide a list of available questionnaires used to quantify bilingual experience in children. We identified 81 questionnaires, 33 of which had to be excluded from analysis (as shown in Figure 1). Reasons for exclusion included: being designed for quantification of bilingualism in adults (n = 25), being designed for use with foreign language learners (n = 2), inquiring mostly about a specific bilingual school (n = 1), extreme brevity (n = 1; this questionnaire included only three language-related questions), focussing on something other than language (n = 1), focus on speech and language problems rather than bilingual language experience (n = 1), the impossibility of translating the questionnaire (from German) within our timeframe (n = 1), being a duplicate of another questionnaire (n = 1). Before the coding procedure started, all authors agreed that questionnaires designed for use with adults would be excluded, even though in practice they might be adapted for use with children. Other reasons for exclusion were raised by the first author after reading through the questionnaires (e.g., whether to exclude the questionnaire for foreign language learners or the one focusing on a specific bilingual school). These decisions were then discussed and approved by all authors. The remaining 48 questionnaires were included in the analysis.

Out of these 48 questionnaires, most were in English (n = 42), 3 had a version in both English and Russian, 1 was in both Spanish and English, 1 was in Spanish only, and 1 was in Dutch only. Note that in practice, some of these questionnaires have translations in other languages. Here we reported the breakdown of the versions collected in our dataset. Five questionnaires were to be administered to teachers, 4 to children themselves, and the other 39 to parents/caregivers. See section A of the supplementary materials for a complete list of questionnaires (Supplementary Materials).

3.2 Documented components across the questionnaires

The second aim of this review was to identify components of bilingual experience documented across questionnaires. We

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

Table 1. An example of an overarching construct, its components and operationalisations

Overarching construct	Component measured	How is it operationalised?
Exposure	Relative frequency of exposure	Mother to child (5-point scale, quantifying adverbs)
		Each caregiver separately: frequency of speaking each language to the child (3-point scale, quantifying adverbs)
		Siblings with child (6-point scale, quantifying adverbs)
		Exposure to languages at home with parents (5-point scale, percentages)
	Language mixing	Rules about mixing languages at home (open-ended, specified by respondent)
		For caregiver 1, when with the child: mark caregiver's frequency of starting a sentence in the societal language and switching to the home language (4-point scale, adverbs)
		For caregiver 1, caregiver 2, additional caregiver, siblings (all together), relatives, friends/peers, community members: do they mix languages when they speak to the child (4-point scale: yes, no, word switch, phrase switch). For each interlocutor, specify languages mixed.

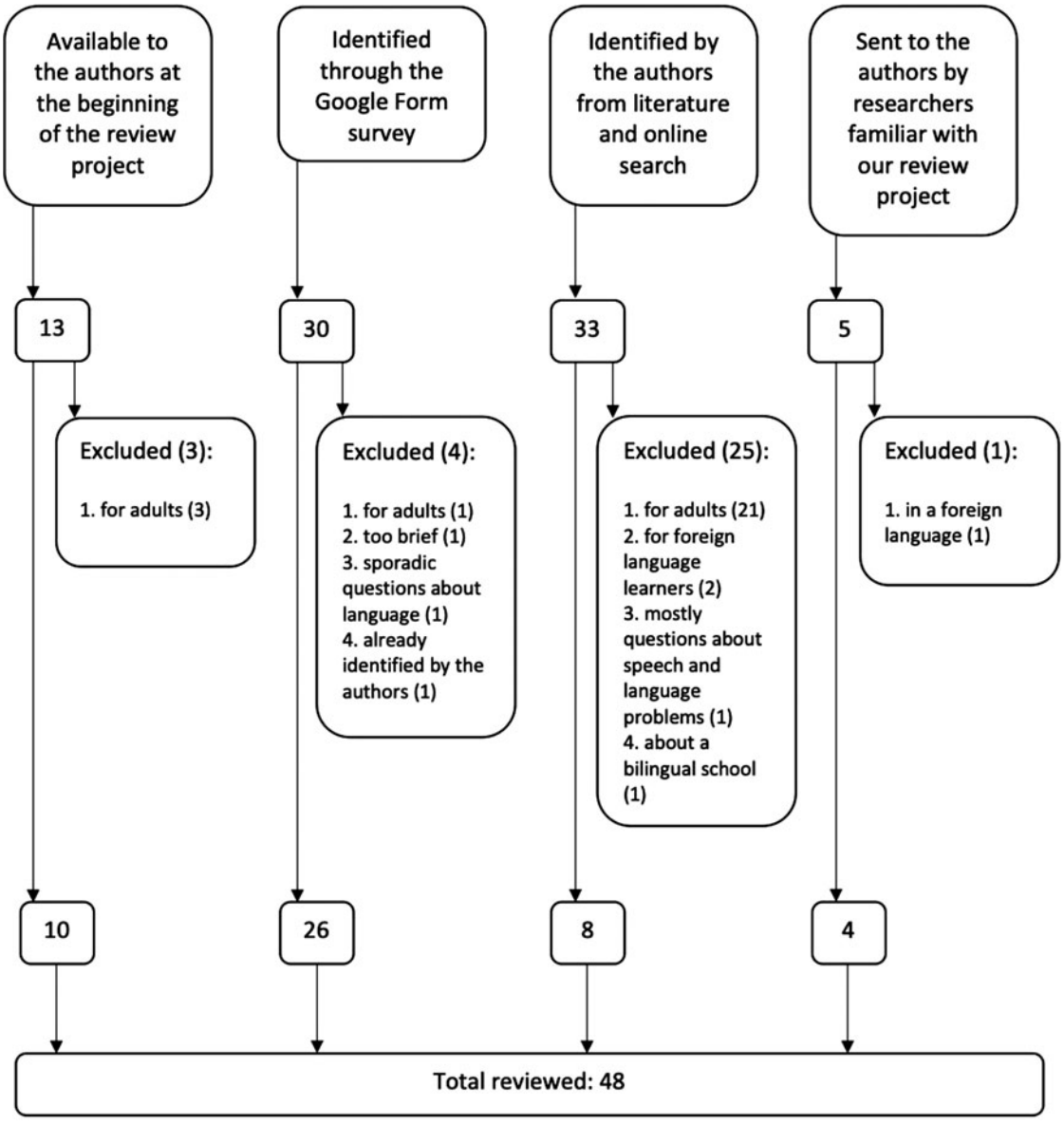


Figure 1. Number of questionnaires included in and excluded from the review

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

identified 32 overarching constructs and their 194 components. Considering the large number of constructs and their components, space limitations prevent us from outlining all of them in the paper. We therefore present those constructs that in our judgement are the most informative of an individual's level of bilingualism. These include language exposure and language use (Table 2), current skills in HL and SL (Table 3), and activities in each language (Table 4). Our decision to focus on these particular constructs was also informed by previous work in the field. Looking at the variables which emerged in Li et al.'s (2006) review of 41 studies that used language history questionnaires, apart from the demographics (e.g., age and years of residence), most other variables were exposure-, use- or language skills-related. Furthermore, input quantity (operationalised through exposure- and use-related variables) and input quality (often estimated through involvement in particular activities) have been shown to play a significant role in language outcomes depending on child's age and the context (see Unsworth, 2016).

Section B of the supplementary material contains a complete list of the overarching constructs and their components, as well as more comprehensive clarifications regarding what each documented component embodies (Supplementary Materials).[1] Section C of the supplementary material (Tables S1-S7) outlines the frequency of constructs/components not discussed here (Supplementary Materials). Those constructs certainly require attention in future work as specific research questions might find them to be relevant factors.

In the first column of each table, next to the name of each overarching construct, we specified what percentage of questionnaires documented that particular construct in at least one way. The second column lists the components of that construct, while the numbers and percentages in the third and the fourth column show how many out of the 48 questionnaires documented each component. Note that the tables do not indicate in how many DIFFERENT WAYS each component was operationalised, as this varied widely across questionnaires and it would be impossible to visualise in a comprehensive way. No matter in how many different ways a questionnaire operationalised a certain component, it was counted only once. For instance, the frequency of relative exposure was documented in 29 out of 48 questionnaires (see Table 2). However, in some questionnaires, it might be documented in one way only (e.g., with two main caregivers), while in others it might be estimated in several ways (e.g., with household members, in school and outside of school, while reading or while listening, etc.).

Table 2 relates to exposure and use. While these two constructs are presented separately in the table below, the distinction was not always made explicit across questionnaires. Where no (clear) distinction was made between exposure and use, the question was counted under the construct 'exposure'. Furthermore, we avoided introducing components such as 'cumulative exposure/use' or 'weighted exposure/use', as the ways in which these composite measures were calculated across some tools varied substantially. Rather, we considered raw estimates only – that is, the precise information gathered by the questionnaires. For instance, the component 'frequency of exposure (relative)' embodies all questions about the target child's language exposure documented in relative terms (e.g., using percentages; frequency adverbs such as 'never', 'sometimes', 'often', 'always'). In some cases, questionnaires asked about the overall relative exposure to a language; in other cases, they did so with respect to a particular type of language experience (e.g., reading). Both were classified here under 'frequency of exposure (relative)'.

Table 3 shows the components documenting the target child's language skills in the HL and SL (leaving aside other potential languages).

Table 4 relates to the target child's activities, as well as explicit questions about the quality of the child's language context.

As presented in the Tables 2–4, as well as the Tables S1-S7 in the supplementary materials (see section C, Supplementary Materials), there is substantial variability as to which constructs

Table 2. Exposure and use variables documented across 48 questionnaires (number and percentage of questionnaires documenting each component)

Overarching construct (%)	Components measured	n	%
Exposure (96%)	Number of interlocutors	15	31%
	Changes in the number of interlocutors	2	4%
	Nativeness of interlocutors	2	4%
	Dialect of interlocutors/ exposure	5	10%
	Early exposure[a]	9	19%
	Changes in exposure	12	25%
	Age/Period of exposure (in relation to interlocutor/ context)	17	35%
	Hours per day	20	42%
	Hours per week	16	33%
	Weeks per year	1	2%
	Sleeping/ Waking hours	6	13%
	Frequency of exposure (relative)	29	60%
	Interlocutor/ Context/ Activities (language used)	31	65%
	Language mixing	6	13%
	Interlocutor other	3	6%
Use (73%)	Early use	1	2%
	Changes in use	1	2%
	Frequency of use (relative)	19	40%
	Hours per day	5	10%
	Interlocutor/ Context/ Activities (language used)	17	35%
	Most spoken (to)[b]	2	4%
	Language mixing	8	17%

Note.
[a]This component includes operationalisations about exposure until a certain point in early life, with an exception of one questionnaire, which inquired about exposure until the 12th year in child's life (which we did not consider early life). The same is the case for 'Early use' (see below in the same table).
[b]Most frequent language or interlocutor.

[1]This supplement section also includes additional notes to flag overlaps – that is, instances of operationalisations which were classified under a particular component but might overlap with another component or construct. Readers are invited to browse through the supplements in addition to inspecting the tables.

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

Table 3. Child proficiency variables documented across 48 questionnaires (number and percentage of questionnaires documenting each component)

Overarching construct (%)	Components measured	n	%
Early skills HL (child)[a] (2%)	Speaking	1	2%
	Understanding	1	2%
Early skills SL (child)[a] (2%)	Speaking	1	2%
	Understanding	1	2%
Current skills in the SL (child) (42%)	SL speaking/ production	17	35%
	SL listening/ comprehension	15	31%
	SL reading	6	13%
	SL writing	5	10%
	SL overall proficiency	3	6%
Current skills in the HL (child) (42%)	HL speaking/ production	18	38%
	HL listening/ comprehension	14	29%
	HL reading	6	13%
	HL writing	5	10%
	HL reading and writing	1	2%
	HL overall proficiency	4	8%
	Reasons for dissatisfaction	1	2%
Other language-related skills/ characteristics (child) (4%)	Talkativeness	1	2%
	Relevant language	2	4%
	Questions	1	2%

Note. [a]Up to a certain point in life.

Table 4. Variables related to activities and context quality documented across 48 questionnaires (number and percentage of questionnaires documenting each component)

Overarching construct (%)	Components measured	n	%
Context quality (2%)	Outside home	1	2%
	Holiday	1	2%
	Other sources	1	2%
Activities (44%)	Reading in HL and/or SL	8	17%
	Literacy and other in HL and/ or SL	12	25%
	Preferred activities	2	4%
	Ease of learning new things (how quickly?)	1	2%
	Activity patterns	1	2%
	Extra-curricular activities	3	6%
	Courses/ Classes	7	15%
	Travel/Holidays	7	15%
	Other activities	2	4%

were documented across the 48 questionnaires surveyed. Excluding the demographics of children (in 94% of questionnaires) and the demographics of parents/caregivers/informants (in 75% of questionnaires), only the following constructs were included in more than 50% of reviewed tools: language exposure (96%), language use (73%), type of bilingualism (60%), socio-economic status (58%), and developmental issues/concerns (56%). Components documented in 40-50% of questionnaires include: activities (44%), children's current skills in SL and HL (both in 42% of questionnaires), information about siblings (42%), and SL skills/quality of the first caregiver (42%). Other constructs were included in less than 40% of questionnaires.

We now turn to the third research aim: estimating the comparability of the measures used to operationalise key components of bilingual experience.

3.3 Comparability of measures used across questionnaires

As justified above, we focus on the components presented in Tables 2 through 4 only: language exposure and language use, current skills in HL and SL, and activities in each language. Our discussion of the components related to language exposure and language use is organised thematically below: relative frequency of exposure and use; language used with interlocutors, in contexts and during activities; age/period of exposure in relation to interlocutors/contexts; time-unit-based measures of exposure and use; language mixing; other exposure- and use-related components. Operationalisations of any components other than these, presented in the tables in section C of the supplementary materials, are referred to where relevant (Supplementary Materials).

Exposure and use

Most questionnaires documented language exposure, but not all documented language use (see Table 2). Furthermore, when language use was documented, this was not always separate from language exposure. For instance, informants might be asked to estimate how frequently the child "hears or speaks" each of their languages. In some cases, although the questionnaire appeared to document language use only, both exposure and use were conflated in the response scale. For instance, in the Teacher Questionnaire by Gutiérrez-Clellen and Kreiter (2003), respondents are asked to document the child's 'language use', clearly defined as how much the child uses each language. However, the scale the participants are invited to use contains descriptors which refer to both speaking and hearing: "0 = Never uses the indicated language. Never hears it.", "1 = Never uses the indicated language. Hears it very little.", "2 = Uses the indicated language a little. Hears it sometimes.", "3 = Uses the indicated language sometimes. Hears it most of the time.", "4 = Uses the indicated language all of the time. Hears it all of the time.", "DK = Don't know". This overlap potentially compromises the accuracy of such measures as it is unclear which dimension the informants will have considered. In turn, this can affect the comparability of measures collected with different questionnaires. In this review, we only classified under 'language use' those questions which explicitly referred to the child's output and only if exposure was documented separately. In other words, when a question about 'use' was contrasted with a question about 'exposure', we assumed that 'use' could be interpreted as a synonym of 'speak'. When there was just one question targeting 'use' only, we assumed that the questionnaire designer intended it to mean

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

'interact'. In this latter case, the relevant questions were classified under 'language exposure', as a construct necessary to document.

Relative frequency of exposure and use

When a questionnaire asked about language exposure or use, it was usually in relative terms – that is, comparing frequency of exposure/use of one language compared to the other. This was the case in 60% of questionnaires for exposure, and 40% for use. Relative estimates were based on different ways of apprehending language experience: as overall exposure/use (e.g., an estimation of overall exposure to each language), as exposure/use with particular interlocutors (e.g., "Mark how often mother uses each language with the child"), in specific contexts (e.g., exposure to each language at school), or during certain activities (e.g., frequency of reading in each language).

Within each of these approaches to language experience, there was also substantial variability in the approach to quantification. For instance, relative exposure to languages in interactions with specific interlocutors was documented in a number of ways. These included 3- to 6-point scales using quantifying adverbs/descriptors, percentage-based questions (e.g., caregiver 1 to child: HL 100% and SL 0%, HL 90% and SL 10%, etc.), a combination of frequency adverbs and percentage scales (e.g., father to child: hardly ever SL and almost always HL = 0% SL, seldom SL and usually HL = 25% SL, etc.), or open-ended questions (e.g., "How often does a sibling speak each language to the target child?"). Even when scales included an identical number of points (e.g., 5-point scales), these also varied in how they were labelled (e.g., De Cat, 2020: always, usually, half the time, rarely, never; Wilson, 2017: never, rarely, sometimes, usually, always; Gunning & Klepousniotou [child questionnaire], n.d.: never, rarely/a bit, half and half, usually/a lot, always). In some cases, these label differences affected the precision of the scale. For instance, the comparability of scales in (1) and (2) could be limited across studies as the highest point in (1) is distributed over the highest two points in (2), thereby leaving the one but highest point in (1) without an equivalent in scale (2).

(1) never, rarely, sometimes, usually, very often/always (PaBiQ, Tuller, 2015)
(2) never, rarely, sometimes, quite often, always (Parent Questionnaire, Arredondo, 2017)

Potential issues with data comparability also arose with percentage scales, which in principle would be expected to ensure more equivalence between the tools. First, the numerical distance between the points on the scale varied across questionnaires (e.g., 10% or 20% in Lyutykh, 2012, vs. 25% in Unsworth, 2013 or in Cattani, Abbot-Smith, Farag, Krott, Arreckx, Dennis & Floccia, 2014). Second, some scales combined percentages and adverbs (Unsworth, 2013 or Cattani et al., 2014), while others do not (Lyutykh, 2012). Third, as questions about each of the child's languages were usually kept separate, there is a possibility that the sum of percentages could add up to more than 100% (of total language experience). For instance, in documenting languages used between the child and a caregiver, the caregiver might list 50% for one language, 30% for another and 30% for a third one. Some questionnaires avoid this by combining the two languages in the scale (e.g., Lyutykh, 2012 uses the following scale: HL 100% and SL 0%, HL 90% and SL 10%, etc.), or by automatically calculating one on the basis of the other (e.g., Unsworth, 2013). Some explicitly state that the percentage of language exposure for each interlocutor should add up to 100% (e.g., Blume & Lust, 2017). The same observations also applied to estimates of the relative frequency of language use.

Finally, variability across questionnaires increased even more when we consider how relative exposure and/or use was operationalised in different contexts (e.g., in school, at home) or during certain activities (e.g., playing games, watching TV, when reading or when being read to), rather than with specific interlocutors alone.

Language used with interlocutors, in contexts and during activities

Another frequently documented component included languages used in interactions with interlocutors, in specific contexts, and during specific activities (for exposure in 65% of questionnaires and for use in 35% of cases). This component differed from the previous one in the sense that it was not estimated in a relative way. Here, for each interlocutor, context or activity, the informant has to provide details on which language or languages are used. In principle, this component was therefore documented more consistently across the questionnaires. There are, however, two notable points of divergence. The first is that the number of interlocutors, contexts and activities is either pre-specified (e.g., de Diego-Lázaro, 2019; De Houwer, 2009) or open-ended (e.g., DeAnda, Bosch, Poulin-Dubois, Zesinger & Friend, 2016). While the pre-specified approach offers more comparability between the datasets collected with the same tool, the open-ended approach offers more flexibility, and consequently less comparability. The second difference relates to the ways of documenting languages used with or by interlocutors. This varied in terms of whether the interlocutors were observed separately or as a group. For instance, De Houwer (2009) documents the language that each sibling speaks to the child, while De Cat (2020) asks about the language used with siblings as a group. Similarly, in the measures of language use (i.e., the languages specifically used by the child when communicating with interlocutors), Marinis ([parent questionnaire], 2012) asks about siblings as a group, whereas de Diego-Lázaro (2019) allows a separate entry for each sibling.

Age/Period of exposure in relation to interlocutors/contexts

Just over a third of questionnaires documented children's language experience history (or part of it) in relation to specific interlocutors and/or contexts. This was in addition to the 29% of questionnaires that enquired about children's language experience history as a whole (see Table S2, 'Date/Age of exposure/acquisition'). Looking at the way in which the age/periods of exposure to languages are operationalised in relation to interlocutors and contexts, we observe diverse levels of granularity. A large set of questionnaires asked about the length of exposure (across different interlocutors or contexts) from a specific age, which is either pre-specified or to be determined by the informant (Antonijevic-Elliott, n.d.; Antonijevic-Elliott, Lyons, O' Malley, Meir, Haman, Banasik, Carroll, McMenamin, Rodden & Fitzmaurice, 2020; Scharff-Rethfeldt, 2012a; Arredondo, 2017; Gagarina, Klassert & Topaj [questionnaires for preschool and school children], 2010; Prentza, Kaltsa, Tsimpli & Papadopoulou [questionnaires for children and parents], 2017; De Houwer, 2002; Byers-Heinlein, Schott, Gonzalez-Barrero, Brouillard, Dubé, Laoun-Rubsenstein, Morin-Lessard, Mastroberardino, Jardak, Pour Iliaei, Salama-Siroishka & Tamayo, 2019b; Blumenthal & Julien, 2000; de Diego-Lázaro, 2019). Another set of questionnaires

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

documented language experience separately for each year of life, with the important difference that the number of years varies across the questionnaires (first eight years of life in Peña, Gutiérrez-Clellen, Iglesias, Goldstein & Bedore [BESA:BIOS, Parent], 2018; every year until the child's age at time of testing in Unsworth, 2013; three years of preschool and five years of primary school in Cohen, 2015a). Finally, other questionnaires inquired about this information using age bands (e.g., from birth until 1, between 1 and 3, etc.), either to be specified by the informant (DeAnda et al., 2016; Blume & Lust, 2017) or pre-specified (De Houwer, 2002; Prentza et al. [parent questionnaire], 2017). In addition to these diverse levels of granularity, across questionnaires information about age/period of exposure was not always documented for all languages of the child or with the same interlocutors and in identical contexts.

Time-unit-based measures of exposure and use

Questionnaires varied substantially in time-unit measures. For instance, language exposure could be documented as: hours per day (in 42% of questionnaires),[2] hours per week (in 33% of questionnaires), weeks per year (in 2% of questionnaires). The average number of waking hours was rarely documented (in 13% of questionnaires). Questionnaires also varied as to whether they asked about the time that the child spends with specific interlocutors (e.g., Gutiérrez-Clellen & Kreiter [parent questionnaire], 2003), the time that the child is in contact with a specific language (e.g., Cohen, 2015a, 2015b), or the time spent in each language with an interlocutor or doing an activity (e.g., de Diego-Lázaro, 2019). Furthermore, daily estimates were documented in different ways. Some questionnaires ask about each day of the week (e.g., Gutiérrez-Clellen & Kreiter [parent questionnaire], 2003; Blume & Lust, 2017), while others further distinguish term-time weeks from holiday weeks (e.g., Cohen, 2015a, 2015b). One questionnaire documents hours per day for each day of the week, but each day is recorded in a separate week (De Houwer & Bornstein, 2003). Other daily estimates included hours per day on an average weekday and on an average weekend day (e.g., De Cat, 2020) or average hours per day of doing an activity on school days and separately on weekends (e.g., watch TV, De Houwer, 2002).

Language mixing

We adopt a broad interpretation of the term 'language mixing', encompassing what is variously referred to as 'code-switching', 'code-mixing', 'language mixing', 'language switching', 'borrowing', etc. While language mixing is gaining more attention in the child bilingualism literature, only 13% of questionnaires asked about it with reference to children's language exposure, and 17% with reference to their language use. How it is documented varied substantially across questionnaires.

For exposure, this component was documented through questions targeting: interlocutors' mixing when speaking with the child (each interlocutor separately, Blume & Lust, 2017), parents using both languages in conversation with the child (De Houwer, 2017), relative frequency of code-switching by each parent with the child (Wilson, 2017), mixing rules in the home (Read, Contreras, Rodriguez & Jara, 2020), caregiver's frequency of borrowing, switching, translating a word/phrase when with the child (Read et al., 2020), caregiver responding in a language different from the one in which the child speaks (De Houwer, 2017; Prentza et al. [parent questionnaire], 2017), or about the frequency of caregiver switching languages in a sentence, mixing languages in general, and borrowing a word from the other language (Byers-Heinlein, 2013).

For language use, there was even more variability in how language mixing was documented. Questions targeted: language mixing with each interlocutor separately (Blume & Lust, 2017), types of language mixing (Blume & Lust, 2017; Scharff-Rethfeldt, 2012a), the relative frequency of the child code-switching with each parent separately (Wilson, 2017), child language mixing when they speak (Özturk, n.d.), frequency of children's mixing in speaking and writing (Lyutykh, 2012), child mixing in the company of bilinguals or monolinguals (Scharff-Rethfeldt, 2012a), starting a sentence in one language and finishing it in another (Gunning & Klepousniotou [child questionnaire], n.d.), frequency of using both languages in a single sentence (De Houwer, 2017), the ease of language switching (De Houwer, 2017), frequency of responding in one language when being asked in another (in both speaking and writing, Lyutykh, 2012), frequency of translating words from one language to another (Lyutykh, 2012), frequency of using words from one language when speaking another (Prentza et al. [parent questionnaire], 2017), not using the same language that the parents speak to them (De Houwer, 2017).

Other exposure- and use-related components

A number of additional components relating to language exposure and use were also documented. Some of these were more frequent, such as the following exposure-related components: number of interlocutors (31%), early exposure (19%), and changes in exposure (25%). Questionnaires varied in the range of interlocutors targeted: for example, in the home, at (pre)school (Gagarina et al. [questionnaire for school children], 2010), or frequent interlocutors in other settings (e.g., Blume & Lust, 2017). Questions about early exposure mostly differed in terms of the point in the child's life until which the questionnaire documents information: for instance, before the age of 4 (e.g., DeAnda et al., 2016; Tuller, 2015), before preschool (e.g., Gagarina et al. [questionnaires for preschool and school children], 2010), within the first two years (e.g., Scharff-Rethfeldt, 2012a), until the age of eight (e.g., Gutierrez-Clellen & Kreiter [parent questionnaire], 2003), etc. Changes in exposure were also documented in a variety of ways, such as changes over the years in the languages used by the adults or children in the household (e.g., Antonijevic-Elliott et al., 2020), changes in language situation before entry into preschool (e.g., Gagarina et al. [questionnaires for preschool and school children], 2010), or languages that the child does not hear or need anymore (e.g., Scharff-Rethfeldt, 2012a).

Finally, there were several less frequently documented components of exposure and use. For language exposure, these included: changes in the number of interlocutors (2%), nativeness of interlocutors (4%), dialect of interlocutors/exposure (10%), and other interlocutor related questions mostly about friends' demographics (6%). For language use, less frequently documented components included the questions about early use (2%), about changes in the use (2%), and questions about the language or the interlocutor that the child speaks the most (in)to (4%).

[2]Note that when a questionnaire asks about hours per day for each day of the week, it is possible to calculate hours per week that the child spends with a specific interlocutor, in a certain context or doing an activity. Therefore, some of the operationalisations of hours per day could have been classified under the component hours per week. We decided to keep them under the hours per day, as that was the smallest unit of measure.

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

In the following section, we present components related to the documentation of activities, some of which have already been included in the exposure/use section.

Activities

Under the overarching construct 'activities', nine different components emerged. As Table 4 illustrates, these components were operationalised in just 2-25% of questionnaires. This low number is mostly due to our approach in classifying activity-related questions. Specifically, as explained in a previous section, given that involvement in activities often reflects a child's exposure and use, many activity-related questions were classified under exposure- and use-related components. This was done whenever the frequency of doing a specific activity was operationalised in one of the following ways: in relative terms (e.g., in HL always, in SL never), in terms of hours per day, hours per week, or in terms of language used during the activity (with an interlocutor or in a context). These patterns of documenting frequency emerged quite clearly under the constructs of exposure and use during the inductive coding procedure. Hence, we classified activity-related questions of this kind under those categories. The purpose of this approach was to ensure consistency in classifying measures in their raw form (i.e., in the manner in which they were documented or asked about) rather than relying on whether a specific questionnaire considered involvement in a given activity as a part of the exposure/use quantification (in one of the ways listed above) or as an estimate of another domain (e.g., exposure/use richness, exposure/use quality). Therefore, operationalisations listed under the activity-related components presented in Table 4 include any questions which (a) queried the frequency of activities in a way different from operationalisations under the constructs of exposure and use (listed above) or (b) any other activity-related questions, such as the language of the written material in the house, attending HL/SL classes, preferred activities, etc.

The most commonly documented component among the activities was 'literacy and other activities in HL and/or SL' (in 25% of questionnaires). Some of the operationalisations encountered under this component included: times per week of reading, using computer, TV, watching movies, storytelling, or singing songs separately for SL and HL (Paradis, 2011), specification of languages in which the written material (e.g., books, newspapers, periodicals) is available at home (Blume & Lust, 2017), number of books in the home (if any) for SL and HL separately (either less than 10 or more than 10, McKendry & Murphy, 2011), frequency of parents borrowing or buying books separately in each language to be rated on a scale consisting of never, a few times a year, 1-2 times a month, 1-2 times a week, 4-5 times a week, usually daily (Cohen, 2015a, 2015b), using computers outside of school for games, for staying in touch with family, for staying in touch with friends, or for looking at websites (McKendry & Murphy, 2011), or specifications of the alphabet used when texting or emailing in HL (HL alphabet, Latin alphabet, both, Prentza et al. [child questionnaire], 2017). In addition, there were other frequency-related questions using diverse scales and inquiring about the number of times per week/month/year (or a combination of these) that the activity was conducted. Furthermore, there were various operationalisations about the written material in the home.

Reading in HL and/or SL was the next most frequently documented activity (in 17% of the questionnaires). Questions included asking parents to specify if they read to their child at least twice a week (completely agree, more or less agree, not quite agree, entirely disagree, De Houwer, 2017), or about the frequency of reading activities for each language separately (never or almost never, at least once a week, every day, Tuller, 2015). Sometimes the frequency of reading was documented separately for the child reading alone and for the parents/caregivers reading with the child (e.g., Cohen, 2015a, 2015b, on a 6-point scale; Arredondo, 2017, on a 4-point scale). There were also questions about who reads the most with the child and whether the comments/questions by the child during a story depend on the language of reading (Read et al., 2020). Other examples included questions about the place where the child started learning to read in each language (Arredondo, 2017), whether anyone reads to the child, since what age, and how often during preschool age (De Houwer, 2002), whether the child is learning to read in HL, who teaches them and how often (McKendry & Murphy, 2011). There was also a case when the informants were given a choice between five age bands and they needed to select when they started and stopped reading to the child (language not specified, Arredondo, 2017).

Courses/classes and travel/holidays were documented in 15% of the questionnaires. Operationalisations of classes/courses mostly included questions referring to instruction in HL/SL, or about any other instruction or classes that the child is attending. For instance, Lyutykh (2012) asked about SL classes/courses at school, gifted/advanced placement classes, weekend school, formal instruction in HL reading, writing or language study, textbooks used in HL study, benefits of HL weekend school, reasons for not attending the HL weekend school, attendance in any other weekend, ethnic, or religious school other than the HL one, or things about the HL weekend school not liked by the parents. Other questions targeted: whether there was a place of formal instruction in HL (Arredondo, 2017), reasons for choosing additional HL or SL lessons (Prentza et al. [parent questionnaire], 2017), whether children attended a bilingual school and parents' opinions of this (Özturk, n.d.), the person teaching the child a specific language for religious purposes, as well as frequency of this activity (McKendry & Murphy, 2011), or whether children attended a HL school (Antonijevic-Elliott, n.d.). Travel/holidays was mostly operationalised as the frequency of visiting the HL country or the frequency of visitors from the HL country. It should be noted that one holiday-related operationalisation (in Wilson, 2017) was classified under the construct 'Socioeconomic status' (SES), as this question was a part of the child's SES estimation (see Table S4).

Several other activity-related components were documented across the questionnaires, although with such low frequency that we do not discuss them in any detail here. These included, preferred activities (4%), ease of learning new things (2%), activity patterns (2%), extra-curricular activities (6%), other activities (4%).

To sum up, the various components documenting a child's activities were operationalised in diverse ways. This illustrates how difficult it is to compare data collected with different tools. This diversity further increases when we factor in the various ways in which activities classified under the constructs 'exposure' and 'use' were operationalised (see section *Exposure and use*).

Current skills in HL and SL

As shown in Table 3, when it comes to children's current skills in HL and in SL (each estimated by 42% of questionnaires), the most frequently documented components included speaking/

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

production (in 35% of questionnaires for SL, and in 38% for HL) and listening/comprehension (in 31% of questionnaires for SL, and in 29% for HL). Speaking/production was assessed on a variety of scales across the questionnaires. These ranged from 3-point scales to 10-point scales for both HL and SL, and they also included binary questions such as the one in Blume and Lust (2017) about whether the production/comprehension in all languages was appropriate. This diversity makes it highly unlikely that the data collected with these tools will be readily comparable.

Even when scales of the same length were used, the point descriptors could vary widely between the tools. An illustration of this can be seen in comparing the 5-point scales used to assess SL speaking/production. Some scales were based on general qualifiers of proficiency (e.g., de Diego-Lázaro, 2019: very poor, poor, acceptable, good, very good; and Prentza et al. [parent questionnaire], 2017: not at all, a little, adequately, well, very well), and could therefore be considered rather similar. However, data collected with these tools are much less comparable with the SL speaking proficiency documented by Özturk's (n.d.) questionnaire, which used the following descriptors and for each of these provided real-life illustrations: not fluent, limited fluency, somewhat fluent, quite fluent, very fluent. Apart from questions about speaking/production skills in general, this component also included questions targeting more specific skills, such as vocabulary proficiency, speech proficiency, grammatical proficiency (Peña et al. [BESA:ITALK at school and at home], 2018), the ability to produce sentences of a certain complexity (De Houwer, 2017), or expressive language skills in relation to specific interlocutor groups or specific contexts (Restrepo, 1998).

When it comes to SL listening/comprehension, most operationalisations included rating understanding skills in general. However, there were also examples of rating the understanding ability or receptive language in specific situations, contexts or with certain interlocutor groups (e.g., Blume & Lust, 2017; Prentza et al. [parent questionnaire], 2017; Marinis [parent questionnaire], 2012; Restrepo, 1998). The points regarding variability illustrated above also apply to the HL speaking/production and HL listening/comprehension assessment.

Reading and writing skills were not frequently documented – for both languages separately, reading was documented in 13% of questionnaires, and for writing in only 10%. One further operationalisation asked jointly about reading and writing in HL. Despite only a few questionnaires documenting these skills, the assessment scales varied considerably, as they included 3- to 6-point scales, as well as binary choice questions, such as whether the child can read/write.

Finally, 6% of questionnaires asked about overall SL proficiency (4- to 6-point scales), while 8% of questionnaires documented overall HL proficiency (4- to 6-point scales). Furthermore, one questionnaire contained an open-ended question about the reasons for parental dissatisfaction with a child's HL and why the child might be different from children in the HL country (Paradis, Emmerzael & Sorenson Duncan, 2010) – this question was classified under a separate component 'reasons for dissatisfaction'.

It is important to mention that some of the ways in which the component 'language problems' was operationalised (classified under the construct 'developmental issues/concerns', see Table S5) included questions about children's current language skills. Examples include questions about: how often children struggled with reading/writing/listening/speaking activities (Gunning & Klepousniotou [teacher questionnaire], n.d.), difficulties in understanding the child because of the way he or she says words (Antonijevic-Elliott, n.d.; Antonijevic-Elliott et al., 2020), concerns and descriptions of the way in which the child talks (Peña et al. [BESA:ITALK home], 2018), etc. However, as these operationalisations focused on language-related issues, delays, struggles, worries, concerns, frustration, problems, difficulties, or diagnosis of language impairment/delay, they were classified under language problems rather than under current skills. Nevertheless, they can provide a useful insight into a child's language abilities or lack thereof.

4. Discussion

In this paper, we have reviewed the extent to which the descriptive tools used in bilingualism research tap into the same constructs and yield comparable measures, focusing on questionnaires designed to document bilingualism in children (age range: 0-18 years). This is the first comprehensive review of the components of bilingual experience across questionnaires that has considered the commonality of the components documented as well as how they are operationalised. As per our three aims, we have (i) surveyed available questionnaires used to quantify bilingual experience in children; (ii) identified (and classified) the components of bilingual experience that are documented across questionnaires; and (iii) discussed the comparability of the measures used to operationalise some of the most important components.

Most of the tools surveyed have been used to inform research in developmental psycholinguistics. This means that, given the practical impossibility of a truly systematic literature review to inform our survey and in spite of our attempts to be fully comprehensive, there remains the possibility of a bias in the sample of questionnaires we were able to find. The importance attributed to language exposure and use in our comparability analysis might therefore be a reflection of such a bias. Nevertheless, we trust that our comprehensive review will be beneficial for bilingualism research in its broadest definition.

With these caveats in mind, our review has highlighted the components most frequently documented across questionnaires (expressed via percentages in our summary tables). This revealed that, in spite of the different research aims across studies, there is a broad common ground in what manifestations of bilingualism are documented. Of particular note are the quantification of language exposure and use, and (albeit less frequently) the documentation of language difficulties possibly experienced by the child. Many factors contribute to explaining the considerable variability in how these components are documented, including, for instance, the broad age range the various questionnaires are intended for, and the different levels of heterogeneity of the populations targeted. Furthermore, reviewed questionnaires were of different lengths. This implies that those constructs and their components which were included more frequently across the tools (even in the shorter ones) might be considered more important by the questionnaire designers.

Other components of bilingualism are less frequently documented across questionnaires, but we do not think that this can be meaningfully interpreted. Such differences are not necessarily indicative of frequency of implementation across studies, as some questionnaires are used more frequently than others. Importantly, we do not mean to argue for any ranking of measures due to a potential bias in questionnaire selection for this review as pointed out above. Furthermore, the tools used to document bilingualism evolve along with research questions in the

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

field, and components such as 'language mixing' (see Byers-Heinlein, 2013) or 'the quality of language exposure' (see, for instance, the double special issue dedicated to this topic by the Journal of Child Language - Blom & Soderstrom, 2020) might become more prevalent in the future.

Importantly, this review has unveiled differences in the operationalisation of apparently similar constructs. Such differences can jeopardise the comparability of the resulting measures (Marian & Hayakawa, 2020). For instance, careful scrutiny has revealed the difficulty of documenting language exposure and language use separately: even when a questionnaire is designed to document one of them, the relevant questions could be insufficiently clear in their formulation, making the resulting measure a conflation of both exposure and use. Another case in point is the degree of measurement error associated with the choice of response scales. For instance, there is a lot of variability in the way questionnaires collect language exposure measures: not all break this construct down by interlocutor and/or by context, and the number of interlocutors and contexts considered is also variable. Furthermore, the amount of time spent with each interlocutor or in each context is not always documented itself. Importantly, there is no consensus on how to deal with overlap between the interlocutors. Does the proportion of overall language exposure get split between interlocutors, and if so, should equal weight be assigned to each interlocutor? What if different languages are used with different interlocutors present at the same time? There are also differences in the types of scales used in terms of measure (e.g., percentages, frequency adverbs) and in terms of granularity (e.g., number of points on a Likert scale). The precision of the scale should in principle affect the precision of the resulting measure, but we have to bear in mind that what is being documented is based on recollections, filtered through the respondent's generalisation and quantification abilities. The impact of scale precision on the degree of measurement error is an empirical question we aim to address in future research. For now, we agree with Kremin and Byers-Heinlein (2020) that the field of bilingualism could benefit from advances in psychometric research methods. That is, by using models such as the factor mixture model or the grade-of-membership model, a better solution might emerge on how to define bilingualism. As explained by Kremin and Byers-Heinlein (2020), the factor mixture model allows capturing variation within categories such as bilingual and monolingual. In addition, the grade-of-membership model allows individuals to belong to different categories to varying degrees. Based on the work presented here, we argue that the building blocks of those models (i.e., raw measures) require further scrutiny.

Not all components of bilingualism are quantifiable, and in some cases, what is crucial is not the quantity or frequency of a component but simply its existence, alongside other relevant components. This information can be used to derive complex indices, used as predictors in a particular field of enquiry. For instance, language impairment is strongly associated with a constellation of indicators (including, among others, early language milestones and parental concerns) which can be translated into a composite score expressing the level of risk of atypical development (Tuller, 2015). The richness of the language environment (e.g., Paradis, 2011) is another dimension that might be best captured by the use of a composite index. The set of components that such composite indices are based on currently varies across studies. For instance, the set of factors informing the level of 'richness of the language environment' may include – among others – the number of interlocutors in each language, whether the child has older siblings, the number and types of activities in each language, levels of literacy, etc. Composite scores may assign a weight to each of their components (e.g., if early language milestones are scored out of 4 and parental concerns on a scale from 0 to 5, a cumulative composite score based on these measures would automatically result in a heavier weight attributed to parental concerns), and this can induce variability across studies. If the components are documented differently in this way, this can result in an additional hidden level of variability.

Reviewing the manifestations of bilingualism documented across questionnaires has required us to classify them as components of overarching constructs. This classification highlighted a number of challenges. One challenge was the choice of terminology. We aimed to be strictly descriptive and neutral, but this was not always possible. For instance, our choice of 'home language' vs. 'societal language' to label the child's languages is not adequate for all types of bilingualism (such as in households featuring two majority languages) but we estimated these labels would be easy to translate into the realities of most bilingual children, and that they were more broadly encompassing and less controversial than 'heritage language' or 'minority/majority languages'. Another challenge arose from the mapping of questions onto components of bilingualism. We adopted the principle that any one question could only be mapped onto one component (and hence one overarching construct). For instance, there was a questionnaire in which a scale documenting relative frequency of exposure included an option about language mixing. We classified this under the relative frequency of exposure only, as that was the main aim of the question, rather than additionally including this item under the component of language mixing.

While we focused on questionnaires used with children, we also expect our findings to extend to the adult population. Specifically, the overarching constructs which we discussed in more detail (exposure and use, activities, and current skills in HL and SL) are most likely to also be relevant for adults. Other constructs (not discussed in the paper, but outlined in the supplements, see section C) could also be pertinent, depending on the aim of individual studies (Supplementary Materials). We might also expect that questionnaires with adults would also show similar variability as the studies with children in how specific variables are operationalised. At the same time, there are also a number of differences between adults and children which necessitate alternative approaches with adults. The most obvious difference is the fact that for adults, the questionnaires will likely always be distributed to the participants themselves rather than to their caregivers or teachers. Furthermore, some contexts of language exposure and use, such as (pre)school, do not apply to adults; other contexts, such as college/university setting and/or the work environment, are more relevant to the adults. The adults' own education, occupation or income might be used as estimates of their socioeconomic status. The home environment is also likely to be different in cases of adults not living with their caregivers/siblings. This will have an effect on how child questionnaires can be adapted for use with adults (or vice versa) and the implications of these factors should be considered when designing/choosing questionnaires to be used with both populations.

In practical terms, bilingualism can only be measured indirectly, and its operationalisation necessarily builds on (the combination of) several phenomena or dimensions. Our aim was not to address the operationalisation of bilingualism itself (as a latent construct), but to assess the comparability of the 'building blocks'

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

researchers exploit to generate the calculation of their choice. Without direct comparability of those building blocks, there is no possible comparability even across studies that rely on the same formula for an overall bilingualism score (e.g., language entropy, see Gullifer & Titone, 2020).

To ascertain the extent to which there is a consensus regarding what components of bilingualism are desirable or essential to document, our team has recently carried out an international Delphi Consensus Survey (De Cat et al., 2021). This will inform the creation of a modular questionnaire and associated calculator of language exposure and use.

5. Concluding remarks

We agree with Marian and Hayakawa (2020) that the field of bilingualism is in urgent need of greater transparency in how we operationalise bilingualism, as well as greater comparability of measures. We hope that the taxonomy delineated in this paper will provide a framework within which scholars/researchers can collaborate towards a consensus on a core set of measures to report across studies, in order to enhance transparency and reproducibility. Ideally, a consensus should be reached across bilingualism researchers, leading to the adoption of a common set of tools. However, the use of different core sets of measures in different subfields of enquiry is not precluded. Specifically, since bilingualism research is a fast-moving field, the tools we use will necessarily have to adapt to account for particular research aims and novel discoveries. However, as a minimum, we believe it is necessary for all studies to report transparently how the relevant components of bilingualism were measured, and to publish the relevant questionnaire(s) in online supplements. Ultimately, no matter which questionnaires are used or devised, a common understanding of the raw measures (i.e., building blocks) should remain a basic requirement across studies.

Methods in bilingualism research are evolving fast, and the time is ripe for a more critical approach to the documentation of bilingualism. This includes the validation of questionnaires and their methods of administration, and the critical appraisal of fitness-for-purpose when selecting a tool to inform research or practice.

Acknowledgements. We are thankful to the colleagues who shared their questionnaires with us, as well as to those who directed us to tools which they used or adapted. We are also thankful to the reviewers whose comments improved the quality of this paper. Finally, we thank the Economic and Social Research Council for funding this project (grant reference: ES/S010998/1).

Supplementary Material. For supplementary material accompanying this paper, visit http://dx.doi.org/10.1017/S1366728921000390

Competing interests. The authors declare none.

References

Adescope OO, Lavin T, Thompson T and Ungerleider C (2010) A Systematic Review and Meta-Analysis of the Cognitive Correlations of Bilingualism. *Review of Educational Research*, **80**(2), 207–245.

Anderson JAE, Mak L, Keyvani Chahi A and Bialystok E (2018) The language and social background questionnaire: Assessing degree of bilingualism in a diverse population. *Behavior Research Methods*, **50**, 250–263.

Antonijevic-Elliott S (n.d.) *PaBiQ adaptation 1*. Retrieved from the author.

Antonijevic-Elliott S, Lyons R, O' Malley MP, Meir N, Haman E, Banasik N, Carroll C, McMenamin R, Rodden M and Fitzmaurice Y (2020) Language assessment of monolingual and multilingual children using non-word and sentence repetition tasks. *Clinical Linguistics & Phonetics*, **34**(4), 293–311. https://doi.org/10.1080/02699206.2019.1637458

Arredondo MM (2017). *A Bilingual Advantage? The Functional Organization of Linguistic Competition and Attentional Networks in the Bilingual Developing Brain* [Unpublished doctoral dissertation]. University of Michigan.

Bak TH (2016) Cooking pasta in La Paz: Bilingualism, bias and the replication crisis. *Linguistic Approaches to Bilingualism*, **6**(5), 699–717.

Blom E and Soderstrom M (2020) Introduction. *Journal of Child Language*, **47**(1), 1–4. https://doi.org/10.1017/S0305000919000862

Blume M and Lust B (2017). Virtual Linguistics Lab Child Multilingualism Questionnaire. Prepared by the Cornell Language Acquisition Lab and Virtual Center for Language Acquisition Members and Affiliates. In M Blume and B Lust (eds.) with Y Chien, C Dye, C Foley & Y Kedar, *Research Methods in Language Acquisition: Principles, Procedures and Practices*. Mouton de Gruyter and American Psychological Association.

Blumenthal M and Julien M (2000) Diagnostiek van spraak-en taalproblemen bij meertalige kinderen, Geen diagnose zonder anamnese meertaligheid. *Logopedie en Foniatrie*, **1**, 13–17.

Byers-Heinlein K (2013) Parental language mixing: Its measurement and the relation of mixed input to young bilingual children's vocabulary size. *Bilingualism: Language and Cognition*, **16**, 32–48. https://doi.org/10.1017/S1366728912000120

Byers-Heinlein K, Esposito AG, Winsler A, Marian V, Castro DC and Luk G (2019a) The Case for Measuring and Reporting Bilingualism in Developmental Research. *Collabra: Psychology*, **5**(1): 37. https://doi.org/10.1525/collabra.233

Byers-Heinlein K, Schott E, Gonzalez-Barrero AM, Brouillard M, Dubé D, Laoun-Rubsenstein A, Morin-Lessard E, Mastroberardino M, Jardak A, Pour Iliaei S, Salama-Siroishka N and Tamayo MP (2019b) MAPLE: A Multilingual Approach to Parent Language Estimates. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S1366728919000282.

Cattani A, Abbot-Smith K, Farag R, Krott A, Arreckx F, Dennis I and 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 and Communication Disorders*, **49**(6), 649–671.

Cohen C (2015a) *INEXDEB parent questionnaire*. Laboratoire Interactions, Corpus, Apprentissages & Représentations. Retrieved from the author.

Cohen C (2015b) *INEXDEB short parent questionnaire*. Laboratoire Interactions, Corpus, Apprentissages & Représentations. Retrieved from the author.

DeAnda S, Bosch L, Poulin-Dubois D, Zesinger P and Friend M (2016) The Language Exposure Assessment Tool: Quantifying Language Exposure in Infants and Children. *Journal of Speech, Language, and Hearing Research*, **59**, 1346–1356.

de Bruin AMT (2019) Not all bilinguals are the same. A call for more detailed assessments and descriptions of bilingual experiences. *Behavioral Sciences*, **9**(3), 33. https://doi.org/10.3390/bs9030033

De Cat C (2020) Predicting language proficiency in bilingual children. *Studies in Second Language Acquisition*, **42**, 279–325.

De Cat C, Kašćelan D, Prévost P, Serratrice L, Tuller L, Unsworth S and The Q-BEx Consortium (2021). *Delphi consensus survey on how to document bilingual experience*. https://doi.org/10.31219/osf.io/ebh3c

de Diego-Lázaro B (2019). *Novel word learning in bilingual children with hearing loss* (Publication No. 13901537) [Doctoral dissertation, Arizona State University]. ProQuest Dissertations Publishing. https://bit.ly/2PEsIqF

De Houwer A (2002). *PRELAR, The preschool language context and activity record*. [Unpublished manuscript]. Retrieved from the author.

De Houwer A (2009). *Bilingual First Language Acquisition*. Multilingualism Matters.

De Houwer A (2017) *PEGEBOS-3. Polish-German and English-German Early Bilingualism Online Survey Wave 3 @ child age 3;6*. [Unpublished manuscript]. Retrieved from the author.

De Houwer A and Bornstein MH (2003, April 30-May 3). *Balancing on the tightrope: language use patterns in bilingual families with young children* [Conference presentation]. Fourth International Symposium on Bilingualism, Tempe, Arizona, USA.

Gagarina N, Klassert A and Topaj N (2010) Russian language proficiency test for multilingual children. *ZAS Papers in Linguistics*, **54**. ZAS.

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press

Gullifer JW and Titone D (2020) Characterizing the social diversity of bilingualism using language entropy. *Bilingualism: Language and Cognition*, **23** (2), 283–294. https://doi.org/10.1017/S1366728919000026

Gunning L and Klepousniotou E (n.d.). *LeBLEQ-C (Leeds-Bradford Language Exposure Questionnaire for Children)*. Retrieved from the authors.

Gunning L and Klepousniotou E (n.d.). *LeBLEQ-T (Leeds-Bradford Language Exposure Questionnaire for Teachers)*. Retrieved from the authors.

Gutiérrez-Clellen VF and Kreiter J (2003) Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics*, **24**, 267–288.

Hulstijn JH (2012) The construct of language proficiency in the study of bilingualism from a cognitive perspective. *Bilingualism: Language and Cognition*, **15**(2), 422–433.

Kremin LV and Byers-Heinlein K (2020) *Why not both? Rethinking categorical and continuous approaches to bilingualism*. PsyArXiv. https://doi.org/10.31234/osf.io/nkvap

Lehtonen M, Soveri A, Laine A, Järvenpää J, de Briun A and Antfolk J (2018) Is Bilingualism Associated With Enhanced Executive Functioning in Adults? A Meta-Analytic Review. *Psychological Bulletin*, **144**(4) 394–425.

Li P, Sepanski S and Zhao X (2006) Language history questionnaires: A web-based interface for bilingual research. *Behavior Research Methods*, **38**(2), 202–210.

Li P, Zhang F, Tsai E and Puls B (2014) Language History Questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, **17**(3), 673–680. https://doi.org/10.1017/S1366728913000606

Li P, Zhang F, Yu A and Zhao X (2019) Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 1–7. https://doi.org/10.1017/S1366728918001153

Luk G, De Sa E and Bialystok E (2011) Is there a relation between onset age of bilingualism and enhancement of cognitive control? *Bilingualism: Language and Cognition*, **14**, 588–595.

Lyutykh E (2012) *Questionnaire for adult speakers and child learners of English and Russian*. Retrieved from: https://nhlrc.ucla.edu/nhlrc/research#researchproficiency

Marian V, Blumfield HK and Kaushanskaya M (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research*, **50**, 940–967.

Marian V and Hayakawa S (2020) Measuring bilingualism: the quest for a "bilingualism quotient". *Applied Psycholinguistics*, 1–22. https://doi.org/10.1017/S0142716420000533

Marinis T (2012) *Language background questionnaire for Turkish/Kurdish speakers*. Retrieved from: https://nhlrc.ucla.edu/nhlrc/research#researchproficiency

McKendry M and Murphy VA (2011) A comparative study of listening comprehension measures in English as an additional language and native English-speaking primary school children. *Evaluation and Research in Education*, **24**, 17–40.

Özturk Ö (n.d.). *Parental Questionnaire*. Retrieved from the author.

Paap KR, Johnson HA and Sawi O (2014) Are bilingual advantages dependent upon specific tasks or specific bilingual experiences? *Journal of Cognitive Psychology*, **26**, 615–639.

Paradis J (2011) Individual Differences in Child English Second Language Acquisition: Comparing Child-Internal and Child-External Factors. *Linguistic Approaches to Bilingualism*, **1**(3), 213–237.

Paradis J, Emmerzael K and Sorenson Duncan T (2010) Assessment of English Language Learners: Using Parent Report on First Language Development. *Journal of Communication Disorders*, **43**, 474–497.

Pelham SD and Abrams L (2014) Cognitive advantages and disadvantages in early and late bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **40**, 313–325.

Peña E, Gutiérrez-Clellen V, Iglesias A, Goldstein B and Bedore L (2018). *Bilingual English-Spanish Assessment (BESA)*. Brookes Publishing.

Prentza A, Kaltsa M, Tsimpli IM and Papadopoulou D (2017) The acquisition of Greek gender by bilingual children: The effects of lexical knowledge, oral input, literacy and bi/monolingual schooling. *International Journal of Bilingualism*, **23**(5). https://doi.org/10.1177/1367006917733066

Read K, Contreras PD, Rodriguez B and Jara J (2020) ¿Read conmigo?: The effect of code-switching storybooks on dual-language learners' retention of new vocabulary. *Early Education and Development*, 516–533. https://doi.org/10.1080/10409289.2020.1780090

Restrepo MA (1998) Identifiers of Predominantly Spanish-Speaking Children With Language Impairment. *Journal of Speech, Language, and Hearing Research*, **41**, 1398–1411.

Rodriguez-Fornells A, Krämer UM, Lorenzo-Seva U, Festman J and Münte TF (2012) Self-assessment of individual differences in language switching. *Frontiers in Psychology*, **2**(388), 1–15. https://doi.org/10.3389/fpsyg.2011.00388

Scharff-Rethfeldt W (2012a) *Fragebogen zum Erhalt von Informationen zum Bilingualen Patientenprofil - Kinder / Bilingual Patient's Profile - Children (ENGLISH)*. Retrieved from: http://www.logo-mobil.net/logo-mobil.net/Publications.html

Surrian S and Luk G (2017) Describing Bilinguals: A Systematic Review of Labels and Descriptions Used in the Literature Between 2005–2015. *Bilingualism: Language and Cognition*, 1–15. https://doi.org/10.1017/S1366728917000682

Tao L, Marzecová A, Taft M, Asanowicz D and Wodniecka Z (2011) The efficiency of attentional networks in early and late bilinguals: The role of age of acquisition. *Frontiers in Psychology*, **2**(123). 1–19.

Tuller L (2015). Clinical use of parental questionnaires in multilingual contexts. In S Armon-Lotem, J de Jong and N Meir (eds), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 301–330). Multilingual Matters.

Unsworth S (2013) Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition*, **16**(1), 86–110.

Unsworth S (2016). Quantity and quality of language input in bilingual language development. In E Nicoladis and S Montanari (eds.), *Bilingualism Across the Lifespan: Factors Moderating Language Proficiency* (pp. 103–121). De Gruyter Mouton.

Valian V (2015) Bilingualism and cognition. *Bilingualism: Language and Cognition*, **18**(1), 3–24.

Wilson EA (2017). *Children's development of Quantity, Relevance and Manner implicature understanding and the role of the speaker's epistemic state* [Unpublished doctoral dissertation]. University of Cambridge. https://elspethwilson.uk/wp-content/uploads/elspethwilson.uk/2017/12/E_Wilson_thesis_to_print.pdf

https://doi.org/10.1017/S1366728921000390 Published online by Cambridge University Press