Title Unpacking the richness of language experience as a predictor of bilingual children's language proficiency

Authors Sharon UNSWORTH[1], Arief GUSNANTO[2], Draško KAŠĆELAN[3], Philippe PRÉVOST[4], Ludovica SERRATRICE[5,6], Laurie TULLER[4] & Cécile DE CAT[7]

Affiliation
[1] Radboud University, Nijmegen, The Netherlands
[2] University of Leeds, UK
[3] University of Essex, UK
[4] University of Tours, France
[5] University of Reading, UK
[6] Arctic University of Norway, Tromsø, Norway
[7] University of Leeds, UK

Corresponding author sharon.unsworth@ru.nl

--- Accepted for publication in *Journal of Child Language*, August 27th, 2025 ---

Abstract

The richness of bilingual children's language experience is typically expressed as a composite score using parental questionnaire data. This study unpacks the concept of input richness by examining one such composite score (Q-BEx; De Cat et al., 2022b) to determine whether it reliably predicts children's language abilities, is no more complex than required, and as user-friendly as possible. Data were collected from 173 bilingual children aged 5 to 8 across three countries (France, Netherlands, UK) with various heritage languages in each. Parents completed the Q-BEx questionnaire and children proficiency tasks in their societal language. We analysed the predictive power of the original score compared to several alternative scoring approaches. Results showed (i) these alternatives were not more informative, (ii) scores including qualitative aspects of richness fared better than those with only quantitative variables, (iii) the latent variables underlying richness were comparable across languages, and (iv) whether SES was included made little difference.

Introduction

Bilingual language development is characterised by individual differences, both between and within children. Variation is observed in children's language learning environments, the rate at which they acquire their two or more languages, the extent to which they actively use these languages, as well as the level of proficiency they attain in each. There is robust evidence that individual variation in children's language abilities is predicted by properties of their language learning environment (see Paradis, 2023 for recent overview). More specifically, quantity of input measured as current or cumulative exposure has been shown to predict children's language abilities across a range of language combinations and target language properties in both the societal – SL – and heritage language – HL (e.g., Chondrogianni & Marinis, 2011; Thordardottir, 2011; Unsworth, 2013).

In addition to amount of input, type of input – or input quality – is also a source of variation in bilingual children's language experience, as it is for monolinguals. According to Rowe and Snow (2020), input quality encompasses three interrelated dimensions, namely the interactive, conceptual and linguistic. The interactive component of input quality focuses on the extent to which caregivers engage children in conversational turn-taking, the conceptual on whether input from caregivers focuses on abstract versus concrete concepts, and the linguistic on properties of caregivers' speech, most commonly in terms of lexical and morphosyntactic complexity. Whilst problematic as a concept (MacLeod & Demers, 2023), there is emerging evidence that input quality also predicts variation in children's (emerging) language abilities (e.g., Paradis, 2011; Place & Hoff, 2015; Unsworth et al., 2019).

In their recent Systematic Concept Analysis, MacLeod and Demers (2023) found that input quality has predominantly been operationalised as – in their terms – linguistic complexity (e.g., Rowe, 2012, a study on monolingual children but nonetheless relevant to bilinguals), the use of language-evoking strategies (e.g., Dixon et al., 2012), parental language competency (e.g., Daskalaki et al., 2020) and enrichment activities (Scheele et al., 2010). Studies may focus on specific sources of input, such as media (e.g., Sun & Yin, 2020) or diversity of interlocutors (e.g., Gollan et al., 2015), or aggregate sources into a composite measure (e.g., G. Jia & Aaronson, 2003). Whilst zooming in on distinct characteristics of input quality facilitates a more nuanced understanding of their role in bilingual language development and of the mechanisms involved in language learning more generally, global measures of input quality offer a useful resource for practitioners in their assessment of bilingual children, as long as they are informed by research. The aim of the present study was to explore and evaluate one such global measure of input quality, namely the composite richness score which is part of the recently developed Q-BEx questionnaire (Quantifying Bilingual Experience; De Cat et al., 2022a), in order to determine

whether it is fit for purpose, that is, a reliable predictor of bilingual children's (emerging) language abilities in the early school years. In doing so, we hoped to gain a better understanding of the role of richness in the language development of bilingual children and how best to measure it.

In this paper, we follow much of the research on bilingual children and use the term *richness* instead of input quality. Richness is defined as "the diverse and complex language children experience through certain activities and interactions" (Paradis, 2023, p. 803). This includes (aspects of) each of the three dimensions of input quality outlined by Rowe and Snow (2020; i.e., interactive, conceptual and linguistic) but is also broader, for example, because it includes characteristics of language input from other modalities. Whilst we acknowledge that like "input quality", the term "richness" is in a certain sense also value-laden (Carroll, 2017; MacLeod & Demers, 2023) and should be used with caution, our intention here is not to pass judgement on the input which parents of bilingual children provide to their children but to accurately describe the specific types of (diversity in the) input and their relation to children's outcomes as best we can with the tools at our disposal. The deficit framing of bilingual language development highlighted in several recent publications (De Houwer, 2022; MacLeod & Demers, 2023) is certainly problematic but in our view is partly an issue of (science) communication. There is a tension between accurately and objectively describing the relevant characteristics of children's language environments and doing so in such a way that it is readily understandable for both researchers and practitioners. It is difficult to achieve this clarity whilst avoiding connotations which are unintended. For the want of a credible alternative over-arching term at this stage, we continue to use "input richness" in the present paper.

Measuring richness

How input richness is operationalised varies. Some studies focus on specific aspects of the language environment, such as multimedia (e.g., Sun & Yin, 2020), home literacy practices (e.g., Prevoo et al., 2014), interlocutor diversity (Gollan et al., 2015), and different types of home-based activities (Cheung et al., 2019), but most use a global or composite measure (e.g., Jia & Aaronson, 2003; Jia & Fuse, 2007; Paradis 2011; Paradis et al., 2020; Unsworth et al., 2019), and all make use of questionnaires. In a recent study, Verhoeven et al. (2024) compared parental questionnaire data with what might be considered a more direct measure of children's language input, namely recordings of child-directed speech. The authors found both to correlate with vocabulary outcomes to the same extent (in line with Marchman et al., 207; Orena et al., 2020; cf. Cyzchosz et al., 2021). They conclude that both are reliable ways of measuring input quantity. To the best of our knowledge, a similar study making such a comparison for

input richness has yet to be conducted. Aside from the methodological and ethical challenges involved in recording children's language environments in this way, it seems unlikely that such a study would in fact be able to capture full extent of the variation in input richness. This is because variation in the richness of bilingual children's language experience often relates to activities which are not amenable to recording (e.g., reading) or to sources of input outside the home (e.g., heritage language education). To measure input richness, it seems that – for now at least – we must rely on parental questionnaire data.

Which variables are included in composite measures derived from parental questionnaires and how they are combined with each other also differs depending on the questionnaire. For example, the ALEQ-4 (Alberta Language Environment Questionnaire; Paradis et al., 2020-AP) incorporates reading and writing, speaking and listening, extra-curricular activities, playing with friends, and HL classes. Parents are asked to indicate the frequency of each using a scale from 1 to 5 (1 = 0 to 1 hours, 2 = 1 to 5 hours, 3 = 5 to 10 hours, 4 = 10 to 20 hours, 5 = 20+ hours), the ratings for each are summed and subsequently divided by the total number of scales answered to arrive at a proportion score (0-1), with 1 indicated a higher frequency of language-rich activities. Similarly, in the PABiQ (Questionnaire for Parents of Bilingual Children; Tuller, 2015) parents indicate which activities (reading, tv, storytelling) children engage in each week and points are assigned based on frequency (0 for never or almost never, 1 for at least once a week and 2 for every day). These are summed to arrive at composite measure with a maximum of 18. Finally, the BiLEC (Bilingual Language Exposure Calculator; Unsworth, 2013) asks parents to specify the number of hours per week children engage in various language-related activities outside school (e.g., sports and clubs, friends, reading or being read to, and media) and to estimate which language or languages are used during each, as a percentage. These two values are multiplied to obtain the number of hours per language per activity, the number of hours per language added up, and this value is divided by the total number of hours spent on all activities together to arrive at an overall percentage for each language.

Whilst the ways in which each of these global measures of richness is derived and the type of score obtained differ, they are all based on frequency data only. More qualitative aspects of richness such as interlocutor diversity and proficiency are typically not included (but see e.g., Jia & Fuse, 2007, who included interlocutor diversity), even though both have been shown to predict bilingual children's developing proficiency in the societal language (Jia & Fuse, 2007) and heritage language (e.g., Place & Hoff, 2015). For example, in a study on bilingual preschoolers in the Netherlands, Unsworth et al., (2019) found that the degree of non-native input (i.e., the proficiency level of non-native input providers) rather than the amount (i.e., the proportion of input from non-native speakers) was a significant

predictor of children's morphosyntactic skills. The Q-BEx questionnaire, which is the focus of the present study, incorporates both frequency-based and more qualitative aspects of input richness.

Effects of input richness on bilingual language development

There is wide-ranging evidence that bilingual children's (emerging) language abilities are influenced by the richness of their language experience. A positive relation between richness and language proficiency has been observed across several linguistic domains including vocabulary in the SL (e.g., Paradis, 2011), HL (e.g., Sun & Yin, 2020) or both languages (e.g., Kaltsa et al., 2020); morphosyntax (e.g., Kaltsa et al., 2020; Paradis et al., 2020; Unsworth et al., 2019 - all studies on the SL), and narrative skills (in the HL; R. Jia & Paradis, 2015), as measured by standardized tasks tapping into overall (e.g., Sun & Yin, 2020) and specific abilities (e.g., G. Jia & Fuse, 2007), as well as non-standardized tasks focussing on specific morphosyntactic structures (e.g., Kaltsa et al., 2020). For example, in a study on preschoolers with a range of heritage or home languages growing up in the Netherlands, Unsworth et al. (2019) found that engagement with language-rich activities in the SL such as shared book-reading was a significant predictor of children's semantic fluency in the same language (i.e., Dutch). Similarly, Kaltsa et al. (2020) found that home literacy practices were associated with sequential bilingual Albanian-Greek children's scores on a sentence repetition task in Greek, their SL. In a number of studies focussing on the SL, experiential factors such as richness have been found to account for more variance in vocabulary than morphosyntactic outcomes (Paradis, et al., 2020; Paradis, 2011; Chondrogianni & Marinis, 2011).

Effects of input richness do not appear to be restricted to certain age groups, having been observed in bilingual children of different ages, including toddlers (e.g., Place & Hoff, 2015 - HL), preschoolers (e.g., Leseman & Boom, 1999 - SL), primary-school children (e.g., Daskalaki et al., 2020 - HL) and adolescents (e.g., Soto-Corominas et al., 2020 - SL). For example, Place and Hoff (2015) found that interlocutor diversity predicted bilingual Spanish-English toddlers' morphosyntax and vocabulary in their HL, Spanish, but not in English, their SL. More specifically, a greater number of different speakers providing input to the children in Spanish was associated with longer utterances and higher scores on a standardized expressive vocabulary task. Importantly, in this as in many other studies (e.g., Paradis, 2011), the observed effects of input richness remained after controlling for variation in input quantity.

The richness of children's language environments has also been found to change over time and this may vary depending on the language in question (SL vs. HL) and, for sequential bilingual children, age of onset to SL. For example, in a cross-sectional study of bilingual Mandarin-English children in the US from age 5 through 18, Jia and colleagues (2014) observed an increase in use of the SL, English, whilst

reading for leisure as children grew older. In contrast, language use whilst watching TV varied, with less English in the younger (< 8 years) and older (> 12 years) groups relative to the age groups in between. In an earlier study with children from the same HL community, Jia and Aaronson (2003) found that a younger age of onset was associated with a richer SL environment compared to the HL environment. In one of the few longitudinal studies examining the role of richness, on Syrian refugees in Canada, Paradis et al. (2021) observed that as children (average age at time 1 = 9.5 years, 24 months after immigration) grew older, the richness of their HL, Arabic, remained stable, whereas their SL (i.e., English) environment became richer. Interestingly, there was a marginal effect of richness on children's scores on an English sentence repetition task at time 1, but this disappeared at time 2, approximately one year later. Paradis and colleagues noted that this limited effect of richness (as well as other environmental factors) may be due to a kind of "ceiling" effect in under-resourced families such as newly arrived refugees. In other words, the level of richness available in the home environment and its potential effect on their L2 development is restricted due to family circumstances. An alternative explanation for this change over time is that morphosyntax is less sensitive to variation in input richness than other linguistic domains, although this may depend on the task used as findings vary in this regard. However, as the authors point out, other studies have shown that such a relation *does* exist (e.g., Jia & Aaronson, 2003, Paradis, 2011). In short, then, richness may vary for individual bilingual children across their two languages and over time.

A positive relation between input richness and language outcomes has been found for both the societal (e.g., Paradis et al., 2020; Unsworth et al., 2019) and heritage (e.g., G. Jia & Aaronson, 2003; Pham & Tipton, 2018) language, with potentially a more crucial role for the HL (Jia, 2008; Paradis, 2023). The extent to which richness effects depend on language status is however difficult to ascertain given that comparisons of language development in the HL and SL within the same group of children are relatively infrequent (a problem in the field more generally – see De Houwer, 2022 for relevant discussion). The available studies containing such a comparison obtained mixed results. For example, in the series of studies on recently arrived Syrian refugees in Canada aged 6-13 years mentioned above, Paradis and colleagues (2020, 2021) observed a positive relation between richness and vocabulary and morphosyntactic development in the SL but not in the HL. They speculated that variation in input richness may have a greater impact on the language being learned (i.e., English), particularly in the case of new arrivals for whom the HL (i.e., Arabic) is more established.

The opposite pattern was observed by Sun and Yin (2020) in their study on bilingual English-Mandarin 4- and 5-year-olds growing up in Singapore: the diversity in multimedia sources was positively

related to children's proficiency in their HL, Mandarin, but English multimedia exposure at home bore little relation with proficiency in that language. Yet a different pattern was found by Cheung et al. (2019). They examined the vocabulary skills of bilingual Cantonese-English children in the US and found that language use during dinner-table interactions and book reading predicted children's outcomes in the same language. In addition, the amount of Cantonese used during book reading was also a (positive) predictor of vocabulary scores in English.

There are some studies that have failed to identify an association between input richness and language outcomes. In those studies, specific characteristics of the language or language community (e.g., Scheele et al., 2010) and age have been put forward as explanations. For example, in another Canadian study, Soto-Corominas et al. (2020) found no effect of SL richness across several linguistic domains in a large and diverse group of bilingual adolescents. The authors speculated that with more than 7 years' exposure, the adolescents in their study may no longer have been at a stage in their language development where richness predicted individual differences.

To summarise, effects of input richness on bilingual children's language outcomes are by and large quite robust. The richness of children's language environment has been found to predict proficiency in both the HL and SL and across linguistic domains. At the same time, not many studies have investigated effects of richness for both languages within the same group of children and when they have, results were mixed. Furthermore, as yet, little attention has been paid to the different ways in which richness is indexed and the extent to which these are comparable. Before considering this question in more detail, we first consider another important factor relating to richness, namely socio-economic status (SES). In the language acquisition literature, this variable has been operationalised as parental education, family affluence or level of deprivation (De Cat, 2021).

On the relation between richness and SES

The relation between richness, SES and children's language development is complex. Richness and SES each predict children's (emerging) language abilities (although not always), and they are also related to each other.

SES has been shown to correlate with bilingual children's vocabulary size (e.g., Gathercole et al., 2016) and their morphosyntactic abilities (e.g., Meir & Armon-Lotem, 2017). These effects may be related to outcomes in one language but not the other. For example, in a study on bilingual Turkish-Dutch six-year-olds in the Netherlands, Prevoo et al. (2014) found that maternal education, as a proxy for SES, was an indirect predictor of vocabulary scores in the SL mediated by frequency of reading by

parents, but no such relation between SES and vocabulary, indirect or direct, was observed for the HL. In contrast, Lauro et al (2020) found the opposite pattern: in their study on bilingual Spanish-English preschoolers in the US, maternal education was a predictor of vocabulary scores in the HL but not in the SL.

Effects of SES may also differ depending on whether maternal education is measured in the HL or SL: Hoff et al., (2018) observed a language-specific effect of maternal education on children's outcomes. In other words, maternal education level in English was significantly related to their children's emerging language abilities in English but not in Spanish, and vice versa for maternal level of education in Spanish. The authors argue that this difference is due to education in a given language changing mothers' use of that same language.

Properties of parental input have indeed been found to vary with SES, so much so that SES is sometimes used as a proxy for input quality (e.g., De Cat, 2021; MacLeod & Demers, 2023). SES is a broad construct, however, and its effects encompass more than differences in parental language use alone. For example, Paradis et al. (2022) found that SES (operationalised as maternal education and maternal employment) was a predictor of bilingual Arabic-English children's abilities in the SL (i.e., English) despite the fact that the mothers were educated in Arabic only and they interacted with their children exclusively in the same language. High SES is not only associated with different patterns of language use but with other, distal factors, such as attitudes towards education (Scheele et al., 2010) and degree of assimilation, which in turn may be associated with access to literacy-related activities (Pearson, 2007), a variable which is often incorporated into richness measures (e.g., Kalia & Reese, 2009). SES has also been found to interact with parental attitudes to specific languages. For example, Saravanan (2001) found that family SES and HL abilities were inversely related, likely due to parents opting for the more prestigious SL language in interaction with their child (see also Oller & Eilers 2002).

In short, SES is an important predictor of bilingual children's (emerging) language abilities and its relation with input richness is complex. Given this complex relation, establishing effects of richness and SES on children's outcomes is often challenging.

The present study

The goal of the Q-BEx project (www.q-bex.org) was to create a user-friendly questionnaire in multiple languages (27 available at the time of writing), which would increase comparability across studies, labs and contexts, and which would be accessible for practitioners (i.e., speech language therapists and teachers). Its design was informed by a review of existing parental questionnaires (see Kašćelan et al.,

2022) and a Delphi consensus study (De Cat et al., 2023), as well as the psychometric literature (e.g., Dillman et al., 2014). This resulted in a questionnaire consisting of seven different modules to select from, including one on the richness of children's language experience. As part of the validation process of the Q-BEx questionnaire, the goal of the present study was to determine whether the composite richness score was fit for purpose.

This score includes several components, and it is not clear whether these tap into the same latent construct, or whether certain components (or combinations thereof) are more important or more informative than others. Given the robust evidence available showing the effect of input quality/richness on the language development of bilingual children, we considered that the score would be fit for purpose if it was shown to be a reliable predictor of children's language outcomes and if it was no more complex than required (in terms of its composition), and as user-friendly as possible (in terms of interpretability).[1]

The aim of the Q-BEx questionnaire is to deliver an index which predicts bilingual children's language outcomes rather than one which only describes the richness of their language experience. We first explored a dimension-reduction approach to richness by using a Principal Component Analysis (PCA; Jolliffe & Cadima, 2016) to account for the observed variance in parents' responses to the questions in the richness module. We then used the resulting components to derive alternative scores to the original composite score. Our first research question concerns whether these alternative scores, as follows:

1. Are alternative data-driven composite scores more informative than the original Q-BEx composite score as predictors of bilingual children's proficiency in the societal and heritage languages?

As noted above, the richness measure in Q-BEx is more comprehensive than the composite scores in other popular parental questionnaires: alongside information about the frequency of engagement in language- and literacy-related activities, it also includes questions about interlocutor diversity and SES. We operationalised SES as the highest caregiver level of education in any language; for this reason, we use 'parental education' as shorthand henceforth). Whilst there are good reasons to believe that both these variables will provide relevant and potentially additional information about the richness of bilingual children's language experience, this remains an empirical question. Furthermore, as outlined

---

[1] We report elsewhere on the relation between individual component parts of the richness score and children's outcomes in the SL and HL (Prévost et al., in prep; Serratrice et al., in prep).

above, there are theoretical and practical objections to including a proxy for SES in such a measure. For this reason, our second research question asked:

2. To what extent do parental education (as a proxy for SES) and interlocutor diversity and proficiency contribute to the predictive power of the composite richness score?

To answer this question, we considered a number of variations on the original Q-BEx score, in part informed by the PCA. More specifically, we compared (i) the original Q-BEx composite score with (ii) a composite score excluding parental education, (iii) composite scores excluding parental education but including interlocutor diversity and interlocutor proficiency, (iv) composite scores excluding parental education as well as interlocutor diversity and interlocutor proficiency. In each of these comparisons, we fitted models including parental education as covariate and models where parental education was excluded altogether. We did not have any specific predictions about which model(s) would be the best fitting. If parental education explained variance over and above the measures which excluded it, this could be interpreted in (at least) two ways, which are not mutually exclusive: either as evidence for parental education capturing aspects of richness which the composite score did not, or as a proxy for environmental variables which go beyond language experience. If parental education did not explain any additional variance, then we could conclude that the richness index successfully captures what SES would otherwise be a proxy for.

Like comparable parental questionnaires, Q-BEx was designed for use in different contexts and with different bilingual populations. A tacit assumption in this approach is that the same latent variables underlie richness irrespective of context or languages involved. Our final research question, which we addressed using a two-way orthogonal partial least square (O2-PLS) analysis, aimed at determining whether this was the case. We asked:

3. To what extent do the same latent variables underlie richness across the heritage and societal language?

In line with the assumption underpinning the questionnaire's design, we expected that the same type of information would be relevant to richness in both languages and that any differences would reflect the language's status (i.e., societal vs. heritage). For example, HL-specific information might be related to the availability of resources (in certain HLs), or to some other context-specific circumstance. Given that

the SL and broader linguistic context differs across the three countries included in this study (i.e., French in France, Dutch in the Netherlands and English in the UK), we furthermore explored the extent to which these latent variables were comparable across countries. We expected that the latent variables would be consistent across countries of residence. We did not include parental education in the O2-PLS analysis, because this variable was operationalised as the highest level of parental education *across* languages rather than *per* language. Rather, as a follow-up to our investigation of the relation between richness and parental education as part of RQ2, we explored how parental education related to the latent variables unveiled by the O2-PLS, expecting that it would correlate with some of them.

Method

Participants

Participants were bilingual (n = 135) and trilingual (n = 38) children growing up in France (n = 42), the Netherlands (n = 76) or the United Kingdom (UK; n = 55) aged between 5 and 9 years old. Just under half (n = 82) started acquiring each of their languages at birth; the others (n = 91) were exposed to their HL(s) from birth and the SL sequentially between the ages of 1 and 84 months (M_{France} = 32.0, SD_{France} = 15.9; $M_{Netherlands}$ = 39.0, $SD_{Netherlands}$ = 22.; M_{UK} = 27.6, SD_{UK} = 19.8). Children with a known language impairment were excluded from the study. Biographical details, including children's heritage languages, are provided in Table 1.

Table 1. Background information for children in three countries

	France (n = 42)	Netherlands (n = 76)	UK (n = 55)
Age (in months)	*M* = 32, *SD* = 16	*M* = 39, *SD* = 22	*M* = 28, *SD* = 20
HLs and no. of families where language is spoken (as either HL1 or HL2)	Arabic (16), Chinese (1), Creole (6), Dutch (1), English (7), Italian (2), Lingala (2), Mande (1), Polish (1), Portuguese (4), Romani (1), Romanian (2), Russian (1), Turkish (1)	Afrikaans (1), Arabic (6), Bulgarian (4), English (25), Filipino (1), French (6), Frisian (2), German (10), Greek (1), Hindi (1), Italian (2), Kurdish (1), Papiamento (1), Polish (3), Portuguese	Albanian (2), Arabic (3), Bangla (5), Chinese (4), Czech (1), Dutch (1), French (2), German (3), Greek (5), Gujarati (4), Hindi (4), Hindko (1), Italian (4), Japanese (1), Latvian/Lithuanian

		(5), Punjabi (1), Romanian (3), Russian (2), Spanish (4), Turkish (13), Ukrainian (3)	(4), Marathi (1), Mirpuri (1), Polish (2), Portuguese (2), Punjabi (2), Romanian (3), Russian (1), Spanish (4), Telugu (2), Urdu (3), Zulu (1)
Current (weighted) exposure to SL (%)	$M = 72, SD = 19$	$M = 46, SD = 22$	$M = 60, SD = 20$
Parents' highest level of education			
up to primary school	n = 2 (17%)	n = 6 (9%)	n = 4 (7%)
secondary school	n = 13 (31%)	n = 10 (13%)	n = 5 (9%)
post-secondary	n = 7 (17%)	n = 13 (17%)	n = 8 (14%)
university degree	n = 20 (48%)	n = 47 (62%)	n = 38 (69%)

Instruments

We used the aforementioned Q-BEx questionnaire to estimate the richness of children's language experience and their proficiency in the HL. Their proficiency in the SL was measured using a sentence repetition task and two standardized vocabulary tasks. Given the diversity in HLs spoken, collecting objective measures of HL proficiency was not practical. Our measure of HL proficiency was therefore based on parental estimation as part of the questionnaire.

– Parental questionnaire

Parents answered all questions in each of the seven modules in the Q-BEx questionnaire (i.e., background information, risk factors, language exposure and use, language proficiency, richness, attitudes, and language mixing).[2] For various logistical reasons (including the availability of the Q-BEx questionnaire in another language at the time of testing), all the parents in the UK completed the questionnaire in English. In France and the Netherlands, the vast majority also opted for the societal language, and the remaining parents did so in a different language (i.e., Arabic (n = 2), English (n = 2) or

---

[2] With the exception of France, where the language mixing module was not distributed.

Romanian (n = 2) in France, and in the Netherlands Arabic (n = 3), Polish (n = 3), English (n = 15), French (n = 3), German (n = 1), Italian (n= 1), Russian (n = 1), Spanish (n= 1) or Turkish (n = 7)).

In the richness module, parents are asked to indicate the frequency with which their child participates in various literacy- and language-related activities (i.e., reading or being read to, doing homework, following language classes inside and outside regular school or daycare, multimedia such as tv, apps and online games, playing with friends and organised activities such as sport and music) on the following scale: (almost) never, once or twice a month, once or twice a week, several times a week, and every day. In addition, parents indicate the number of people who speak in each language to the child at least once a week, how many of these speak the language very well, as well as the highest education level completed for each caregiver in any language. In total, the richness module contains 19 questions. Following the ALEQ, all of these responses are assigned a numerical value ranging from 0 to 4 (see Appendix for details), these are summed and subsequently divided by the highest possible score to arrive at a proportion ranging from 0 to 1. This composite score features in the individual reports, which are generated automatically and can be downloaded in the questionnaire's online interface. Aimed at both practitioners and researchers, the composite richness score can easily be used by researchers needing a single richness variable (e.g., because they do not have enough participants to include multiple richness-related variables in one model). For further details about the individual reports as well as a complete list of questions and answer scales, the reader is referred to the Q-BEx website (www.q-bex.org).

In addition to the richness score, two more composite measures were derived from parents' responses, namely current exposure and language entropy (Gullifer & Titone, 2020), as part of the analysis for RQ2. Language entropy is a measure of the extent to which multiple languages are engaged across individuals and contexts (see estimates individual- and contextual-level differences in the extent to which multiple languages are engaged (see Gullifer & Titone, 2020 and Serratrice et al., in prep for more details). As part of the language proficiency module, parents were asked how well their child could speak and how well they could understand the HL for their age. Answer options included: hardly at all / not very well / pretty well / very well. We used parents' responses to these two questions as our measure of HL proficiency. For the trilingual children, the first HL listed by parents was always the one to which children were exposed more frequently. For the sake of simplicity, we included this HL only in the analysis.

– *Sentence repetition task*

Morphosyntactic abilities in the SL were measured using the LITMUS Sentence Repetition Task (SRT; Marinis & Armon-Lotem, 2015) in the three societal languages, French, Dutch and English. The SRT consisted of 30 sentences in the English and Dutch versions and 16 sentences in the French version varying in complexity, from less (short sentences in present simple) to more (object relative clauses) complex (see Marinis & Armon-Lotem, 2015). The sentences were presented auditorily using headphones in a fixed order and children's responses recorded. Responses were given 1 point if they included a verbatim repetition of the target sentence and 0 points for non-verbatim repetitions.

– Vocabulary breadth

Vocabulary breadth, which corresponds to vocabulary size, was assessed using the receptive Peabody Picture Vocabulary Task (BPVS-3 for English, I. M. Dunn et al., 2009; EVIP for French, L. M. Dunn et al., 1993, PPVT-III-NL for Dutch, 2005). In this task, children are presented with four pictures, they hear a single word and are asked to point to the corresponding picture. Administration followed the instructions in the manual, starting at the age-appropriate starting set and moving up (and if necessary, down) until the required number of errors was met. The total number of correct responses (i.e., the raw score) was included as dependent variable in the analyses. We refrained from using standard scores as these are inaccurate for bilingual children given that they are not adjusted for reduced experience in the SL.

– Vocabulary depth

Vocabulary depth, which corresponds to how well words are known, was assessed using the Word Classes sub-test of the CELF-5 in English (Semel et al., 2017) and CELF-4 for Dutch (Kort et al., 2008) and French (Wiig et al., 2019). In this task, children hear words and are asked to indicate which words belong together. As the task progresses, the number of words from which children need to make a selection increases from three to four and visual support in the form of pictures is removed. Administration followed the instructions in the manual until children reached the end or failed to provide a correct response to the required number of consecutive items (four in English, five in Dutch and French). The proportion of correct responses out of the total number of items answered was included as dependent variable in the analyses.

Procedure

Ethics approval was obtained from the research institutions in each country [blinded for review]. Informed written consent was obtained from all parents. Children were tested individually in their home or at school by a research assistant proficient in the respective SL. Because part of the study took place during the Covid-19 pandemic, some children (n = 15 in the Netherlands and n = 6 in the UK) were tested online via Zoom. Most parents completed the questionnaire by themselves on their mobile phone, tablet or computer, but for some, the questionnaire was administered in an interview or parents were assisted by a research assistant or teaching assistant, either in the SL (n = = 30 in France) or the HL (n = 10 in the Netherlands).

Analysis

As a first step in answering our first research question, we conducted a Principal Component Analysis using normalised data and inspected the resulting components to discover the latent structure of the Q-BEx composite measure of richness. Adopting an approach incorporating both science (the data and the loadings) and "art" (the partly subjective judgement of conceptual meaning), we examined the factor loadings for all the components which were needed to account for 80% of the variance in the data. We repeated this analysis separately for the HL and SL.

We subsequently used the results of the PCA to derive new, alternative data-driven composite scores for each language and compared the predictive power of these alternative scores with the original score. For the SL, our dependent variable was children's scores on the two vocabulary tasks and the sentence repetition task. For the HL, this was their understanding and speaking skills as evaluated by the parents. Factor loadings higher than .3 were interpreted. To address our second research question, we included in this comparison a version of the original composite score excluding parental education (but including interlocutor diversity and proficiency) as well as version excluding both parental education and interlocutor diversity and proficiency (i.e., including only the data about the frequency of various activities in each language). We fitted models with parental education as a covariate and without.

We started with the optimal model identified in our previous papers (Prévost et al., in prep; Serratrice et al., in prep). In those analyses, the variables used to compile the composite richness score were included in the model separately. We refitted the best-fitting model from those analyses by replacing all the individual richness variables retained in that optimal model with each of the alternative (combinations of) scores listed above in turn. We then compared the goodness-of-fit for the resulting models by examining the AIC scores: the model with the lowest AIC was considered best-fitting.

Throughout, all the ordinal predictor variables were transformed into numeric variables to optimise the readability of model summaries.

We answered our final research question using a two-way orthogonal partial least square analysis (O2-PLS) (Trygg, 2002). This is a method that can identify shared information between two sets of variables (i.e., richness in SL vs richness in HL) while separating the unique information from each set (i.e., what is orthogonal to the shared information). In addition to evaluating the amount of joint vs. orthogonal information, we interpreted the resulting latent variables to determine which individual richness variables from our original composite score were relevant to both the HL and SL (i.e., the shared variance), and which were relevant for specific languages (i.e., the orthogonal variance). We furthermore conducted a linear regression analysis of factor loadings with Country as predictor and each of the latent variables defined by the O2-PLS analysis to determine whether the variability in the factor loadings varied between France, the Netherlands and the UK. Finally, we carried out correlational analyses between parental education and each of these latent variables to further explore the relationship between parental education and richness.

All analyses were conducted using R (version 4.3.2), using the following packages: betareg (3.1-4) (Cribari-Neto & Zeileis, 2010), ordinal (12-4) (Christensen, 2023) , factoextra (1.0.7) (Kassambara & Mundt, 2020), FactoMineR (2.10) (Kassambara & Mundt, 2020), CCA (1.2.2) (González et al., 2008), o2plsda (0.0.18) (Guo et al., 2022).

**Results**

Descriptives for the variables which form the basis of the Q-BEx richness score (except parental education) are provided for all children together for both languages in Figures 1 and 2. Figure 1 presents the frequency-based variables (activities in each language) and Figure 2 presents interlocutor diversity and proficiency. For the same data per country, see Supplementary Materials, S1.

Figure 1. Descriptives for individual richness variables for all children in the HL and SL.

Panel A. Frequency of literacy activities (i.e., reading and writing) in each language, all children together.
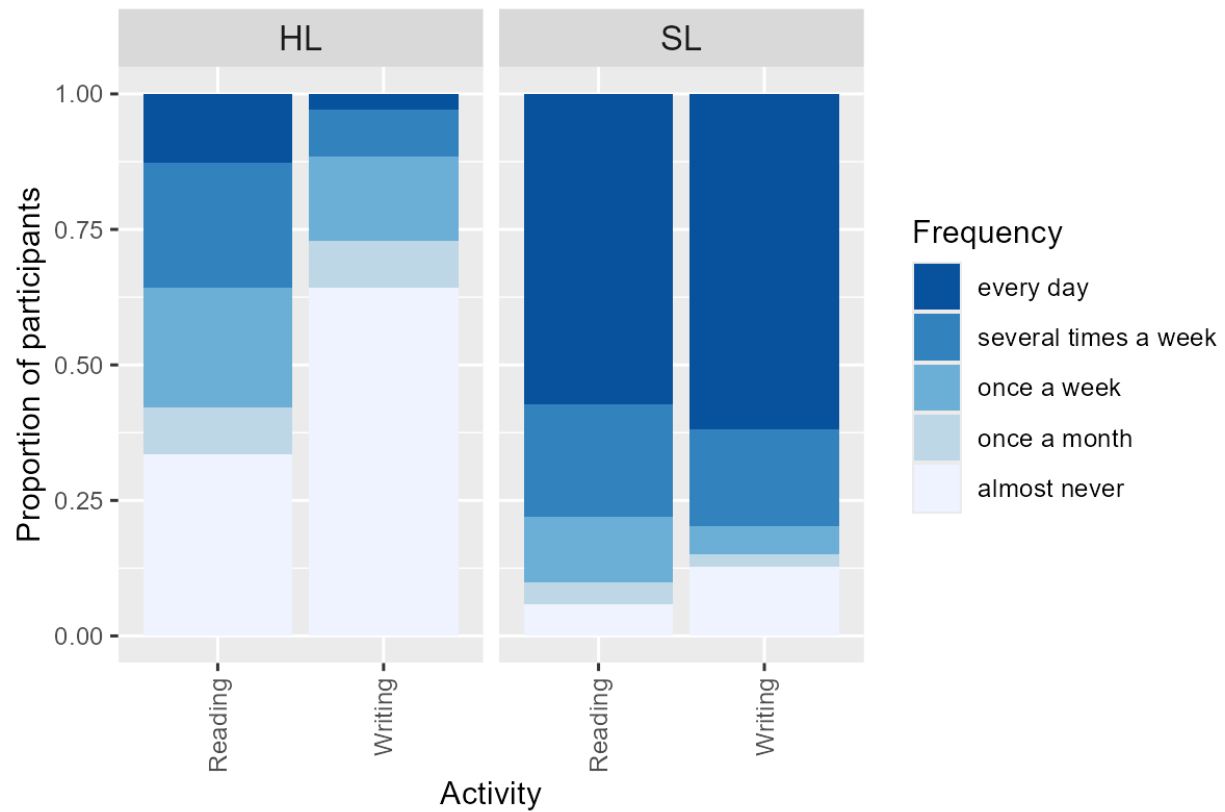
Panel B. Frequency of education-related activities (i.e., language lessons at mainstream school, language lesson outside mainstream school, time spent doing homework) in each language, all children together.
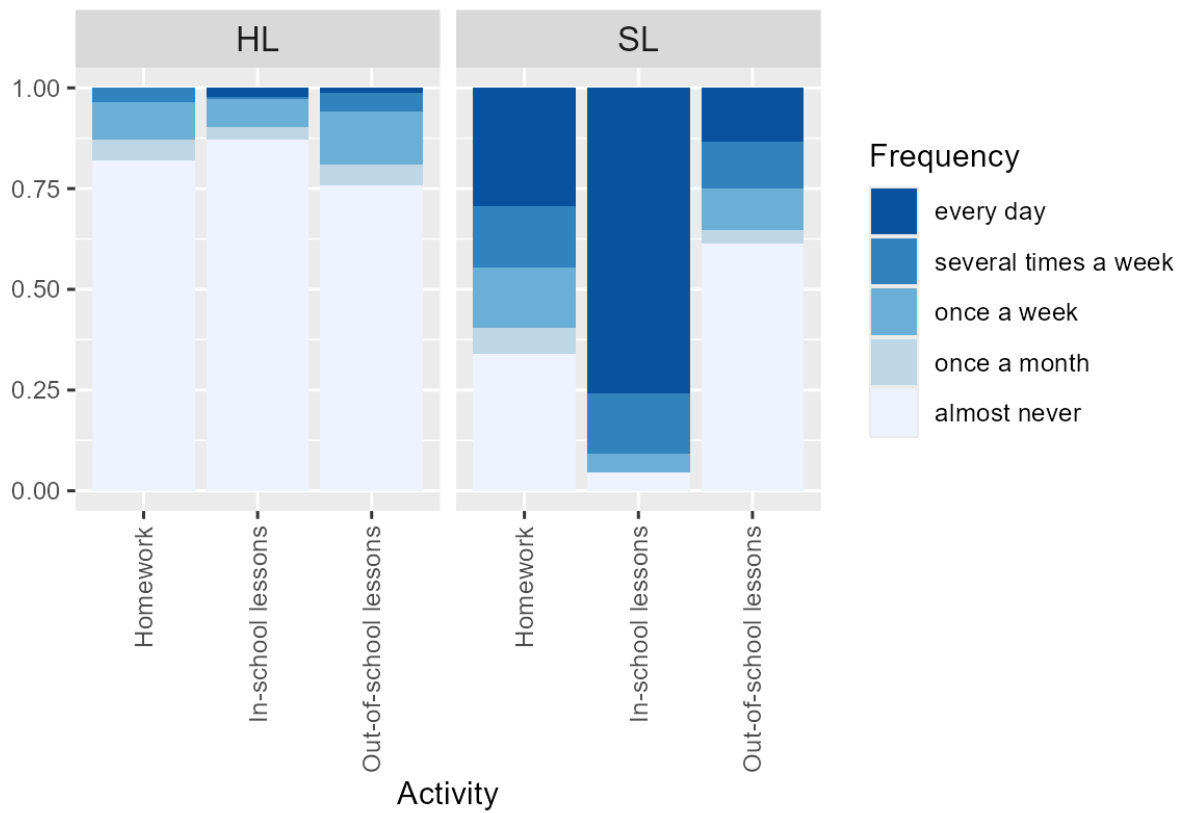
Panel C. Frequency of time spent with friends and on organised and tech-related activities, in each language, all children together.
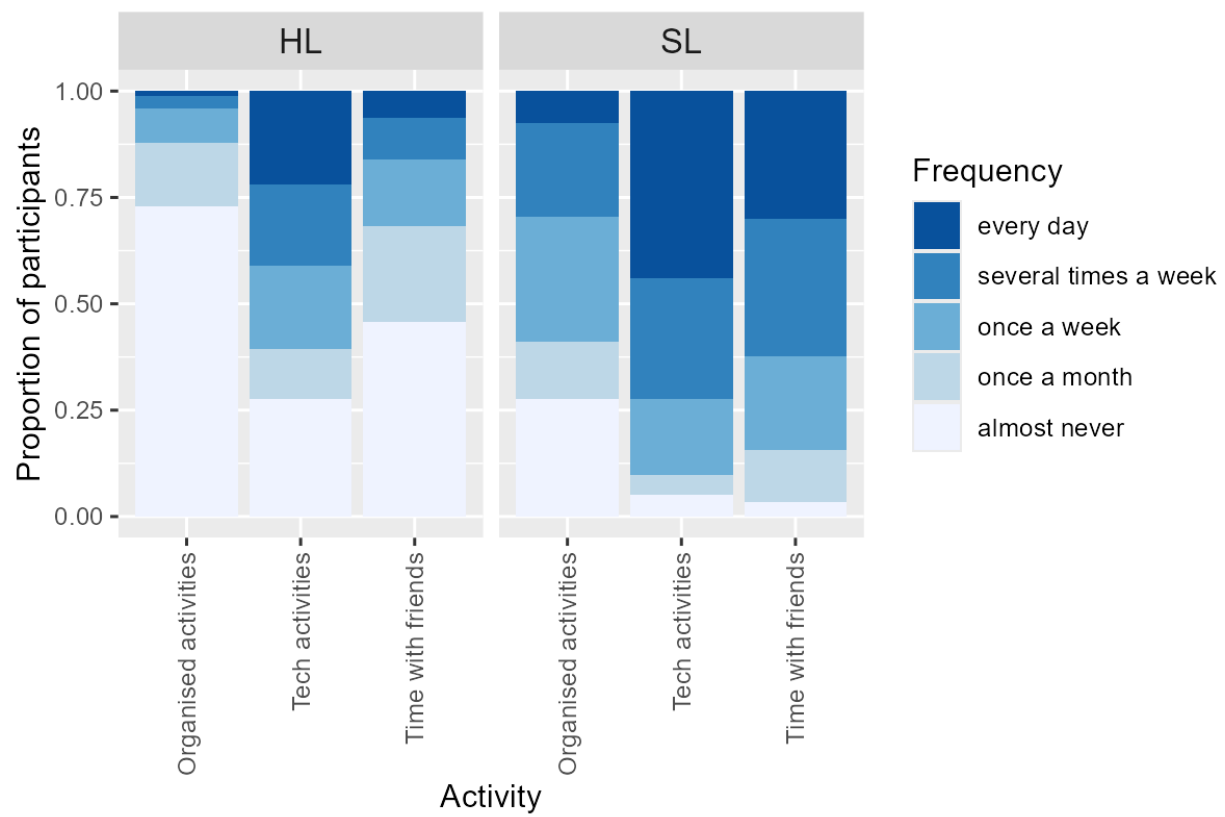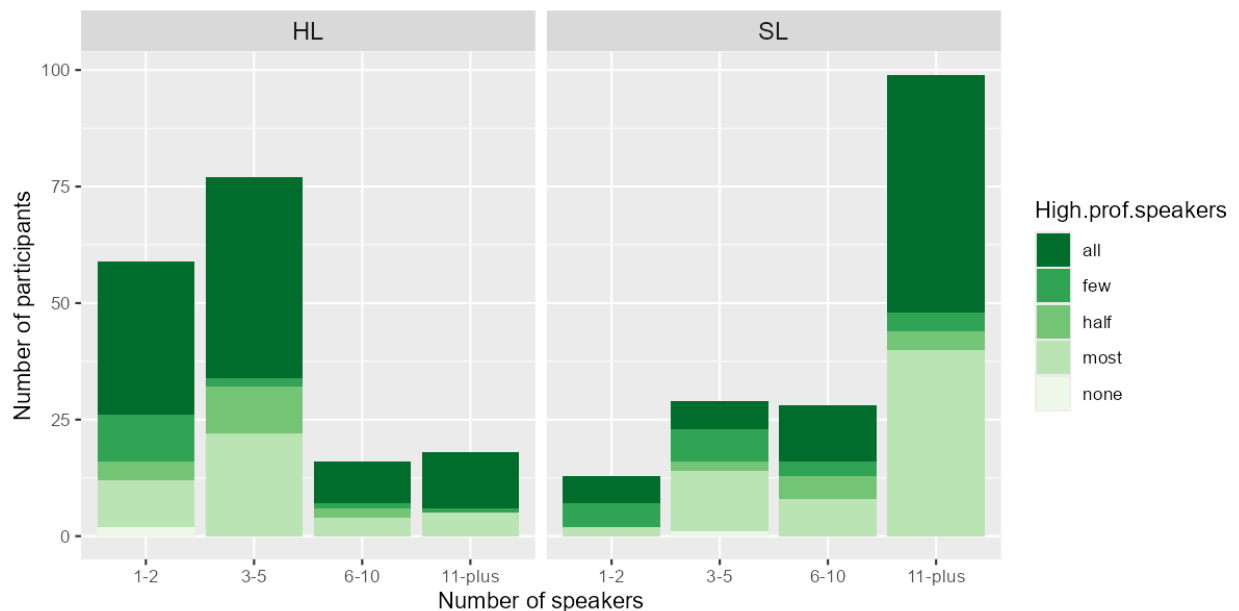
Figure 2. Number of different people who speak the language to the child at least once a week and how many of these speak the language very well, all children together (left panel: HL, right panel: SL).

The frequency of most activities (Figure 1) is greater in the SL than the HL. This difference is particularly striking for writing (Panel A) and education-related activities (Panel B), whereas for reading (Panel A), tech activities and time spent with friends, there are also many children who frequently engage in such activities in the HL. Most children hear SL input from many (10+) different speakers and on the whole, most, and in many cases all, of these speak the language very well (Figure 2). At the same time, there are also many children for whom exposure to the SL comes from a more limited number of speakers and from less proficient speakers. For the HL, the pattern is somewhat different: the vast majority of children interact with a (≤ 5) more limited number of speakers in that language and in several cases, these are not considered to speak the language very well.

Whilst these responses provide a useful insight into the diversity of children's bilingual experience, we note that some of the parents' responses to the two questions about interlocutor diversity and interlocutor proficiency were implausible. More specifically, it is not clear what it means when parents indicated that a child heard a language from 1 to 2 people and *most* of them speak the language very well. Luckily, the frequency of such implausible answer combinations was limited, and the vast majority of parents' responses make sense (see [blinded for review] for more details).

Deriving alternative composite scores using principal components analysis (PCA)

We first analysed parents' responses to the questions about richness for the HL (excluding parental education). Correlational analyses between the individual variables revealed moderate to strong positive relationships between time spent in language lessons (inside and outside school), writing and homework, and between interlocutor diversity and time spent with friends (full details are provided in Supplementary Materials, S2). The results of the PCA indicated that six components were needed to account for 80% of variance in the data (see S3 for the screeplot). The factor loadings for each of these components is given in Table 2.

Table 2. Factor loadings for different components in PCA of richness variables for HL

Richness variable	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Reading.freq.HL1	0.34	0.05	0.53	0.02	0.05	0.29
Writing.freq.HL1	0.46	0.09	0.07	0.17	-0.05	0.32
Homework.freq.HL1	0.45	0.30	-0.21	-0.06	-0.01	-0.03
In.school.lessons.freq.HL1	0.37	0.15	-0.14	0.08	-0.15	-0.79
Out.school.lessons.freq.HL1	0.45	0.18	-0.13	-0.21	0.03	0.21
Tech.activities.freq.HL1	0.20	-0.34	0.36	0.07	0.72	-0.30
Time.friends.freq.HL1	0.18	-0.53	-0.14	0.36	-0.14	0.11
Org.activities.freq.HL1	0.11	-0.30	-0.47	-0.63	0.32	0.12
HL1.speakers.n	0.20	-0.51	-0.23	0.24	-0.27	0.03
HL1.high.prof.speakers.n	0.09	-0.31	0.46	-0.57	-0.50	-0.16

The first principal component accounted for 29.2% of the total variation and was mainly constructed by the first five variables. This suggests that formal education and literacy captured most of the variation we observed in HL richness. The second component (17.9% variance) showed the opposite pattern, with the highest loadings for the final five, social/leisure-related variables. The third component (13.2% variance) contrasted literacies (i.e., reading, use of tech such as the internet, and high-proficiency interlocutors) with organised activities. Similar contrasts were also captured in components 4 (8.7% variance; high-proficiency interlocutors vs. organised activities) and 5 (7.5% variance; tech-related activities vs. high-proficiency interlocutors), whereas component 6 (7% variance) represented language lessons in school.

Correlational analyses between the individual variables for the SL revealed moderate to strong relationships between interlocutor diversity, time spent on organised activities, with friends and on tech-related activities, as well as between SL lessons outside school, reading, writing and homework. There was also a negative correlation between interlocutor diversity, number of high-proficiency interlocutors and SL lessons outside of school. (Full details are provided in S2.) Similar to the HL, the results of the PCA for the SL indicated that six components were needed to account for 80% of variance in the data (see S3 for the screeplot). The factor loadings for each of these components is given in Table 3.

Table 3. Factor loadings for different components in PCA of richness variables for SL

Richness variable	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
Reading.freq.SL	0.44	0.03	0.19	0.31	0.01	0.51
Writing.freq.SL	0.42	0.29	0.30	0.17	-0.04	0.09
Homework.freq.SL	0.29	0.51	0.06	0.11	0.15	-0.13
In.school.lessons.freq.SL	0.25	0.19	0.21	-0.81	-0.13	0.09
Out.school.lessons.freq.SL	0.09	0.48	-0.40	0.09	0.05	-0.45
Tech.activities.freq.SL	0.33	-0.27	-0.41	-0.27	-0.02	0.25
Time.friends.freq.SL	0.35	-0.12	-0.48	-0.13	0.46	-0.05
Org.activities.freq.SL	0.32	-0.16	-0.18	0.11	-0.81	-0.29
SL.speakers.n	0.30	-0.44	0.04	0.27	0.20	-0.18
SL.high.prof.speakers.n	0.21	-0.28	0.49	-0.16	0.21	-0.57

There is no variable that stands out in the first component (22.5% variance). Rather, this component captured the correlation between the variables (with the exception of SL lessons outside school), indicating an averaging effect. The second component (18.9% variance) represented formal education (i.e., SL lessons outside school, writing and homework) in contrast with interlocutor diversity, whereas the third component (13.2% variance) tapped into time spent with friends and on tech-related activities in contrast with the proportion of high-proficiency interlocutors. Component 4 (9.4% variance) reflected SL lessons in school, and component 5 (8.6% variance) contrasted organised activities and time spent with friends. Finally, component 6 (6.9% variance) represented the proportion of high-proficiency interlocutors versus time spent reading.

The groupings of variables which stand out the most across both languages (HL and SL) relate to literacy and formal aspects of language learning (i.e., components 1, 3 and 6 for HL, components 2 and 6 for SL) or to social or leisure-related aspects of language learning (i.e., components 2, 3 and 5 for HL, components 3 and 5 for SL).

We used the clusters of variables emerging from the PCA to inform the creation of new composite scores. These two data-driven alternatives to the original richness measure involved a score focusing on literacy and formal education (i.e., reading, writing, homework, language lessons in school and outside school) and a score focusing on social and leisure-related activities (i.e., tech, organised activities, friends, interlocutor diversity and high-proficiency interlocutors). Note that the first of these two alternatives was based on frequency data only, whereas the second included the more qualitative aspects of interlocutor diversity and proficiency. Figures 3 and 4 provide an overview of these scores, alongside the original Q-BEx richness measure which included parental education (i.e., highest level of parental education in any language) and an alternative where parental education was excluded. There was no significant correlation between parental education and the alternative score excluding parental education, either for HL (r(168) = 0.03, p = .699) or SL (r(170) = 0.08, p = .305).

Figure 3. Comparison of original Q-BEx richness scores (converted back to a 0-4 scale) including parental education (QB.original) and excluding parental education (QB.no.SES), and two data-driven alternative scores based on literacy/formal variables (Literacy) and social/leisure variables (Social) for the HL.
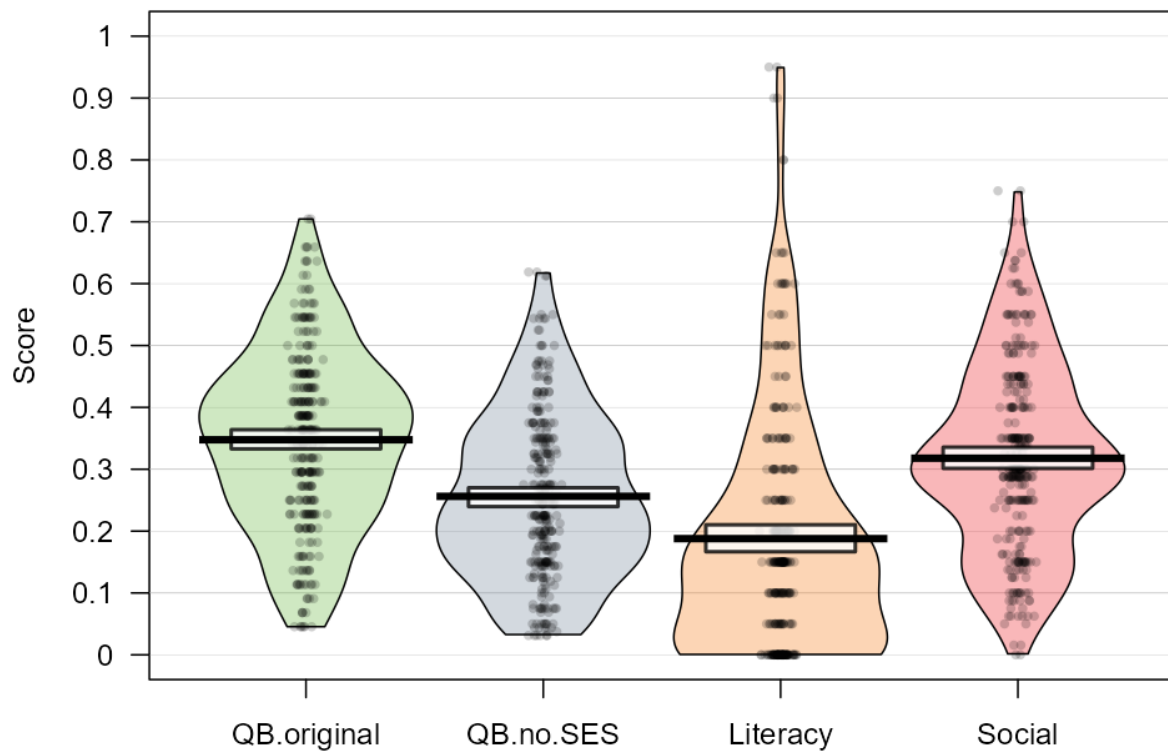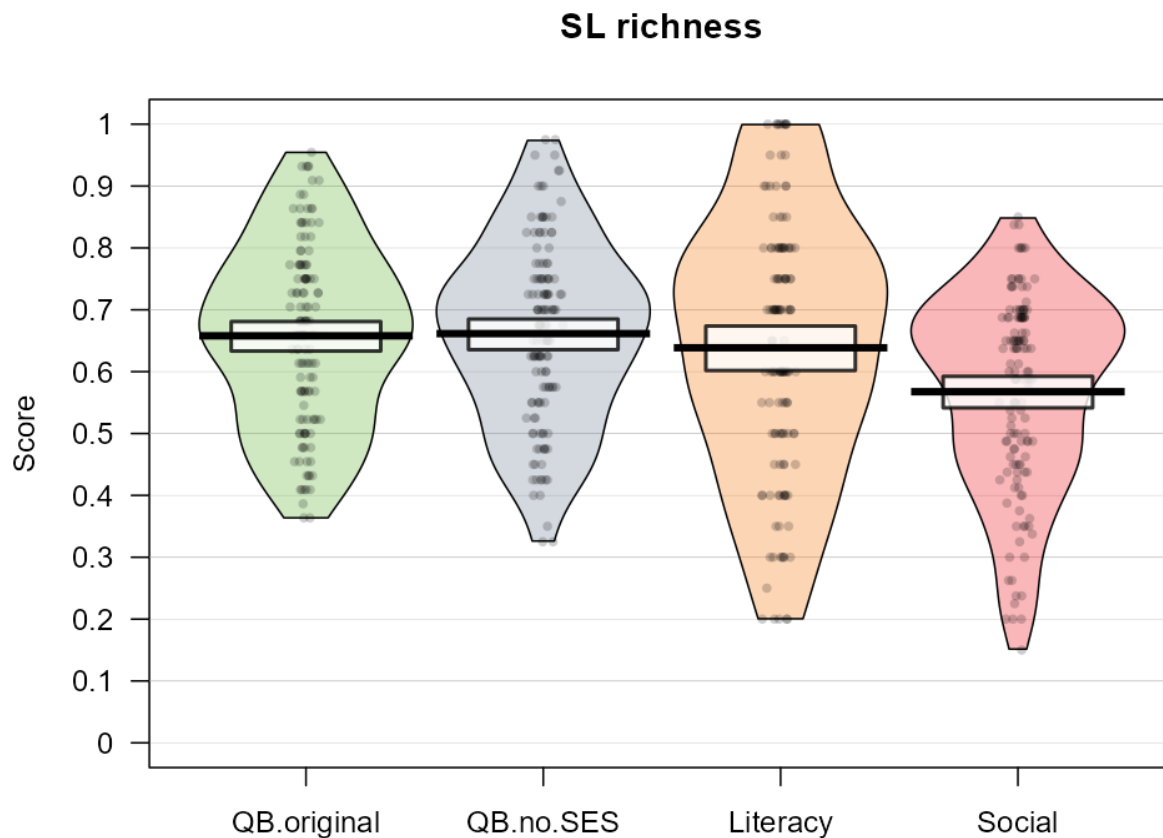
Figure 4. Comparison of original Q-BEx richness scores (converted back to a 0-4 scale) including parental education (QB.original) and excluding parental education (QB.no.SES), and two data-driven alternative scores based on literacy/formal variables (Literacy) and social/leisure variables (Social) for the SL.

Comparing the richness scores using an information-theoretic approach

Descriptives for the outcome variables are provided for all children in Figure 5 for HL and SL proficiency based on parental estimate and in Table 4 for SL proficiency based on objective measures. NB: Parental estimates for SL proficiency are included here for reference. We do not use these data in the analyses because some of the parents were not proficient in the SL and hence may not have been able to provide accurate estimations of their child's abilities in that language. Note that this is much less a problem for the parental estimates of HL proficiency as in the vast majority of cases, it was (one of) the HL-speaking parent(s) who completed the questionnaire.

Figure 5. Number of children at each proficiency level for HL and SL outcomes (parental estimates of children's understanding, speaking, reading and writing skills).
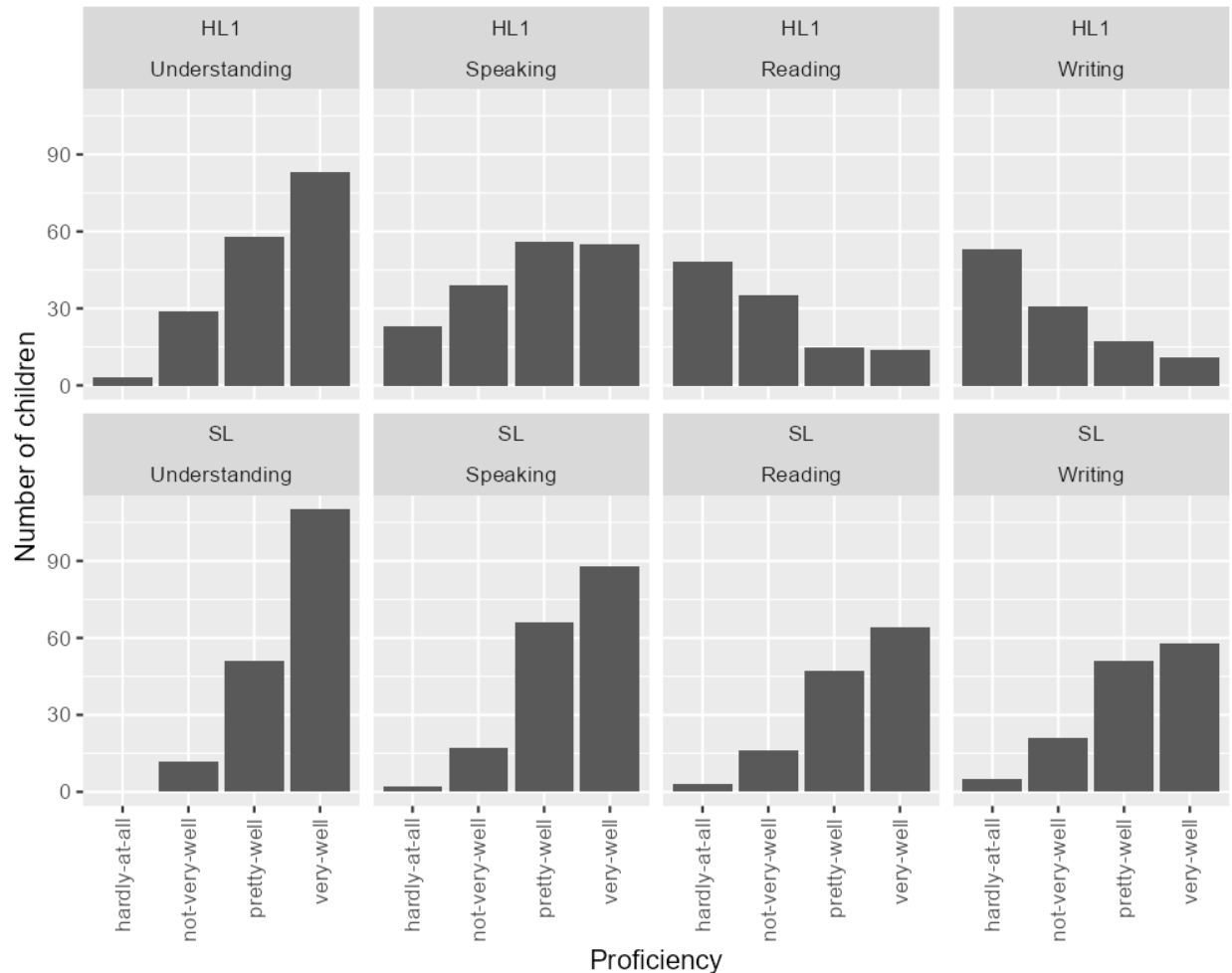
Table 4. SL outcomes (scores on vocabulary depth, vocabulary breadth and sentence repetition tasks) for all children.

Outcome	M (SD)
Vocabulary breadth (raw score)	89.9 (24.7)
Vocabulary depth (% correct)	45 (18)
Sentence repetition (% correct)	56 (35)

For each outcome, we compared the four richness scores by re-fitting the best-fitting model from our previous analyses (Prévost et al., in prep; Serratrice et al., in prep) with each of these scores in turn. As

the models differed only with respect to richness measure, the one with the best fit can be considered

the one with the most informative predictor of language proficiency out of our candidate set. The

results of this comparison are summarized for each outcome in Table 5.

Table 5. Goodness-of-fit comparison for models including different measures of richness on bilingual children's HL and SL outcomes. Best-fitting model based on Akaike's information criterion (AIC) is highlighted.

| | | HL | | | | SL | | | |
| | | Parental estimates | | Vocabulary breadth | | Vocabulary depth | | Sentence repetition | |
Richness score[a]	Parental education as covariate?	df	AIC	df	AIC	df	AIC	df	AIC
Original		9	532.0809	13	1010.191	6	-119.0464	8	-214.2713
– excl. parental education	Yes	10	523.1939	14	1010.411	7	-117.7203	9	-213.9947
– excl. parental education	No	9	523.9689	13	1010.504	6	-119.1225	8	-212.6130
Alternatives	Yes	11	535.8111	15	1012.410	7	-117.9772	9	-215.3010
Alternatives	No	10	537.1205	14	1012.503	8	-117.1312	10	-215.4800

[a] Original = Q-BEx score, including parental education, unless stated otherwise; Alternatives = data-driven scores based on literacy/formal variables (Literacy) and social/leisure variables (Social)

For HL proficiency (as estimated by the parents), the best-fitting model was the one containing the original version of the richness score which excluded parental education from the score itself, but which included it as a covariate instead. Note, however, that there was little difference between this model and the one with the same richness score but excluding parental education as covariate. Furthermore, even though the models with the alternative scores had a worse fit, these scores were nevertheless significant in the resulting model. See Table 1 in S4 for full details of the best-fitting model.
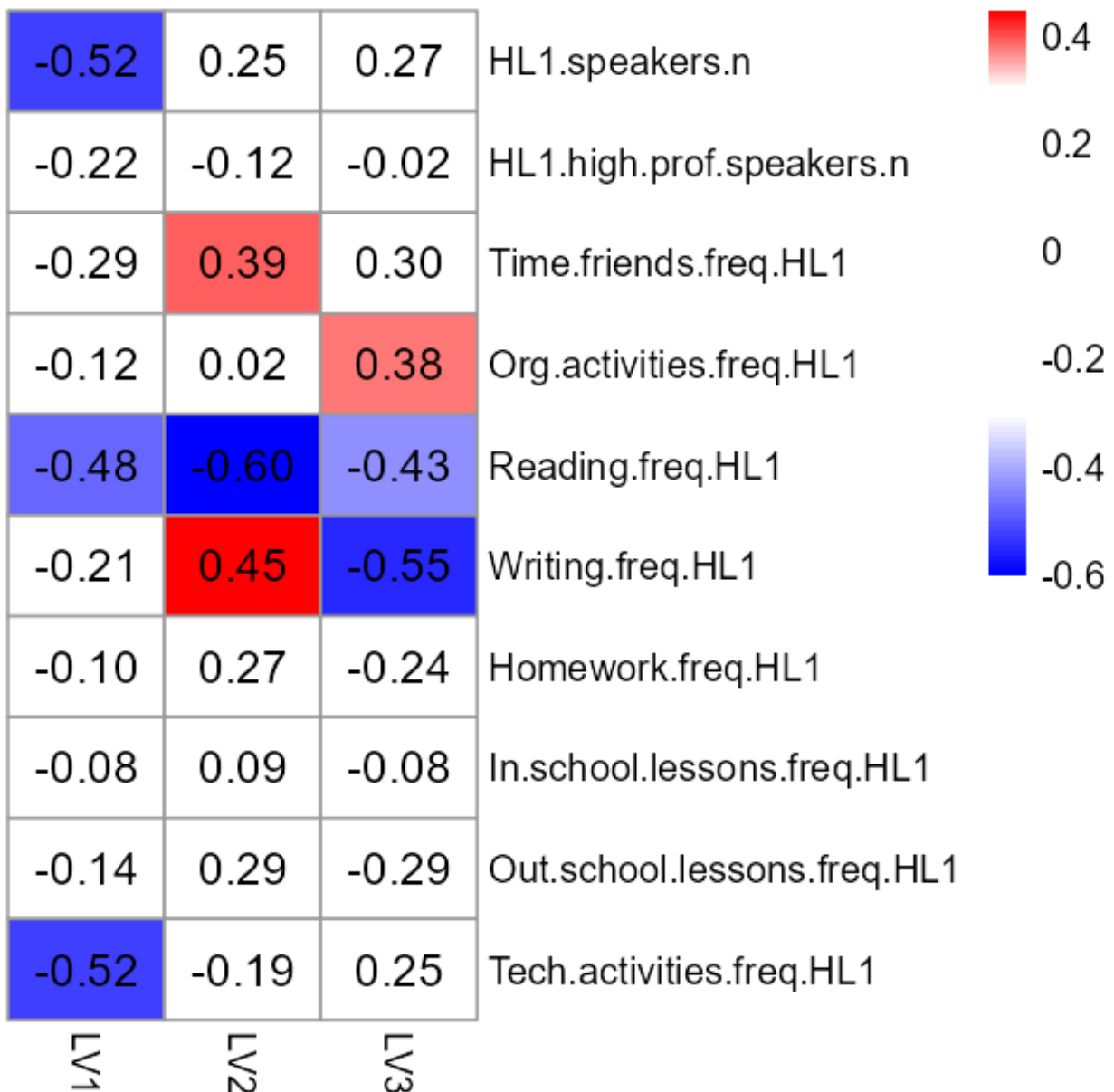
For SL proficiency, the best-fitting model for vocabulary breadth contained the original Q-BEx score (i.e., including parental education), though once again, the difference between this model and the model containing the original Q-BEx score excluding parental education was minimal, irrespective of whether parental education was included as covariate. For vocabulary depth, the best-fitting model contained the original Q-BEx score excluding parental education and without parental education as covariate, although the AIC is almost indistinguishable from the model containing the original Q-BEx score. The best-fitting model for sentence repetition contained the alternative scores without parental education as covariate, although this was barely different from the model including parental education as covariate. In this model, only the alternative score based on social/leisure variables was a significant predictor of children's scores. See Tables 2 through 4 in S4 for full details for the best-fitting models.

Comparing latent variables underlying richness across languages (HL vs. SL) and countries
The purpose of this analysis was to identify which information was shared between the richness variables in the SL and the HL, while at the same time separating the unique information from each set. Included in this analysis were 39 children in France, 76 children in the Netherlands and 54 children in the UK. The results of the O2-PLS analysis revealed that the common covariance between SL and HL richness was explained by 84% of the variance in HL and 77% of that in the SL, in three latent variables. Figure 6 presents heatmaps showing the associations between the latent variables and the individual richness variables for the HL (Panel A) and SL (Panel B), respectively.

1

Figure 6. Heatmap showing (by column) the three latent variables (LV) capturing shared variance in richness across the HL (panel A) and SL (panel B). The colours (blue vs red) highlight the contrast captured by each component. The intensity (light vs dark) reflects the value of the loadings (only the darker cells, indicating values below -0.3 or above 0.3, are interpreted). Non-interpreted values are not coloured.

Panel A. HL

LV1	LV2	LV3	
-0.52	0.25	0.27	HL1.speakers.n
-0.22	-0.12	-0.02	HL1.high.prof.speakers.n
-0.29	0.39	0.30	Time.friends.freq.HL1
-0.12	0.02	0.38	Org.activities.freq.HL1
-0.48	-0.60	-0.43	Reading.freq.HL1
-0.21	0.45	-0.55	Writing.freq.HL1
-0.10	0.27	-0.24	Homework.freq.HL1
-0.08	0.09	-0.08	In.school.lessons.freq.HL1
-0.14	0.29	-0.29	Out.school.lessons.freq.HL1
-0.52	-0.19	0.25	Tech.activities.freq.HL1

Panel B. SL

	LV1	LV2	LV3	
	-0.35	-0.29	-0.31	SL.high.prof.speakers.n
	-0.37	-0.18	-0.63	Reading.freq.SL
	-0.35	0.18	-0.22	Writing.freq.SL
	-0.23	0.64	-0.13	Homework.freq.SL
	-0.41	0.06	0.16	In.school.lessons.freq.SL
	-0.13	0.58	0.08	Out.school.lessons.freq.SL
	-0.34	-0.14	0.50	Tech.activities.freq.SL
	-0.31	0.02	0.27	Time.friends.freq.SL
	-0.19	-0.17	0.13	Org.activities.freq.SL
	-0.37	-0.20	0.26	SL.speakers.n

The first (leftmost) latent variable (LV1) captured most of the shared information across languages (85% for SL and 62% for HL). Three richness variables stood out in LV1 for the HL (i.e., interlocutor diversity, reading, and tech-related activities); they were thus sufficient to capture the joint information in this first latent variable, reflecting factors which depend on the presence of a HL-speaking community and resources in the HL. LV2 for the HL was more difficult to interpret, contrasting reading, on the one hand, with writing and time spent with friends, on the other. The association between writing and friends may
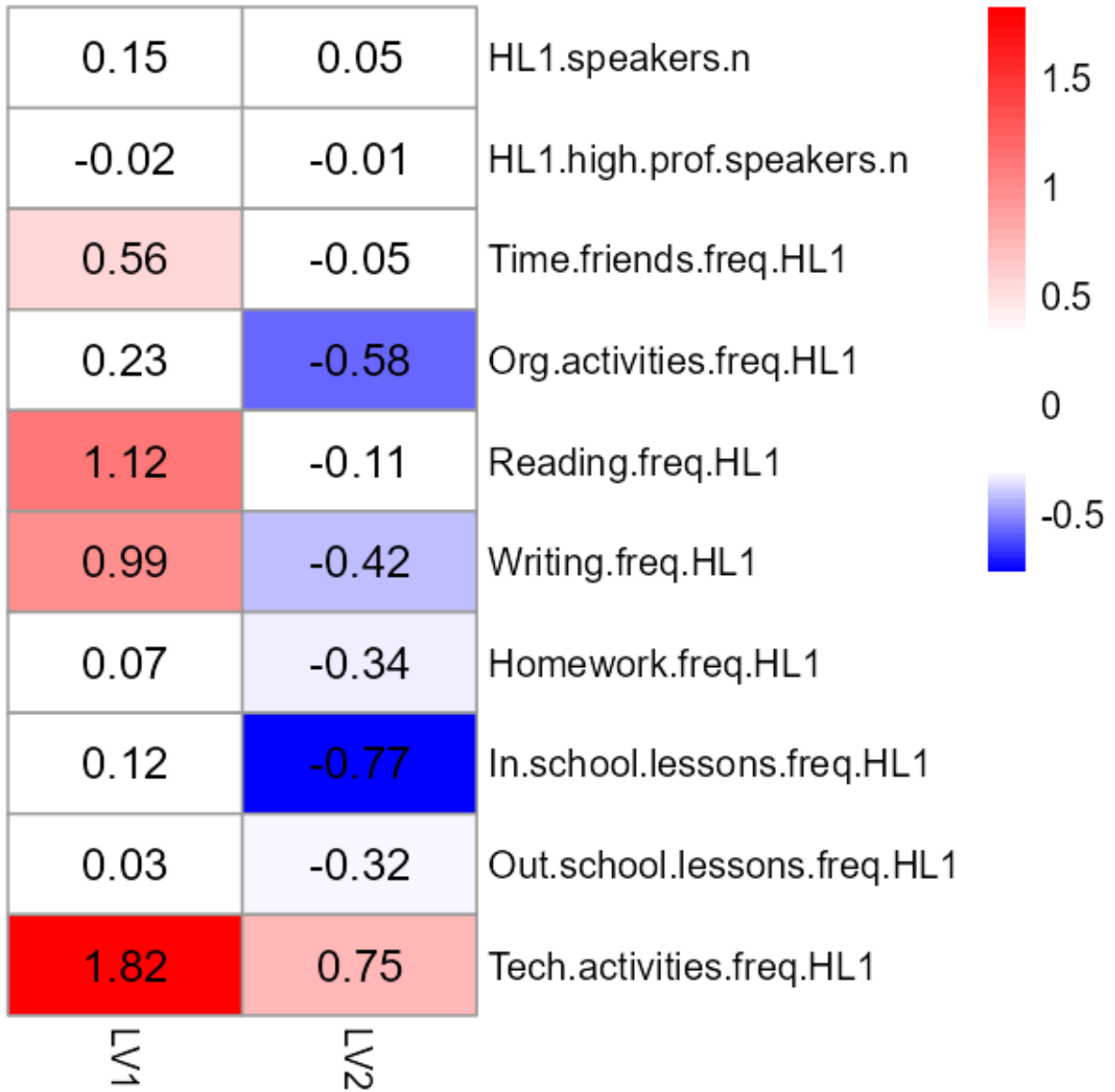
reflect the availability of HL schooling. Finally, in LV3 reading and writing (i.e., literacy) were contrasted with organized (i.e., social) activities.

For the SL (Panel B), the first latent variable included everything except homework and out-of-school lesson (picked up in the second latent variable), and organised activities (not picked up in any of the shared variance in the SL). The second latent variable (LV2) involved out-of-school activities relating to formal learning (i.e., language lessons outside school and homework), whereas the third latent variable (LV3) contrasted reading with tech-related activities. These two latent variables thus reflect what we might dub formal and informal literacies, respectively.
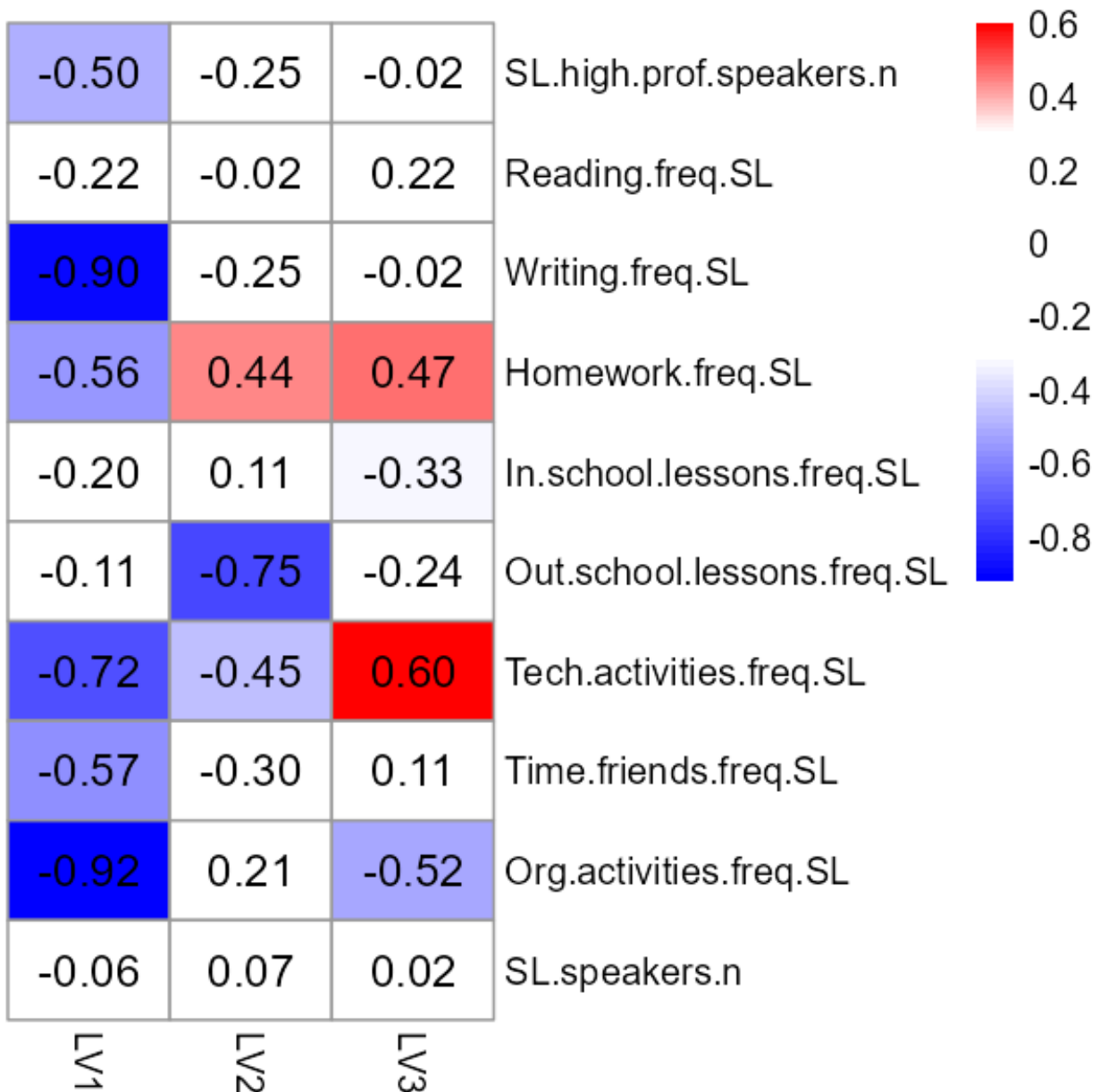
We turn now to the orthogonal variance between the HL and SL, that is, the richness information which is specific to each language - see Figure 7.

Figure 7. Heatmap showing the components capturing orthogonal variance in richness across the HL (panel A) and SL (panel B) alongside correlations between these components and the individual richness variables.

Panel A. HL

Panel B. SL

	LV1	LV2	LV3	
	-0.50	-0.25	-0.02	SL.high.prof.speakers.n
	-0.22	-0.02	0.22	Reading.freq.SL
	-0.90	-0.25	-0.02	Writing.freq.SL
	-0.56	0.44	0.47	Homework.freq.SL
	-0.20	0.11	-0.33	In.school.lessons.freq.SL
	-0.11	-0.75	-0.24	Out.school.lessons.freq.SL
	-0.72	-0.45	0.60	Tech.activities.freq.SL
	-0.57	-0.30	0.11	Time.friends.freq.SL
	-0.92	0.21	-0.52	Org.activities.freq.SL
	-0.06	0.07	0.02	SL.speakers.n

Orthogonal variance in the HL was captured by two latent variables only. In the first (LV1), four richness variables were important (i.e., friends, reading, writing and tech-related activities), with tech-related activities standing out most. LV1 thus reflects (the availability of) resources and friends in the HL. The second latent variable reflected all the frequency-based richness variables, except time spent with friends, reading, and interlocutor diversity and proficiency. For the SL, three latent variables were needed to capture the orthogonal variance. The first included interlocutor diversity plus all frequency-based richness measures except reading and those relating to SL lessons, with writing and organised

activities having the highest loadings. This variable thus broadly reflects activities which occur outside the home or – in the case of homework – most likely at home but focussing on content relating to school. Homework was also important in the other two latent variables. In LV2, language lessons outside school and tech-related activities were the most important variables, reflecting engagement with resources and/or the SL community outside of school. Finally, in LV3 homework and tech-related activities contrasted with organised activities outside the home.

Summarising the results of the O2-PLS analysis so far, the patterns of associations between variables were different in the two joint information matrices. For the HL, the most important richness indicators were tech-related activities, time spent with friends, organised activities, writing and reading. For the SL, these were homework, language lessons outside school, and reading. Note, however, that reading clustered differently in the SL vs. HL. In the orthogonal information matrices, the variable with the highest factor loading in the first component differed for the HL (tech-related activities) and the SL (i.e., organised activities, closely followed by writing).

Next, we investigated whether the mean of the loadings in each component of the shared and orthogonal information between the HL and SL differed across countries. Mostly there was no significant difference across countries, except in the first latent variable of the joint information contributed by the HL (see model and plots in the Supplementary Materials, S5). In that component, the Netherlands patterned differently from the UK and France.

Finally, we examined the extent to which each component in the joint and orthogonal information was related to parental education. Parental education was significantly correlated with each of the three components in the shared information for both the HL (r = .133 for the first component, r = -.400 for the second, r = -.282 for the third, p <.001 for all components) and the SL (r = -.289 for the first component, r = -.498 for the second, r = -.184 for the third, p <.001 for all components). (NB: the sign of the coefficient cannot be interpreted.) In the orthogonal information, the first two components were also significantly correlated with parental education (HL: r = .112 for the first component, r = .157 for the second, p < .001 for both; SL: r = .108, p < .001 for the first component, r = .020, p = .020 for the second). It was only in the third SL component that the relationship was not significant (r = -.008, p = .417). These results need to be interpreted in light of the latent variables defined by each component. Highest parental education in any language (i.e., the Q-BEx proxy for SES) was significantly associated with formal and informal literacies in both languages (as shown by the aspects that contribute the most to the shared information and some of the orthogonal information), and with social and tech activities, especially in the HL. While the sign of the correlation coefficient cannot be interpreted in this analysis

(as the sign of the loadings is not numerically interpretable), our other analyses showed this correlation to be positive, such that children with more educated parents tended to have a richer language experience across both their languages. Full details are provided in Supplementary Materials, S6.

Discussion

The aim of this study was to unpack the richness of bilingual children's language experience as a predictor of their language proficiency in both the heritage and the societal language. Specifically, using data from 5- to 9-year-old children across three different countries, we investigated whether the composite richness score in the Q-BEx questionnaire was fit for purpose by examining its predictive power in comparison with alternative data-driven scores and by determining whether it involved the same latent variables across the societal and heritage languages and across countries. We furthermore investigated whether parental education (as a proxy for SES) and interlocutor diversity and proficiency should be included as part of the richness score, as intended, or whether a score without one or both of these variables was sufficient to predict children's language outcomes.

Comparing the original Q-BEx score to alternative data-driven measures
The Principal Component Analysis revealed five (SL) or six (HL) components that were needed to account for 80% of the variance in the data. Two main contrasts emerged, namely between literacy and formal aspects of language learning and social-/leisure-related variables. This held for both languages. Interestingly, time spent reading did not always clearly cluster with more formal aspects of literacy such as writing and homework, probably because in the context of shared book-reading, reading also includes a leisure/social dimension. This was most clearly the case for the HL where children who could not read in the HL may have been exposed to written texts by being read to. In contrast, the contribution of time spent writing was quite limited, possibly because the children in our sample were quite young and because very few children could write in their HL. For older children, writing will likely be more relevant and for this reason, we do not advocate removing this question even. The same holds for the question about homework. Results may furthermore when language outcomes involve reading skills.

Informed by the PCA, we created two alternative data-driven scores, one incorporating literacy/formal learning variables and the other using leisure-/social-related variables. For the two vocabulary measures in the SL and for HL proficiency, the model containing the original Q-BEx score fared best. On the sentence repetition task in the SL, however, the model containing the alternative scores was a better fit, although one of these was a significant predictor of children's scores, namely the

one based on social-/leisure-related variables. In other words, for the children in our study, morphosyntactic outcomes were better captured by factors such as time spent with friends, engagement with tech-related and organised activities, and the number of high-proficiency speakers better captured than by writing, homework and reading.

In sum, we do not have robust evidence to suggest that alternative measures based on the PCA were more informative in predicting children's outcomes than the original Q-BEx composite score of input richness. For this reason, and because we would need more research evidence to inform practitioners as to how to interpret the two composite scores, we conclude, for now at least, that it is better to adopt a conservative approach and leave the richness score in the Q-BEx questionnaire as is. At the same time, we acknowledge that this conclusion is based on one dataset, albeit reasonable in size and diversity in terms of HLs, types of bilinguals and proficiency levels. Full details including the R script for the PCA analysis and the calculation of the alternative scores are available at [LINK] so that other researchers can determine whether different results may arise from different datasets.

SES and its relation to richness and to language outcomes
Our analyses also included comparisons with a variant of the original Q-BEx score excluding parental education, our proxy for SES, and we ran models both with and without parental education as a covariate. Overall, we found there was very little difference between the models containing (scores including) parental education and those without.

We found that parental education, either as part of a composite score, or as a covariate, was a predictor of children's outcomes in both their HL and their SL, in line with previous studies showing parental education to predict bilingual children's outcomes (e.g., Hoff et al., 2018; Paradis et al., 2022). At the same time, they also contrast with some earlier findings where SES was related to children's proficiency in the HL but not their SL (Lauro et al. 2020), or in their SL but not their HL (Prevoo et al., 2014). Given the difference in number and type of proficiency measures across languages in our study, this comparison should be interpreted with caution, however. Variation in how SES is operationalised may furthermore contribute to these mixed findings, and parental education may be a problematic proxy for SES in a migration context (Kašćelan & Parafita Couto, 2024).

Parental education was significantly correlated with the (latent) components of richness, suggesting that it can be used as a proxy for input richness (e.g., De Cat, 2021). This aligns with previous research showing that factors relating to richness such as attitudes towards education (e.g., Scheele et al., 2010) and home literacy practices (e.g., Kaltsa et al., 2020) correlate with SES. The results of the O2-

PLS analysis showed that some of these factors, for example those relating to formal and informal literacies, were associated with the shared and orthogonal variance in both the HL and SL, and that in turn, this variance was associated with parental education. In short, children from more privileged backgrounds (i.e., those with parents who have higher education levels) tended to have a richer language experience across both their languages, where 'richer' means – amongst other things – more diverse, more likely to include input from more proficient speakers, more access to tech-related activities and more frequent literacy practices. Once richness (without parental education) was included in the model, parental education did not capture (much) additional variability in children's proficiency scores. Richness was likely a better estimate than parental education because it measures the relevant dimensions more directly than SES.

Given that there was little difference in the models including parental education as part of the richness score or as a covariate and the models where parental education was excluded from the score or as a covariate, we can conclude that when there are concerns about collecting information from parents about their level of education or objections to including this variable for other reasons, a richness measure which does not incorporate parental education should still function as a good predictor of bilingual children's language outcomes. (Note that Q-BEx can calculate a richness score irrespective of whether the chooses to ask parents about their level of education.)

Whether we should refer to these specific characteristics of children's language experience directly, rather than using an overarching term such as "richness" or "input quality", as argued by MacLeod and Demers (2023), is an important discussion but we leave this question open here. Our Delphi consensus study ([anonymised for review]) indicated agreement about the importance of input richness/quality as an overarching construct to document, although the voices we included (from 29 countries) were predominantly from the European and North American context, which can have consequences on how input quality/richness is conceptualised and labelled.

Interlocutor diversity and proficiency

The alternative richness score derived using the PCA was based on literacy/formal learning variables used frequency data only. Comparing the predictive adequacy of this score with (one of) the original Q-BEx score(s), which included interlocutor diversity and interlocutor proficiency, allowed us to determine the contribution of these qualitative aspects of input richness in predicting children's language outcomes. As noted above, there was only one instance where the best-fitting model was the one containing the alternative richness measures and in that model, it was the score based on the social-

/leisure-related variables which was a significant predictor, not the literacy/formal learning score. Our findings thus underscore the importance of interlocutor diversity and proficiency as qualitative aspects of input richness which predict bilingual children's language outcomes (in line with e.g., Gollan et al., 2015; Unsworth et al., 2019), furthermore suggesting that composite measures including these variables are preferable to those based on frequency-based variables only.

Latent variables underlying richness across languages
In addition to examining which (component parts of) measures of input richness predicted children's language outcomes in their HL and SL, we also examined the extent to which the latent variables underlying richness were comparable across languages (i.e., across HL and SL, and across the three SLs). Studies on the role of input richness which directly compare bilingual children's two languages have almost exclusively focussed on the extent to which richness predicts language outcomes in the HL vs. SL (e.g., Sun & Yin, 2020) rather than on the composition of the richness measure itself (but see Paradis et al., 2020). This is important, however, because comparability across languages (HL and SL) and countries (i.e., across different SLs) is to a certain extent presupposed when the results of studies examining overall richness effects are compared with each other.

We found that the proportion of shared variance between the HL and SL was very high, demonstrating that the same type of information is relevant to richness in both languages. This is reassuring given that a certain degree of comparability across different contexts is assumed in parental questionnaires such as the Q-BEx. They are designed to elicit joint information about both HL(s) and SL and encourage users to compare results for the two. What stands out in the joint information is literacies (formal and informal) and social activities (especially in the HL).

There was also orthogonal variance, showing that some aspects of richness vary according to language status (i.e., SL vs. HL). More specifically, this language-specific variation varied in terms of the richness variables that constructed the latent variables. In the orthogonal information matrices, the variable with the highest factor loading in the first latent variable differed for the SL (i.e., organised activities, closely followed by writing) and the HL (i.e., tech-related activities). This contrast likely reflects variation in writing skills between the SL, which all children (eventually) acquire, and the HL, which many children may not, as well as differences in access to technology in specific HLs. As Paradis and colleagues note (2020), there may also be an economic component here, as limited financial resources may "play a role in determining richness of the home language environment" (p. 1272).

Generally, languages and language communities may also differ in terms of the available opportunities to engage in various language- and literacy-related practices (Scheele et al., 2010, p. 135) and this will contribute to the variation between children by constraining the range of possible answers parents can give for certain questions. For example, factors such as overall literacy levels in the community (of origin) and the availability of HL education may affect parents' responses to questions relating to children's reading and writing behaviours. Similarly, the frequency with which children might do homework in the HL will depend on a number of circumstances. Children who have access to HL schools may get homework in the HL as part of their education and hence parents' responses will in part depend on the specific HL and the availability of HL programmes for that language. This, in turn, may depend on the size of the HL community and where the family lives, something which may also impact the availability of organised activities and (high-proficiency and/or attritted) speakers in that language. Other circumstances under which bilingual children might complete homework in their HL include parents helping them with homework from their mainstream (SL) school and using the HL to do so, either because they are not proficient enough in the SL or because they prefer the HL. In this case, parents' responses will depend on their own SL proficiency and/or attitudes rather than the specific HL in question.

For some HLs (e.g., Turkish and Chinese), there are simply more resources available than in others (e.g., Berber and Tigrinya), digital and analogue, and certain HLs (e.g., English in the Netherlands) are afforded a certain status in society more broadly as well as being studied in school and may not only be more prevalent but also more positively valued than others. This, too, may have an indirect effect on the richness of children's environments when acquiring these languages. In short, then, there are several considerations, often beyond parents' control, which can affect how richness-related questions are answered by parents, which may consequently lead to differences between children in this regard.

In addition to examining the latent variables underlying richness across the HL vs. the SL, we also explored the extent to which these were comparable across the three countries where data were collected. For the most part, there were no significant differences between countries. The only exception was for the first latent variable in the joint information contributed by the HL, where the data for children in the Netherlands patterned differently from the data for children in France and the UK. Recall that this latent variable incorporated factors which depend on the presence of a sizeable HL-speaking community and resources in the HL (i.e., interlocutor diversity, reading and tech-related activities). The distinctiveness of the Netherlands in these areas is not surprising, as the most frequently occurring HLs of the children recruited in that country were English, German and Turkish. These are all

languages which are well resourced (especially English), spoken by communities which are well represented in the Netherlands (especially Turkish) and in countries which are close by (especially German), allowing better access to a more diverse and more proficient group of HL speakers and to additional HL resources.

Implications

The Q-BEx questionnaire is intended for use not only by researchers but also by practitioners working in educational and clinical settings. Our results indicate that it is acceptable for practitioners to use the original composite richness scores to inform their expectations as to the child's proficiency in each language. Higher richness scores predict better proficiency, and richness can be interpreted similarly across languages. Indeed, we have demonstrated that this was the case in spite of differences between different languages and language communities impacting on the existence of resources in each. It is an empirical question whether this will be replicated in studies when other language combinations/communities are studied by using Q-BEx. Our data suggest that the same richness measure can be used across age groups, even if there is not yet much formal literacy and for each language even if resources are limited or non-existent for some HLs.

Conclusion

The goal of the Q-BEx richness measure, and of the questionnaire more generally, is maximal comparability across languages, children, and countries, to gain a better understanding of the role of language experience in the language development of bilingual children across these varying contexts. We do not consider the use of a parental questionnaire such as Q-BEx to estimate richness as deficit-framing (see MacLeod & Demers, 2023 for relevant discussion), as our goal is to describe bilingual children's language environments as systematically and objectively as possible rather than labelling them or their caregivers' linguistic practices as (in)adequate. We acknowledge that the use of the terms "richness" and "input quality" can be problematic, but we note that the term "input quality" is no less loaded than "input quantity". Arriving at a consensus about which of the over-arching terms for the characteristics of bilingual children's language environments discussed here are more acceptable is a challenge for the field moving forward. These issues require further attention in future adaptations of tools documenting bilingual experience, such as the Q-BEx, among many others. Making these tools fully available is a requirement for enabling their scrutiny and future improvement.

Appendix

Table 1. Answer options to different types of questions in richness module

Example question	Answer options	Points assigned
Outside of regular school/day care, how often does the child participate in any organised activities in [Language]? For example: religious practice, sports, music, cultural activities, etc.	(almost) never once or twice a month once or twice a week several times a week every day	0 1 2 3 4
How many people speak in [Language] to the child at least once a week? Think about caregivers, brothers and/or sisters, other people in your home, family outside your home, friends/playmates, teachers, other important people in child's life.	0 1-2 3-5 6-10 more than 10	0 1 2 3 4
How many of those [Language] speakers speak [Language] very well?	none of them a few of them half of them most of them all of them	0 1 2 3 4
Select the highest education level that [caregiver] completed in [Language].	none primary school secondary school or equivalent post-secondary school training, but not a university degree university degree	0 1 2 3 4

References

Carroll, S. E. (2017). Exposure and input in bilingual development. *Bilingualism: Language and Cognition*, *20*(01), 3–16. https://doi.org/10.1017/S1366728915000863

Cheung, S., Kan, P. F., Winicour, E., & Yang, J. (2019). Effects of home language input on the vocabulary knowledge of sequential bilingual children. *Bilingualism*, *22*(5), 986–1004. https://doi.org/10.1017/S1366728918000810

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism*, *1*(3), 318–345. https://doi.org/10.1075/lab.1.3.05cho

Christensen, R. (2023). *ordinal—Regression Models for Ordinal Data. R package version 2023.12-4*.

Cribari-Neto, F., & Zeileis, A. (2010). Beta Regression in R. *Journal of Statistical Software*, *34*(2 SE-Articles), 1–24. https://doi.org/10.18637/jss.v034.i02

Cychosz, M., Villanueva, A. & Weisleder, A. (2021). Efficient estimation of children's language exposure in two bilingual communities. *Journal of Speech, Language, and Hearing Research, 64,* 3843–66.

Daskalaki, E., Blom, E., Chondrogianni, V., & Paradis, J. (2020). Effects of parental input quality in child heritage language acquisition. *Journal of Child Language*, *47*(4), 709–736. https://doi.org/10.1017/S0305000919000850

De Cat, C. (2021). Socioeconomic status as a proxy for input quality in bilingual children? *Applied Psycholinguistics*, *42*(2), 301–324. https://doi.org/10.1017/S014271642000079X

De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2022a). *Quantifying Bilingual EXperience (Q-BEx): questionnaire manual and documentation.* https://doi.org/10.17605/OSF.IO/V7EC8

De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2022b). *Quantifying Bilingual EXperience (Q-BEx): questionnaire manual and documentation*. https://doi.org/10.17605/OSF.IO/V7EC8

De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., & Unsworth, S. (2023). How to quantify bilingual experience? Findings from a Delphi consensus survey. *Bilingualism*, *26*(1), 112–124. https://doi.org/10.1017/S1366728922000359

De Houwer, A. (2022). *The danger of bilingual – monolingual comparisons in applied psycholinguistic*. 1–15. https://doi.org/10.1017/S014271642200042X

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *nternet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method (4th ed.)*. John Wiley & Sons, Ltd.

Dixon, L. Q., Zhao, J., Quiroz, B. G., & Shin, J.-Y. (2012). Home and community factors influencing bilingual children's ethnic language vocabulary development. *International Journal of Bilingualism*, *16*(4), 541–565. https://doi.org/10.1177/1367006911429527

Dunn, I. M., Dunn, D. M., Sewell, J., Styles, B., Brzyska, B., & Shamsan, Y. (2009). *The British Picture Vocabulary Scale, 3rd Edition.* GL Assessment.

Dunn, L. M., Dunn, L. M., & Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Hartcourt Test Publishers.

Dunn, L. M., Thoriault-Whalen, C. M., & Dunn, L. M. (1993). *Échelle de Vocabulaire en Images Peabody (EVIP)*. Psychan.

Gathercole, V. C. M., Kennedy, I., & Thomas, E. M. (2016). Socioeconomic level and bilinguals' performance on language and cognitive measures. *Bilingualism: Language and Cognition*, *19*, 1057–1078. https://doi.org/10.1017/S1366728915000504

Gollan, T. H., Starr, J., & Ferreira, V. S. (2015). More than use it or lose it: The number of speakers effect on heritage language proficiency. *Psychonomic Bulletin & Review*, *22*, 147–155.

González, I., Déjean, S., Martin, P. G. P., & Baccini, A. (2008). CCA: An R Package to Extend Canonical Correlation Analysis. *Journal of Statistical Software*, *23*(12 SE-Articles), 1–14. https://doi.org/10.18637/jss.v023.i12

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

Guo, K., Hur, J., & Feldman, E. (2022). *o2plsda: Multiomics Data Integration. R Package Version 0.0.18*.

Hoff, E., Burridge, A., Ribot, K. M., & Giguere, D. (2018). Language specificity in the relation of maternal education to bilingual children's vocabulary growth. *Developmental Psychology*, *54*(6), 1011–1019. https://doi.org/10.1037/dev0000492

Jia, G., & Aaronson, D. (2003). A longitudinal study of Chinese children and adolescents learning English in the United States. In *Applied Psycholinguistics* (Vol. 24, Issue 01). https://doi.org/10.1017/S0142716403000079

Jia, G., Chen, J., Kim, H., Chan, P. S., & Jeung, C. (2014). Bilingual lexical skills of school-age children with Chinese and Korean heritage languages in the United States. *International Journal of Behavioral Development*, *38*(4), 350–358. https://doi.org/10.1177/0165025414533224

Jia, G., & Fuse, A. (2007). Acquisition of english grammatical morphology by native mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing*

Research, *50*(5), 1280–1299. https://doi.org/10.1044/1092-4388(2007/090)

Jia, R., & Paradis, J. (2015). The use of referring expressions in narratives by Mandarin heritage language children and the role of language environment factors in predicting individual differences. *Bilingualism*, *18*(4), 737–752. https://doi.org/10.1017/S1366728914000728

Jolliffe, I.T. & Cadima, J. (2016). Principle Components Analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A, 374*: 20150202. https://doi.org/10.1098/rsta.2015.0202

Kalia, V., & Reese, V. (2009). Relations between Indian children's home literacy environment and their English oral language and literacy skills. *Scientific Studies of Reading*, *13*, 122–45.

Kaltsa, M., Prentza, A., & Tsimpli, I. M. (2020). Input and literacy effects in simultaneous and sequential bilinguals: The performance of Albanian–Greek-speaking children in sentence repetition. *International Journal of Bilingualism*, *24*(2), 159–183. https://doi.org/10.1177/1367006918819867

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*(1), 29–41. https://doi.org/10.1017/S1366728921000390

Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R Package Version 1.0.7.*

Kort, W., Schittekatte, M., & Compaan, E. (2008). *CELF-4-NL: Clinical evaluation of language fundamentals-vierde-editie.* Pearson Assessment and Information B.V.

Lauro, J., Core, C., & Hoff, E. (2020). Explaining Individual Differences in Trajectories of Simultaneous Bilingual Development: Contributions of Child and Environmental Factors. *Child Development*. https://doi.org/10.1111/cdev.13409

Leseman, P. P. M., & Boom, van den D. C. (1999). Effects of quantity and quality of home proximal processes on Dutch, Surinamese–Dutch and Turkish–Dutch pre-schoolers' cognitive development. *Infant and Child Development*, *8*, 19–38.

MacLeod, A. A. N., & Demers, C. (2023). Transmitting white monolingual Anglo-American norms: A concept analysis of quality of language in parent-child interactions. *Applied Psycholinguistics*, *44*(4), 431–459. https://doi.org/10.1017/S014271642300005X

Marchman, V.A., Martínez, L.Z., Hurtado, N., Grüter, T., & Fernald, A. (2017). Caregiver talk to young Spanish-English bilinguals: Comparing direct observation and parent-report measures of dual-language exposure. *Developmental Science*, *20*, e12425.

Marinis, T., & Armon-Lotem, S. (2015). *Sentence repetition. Assessing multilingual children: Disentangling bilingualism from language impairment,.* (pp. 95–124).

Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on children's vocabulary and verbal short-term memory. *Frontiers in Psychology, 8*, 1442. https://doi.org/10.3389/fpsyg.2017.01442

Orena, A.J., Byers-Heinlein, K., & Polka, L. (2020). What do bilingual infants actually hear? Evaluating measures of language input to bilingual-learning 10-month-olds. *Developmental Science*, *23,* e12901.

Paradis, J. (2011). Individual differences in child English second language acquisition. *Linguistic Approaches to Bilingualism*, *1*(3), 213–237. https://doi.org/10.1075/lab.1.3.01par

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793–817. https://doi.org/https://doi.org/10.1017/S0305000922000708`

Paradis, J., Soto-Corominas, A., Chen, X., & Gottardo, A. (2020). How language environment, age, and cognitive capacity support the bilingual development of Syrian refugee children recently arrived in Canada. *Applied Psycholinguistics*, *41*(6), 1255–1281. https://doi.org/10.1017/S014271642000017X

Paradis, J., Soto-Corominas, A., Daskalaki, E., Chen, X., & Gottardo, A. (2021). Morphosyntactic development in first generation arabic—english children: The effect of cognitive, age, and input factors over time and across languages. *Languages*, *6*(1). https://doi.org/10.3390/languages6010051

Paradis, J., Soto-Corominas, A., Vitroulis, I., Al Janaideh, R., Chen, X., Gottardo, A., Jenkins, J., & Georgiades, K. (2022). The role of socioemotional wellbeing difficulties and adversity in the L2 acquisition of first-generation refugee children. *Bilingualism: Language and Cognition*. https://doi.org/10.1017/S136672892200030X

Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics*, *28*(03), 399–410. https://doi.org/10.1017/S014271640707021X

Pham, G., & Tipton, T. (2018). Internal and External Factors That Support Children's Minority First Language and English. *Language, Speech, and Hearing Services in Schools*, *49*, 595–606.

Place, S., & Hoff, E. (2015). Effects and noneffects of input in bilingual environments on dual language skills in 2 ½-year-olds. *Bilingualism: Language and Cognition*, *19*(July), 1–19. https://doi.org/10.1017/S1366728915000322

Prevoo, M. J. L., Malda, M., Mesman, J., Emmen, R. A. G., Yeniad, N., Van Ijzendoorn, M. H., & Linting, M.

(2014). Predicting ethnic minority children's vocabulary from socioeconomic status, maternal

language and home reading input: different pathways for host and ethnic language. *Journal of Child Language*, *41*(05), 963–984. https://doi.org/10.1017/S0305000913000299

Prévost, P., Gusnanto, A., Kašćelan, D., Serratrice, L., Tuller, L., Unsworth, S. & De Cat, C. (in prep). Predicting accuracy on the LITMUS Sentence Repetition task.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech vocabulary development. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Rowe, M. L., & Snow, C. E. (2020). Analyzing input quality along three dimensions: Interactive, linguistic, and conceptual. *Journal of Child Language*, *47*(1), 5–21. https://doi.org/10.1017/S0305000919000655

Saravanan, V. (2001). The significance of bilingual Chinese, Malay and Tamil children's English network patterns on community language use patterns. *Early Child Development and Care*, *166*, 81–91.

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics*, *31*(01), 117. https://doi.org/10.1017/S0142716409990191

Semel, E., Wiig, E., & Secord, W. A. (2017). *Clinical Evaluation of Language Fundamentals, Fifth edition (CELF-5)*. Harcourt Assessment.

Serratrice, L., Gusnanto, A., Kašćelan, D., Prevost, P., Serratrice, L., Tuller, L., Unsworth, S. & De Cat, C. (in prep). Profiles of vocabulary breadth and depth in the societal language of bilingual children in three European countries.

Soto-Corominas, A., Paradis, J., Rusk, B. V., Marinova-Todd, S., & Zhang, X. (2020). ORAL LANGUAGE PROFILES of ENGLISH SECOND LANGUAGE LEARNERS in ADOLESCENCE. *Studies in Second Language Acquisition*, *42*(4), 697–720. https://doi.org/10.1017/S0272263119000767

Sun, H., & Yin, B. (2020). Multimedia Input and Bilingual Children's Language Learning. *Frontiers in Psychology*, *11*(August 2020), 1–11. https://doi.org/10.3389/fpsyg.2020.02023

Thordardottir, E. T. (2011). The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism*, *15*(4), 426–445.

Trygg, J. (2002). O2-PLS for qualitative and quantitative analysis in multivariate calibration. Journal of Chemometrics, 16(6),. *Journal of Chemometrics*, *16*(6), 283–293. https://doi.org/10.1002/cem.724

Tuller, L. (2015). Clinical Use of Parental Questionnaires in Multilingual Contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language*

impairment (pp. 299–328). Multilingual Matters.

Unsworth, S. (2013). Assessing the role of current and cumulative exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism*, *16*(1), 86–110. https://doi.org/10.1017/S1366728912000284

Unsworth, S., Brouwer, S., De Bree, E., & Verhagen, J. (2019). Predicting bilingual preschoolers' patterns of language development: Degree of non-native input matters. *Applied Psycholinguistics*, *40*(5), 1189–1219. https://doi.org/10.1017/S0142716419000225

Verhoeven, E., Witteloostuijn, van M., Oudgenoeg-Paz, O., & Blom, E. (2024). Comparing different methods that measure bilingual children's language environment: A closer look at audio recordings and questionnaires. *Languages, 9*, 231. https://doi.org/10.3390/ languages9070231

Wiig, E., Semel, E., & Wayne, A. (2019). *CELF 5—Batterie d'évaluation des fonctions langagières et de communication (Adaptation française ECPA).* Pearson.