ReadMe - De Cat et al. (2024)

2024-03-14

1. About the dataset

Title: Quantifying language experience in bilingual and trilingual children

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

Organisations: 1. University of Leeds, 2. University of Essex, 3. University of Tours, 4. University of Reading, 5. Radboud University

Publication Year: 2024

Description : The data in this repository were collected in France, the Netherlands and the UK between 2020 and 2022 to inform the validation of the Q-BEx questionnaire (www.q-bex.org). Children between the ages of 5 and 8 were tested individually to assess their proficiency in the societal language (i.e., French, Dutch or English), as well as their non-verbal intelligence and working memory. One of their parents or caregivers also filled in the full version of the Q-BEx questionnaire. The respository includes data from 299 children (FR: n=78, NL: n=117, UK: n=104), although some measures are not available for all children, as explained below.

Cite as: De Cat, C., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S. (2024): Quantifying language experience in bilingual and trilingual children. [Dataset]. https://doi.org/10.5518/1496

Related publication: De Cat, C., Gusnanto, A., Kašćelan, D., Prévost, P., Serratrice, L., Tuller, L., Unsworth, S. (in preparation) How much questionnaire detail is required to document language experience in bilingual children? (Preprint will be available in June 2024)

2. Terms of use

3. Project and funding information

Title: Quantifying bilingual experience (Q-BEx): optimising tools for educators, clinicians and researchers

Dates: September 2019 - October 2023

Funding organisation: Economic and Social Research Council (UKRI)

Grant no.: ES/S010998/1

4. Contents

File listing

- QBEx_data.csv
- QBEx_data_dictionary.csv
- Backend_calculator_example.xlsx
- QBEx_wrangling_OS.R
- corrections_countries.csv

All the data can be found in QBEx_data.csv (including language experience data collected via the Q-BEx questionnaire, as well as direct measures of language proficiency and cognition - as detailed in the Methods section). The list of variables can be found in the data dictionary file (QBEx_data_dictionary.csv). Some of the variables are derived by the Q-BEx platform, as explained in the back-end calculations spreadsheet (Backend_calculator_example.xlsx). The additional data wrangling procedures are explained in the data dictionary file. The R script used to assign language status (Societal Language, Heritage Language 1, Heritage Language 2) is QBEx_wrangling_OS.R, and the accompanying spreadsheet allowing the manual fix of respondent misspelling of language names is corrections_countries.csv. Additional R scripts for data wrangling are available on the OSF repository for the Q-BEx project (https://osf.io/gb9m7/).

We consistently use NA as missing data code.

5. Methods

Recruitment

In France, children were recruited in ordinary schools and in private clinics for Speech & Language Therapy (37 children were recruited via clinics). The consent form for schools asked parents if the child had previous or current SLT, and the reason. In the Netherlands, recruitment took place in schools and via social media advertisement. Language disorder (reported by parent/teacher/remedial teacher) was an exclusionary criterion. In the UK,

all children were recruited in schools; no exclusionary criteria were applied, and not SLT information was collected.

Language experience data

Language experience data was collected using the full version of the Q-BEx questionnaire (www.q-bex.org), which was completed by one of the child's parents or caregivers. This includes all the following modules (except that Language mixing wasn't included in France):

- Background (languages the child is exposed to, adults and children the child lives with)
- Risk factors (early language milestones, early parental concerns)
- Language exposure and use (current and cumulative estimates; onset of exposure to each language)
- Estimates of proficiency in each language (listening, speaking, reading, writing)
- Richness of experience in each language (activities, diversity of interlocutors, parental education)
- Language mixing
- Attitudes

The questionnaire was administered either in the societal language (French, Dutch or English) or in one of the child's home languages (Arabic, Dutch, English, French, German, Italian, Polish, Romanian, Russian, Spanish, Turkish). The choice of administration language was constrained by the translated versions available at the time of testing. The translation protocol used to create the versions in different languages can be found at https://www.q-bex.org/wp-content/uploads/2022/10/V7-Translation-Protocol-10_August_2021.pdf.

The list of variables obtained from questionnaire data can be found in the data dictionary file (QBEx_data_dictionary.csv). Some of the variables are derived by the Q-BEx platform, as explained in the back-end calculations spreadsheet (Backend_calculator_example.xlsx). The additional data wrangling procedures are explained in the data dictionary file. The R script used to assign language status (Societal Language, Heritage Language 1, Heritage Language 2) is QBEx_wrangling_OS.R, and the accompanying spreadsheet allowing the manual fix of respondent misspelling of language names is corrections_countries.csv. Additional R scripts for data wrangling are available on the OSF repository for the Q-BEx project (https://osf.io/gb9m7/).

Manual data correction: Assuming that all children were exposed to at least one language from birth, if the age of first exposure was above 0 months for all the languages the child is subsequently reported to have experienced, we replaced any value of 12 months or less to 0 as long as the place of exposure is home. Our reasoning is that the caregiver most likely assumed that language exposure before one year of age did not count. Data from 40 children were manually corrected in this way.

Direct outcome measures

We collected measures of language and cognitive outcomes during individual, face-to-face sessions with each child (two sessions per child, lasting approximatively 45 minutes each). Most of the testing was done in the child's school. In France, the children recruited via speech & language therapy clinics were tested in the clinic. In the UK and the NL, some testing sessions took place in a different location (e.g., university premises), and on rare occasions online via Zoom.

Language proficiency

Outcomes in the societal language (i.e., Dutch, English, or French) include phonology, morphosyntax and vocabulary.

Phonology Phonological competence was assessed with the LITMUS Quasi-Universal Non-Word Repetition task. See dos Santos, C., and Ferré, S. (2016) "A Nonword Repetition Task to Assess Bilingual Children's Phonology". *Language Acquisition* 41: 1–14.

Morphosyntax Morphosyntax outcomes were assessed with the LITMUS test in each societal language. See Marinis, T. and S. Armon-Lotem (2015). "Sentence Repetition. Methods for assessing multilingual children: disentangling bilingualism from Language Impairment." in S. Armon-Lotem, J. de Jong and N. Meir, *Methods for assessing multilingual children: disentangling bilingualism from Language Impairment.* Amsterdam: Multilingual Matters).

The English and Dutch versions included 30 items. The French version included 16 items. We created two blocks in the English and Dutch versions so that the first block be comparable to the 16-item French version. Subsequent analyses demonstrated that there was no block effect in EN and NL.

For each child, we report three overall scores: Identical Repetitions, Target Repetitions, and Grammatical Attempts. These overall scores correspond to the mean across items (n=30 in English and Dutch; n=16 in French), excluding NAs (i.e., the mean is calculated on the items for which the child did provide a response).

Vocabulary Vocabulary breadth was measured in the UK with the British Picture Vocabulary Scale (BPVS), in France with the the Échelle de vocabulaire en images Peabody (EVIP), and in the Netherlands with the Peabody Picture Vocabulary Test (PPVT).

Vocabulary depth was measured with the Word Classes component of the Clinical Evaluation of Language Fundamentals CELF-V (in its Dutch, English, or French version).

Cognitive measures

The tasks used to evaluate cognitive skills were administered in the child's societal language.

Memory Short-term memory was assessed through Forward Digit Recall; working memory was assessed through Backward Digit Recall.

Most children were tested using the digit span protocols described in Hill et al (2021), implemented in Psychopy (to allow randomisation of the digit sequences and facilitate the acquisition of detailed data). Children were presented with sequences of numbers (through headphones) and asked to repeat these numbers either in the same order (in the FDR task) or in reverse order (in the BDR task). The length of the sequence increased by one digit after 4 trials, starting with 3 digits in the first block of the FDR task, and 2 digits in the first block of the BDR task. The maximum sequence length was 6 digits in the FDR task, and 5 digits in the BDR task.

Children recruited via the Speech and Language Therapy clinics in France (n=37) experienced more difficulty with this task, so it was decided to use the WISC-V protocol instead for these children, as it included a discontinuation rule.

To allow comparison across groups, we created a WISC-like score for the data collected via the Hill et al (2021) protocol. This consisted in the digit span for which the child had at least one fully accurate response (i.e., all the digits in the span, in the right order) for at least one out of the first two trials for that span (as the WISC-V protocol only features 2 trials per digit span).

- **FDR_overall** and **BDR_overall** correspond to the total accuracy scores, as per the Hill et al (2021) protocol.

- **FDS_Q** and **BDS_Q** correspond to either the WISC-V score (for children recruited via clinics) or the WISC-like score created as explained above, depending on which protocol the child was tested with. All children have a FDR_Q and a BDR_Q score in the dataset.

Non-verbal intelligence The matrices task from either the Wechsler Intelligence Scale for Children–Fifth Edition (WISC–V) or the Wechsler Preschool & Primary Scale of Intelligence - Fourth UK Edition (WPPSI-IV) was used to measure non-verbal intelligence - depending on the age of the child: children below the age of 6 were tested with the WPPSI.

Depending on the age of the child at the time of testing, we used the WPPSI protocol (for children younger than 6 years of age, n= 77) or the WISC-V protocol (for all the other children).

Available and missing data

We have reliable Q-BEx data from 299 children. This excludes 52 children, whose questionnaire data were unreliable for one or more of the following reasons: they were above 8 years of age; the respondent completed the wrong version questionnaire (e.g., monolingual version for a bilingual child); the respondent did not include the societal language in the languages experienced by the child; the box "does not speak" was erroneously ticked for one of the

caregivers, which prevented the questionnaire from including the relevant questions about that caregiver; parts of the questionnaire were not completed.

Table 1: Number of children by multilingualism status and country of residence

	Monolingual	Bilingual	Trilingual
France	36	38	4
Netherlands	41	57	19
United Kingdom	49	40	15

Table 2: Number of children according to their clinical status

	Monolingual	Bilingual	Trilingual
Current.diagnostic	9	10	2
Some.SLT	7	9	0
TD	109	116	36

There are 217 children with a full dataset (including Q-BEx and all outcome measures). Sentence Repetition scores are missing for 68 children; vocabulary scores are missing for 72 children; Non-Word Repetition scores are missing for 68 children; Non-Verbal Intelligence scores aremissing for 70 children; Memory scores are missing for 77 children. This was mostly due to the children not showing up for the planned testing sessions.

At the time of submitting the data to this repository, the publications reporting our analyses of these data were still in preparation. The preprints will be made available as soon as possible on the OSF at https://osf.io/gb9m7/.