

MB3L Longitudinal Language Follow-up

Lab Manual

Thank you for contributing to *ManyBabies*, a large-scale, collaborative network of infancy researchers seeking to provide an empirical basis for discussions of replicability as well as cultural, developmental, and methodological variability in developmental sciences. This project is a follow-up project to the **ManyBabies3 project** examining infants' ability to learn and generalize algebraic rules. This is following work from Marcus et al (1999), where infants were familiarized with 3-syllable sequences and then tested on new sequences that either follow, or don't follow the familiarized pattern. More details about the background, design, and hypotheses of the MB3 study can be found in the [Registered Report](#).

The **MB3L follow-up project** will investigate the relationship between infants' rule learning abilities and their later developing language acquisition skills in a longitudinal design. Therefore, participating labs will assess language abilities of **participants who have already previously been recruited for the ManyBabies3 project** at 24 to 30 months of age using the CDI. We are for example interested in the extent to which rule learning abilities predict later grammar skills, if this relationship changes with age, and if certain characteristics of the first language the infants are exposed to (e.g., complexity) lead to any systematic differences in the strength of this relationship. More details about the background, design, and hypotheses of the MB3L study can be found in the [Pre-Registration](#).

Below we provide instructions on how to implement the experiment in your lab and how to report data back to the project. For any details not explicitly referenced here (e.g. authorship guidelines, code of conduct), please refer to the main [MB3 manual](#) and/or the primary [ManyBabies manual](#).

Roadmap

1. Collect data for MB3 following all the in the [Lab Manual](#).
2. Indicate your interest in joining the longitudinal language follow-up project by emailing Anna Exner (anna.exner@posteo.de)
3. Read this manual from start to finish.
4. Choose an appropriate CDI version..
5. Create demographic questionnaire (see below)
6. Complete the MB3L [data collection survey](#).
7. Collect your data!
8. Complete your participant and trial data files for the retest session in consultation with this manual.
9. Run the sample through the [MB3L data validator](#) (Note that the handling is similar to the MB3 validator).
10. Submit your data.

1. Choose an appropriate CDI version

Infants' language abilities will be assessed through the parent-administered questionnaire "MacArthur-Bates Communicative Development Inventories (CDI): Word and Sentences form" (Fenson et al., 2006), one of the most common assessment tools widely adapted in several languages for studying language development in toddlers aged 16–30 months.

Each lab will pre-select the CDI version or versions (preferably Long Format) that best matches their language community and determine their own data collection procedure. Labs may choose to use a

self-created online version, a paper-pencil version, or the Web-CDI online form (details below), depending on which method aligns with their standard practices. For some labs, it might be appropriate to prepare two or more CDI versions given a diverse language background of a certain sample.

If you find an appropriate language community in the list below, add your lab ID to the last column ("Labs") of this table. If your lab is the first to collect data in a certain language community, please choose a CDI version and add the details to the following list. If you need assistance with that choice, please feel free to contact the lead team.

If you are using a CDI that is not published or is under development, please specify it in the table and put a copy of your CDI form and any associated instructions documents, etc., [in this folder](#).

WebCDI: Access and Collaboration Guide

WebCDI is an online platform that hosts a range of CDI (Communicative Development Inventories) questionnaires available in multiple languages (link: <https://webcdi.org/interface/>). The currently supported languages and formats include:

- Dutch - Long format
- English (American/Canadian) - Long format
- English (American) - Short format
- French (Canadian) - Long format
- French (France) - CAT
- Hebrew - Long format
- Japanese - Short format
- Korean - Long format
- Spanish (Argentinian) - Long format
- Spanish (Mexican) - Long format

Collaborations and New Languages

Labs interested in contributing by adding new languages to the WebCDI platform are highly encouraged to collaborate. Please contact both Michael C. Frank (mcfrank@stanford.edu) and the project organizers at Webcdi-contact@stanford.edu for more information on how to participate.

Subscription and Access Information

WebCDI is a **free resource for researchers**. However, for certain CDI instruments licensed by Brookes Publishing (specifically American English and Mexican Spanish WG, WS, and CAT), there is a subscription fee. If your lab works in one of these languages (American English or Mexican Spanish) and cannot afford the subscription fee, you can still request free access. Please reach out to Michael C. Frank (mcfrank@stanford.edu) and the project organizers (Webcdi-contact@stanford.edu) for assistance in obtaining a free subscription.

Instructions for Completing the CDI for Bilingual/Multilingual Children

The CDI should be administered in the language best reflecting the language community of participants of your lab (if it is common practice in your lab to prepare instructions in different languages because of multiple official languages in your country, you can choose to prepare two (or more) CDI versions and use the version for each child best reflecting their primary language). However, if the child produces words in any other language other than the primary one, please instruct the parents to **mark those words as well, even if they are not part of the primary language**.

| Country | Language | CDI version | Reference | Link | Labs |
|-------------|------------------|---|--|--|--|
| Germany | German | Fragebogen zur frühkindlichen Sprachentwicklung, FRAKIS | Szagun, G., Stumper, B. & Schramm, S.A. (2023). <i>Fragebogen zur frühkindlichen Sprachentwicklung: FRAKIS (Langform) und FRAKIS-K (Kurzform)</i> , 2. Auflage. Universitätsverlag Potsdam. doi.org/10.25932/publishup-58675 | https://www.uni-potsdam.de/de/inklusion/ese/schramm/frakis | epsyRUB |
| UK | English | UK-MB-CDI Words and Sentences Standardisation | Alcock, K. J., Meints, K., Rowland, C. F., Brelsford, V., Christopher, A., & Just, J. (2019). The UK communicative development inventory: Words and gestures. Guildford: J&R Press Ltd. | https://mb-cdi.stanford.edu/adaptations.html https://mb-cdi.stanford.edu/documents/Info2023-English(British).pdf | gertlabLancaster ddlabUEA BrookesBabyLab |
| Switzerland | German | Fragebogen zur frühkindlichen Sprachentwicklung, FRAKIS | Szagun, G., Stumper, B. & Schramm, S.A. (2023). <i>Fragebogen zur frühkindlichen Sprachentwicklung: FRAKIS (Langform) und FRAKIS-K (Kurzform)</i> , 2. Auflage. Universitätsverlag Potsdam. doi.org/10.25932/publishup-58675 | https://www.uni-potsdam.de/de/inklusion/ese/schramm/frakis | weltentdecker Zurich |
| USA | American English | MB-CDI Words and Sentences | Fenson, L., Marchman, V., & Thal, D. (1993). | https://mb-cdi.stanford.edu/eng_long.html | nyullearn, babylabPrinceton |

| | | | | | |
|--|--|--|--|--|--|
| | | | MacArthur-bat es communicative development inventories: words and sentences form MCDI. <i>Baltimore (MD): Brookes.</i> | | |
|--|--|--|--|--|--|

2. Sample size

Contributing labs will be asked to re-recruit their participants with usable data for MB3. Notice that the lab should recruit infants from the existing MB3 sample and will not need to recruit new participants. Even though we aim for a contribution of as many participants as possible from that sample, we are aware that labs are likely to lose infants to attrition due to failure to agree to participate in the follow-up. Thus, we expect sample sizes of around 10 usable infants per lab from that original sample. As always, it is important to ensure that any decisions about sample size/stopping data collection are made independently of the data you collect.

Participating labs will assess language abilities using the CDI when participants are between 24 and 30 months of age. Data collection should begin only once participants are within this age range. We encourage the labs to test infants across the full 24-30 months age range.

3 Create demographic questionnaire

We would like to re-collect some demographic data already collected for MB3. So you can use specific questions from the Family Background Questionnaire that you used for MB3. All questions highlighted yellow in this template should be asked again. Please adapt your questionnaire accordingly: [Family Background Questionnaire](#)

4 Data submission

4.1 Participant Data

This will be a CSV/TSV file with **one** row for each participant [[Excel template](#)]. Note that it is important that the participant IDs are identical between MB3 and MB3L!

4.2 Data dictionary

All variables are identical with the ones used in MB3, except for those in the table below.

Discards and Exclusions: Decisions regarding data discards for experimental reasons will be made centrally and must be reported consistently. Therefore, it is crucial to retain all data from all children who participated in the follow-up, including those whose families accepted participation but did not complete the questionnaire.

Language: When collecting data from multilingual children, please provide an approximate percentage of exposure to each language. The total percentage should add up to 100%.

Participant data
due

| Variable name | Format | Value example(s) | Description |
|----------------------|-----------|---|---|
| lab_id | [string] | 'bllumanitoba' | Identifier chosen for each lab in the main ManyBabies study. Please ensure that the IDs match exactly between both studies for pairing purposes. |
| participant_id | [string] | 'MB3_001' 'MB3_002' | The identifier used for each participant in the main ManyBabies study. Please ensure that the IDs match exactly between both studies for pairing purposes. |
| cdi_version | [string] | 'Fragebogen zur frühkindlichen Sprachentwicklung , FRAKIS' | The version of the CDI your lab chose. It should match the one in the table under paragraph 1. Choose an appropriate CDI version |
| method | [options] | 'online' 'paper pencil' 'web-CDI' | Style of administering the questionnaire in your lab. Choose which one suits your lab practices best. |
| participant_age_days | [integer] | | Participants' age in <i>days</i> when the cdi was filled out. Please use this calculator to compute age in days: https://www.calculator.net/age-calculator.html . Copy and paste errors are common here, so please double check prior to data upload |
| session_exclusion | [options] | 'exclude' 'include' | Sessions marked with 'exclude' will be removed from the MB3L analysis. Use this field if all CDI data from a participant should be excluded (e.g., due to incompleteness). The central analysis team will review all sessions marked 'exclude' to ensure |

| | | | |
|----------------------------|-----------|---|--|
| | | | consistent participant-level exclusions across labs. |
| session_exclusion_reason_1 | [options] | 'outside_age_range', 'dev_disorder', 'preterm', 'incomplete' | NA unless you are reporting that ALL data from this participant should be excluded from analyses. Otherwise enter the first or primary exclusion criterion. |
| session_exclusion_reason_2 | [options] | 'outside_age_range', 'dev_disorder', 'preterm', 'incomplete' | NA unless you are reporting that ALL data from this participant should be excluded from analyses. Otherwise enter the first or primary exclusion criterion. |
| session_exclusion_reason_3 | [options] | 'outside_age_range', 'dev_disorder', 'preterm', 'incomplete' | NA unless you are reporting that ALL data from this participant should be excluded from analyses. Otherwise, enter the first or primary exclusion criterion. |
| session_exclusion_reason_4 | [options] | 'outside_age_range', 'dev_disorder', 'preterm', 'outside_interference', 'experimenter_error', 'equipment_error', 'fuss_out' | NA unless you are reporting that ALL data from this participant should be excluded from analyses. For session exclusions, enter NA or a fourth exclusion criterion. |
| session_exclusion_info | [string] | | Explanation of what caused the session error (e.g. “music audible from another room throughout session” or “child fussed out after entering test room but prior to completing any trials”). Use this field to report a reason that ALL data from this participant should be excluded from analysis. Write NA if no session level error. |
| notes | [string] | The family completed the questionnaire in two sessions, with a two-month interval between them. | NA. Otherwise, provide information about data collection that you believe may affect the quality of the collected data. |
| first_ra_id | [string] | | A research assistant identifier, non-identifying and unique. If there is more than one RA present, please choose a standard role in your lab sessions (e.g. the person who operates the camera) and report that person's ID. Unique RA identifier (anonymous code). We are collecting information about which RA ran a given participant for possible secondary analysis about variability in RA coding. As with all our analyses, the goal is not to assess |

| | | | |
|-------------------------------|-----------------------------|--------------------------------------|--|
| | | | individual performance, but to examine sources of variability. However, you may wish to provide this information as an anonymous code. Like with the participants, the coding scheme is up to you. |
| second_ra_id | [string] | | A second research assistant identifier, non-identifying and unique. Enter NA if only one RA/experimenter/staff member was involved in testing. |
| respondent_relationship | [string] | father mother grandmother | Relationship of the person filling out the CDI to the participant |
| participant_sex | [options] | female, male, another_response | List the infant participant's sex as reported by the family member responding to the form. |
| participant_place_of_birth | [string] | Dayton, OH, USA | List the infant's place of birth (city, state, country) |
| participant_current_residence | [string] | Dayton, OH, USA | The participant's current residence (might be different from the one in the main MB3 study) |
| lang1_name | [string] | 'dutch' | Native or most-heard language |
| lang1_exposure | [integer between 0 and 100] | '70' | Percentage exposure to cdi_lang1_name. Note that exposure to all languages for a child should sum to 100 |
| lang2_name | [string] | 'english' 'NA' | Second most-heard language. Put NA if they do not have a second language |
| lang2_exposure | [integer between 0 and 100] | '20' 'NA' | Percentage exposure to lang2. Exposures to all languages for a child should sum to 100. Put NA if they do not have a second language. |
| lang3plus_name | [string] | 'spanish' 'NA' | Third most-heard language and any additional languages. Put NA if they do not have a third language. |
| lang3plus_exposure | [integer between 0 and 100] | '10' 'NA' | Percentage exposure to languages 3, 4, etc. Put NA if they have 2 or fewer languages. |
| num_adults_household | [integer] | | The number of adults in the household. |
| num_children_household | [integer] | | The number of children in the household. |
| race_construct_response | [string] | | The child's racial identit(ies) as reported by the family respondent. This item varies significantly by lab. |

| | | | |
|---|--|---------------------------------------|---|
| ethnicity_construct_response | [string] | | The child's ethnic identit(ies) as reported by the family respondent. This item varies significantly by lab. |
| preterm | [options] | 'preterm' 'not_preterm' | Whether the participant was born prematurely. Full term is defined as 37 weeks or more of gestation. |
| gestation_weeks | [integer above 0] | | The gestational age at birth of a preterm participant. Note that this should be converted to gestational weeks, even if you ask parents about 'days or weeks early' - different countries define a full term pregnancy differently! This variable can be listed as NA if the participant was born at full term. |
| cdi_medical_issue | [options] | 'med_issue' 'no_med_issue' | Whether the caregiver reports the child having any medical issues or diagnosed conditions. If so, this should be described in the med_issue_info variable. |
| cdi_medical_issue_info | [string] | | Enter the respondent's description if the previous column is 'med_issue'. Write NA if the previous column is 'no_med_issue'. |
| developmental_concern | [options] | 'dev_concern' 'no_dev_concern' | Whether the caregiver reports any concerns about the participant's development. If so, this should be described in the developmental_concern_info variable. |
| developmental_concern_info | [string] | | Enter the respondent's description if the previous column is 'dev_concern'. Write NA if the previous column is 'no_dev_concern'. |
| cdi_hearing_difficulties | [options] | 'hearing-issue' 'no_hearing_issue' | Whether the caregiver reports the child having hearing difficulties. If so, this should be described in the cdi_hearing_difficulties variable. |
| cdi_hearing_difficulties_info | [string] | | Enter the respondent's description if the previous column is 'hearing_issue'. Write NA if the previous column is 'no_med_issue'. |
| caregiver1_education_construct_response | [string] | | Highest level of education for the first parent/caregiver - this item is expected to vary significantly across labs. |
| caregiver2_education_construct_response | [string] | | Highest level of education for the second parent/caregiver - this item is expected to vary significantly across labs. |
| household_income_construct_response | [string] or [positive number] | | Combined household gross income (before taxes) - this item is expected to vary significantly across labs. |

| | | | |
|-----------------|-------------------------------|--|---|
| cdi_ev_raw | [positive number] | | Expressive vocabulary: the number of words reported by the parent |
| cdi_ev_percent | [string] or [positive number] | | Percentile conversion of the expressive vocabulary (ev) for the CDI in your language (if available). N/A if there is no existing percentile conversion for the CDI in your language. |
| cdi_mlu_raw | [positive number] | | Mean length of utterance: The average number of morphemes (or words) per utterance (refer to the CDI manual for calculation instructions). |
| cdi_mlu_percent | [string] or [positive number] | | Percentile conversion of Mean length of utterance (mlu) for the CDI in your language (if available) N/A if there is no existing percentile conversion for the CDI in your language. |
| cdi_gc_raw | [positive number] | | Sentence/Grammar complexity: The number and types of complex sentences used (refer to the CDI manual for calculation instructions) |
| cdi_gc_percent | [string] or [positive number] | | Percentile conversion of Sentence/Grammar complexity (gc) for the CDI in your language (if available) N/A if there is no existing percentile conversion for the CDI in your language. |

4.3 Data upload

Please upload the file using the [data upload form](#).

4.4 Labeling

These two files should be clearly labeled in the filename with the lab ID you identified for your lab in the laboratory questionnaire and the **_language** appendix (e.g.

lab-bllumanitoba_level-participant_data_**language**.csv and

lab-bllumanitoba_level-trial_data_**language**.csv).

5. Authorship

In general, the authorship expectations for MB3L mirror that of the main study (e.g., ideally 2 authors max per lab for data contribution, exceptions to be discussed with the leads, contributions other than data may result in authorship).

It is possible for labs to allocate authorship differently for MB3 and MB3L.

Where not otherwise explicitly stated, the project conforms to the authorship guidelines as outlined in the main ManyBabies manual.