Web-CDI: A system for online administration of the MacArthur-Bates Communicative Development Inventories

Benjamin deMayo[1], Danielle Kellier[2], Mika Braginsky[3], Christina Bergmann[4], Cielke Hendriks[4], Caroline Rowland[4,6], Michael C. Frank[5], & Virginia A. Marchman[5]

[1] Princeton University

[2] University of Pennsylvania

[3] Massachusetts Institute of Technology

[4] Max Planck Institute for Psycholinguistics

[5] Stanford University

[6] Radboud University

Abstract

Understanding the mechanisms that drive variation in children's language acquisition requires large, population-representative datasets of children's word learning across development. Parent report measures such as the MacArthur-Bates Communicative Development Inventories (CDI) are commonly used to collect such data, but the traditional paper-based forms make the curation of large datasets logistically challenging. Many CDI datasets are thus gathered using convenience samples, often recruited from communities in proximity to major research institutions. Here, we introduce Web-CDI, a web-based tool which allows researchers to collect CDI data online. Web-CDI contains functionality to collect and manage longitudinal data, share links to test administrations, and download vocabulary scores. To date, over 3,500 valid Web-CDI administrations have been completed. General trends found in past norming studies of the CDI (e.g., Feldman et al., 2000) are present in data collected from Web-CDI: scores of children's productive vocabulary grow with age, female children show a slightly faster rate of vocabulary growth, and participants with higher levels of educational attainment report slightly higher vocabulary production scores than those with lower levels of education attainment. We also report results from an effort to oversample non-white, lower-education participants via online recruitment (N = 243). These data showed similar age, gender and primary caregiver education trends to the full Web-CDI sample, but this effort resulted in a high exclusion rate. We conclude by discussing implications and challenges for the collection of large, population-representative datasets.

Keywords: vocabulary development, parent report

Word count: X

Web-CDI: A system for online administration of the MacArthur-Bates Communicative Development Inventories

Children vary tremendously in their vocabulary development (Fenson et al., 1994; Frank, Braginsky, Yurovsky, & Marchman, 2021). Characterizing this variability is central to understanding the mechanisms that drive early language acquisition, yet capturing this variation in broad, diverse samples of children has been a significant challenge for cognitive scientists for decades. The MacArthur-Bates Communicative Development Inventories (MB-CDI, or CDI for short) are a set of commonly-used parent report instruments for assessing vocabulary development in early childhood (Fenson et al., 2007) that were introduced in part to create a cost-effective method for measuring variability across individuals.

In this paper, we introduce a web-based tool, Web-CDI, which was developed to address the need for collecting CDI data in an online format. Web-CDI allows researchers to increase the convenience of CDI administration, further decrease costs associated with data collection and entry (particularly with item-level data), and access participant samples that have traditionally been difficult to reach in language development research. Our purpose in this paper is twofold: first, we describe Web-CDI as a platform which streamlines the process of collecting CDI data and collates the data in a way that facilitates the creation of large-scale, multisite collaborative datasets. Second, we profile usage of Web-CDI thus far, with a particular focus on broadening the reach of traditional paper-based methods of collecting vocabulary development data.

The Importance of Parent Report Data

Gaining empirical traction on variation in children's early language requires reliable and valid methods for measuring language abilities, especially in early childhood (8 to 30 months). Parent report is a mainstay in this domain. Parents' reports are based on their daily experiences with the child, which are much more extensive than a researcher or clinician can generally obtain. Moreover, they are less likely to be influenced by factors that may mask a child's true ability in the laboratory or clinic (e.g., shyness). One widely used set of parent-report instruments is the MacArthur-Bates Communicative Development Inventories, originally designed for children learning American English (Fenson et al., 2007). The American English CDIs come in several versions, two of which are Words & Gestures (WG) for children 8 to 18 months, focusing on word comprehension and production, as well as gesture use, and Words & Sentences (WS) for children 16 to 30 months, focusing on word production and sentence structure. Both the WG and WS measures come in short forms with vocabulary checklists of approximately 90-100 words (Fenson et al., 2000), and long forms, which contain vocabulary checklists of several hundred items each. (An additional shorter form of the Web-CDI for children 30-37 months, CDI-III, also exists.) Together, the CDI instruments allow for a comprehensive picture of milestones that characterize language development in early childhood. A substantial body of evidence suggests that these instruments are both reliable and valid (e.g., Fenson et al., 2007, 1994) leading to their widespread use in thousands of research studies over the last few decades. Initial large-scale work to establish the normative datasets for the American English CDI not only provided key benchmarks for determining children's progress, but also documented the extensive individual differences that characterize early language learning during this critical period of development (Bates et al., 1994; Fenson et al., 1994). Understanding the origins and consequences of this variability remains an important empirical and theoretical endeavor (e.g., Bates & Goodman, 2001; Bornstein & Putnick, 2012; see also, Frank, Braginsky, Yurovsky, & Marchman, 2021).

The popularity of CDI instruments has remained strong over the years, leading to extensions of the methodology to alternative formats and cross-language adaptations (Fenson et al., 2000). Many teams around the world have adapted the CDI format to the particular languages and communities (Dale, 2015). Importantly, these adaptations are not simply translations of the original form but rather incorporate the specific features of different languages and cultures, since linguistic variability exists even among cultures that share a native language. As an example of this phenomenon, the word "Cheerios" is more common in the United States than it is in the United Kingdom; as a result, it might be expected that caregivers would report children's knowledge of this word in the U.S. and not the U.K., even though English is the most common language in both countries. To date there are more than 100 adaptations for languages around the globe. Moreover, several research groups have developed shorter versions of the CDI forms by randomly sampling items from the full CDI and comparing participants' responses to established norms (Mayor & Mani, 2019) or by developing computer adaptive tests (CATs) that use item response theory or Bayesian approaches to guide the selection of a smaller subset of items to which participants respond (Chai, Lo, & Mayor, 2020; Kachergis et al., 2021; Makransky, Dale, Havmose, & Bleses, 2016).

While the reliability and validity of the original CDI instruments is well-established for the American English versions of the forms and several others, most existing norming samples are skewed toward families with more years of formal education and away from non-white groups (Fenson et al., 2007). For example, representation in the American English norming samples is generally restricted to families living on the U.S. east and west coasts. Further, although paper survey administration is a time-tested method, increasingly researchers and participants would prefer to use an electronic method to administer and fill CDI forms, obviating the need to track (and sometimes mail) paper forms, and the need to key in hundreds of item-wise responses for each child.

Here, we report on our recent efforts to create and distribute a web-based version of the CDIs in order to address some of the limitations of the standard paper versions. Online administration of the CDI is not a novel innovation – a variety of research groups have created purpose-build platforms for administering the CDI in particular languages. For example, Kristoffersen et al. (2013) collected a large normative sample of Norwegian CDIs using a custom online platform. Similarly, the Slovak adaptation of the CDI uses an online administration format (Kapalková & Slanèová, 2007). And many groups have used general purpose survey software such as Qualtrics and Survey Monkey to administer CDIs and variants online (e.g., Caselli, Lieberman, & Pyers, 2020). The innovation of Web-CDI is to provide a comprehensive researcher management interface for the administration of a wide range of CDI forms, allowing researchers to manage longitudinal administrations, download scores, and share links with parents easily, all while satisfying strong guarantees regarding privacy and anonymity. Moreover, a key benefit of a unified data collection and storage system such as Web-CDI is that data from disparate sources are combined into a single repository. This substantially reduces the overhead efforts associated with bringing together data collected by researchers across the world and allows for the analysis of large comparative datasets with the power to detect general trends in vocabulary development that may emerge across languages. Finally, due to an agreement between the CDI Advisory Board and Brookes Publishing, the publisher of the print versions of the CDI suite, Web-CDI is free of charge for those researchers who agree to contribute their data for the renorming of the long form instruments.

Introducing Web-CDI

Web-CDI is a web-based platform for CDI administration and management. Web-CDI allows researchers to communicate with families by sharing URLs (web links that contain individual users' own administration of the Web-CDI) via email or social media, facilitating access to families in areas distant from an academic institution and eliminating costly mailings and laboratory visits. Web-CDI also standardizes electronic administration and scoring of CDI forms across labs and institutions, making possible the aggregation of CDI data for later reuse and comparison across administrations by different labs. Indeed, researchers who use Web-CDI grant the CDI Advisory Board permission to access and analyze the resulting data on an opt-out basis, providing a path towards continual improvement of CDI instruments. Since 2018, more than 3,500 CDIs have been collected by 15 research groups throughout the U.S. who are using Web-CDI, demonstrating the potential for large-scale data collection and aggregation.

Below, we outline how Web-CDI is used. We begin by detailing the consent obtention process and participant experience. Second, we describe the interface that researchers use to collect data using Web-CDI, specifying a number of common use cases for the platform.

Participant interface

Participants can complete the Web-CDI on a variety of devices, including personal computers and tablets. Web-CDI can be also administered on a smartphone, although the experience is not as ideal for the user due to the length of the survey and the small screen. As Web-CDI moves in the future to incorporate more short forms and computer adaptive tests (CATs) formats (e.g., Chai, Lo, & Mayor, 2020; Makransky, Dale, Havmose, & Bleses, 2016; Mayor & Mani, 2019), smartphone-responsive design will become a priority.

When a participant clicks a URL shared by a researcher, they are directed to a website presenting their own personal administration of the Web-CDI. In some cases, they may be asked to read and accept a waiver of consent documentation, depending on whether the researcher has chosen to use that feature (see also Researcher Interface below).

Instructions. After completing the first demographics page, participants are provided with detailed instructions that are appropriate for either the Words & Gestures or Words & Sentences version (see Figure 1 for an example of the instructions for how to determine whether the child "understands and says" a word, which is pertinent to both the Words & Gestures and Words & Sentences forms.). In addition, there are more detailed instructions for completing the vocabulary checklist. Unlike the traditional paper versions, instructions on how to properly choose responses are provided both in written and pictorial form. The pictorial instructions (Figure 1) aim to further increase caregivers' understanding of how to

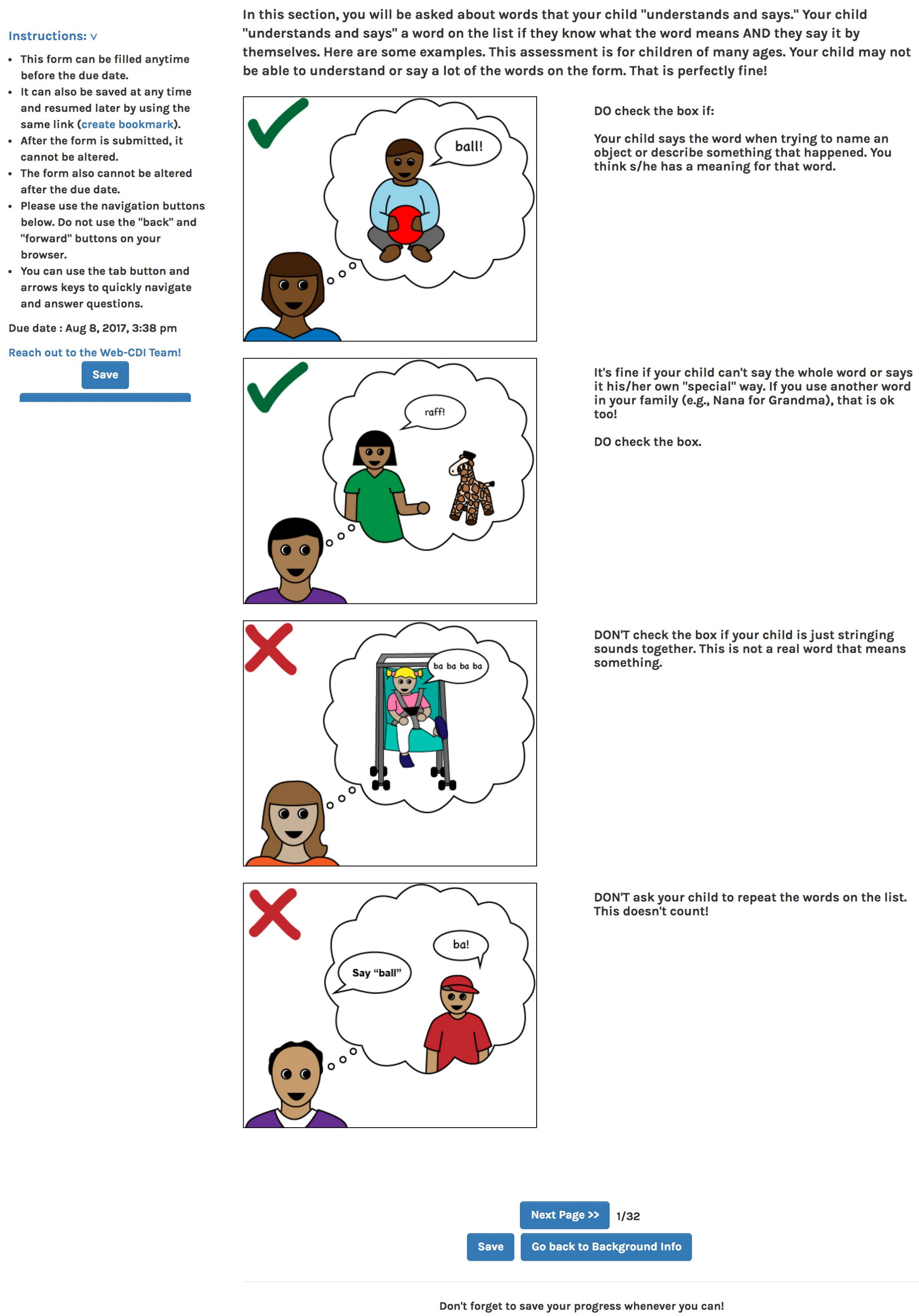

Figure 1. Pictorial instructions indicating how to mark whether a child *understands and says* a word, from the Web-CDI Words and Sentences instrument.

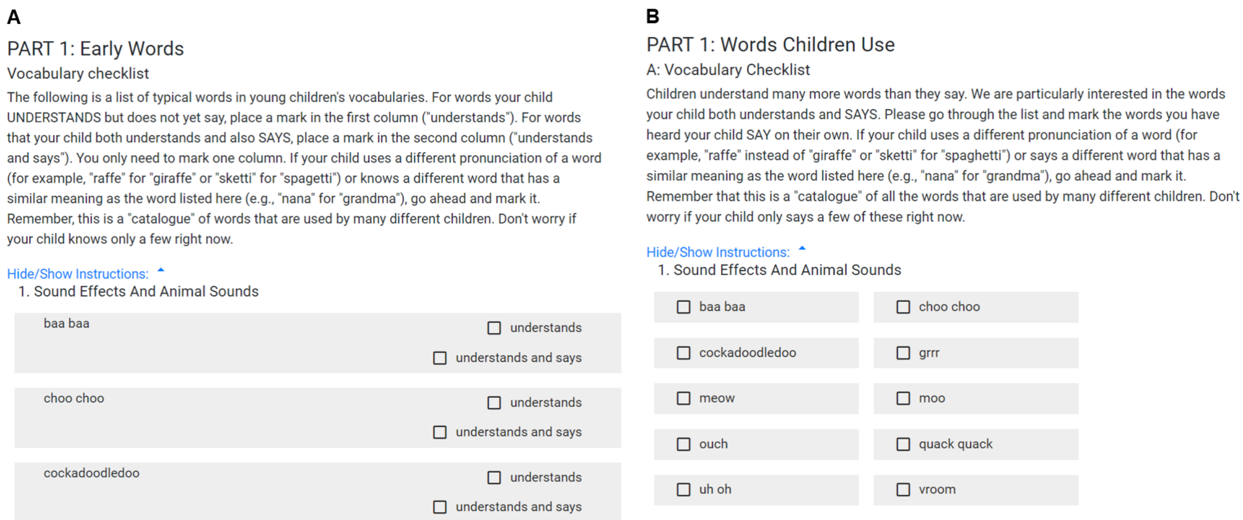

Figure 2. (A) Sample items from the American English Words and Gestures form. (B) Sample items from the American English Words and Sentences form.

complete the checklist. For example, these instructions clarify that the child's understanding of a word requires them to have some understanding of the object that the word refers to or some aspect of the word's meaning. In addition, caregivers are reassured that "child-like" forms (e.g., "raff" for "giraffe") or family- or dialect-specific forms (e.g., "nana" for "grandma") are acceptable evidence. Lastly, caregivers are reminded that the child should be able to produce the words "on their own" and that imitations are not acceptable. These general "rules of thumb" for completing the form should be familiar to researchers who are distributing the forms to caregivers so they can field any questions that may arise. While this is not possible for certain use-cases (e.g., social media recruitment), these instructions should ideally also be reviewed either in writing (e.g., via email) or verbally (e.g., over the phone), so that these pictured instructions serve merely as a reminder to caregivers when completing the form. Pictured instructions are available for download on the MB-CDI website at http://mb-cdi.stanford.edu/about.html.

Completing the instrument. The majority of the participant's time is spent completing the main sections of the instruments. As shown in Figure 2, on the American English Words and Gestures form, the vocabulary checklist portion (396 items) asks caregivers to indicate whether their child can "understand" or "understand and say" each word; they can also indicate that their child neither understands nor says the word by checking neither box. Additionally, gesture communication and other early milestones are assessed. In the American English Words and Sentences form, the vocabulary checklist (680 items) only asks caregivers to indicate which words their child "says." Additional items assess children's production by requesting three of their longest sentences, as well as morphological and syntactic development more broadly. All of these items are broken up across multiple screens for easier navigation through the form.

At the completion of the form, a graph is displayed illustrating how the responses of "understands" or "understands and says" are distributed across the semantic categories on the form. Participants can select to download their own responses. In addition, data from the norming studies are used to estimate the 'hardest' (i.e., most advanced based on previous work on age of acquisition of individual words, Frank, Braginsky, Yurovsky, and Marchman (2021)) word that the child currently understands or produces. This feedback to caregivers is intended to provide caregivers with a fun "thank you" and intentionally avoids any information which frames their child's progress relative to other children or any normative standard, so as to not give the impression that the Web-CDI is a clinical assessment of the child's development. To further underscore this point, the closing page reminds caregivers that their participation does not constitute a clinical evaluation and that they should contact their pediatrician or primary care physician if they have any concerns about their child's development.

Researcher interface

One of the main goals of Web-CDI is to provide a unified CDI platform to the child language research community. To that end, researchers request an account by contacting members of the CDI Advisory Board at `webcdi-contact@stanford.edu`. Once the request is granted, they can design and distribute studies. One rationale for this personalized registration process is that we ask that researchers allow fully anonymized data from their participants to be shared with the CDI Advisory Board, so that it can be added to Wordbank [http://wordbank.stanford.edu/; Frank et al. (2017)] and shared with the broader research community. However, if particular participants indicate in the consent process that they do not want their data to be shared more broadly, then researchers can indicate this in the Web-CDI dashboard to prevent data from specific administrations being contributed to any analyses conducted by the CDI Advisory Board and/or Wordbank. Data currently in Web-CDI, which have not yet been added to the Wordbank repository, will be vetted before being added to ensure that all data being added to Wordbank from Web-CDI are drawn from families with typically-developing children who meet similar inclusion criteria to the ones we describe below in the *Dataset 1* section. Additionally, date of form completion will be preserved when adding Web-CDI data into Wordbank, so that researchers can choose to filter out data that may be affected by the particular point in time at which they were collected (for example, the COVID-19 pandemic, Kartushina et al., 2021).

A study in the context of the Web-CDI system is a set of individual administrations created by a researcher that share certain specifications. Table A1 in the Appendix gives an overview of the customizable features that are available at the study level in Web-CDI. These features are set when creating a study using the “Create Study” tool, and most of the features can be updated continuously during data collection using the “Update Study” tool. While some of these features are only relevant to specific use cases (e.g., longitudinal research and social media data collection, described below), others are relevant to all researchers using Web-CDI.

There are currently several CDI forms available for distribution on Web-CDI, including the English WG and WS forms and forms in other languages (see Cross-linguistic research, below). When creating a study, researchers choose one of the forms that they would like to distribute to participants; only one can be used in a given study. Researchers who wish to send multiple forms to participants simultaneously (e.g., those conducting multilingual research) should create multiple studies, each with a single instrument associated with it.

Researchers can download participant data in two formats. Both formatting options output a comma-separated values file with one row per participant; the full data option includes participant-by-item responses, and allows researchers to explore item-level trends, while the summary data option omits item-level data and only provides summary scores and normative information, including total number of words understood/produced and percentile scores by age in months and gender. Percentile scores based are calculated to a single percentile resolution using norms from Fenson et al. (2007).

Below, we outline several possible use cases of Web-CDI, as well the features which may facilitate them from a researcher's perspective.

Individual recruitment. A first possible workflow using Web-CDI is to send unique study URLs to individual participants. Researchers do so by entering numerical participant IDs or by auto-generating a specified quantity of participant IDs, each with its own unique study URL, using the "Add Participants" tool in the researcher dashboard. New participants can be added on a continual basis so that researchers can adjust the sample size of their study during data collection. Unique links generated for individual participants expire, by default, 14 days after creation, though the number of days before link expiration is adjustable, which may be an important consideration for some researchers depending on their participant populations and specific project timelines. Workflows that involve generating unique links are most suitable for studies which pair the CDI with other measures, or when researchers contact specific participants from an existing database.

Longitudinal studies. Web-CDI also facilitates longitudinal study designs in which each participant completes multiple administrations. Researchers wishing to design longitudinal studies can do so by entering a list of meaningful participant IDs using the "Add Participants" tool in the researcher dashboard. If a specific participant ID is added multiple times, Web-CDI will automatically create multiple unique study URLs in the study dashboard that have that ID. In addition, when creating studies, researchers can select whether they would like the demographics information, vocabulary checklist, or no sections at all to be pre-filled when a participant fills out a repeat administration of the instrument. Unless researchers are interested in cumulative vocabulary counts, it is strongly recommended that they do not use the option to pre-fill the vocabulary checklist portion of the instrument in longitudinal administrations as caregivers should complete the instrument at each time point independently. In the case that researchers do choose this option, this is recorded in the Web-CDI database so that, when the data are added to WordBank, researchers can choose to filter out any pre-filled questionnaires.

Social media and survey vendors. Web-CDI contains several features designed to facilitate data collection from social media recruitment or through third-party crowd-sourcing applications and vendors (e.g., Amazon Mechanical Turk, Prolific). First, rather than creating unique survey links for each participant, researchers can also use a single, anonymous link. When a participant clicks the anonymous link, a new administration with a unique subject ID is created in the study dashboard. Additionally, Web-CDI studies have several customizable features that are geared towards anonymous online data collection. For example, researchers can adjust the minimum amount of time a participant must take to fill out the survey before they are able to submit; with a longer minimum time to completion, researchers can encourage a more thorough completion of the survey. This feature is typically most relevant in research designs in which participants are not vetted by the researcher or those in which there is no direct communication between participants and researchers, as might be the case when recruiting respondents on social media. Responses collected via personal communication with participants show low rates of too-fast responding, mostly removing the need for the minimum time feature. Even in the case of anonymous data collection, however, it is recommended that researchers not raise the minimum completion time higher than 6 minutes, since some caregivers of very young children may theoretically be able to proceed through the measure quickly if their child is not yet verbal. Aside from the minimum time feature, researchers can ask participants to verify that their information is accurate by checking a box at the end of the survey, and can opt to include certain demographic questions at both the beginning and end of the survey, using response consistency on these redundant items as a check of data quality.

Paid participation. If researchers choose to compensate participants directly through the Web-CDI interface, Web-CDI has built-in functionality to distribute redeemable gift codes when a participant reaches the end of the survey. Web-CDI contains several features to facilitate integration with third-party crowdsourcing applications and survey vendors should they choose to handle participant compensation through another platform. For example, when creating studies, researchers can enter a URL to which participants are redirected when they reach the end of the survey. Researchers using the behavioral research platform Prolific can configure their study to collect participants' unique Prolific IDs and pre-fill them in the survey.

Cross-linguistic research. Web-CDI forms are currently available in English (U.S. American and Canadian), Spanish, French (Quebecois), Hebrew, Dutch and Korean. We are looking to add more language forms to the tool, as the paper version of the forms has been adapted into more than 100 different languages and dialects, and further ongoing adaptations have been approved by the MB-CDI board (http://mb-cdi.stanford.edu/adaptations).

System Design

Web-CDI is constructed using open-source software. All of the vocabulary data collected in Web-CDI are stored in a standard MySQL relational database, managed using Django and Python and hosted either by Amazon Web Services or by a European Union (GDPR) compliant server (see below). Individual researchers can download data from their studies through the researcher interface, and Web-CDI administrators have access to the entire aggregate set of data from all studies run with Web-CDI. Website code is available in a GitHub repository at https://github.com/langcog/web-cdi, where interested users can browse, make contributions, and request technical fixes.

Data Privacy and GDPR Compliance

Web-CDI is designed to be compliant with stringent human subjects privacy protections across the world. First, for U.S. users, we have designed Web-CDI based on the United States Department of Health and Human Services "Safe Harbor" Standard for collecting protected health information as defined by the Health Insurance Portability and Accountability Act (HIPAA). In particular, participant names are never collected, birth dates are used to calculate age in months (with no decimal information) but never stored, and geographic zip codes are trimmed to the first 3 digits. Because of the architecture of the site, even though participants enter zip codes and dates of birth, these are never transmitted in full to the Web-CDI server. Since no identifying information is being collected by the Web-CDI system, this feature ensures that Web-CDI can be used by United States labs without a separate Institutional Review Board agreement between users' labs and Web-CDI (though of course researchers using the site will need Institutional Review Board approval of their own research projects).[1]

[1] Issues of de-identification and re-identifiability are complex and ever changing. In particular, compliance with DHHS "safe harbor" standards does not in fact fully guarantee the impossibility of statistical

In the European Union (EU), research data collection and storage is governed by the Generalized Data Protection Regulation (GDPR) and its local instantiation in the legal system of the member states. Some of the questions on the demographic form contain information that may be considered sensitive (e.g., information about children's developmental disorders), and in some cases, the possibility of linking this sensitive information to participant IDs exists, particularly when researchers draw on local databases that contain full names and addresses for recruitment and contacting. As a result, issues regarding GDPR compliance arise when transferring data outside the EU, namely to Amazon Web Services servers housed in the United States. Following GDPR regulations, these issues would make a data sharing agreement between data collectors and Amazon Web Services necessary. In addition, all administrators who can access the collected data would have to enter such an agreement, which needs updating whenever personnel changes occur. To overcome these hurdles, and in consultation with data protection officers, we opted to leverage the local technical expertise and infrastructure to set up a sister site housed on GDPR-compliant servers, currently available at http://webcdi.mpi.nl. This site is updated synchronously with the main Web-CDI website to ensure a consistent user experience and access to the latest features and improvements. This site has been used in 135 successful administrations so far and is the main data collection tool for an ongoing norming study in the Netherlands. We are further actively advertising the option to use the European site to other labs who are following GDPR guidelines and are planning adaptations to multiple European languages, where copyright allows.

Current data collection

We now turn to an overview of the data collected thus far using Web-CDI. First, we examine the full sample of all of the Web-CDI administrations collected as of autumn 2020

re-identification in some cases and if potential users have questions, we encourage them to consult with an Institutional Review Board.

(Dataset 1); we then focus in on a specific subset of Dataset 1 which is comprised of data from recent efforts to oversample non-white, less highly-educated U.S. participants (Dataset 2). Across both datasets, we show that general trends from prior research on vocabulary development are replicated using Web-CDI. Based on this work to date, we then discuss the potential for using Web-CDI to collect vocabulary development data from diverse communities online.

Dataset 1: Full Current Web-CDI Usage

Table 1

Exclusions from Dataset 1: full Web-CDI sample

Exclusion	**WG exclusions**	**% of full WG sample excluded**	**WS exclusions**	**% of full WS sample excluded**
Not first administration	163	5.68%	444	12.35%
Premature or low birthweight	37	1.29%	67	1.86%
Multilingual exposure	449	15.66%	492	13.69%
Illnesses/Vision/Hearing	191	6.66%	203	5.65%
Out of age range	88	3.07%	200	5.56%
Completed survey too quickly	319	11.12%	274	7.62%
System error in word tabulation	1	0.03%	4	0.11%
Total exclusions	1248	44%	1684	47%

In this section, we provide some preliminary analyses of Dataset 1, which consists of the full sample of American English Web-CDI administrations collected before autumn 2020. At time of writing, researchers from 15 universities in the United States have collected over 5,000 administrations of the American English CDI using Web-CDI since it was launched in late 2017, with 2,868 administrations of the WG form before exclusions and 3,594 administrations of the WS form before exclusions. We excluded participants from the subsequent analyses based on the following set of stringent criteria designed for the creation of future normative datasets. We excluded participants if it was not their first administration of the survey; if they were born prematurely or had a birthweight under 5.5 lbs (< 2.5 kg); reported more than 16 hours of exposure to a language other than English per week on average (amounting to approximately > 10% of time during a week that a child hears another language than English); had serious vision impairments, hearing deficits or other developmental disorders or medical issues[2]; were outside of the correct age range for the survey; or spent less time on the survey than a pre-specified timing cutoff. Timing cutoffs were determined by selecting two studies within Dataset 1 that, upon a visual inspection, appeared to contain high-quality responses (i.e., did not contain a disproportionate number of extremely quick responders), and using these to estimate the 5th percentile of completion time by the child's age in months with a quantile regression (following a similar quantile regression method as Bleses, Makransky, Dale, Højen, and Ari (2016)). Thus, for each age on the WG and WS measures, we obtained an estimate of the 5th percentile of completion time and used this estimate as the shortest amount of time participants could spend on the Web-CDI without being excluded from our analyses here.

The exclusion criteria we used were designed to be generally comparable with those used in Fenson et al. (2007), who adopted stringent criteria to establish vocabulary norms that reflect typically developing children's vocabulary trajectories. A complete breakdown of the number of participants excluded on each criterion is in Table 1. Of the completed WG forms, 1,248 were excluded, leading to a final WG sample size of 1,620 administrations, and 1,694 WS administrations were excluded, leading to a final WS sample size of 1,900.

Demographic distribution and exclusions. Figure 3 shows the distribution of participant ethnicities in Dataset 1 as compared with previously reported numbers in the published norming study of the paper-based CDI form by Fenson et al. (2007). Several

[2] Exclusions on the basis of child health were decided on a case-by-case basis by author V.M. in consultation with Philip Dale, Donna Thal, and Larry Fenson.

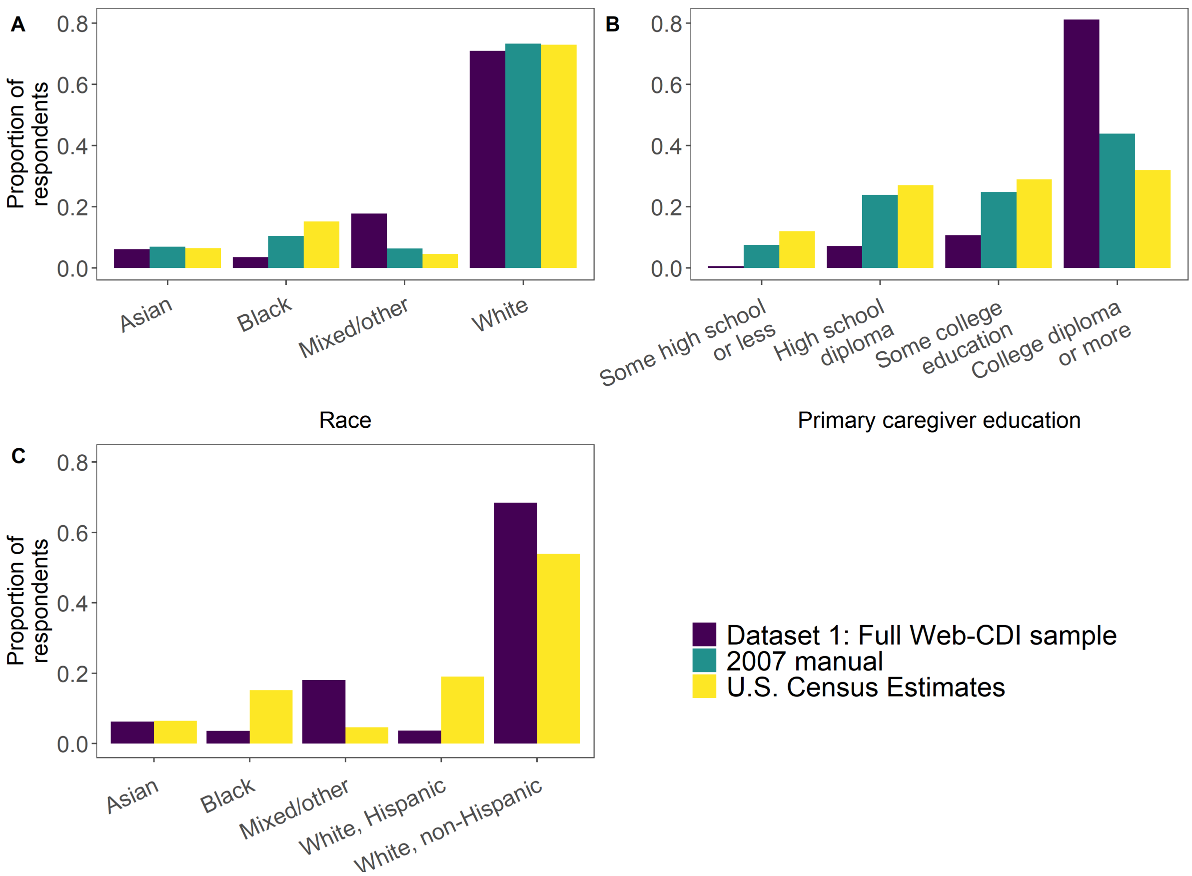

Figure 3. Top row: Proportion of respondents plotted by child race (A) and educational level of primary caregiver (B) from full Web-CDI sample (Dataset 1) to date (N = 3,520), compared with norming sample demographics from Fenson (2007) and U.S. Census data (American Community Survey, 2019; National Center for Education Statistics, 2019). Bottom row (C): Participant breakdown by race in Dataset 1 as compared with U.S. Census data, splitting white participants into those who are Hispanic and those are not.

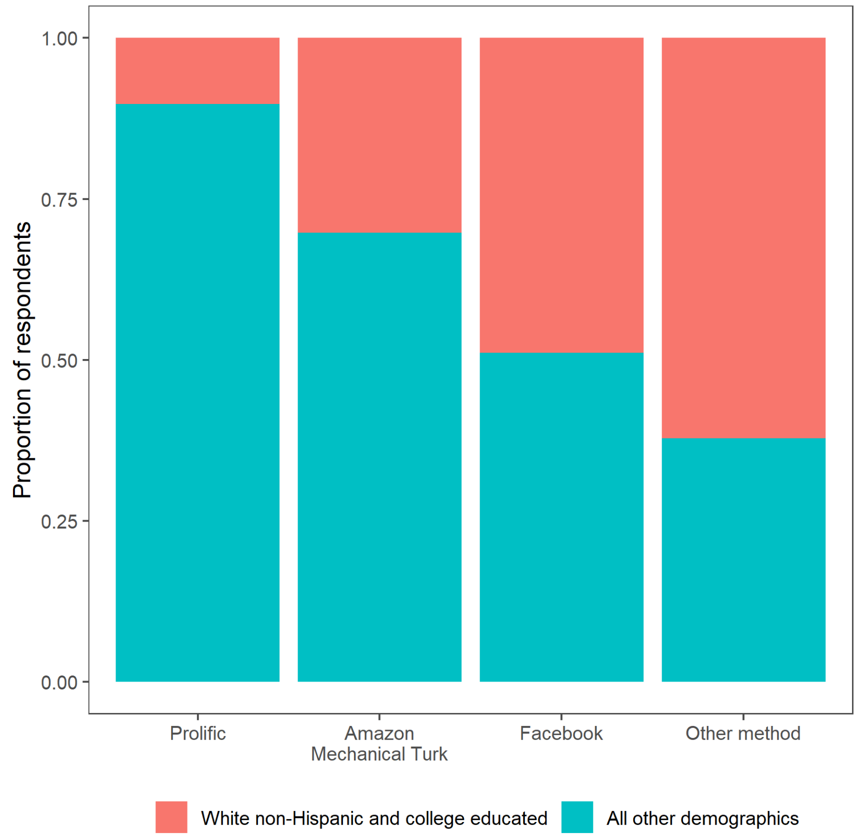

Figure 4. Proportion of participants from Dataset 1 who were white, college educated and not Hispanic, plotted by recruitment method.

issues pertaining to sample representativeness are appreciable. First, as shown in Figure 3A, white participants comprised nearly three quarters of Dataset 1, which is comparable to U.S Census estimates in 2019 of U.S. residents between the ages of 15 and 34 in 2019; however, Figure 3C shows that, compared with U.S. Census estimates, many more white participants in Dataset 1 were non-Hispanic than is true of the U.S. population in general, indicating that Web-CDI is significantly oversampling white, non-Hispanic individuals (the breakdown of white participants into Hispanic and non-Hispanic is not reported in the 2007 norms). Moreover, few participants identified as Hispanic/Latinx: 6.4% of WG participants and 5.2% of WS participants reported Hispanic or Latinx heritage. The low percentage of Hispanic/Latinx participants was due in part to our exclusion of children with substantial exposure to languages other than English: before exclusions, 8.4% of WG participants were Hispanic/Latinx, and 8.1% of WS participants were Hispanic/Latinx. Finally, representation of Black participants is generally lower in Dataset 1 (3.5%) than in the 2007 norms (10.5%), which is in turn lower than U.S. Census estimates (15.2%). This indicates that both Web-CDI data and existing norming samples tend to substantially underrepresent Black participants.

Participants' educational attainment level, as measured by the primary caregiver's highest educational level reached[3], was similarly skewed. In Dataset 1, 81.2% of responses came from families with college-educated primary caregivers compared to 43.8% from the same group in the 2007 norms and 32.0% (Figure 3). Furthermore, less than 1% of participants report a primary caregiver education level less than a high school degree, compared to 7% from the same group in the 2007 norms.

The overrepresentation of white, non-Hispanic Americans and those with high levels of education attainment points to a general challenge encountered in vocabulary development research, which we return to when we detail our efforts to recruit more diverse participants. Figure 4 shows that, of the recruitment methods used in Dataset 1, the studies conducted using the platform Prolific (which we detail in the *Dataset 2* section) contributed the least to the high proportion of white, non-Hispanic, college educated participants. Respondents not known to be recruited through an online channel or crowdsourcing platform (labeled "Other method" in Figure 4) showed the most overrepresentation of white, college educated participants, suggesting that reliance on university convenience samples may be driving the demographic skewness of Dataset 1 most acutely.

[3] Maternal education level is a common measure of family socioeconomic status; we probe *primary caregiver* education level here to accommodate family structures in which child-rearing may not primarily be the responsibility of the child's mother, but we expect that in the vast majority of cases this corresponds to the child's mother.

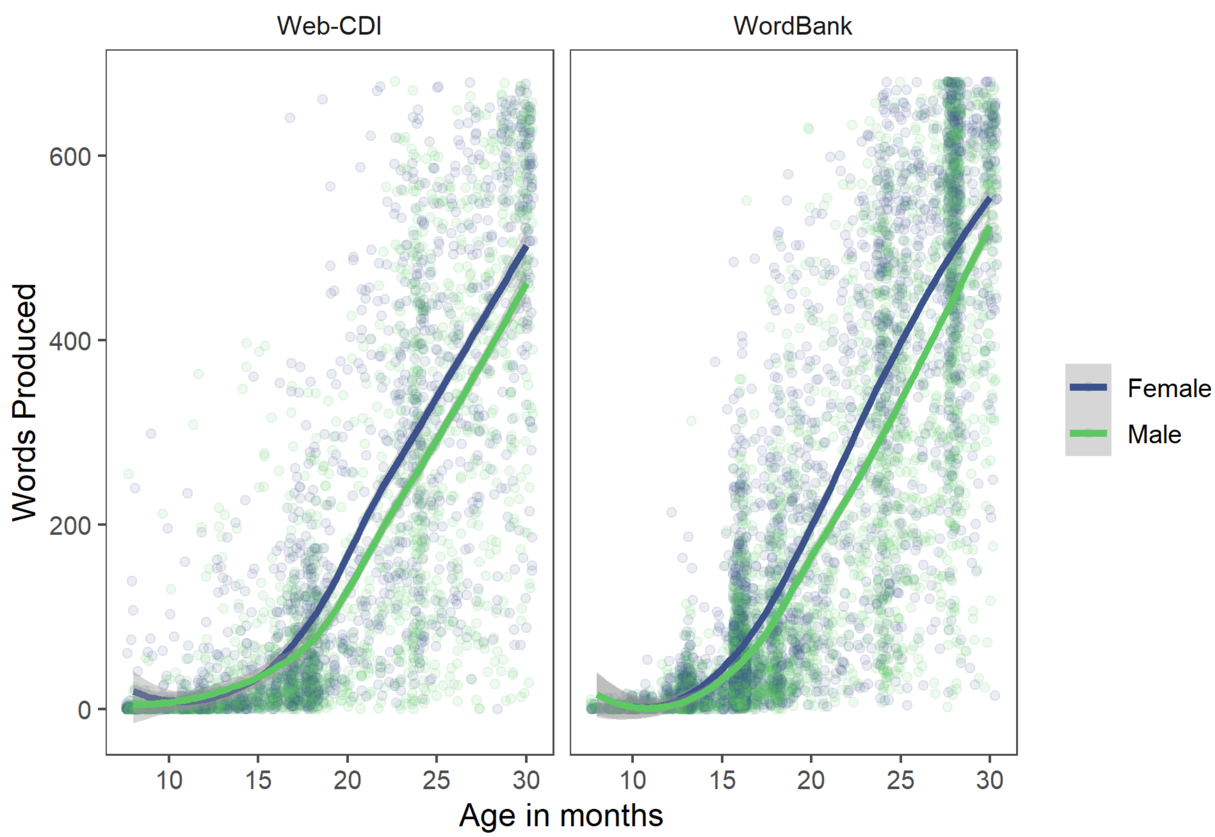

Figure 5. Individual children's vocabulary production scores plotted by children's age and gender (both WG and WS). Left panel: Dataset 1 (full sample of Web-CDI administrations, N = 3,510, with 1,673 girls). Right panel: American English CDI administrations in the WordBank repository (Frank et al., 2021), including only those administrations for which the child's gender was available (N = 6,486, with 3,146 girls). Lines are locally weighted regressions (LOESS) with associated 95% confidence intervals. Children with a different or no reported gender (N = 10) are omitted here.

Results: Dataset 1. Although the CDI instruments include survey items intended to measure constructs other than vocabulary size, such as gesture, sentence production and grammar, we focus exclusively on the vocabulary measures here. We also visualize key analyses from Dataset 1 alongside the analogous analyses on the American English CDI administrations from the WordBank repository (Frank, Braginsky, Yurovsky, & Marchman, 2021) that include the relevant demographic information needed to provide a comparison dataset of traditional paper-and-pencil forms. Across both the WG and WS measures, Dataset 1 shows greater reported vocabulary comprehension and production for older children. Moreover, data from both the WG and WS measures in Dataset 1 replicate a subtle but reliable pattern such that female children tend to have slightly larger vocabulary scores than male children across the period of childhood assessed in the CDI forms (Frank, Braginsky, Yurovsky, & Marchman, 2021), though in these data this difference does not appear until around 18 months (Figure 5).

On the WG form, respondents' reports of children's vocabulary comprehension and production both increased with children's age (Figure 6). We replicate overall patterns found by Feldman et al. (2000) in that, on both the "Words Understood" and "Words Produced" measures, vocabulary scores were slightly negatively correlated with primary caregivers' education level, such that those caregivers without any college education reported higher vocabulary scores on both scales; on the word comprehension scale, this was particularly the case for the youngest infants in the sample. A linear regression model with robust standard errors predicting comprehension scores with children's age and primary caregivers' education level (binned into categories of "High school diploma or less," "Some college education" and "College diploma or more"[4]) as predictors shows main effects of both age ($\beta = 20.05$, $p < 0.001$) and caregiver primary education ($\beta_{highschool} = 21.86$, $p = 0.05$). Similarly, a linear regression model with robust standard errors predicting

[4] "High school diploma or less" corresponds to 12 or fewer years of education; "Some college" corresponds to 13 - 15 years of education; "College diploma or more" refers to 16 or more years of education.

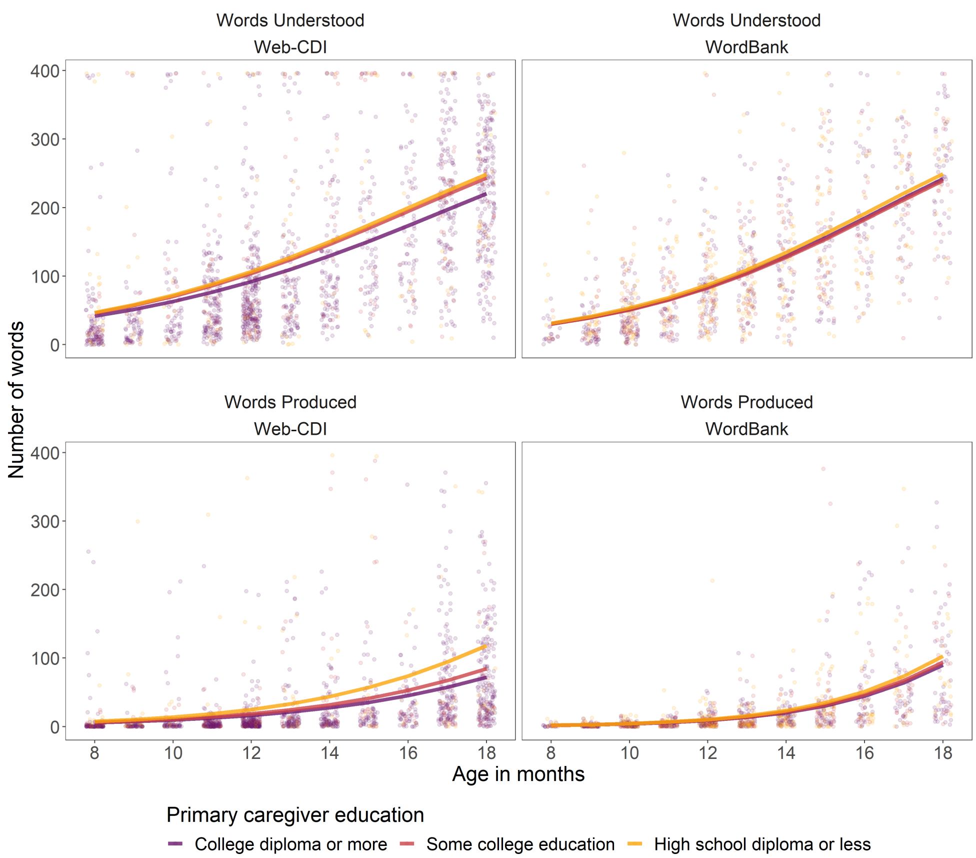

Figure 6. Individual children's word production (top panels) and comprehension (bottom panels) scores from Dataset 1 (full Web-CDI sample) plotted by age and primary caregiver's level of education (binned into "High school diploma or less," "Some college education," and "College diploma or more"). Left panels show results from the sample of Words and Gestures Web-CDI administrations collected as of November 2020 (N = 1,620), and right panels show the subset of American English administrations from Wordbank (Frank et al., 2021) that contain information about caregiver education (N = 1,068) for comparison. Curves show generalized linear model fits.

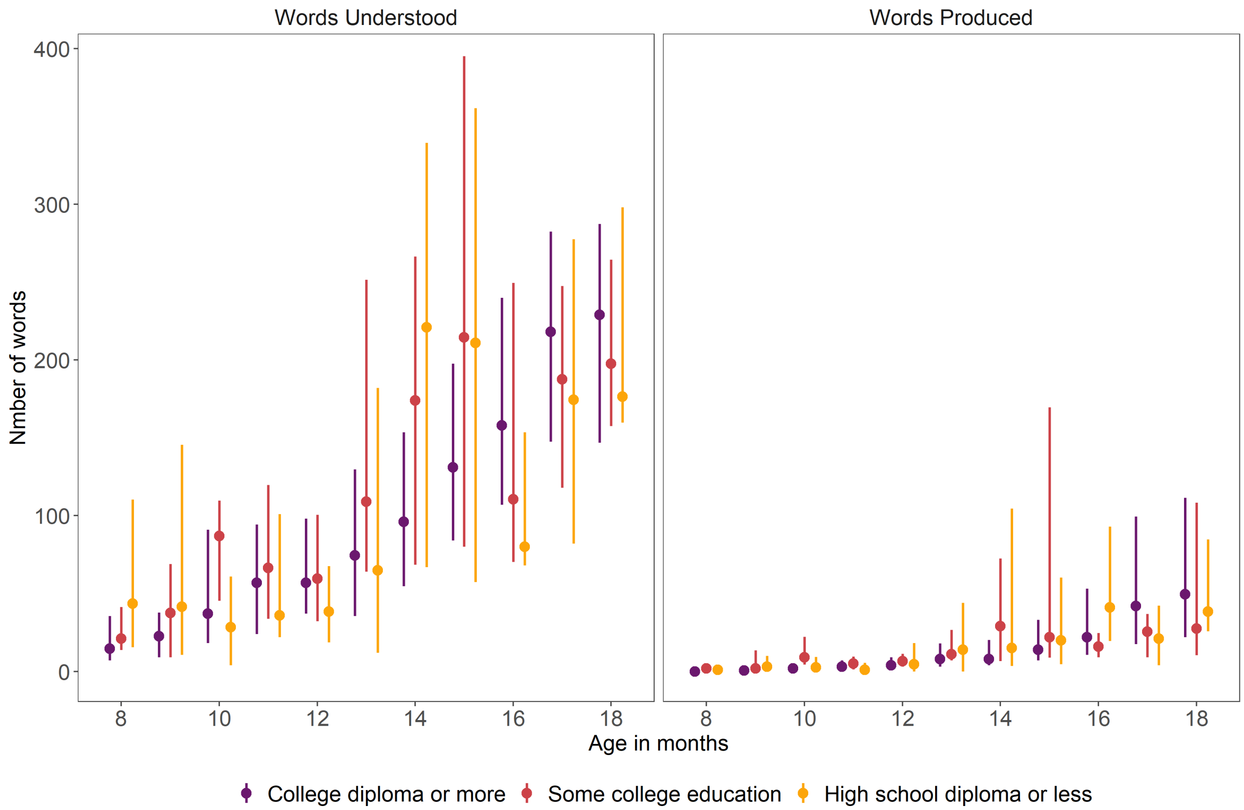

Figure 7. Median vocabulary comprehension (left) and production (right) scores from Dataset 1 (full Web-CDI sample) by age and primary caregiver's level of education attainment on the WG form. Lines indicate span between first and third quartiles for each age.

production scores by children's age and primary caregivers' education level shows main effects of age ($\beta = 7.60$, $p < 0.001$) and caregiver primary education ($\beta_{highschool} = 20.46$, $p = 0.008$). These analyses were not preregistered, but generally follow the analytic strategy in Frank, Braginsky, Yurovsky, and Marchman (2021); additionally, we fit linear models with robust standard errors to account for heteroskedasticity in the data (Astivia & Zumbo, 2019). Generalized linear model predictions for Web-CDI shown in Figure 6 differ somewhat from those for WordBank; prediction curves for caregivers of different education attainment levels diverge slightly more in the Web-CDI sample than in the WordBank sample.

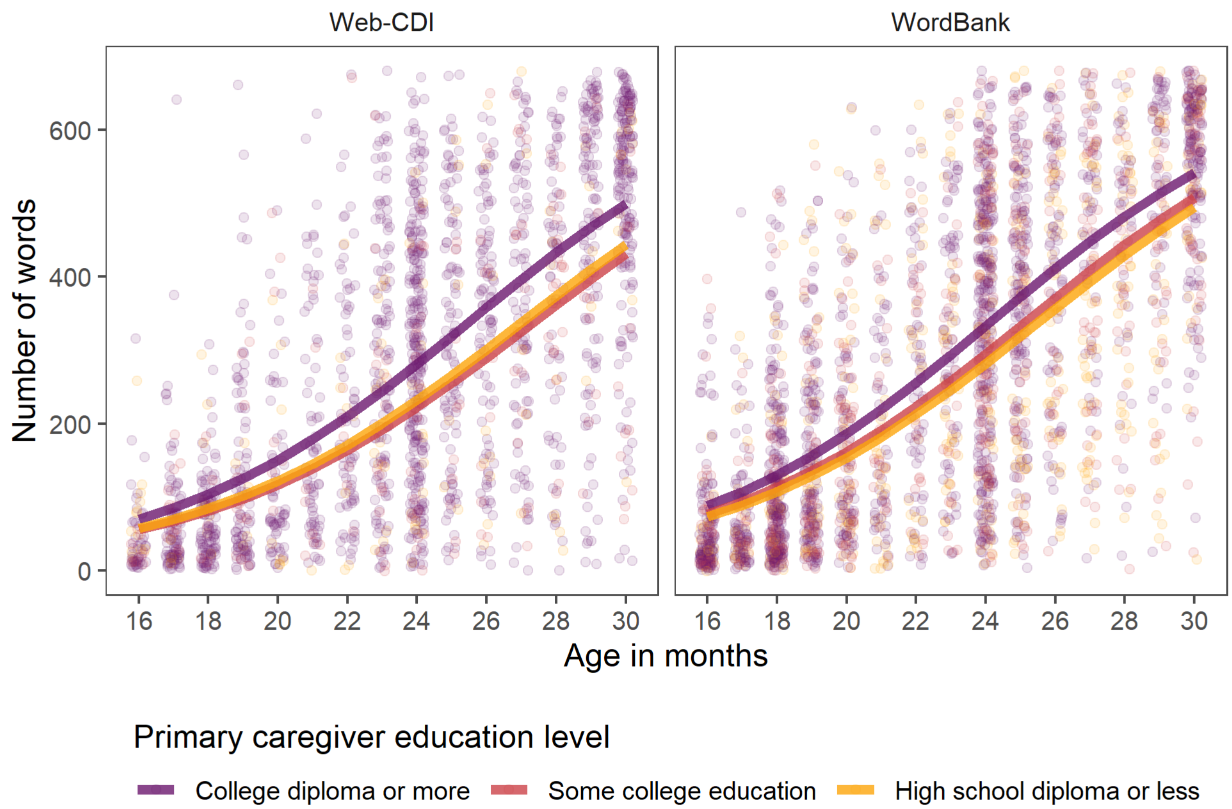

Figure 8. Individual children's vocabulary production scores from Dataset 1 (full Web-CDI sample) plotted by children's age and primary caregiver education level of primary caregiver education on as reported in the sample of Words and Sentences Web-CDI administrations collected as of November 2020 (N = 1,900, left panel) and in the WordBank repository (N = 2,776, right panel). Curves show generalized linear model fits.

The pattern of results seen in the WG subsample of Dataset 1 is consistent with prior findings indicating that respondents with lower levels of education attainment report higher vocabulary comprehension and production on the CDI-WG form (Feldman et al., 2000; Fenson et al., 1994). However, although caregivers with lower levels of education attainment report higher mean levels of vocabulary production and comprehension, median vocabulary scores (which are more robust to outliers) show no clear pattern of difference across primary caregiver education levels (Figure 7). This discrepancy between the regression effects and a group-median analysis suggests that the regression effects described previously are driven in part by differential interpretation of the survey items, such that a few caregivers with lower levels of education attainment are more liberal in reporting their children's production and comprehension vocabulary scores, especially for the youngest children, driving up the mean scores for this demographic group.

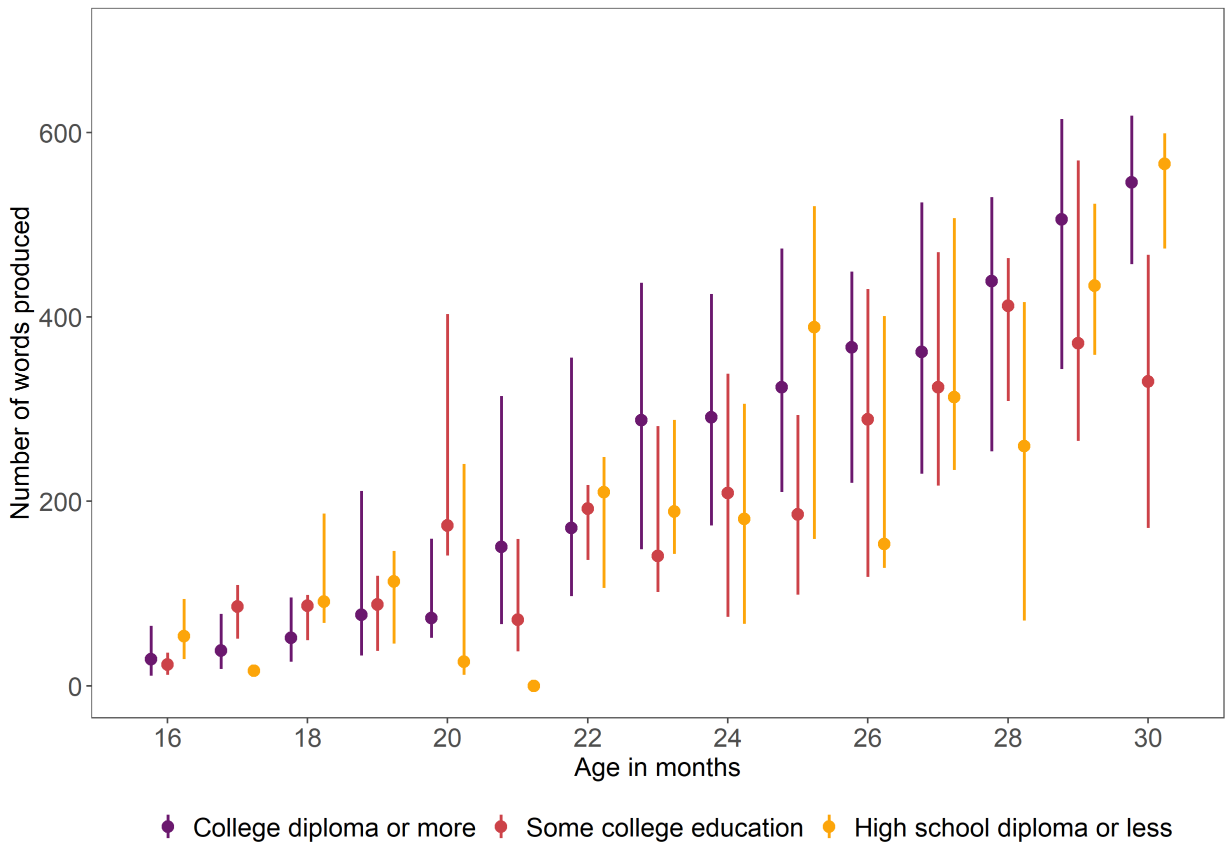

Figure 9. Median vocabulary production scores from Dataset 1 (full Web-CDI sample) by age and primary caregiver's level of education attainment on the WS form. Lines indicate span between first and third quartiles for each age.

Vocabulary production scores on the WS form show the expected pattern of increase with children's age in months; in addition, scores replicate the trend reported in Feldman et al. (2000) and Frank, Braginsky, Yurovsky, and Marchman (2021) such that primary caregiver education is positively associated with children's reported vocabulary size (Figure

8). Because representation of caregivers without a high school diploma is scarce (N = 6 out of a sample of 1,900), interpretation of the data from this group is constrained. Nevertheless, as shown in Figure 8, a small but clear positive association between primary caregiver education and vocabulary score exists such that college-educated caregivers report higher vocabulary scores than those of any other education level. Notably, this association is not the result of outliers and is still appreciable in median scores (Figure 9), unlike the data from the WG measure shown in Figure 7. The implications from these data converge with previous findings which indicate that parental education levels, often used as a metric of a family's socioeconomic status, are related to children's vocabulary size through early childhood.

Discussion: Dataset 1. In general, the full sample of Web-CDI data after exclusions (Dataset 1) replicates previous norming datasets used with the standard paper-and-pencil form of the MB-CDI. We find that vocabulary scores grow with age and that females hold a slight advantage over males in early vocabulary development. Moreover, Dataset 1 replicates a previously documented relationship between primary caregiver education level and vocabulary scores: on the WG form, primary caregiver education shows a slight negative association with vocabulary scores, whereas the trend is reversed in the WS form. Taken together, these data illustrate that Web-CDI and the standard paper-and-pencil form of the CDI give similar results, and thus that Web-CDI can be used as a valid alternative to the paper format.

The data discussed above have resulted from efforts by many researchers across the United States whose motivations for using the Web-CDI vary. As a result, they reproduce many of the biases of standard U.S. convenience samples. In the next section, we describe in more detail our recent efforts to use the Web-CDI to collect vocabulary development data from traditionally underrepresented participant populations in the United States, attempting to counteract these trends.

Dataset 2: Using Web-CDI to Collect Data from Diverse U.S.-based Communities

Despite the large sample sizes we achieved in the previous section, Dataset 1 is, if anything, even more biased towards highly-educated and white families than previous datasets collected using the paper-and-pencil form. How can we recruit more diverse samples to remedy this issue? Here, we discuss and analyze Dataset 2, which consists of those administrations from Dataset 1 which were part of recent data-collection efforts (within the past year and a half) that were specifically aimed towards exploring the use of online recruitment as a potential way to collect more diverse participant samples than are typical in the literature. In other words, the following data from Dataset 2 were included in the previous discussion and analysis of Dataset 1, but we examine them separately here to give special attention to the issue of collecting diverse samples online.

While understanding that the performance of standard measurement tools like the CDI among multilinguals is of immense import to the field of vocabulary development research [Gonzalez et al., in prep; Floccia et al. (2018); De Houwer (2019)], we focused in Dataset 2 only on vocabulary development in monolingual children, because collecting data from multilingual populations introduces additional methodological considerations (e.g., how to measure exposures in each language) that are not the focus of our work here. However, it will be imperative in future to collect large-scale datasets of vocabulary data in bilingual children, both to better calibrate standard tools such as the CDI, as well as to reduce the bias towards monolingual families in the existing literature on measuring vocabulary development.

Online data collection. Online recruitment methods, such as finding participants on platforms such as Amazon Mechanical Turk, Facebook and Prolific, represent one possible route towards assembling a large, diverse sample to take the Web-CDI. These methods allow researchers to depart from their typical geographical recruitment area much more easily than with paper-and-pencil administration. Online recruitment strategies for vocabulary development data collection have been used in the United Kingdom (Alcock, Meints, & Rowland, 2020), but their usage in the U.S. context remains, to our knowledge, rare. In a series of data collection efforts, we used Web-CDI as a tool to explore these different channels of recruitment.

Figure 10. Example Facebook advertisement in Phase 1 of recent data collection.

Dataset 2 consists of data that were collected in two phases. In the first phase, we ran advertisements on Facebook which were aimed at non-white families based on users' geographic locations (e.g., targeting users living in majority-Black cities) or other profile features (e.g., ethnic identification, interest in parenthood-related topics). Advertisements consisted of an image of a child and a caption informing Facebook users of an opportunity to fill out a survey on their child's language development and receive an Amazon gift card

Table 2

Exclusions from Dataset 2: recent data collection using Facebook and Prolific.

Exclusion	**WG exclusions**	**% of full WG sample excluded**	**WS exclusions**	**% of full WS sample excluded**
Not first administration	0	0.00%	0	0.00%
Premature or low birthweight	7	2.53%	1	0.33%
Multilingual exposure	18	6.50%	23	7.62%
Illnesses/Vision/Hearing	4	1.44%	4	1.32%
Out of age range	1	0.36%	26	8.61%
Completed survey too quickly	119	42.96%	133	44.04%
System error in word tabulation	0	0.00%	0	0.00%
Total exclusions	149	54%	187	62%

(Figure 10). Upon clicking the advertisement, participants were redirected to a unique administration of the Web-CDI; they received $5 upon completing the survey. This open-ended approach to recruitment offered several advantages, namely that a wide variety of potential participants from specific demographic backgrounds can be reached on Facebook. However, we also received many incomplete or otherwise unusable survey administrations, either from Facebook users who clicked the link and decide not to participate, or those who completed the survey in an extremely short period of time (over half of all completed administrations, Table 2).

In the second phase, we used the crowdsourcing survey vendor Prolific (http://prolific.co) in the hopes that some of the challenges encountered with Facebook recruitment would be addressed. Prolific allows researchers to create studies and post them to individuals who are in the platform's participant database, each of whom is assigned a unique alphanumeric "Prolific ID." Importantly, Prolific maintains detailed demographic information about participants, allowing researchers to specify who they would like to complete their studies. Prolific further has a built-in compensation infrastructure that handles monetary payments to participants, eliminating the need to disburse gift cards through Web-CDI.

In the particular case of Web-CDI, the demographic information needed to determine whether an individual was eligible to complete our survey (e.g., has a child in the correct age range, lives in a monolingual household, etc.) was more specific than the information that Prolific collects about their participant base. We therefore used a brief pre-screening questionnaire to generate a list of participants who were eligible to participate, and subsequently advertised the Web-CDI survey to those participants. Given that we were interested only in reaching participants in the United States who were not white or who did not have a college diploma, our data collection efforts only yielded a sample that was small (N = 68) but much more thoroughly screened than that which we could obtain on Facebook.

Across both phases (Facebook and Prolific recruitment), we used the same exclusion criteria as in the full Web-CDI sample to screen participants. A complete tally of all excluded participants is shown in Table 2. In both the WG and WS surveys, exclusion rates in Dataset 2 were high, amounting to 58% of participants who completed the survey. The high exclusion rates were notably driven by an accumulation of survey administrations which participants completed more quickly than our time cutoffs allow (Tables A4 and A5). Many of the survey administrations excluded for fast completion also had missing demographic information reported: Among WG participants excluded for too-fast completions, 93% did not report ethnicity, and among WS participants excluded for the same reason, 97% did not report ethnicity. Absence of these data prevents us from drawing conclusions about the origin or demographic profile of administrations that were excluded. After exclusions, full sample size in Dataset 2 was N = 128 WG completions and N = 115 WS completions.

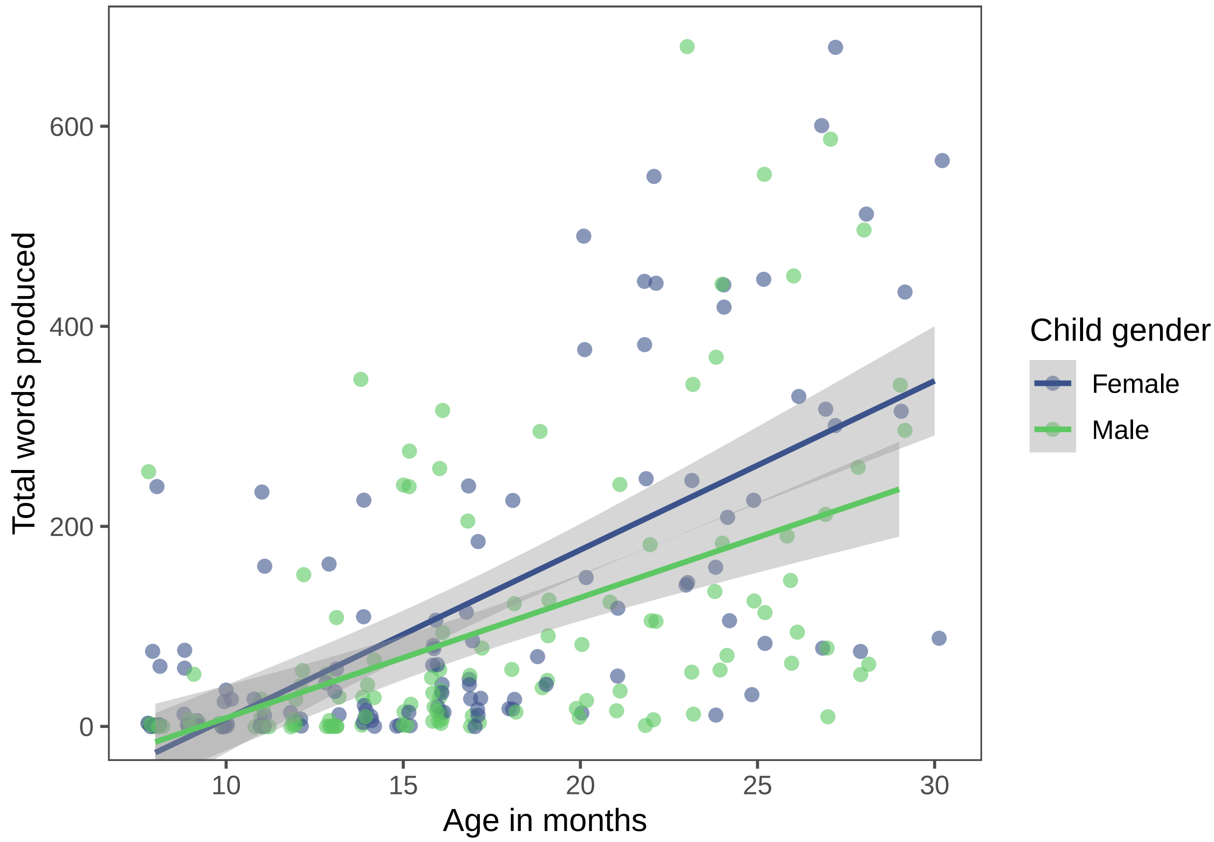

Figure 11. Individual children's vocabulary production scores from Dataset 2 (recent data collection efforts) plotted by children's age and gender (both WG and WS, N = 240, with 114 girls). Lines are best linear fits with associated 95% confidence intervals. Children with a different or no reported gender (N = 3) are omitted here.

The results from Dataset 2 show overall similar patterns to the full Web-CDI sample in several regards. Word production scores from both the WG and WS administrations reflect growing productive vocabulary across the second and third years, with a very small gender effect such that female children's vocabularies are higher across age than males' (Figure 11). The relationship between caregivers' reported levels of education and child's vocabulary score is not as clear as it is in the full Web-CDI sample (Figure 12); however, children of college-educated caregivers reported generally higher vocabulary scores across age than did children of caregivers without any college degree. These patterns suggest that

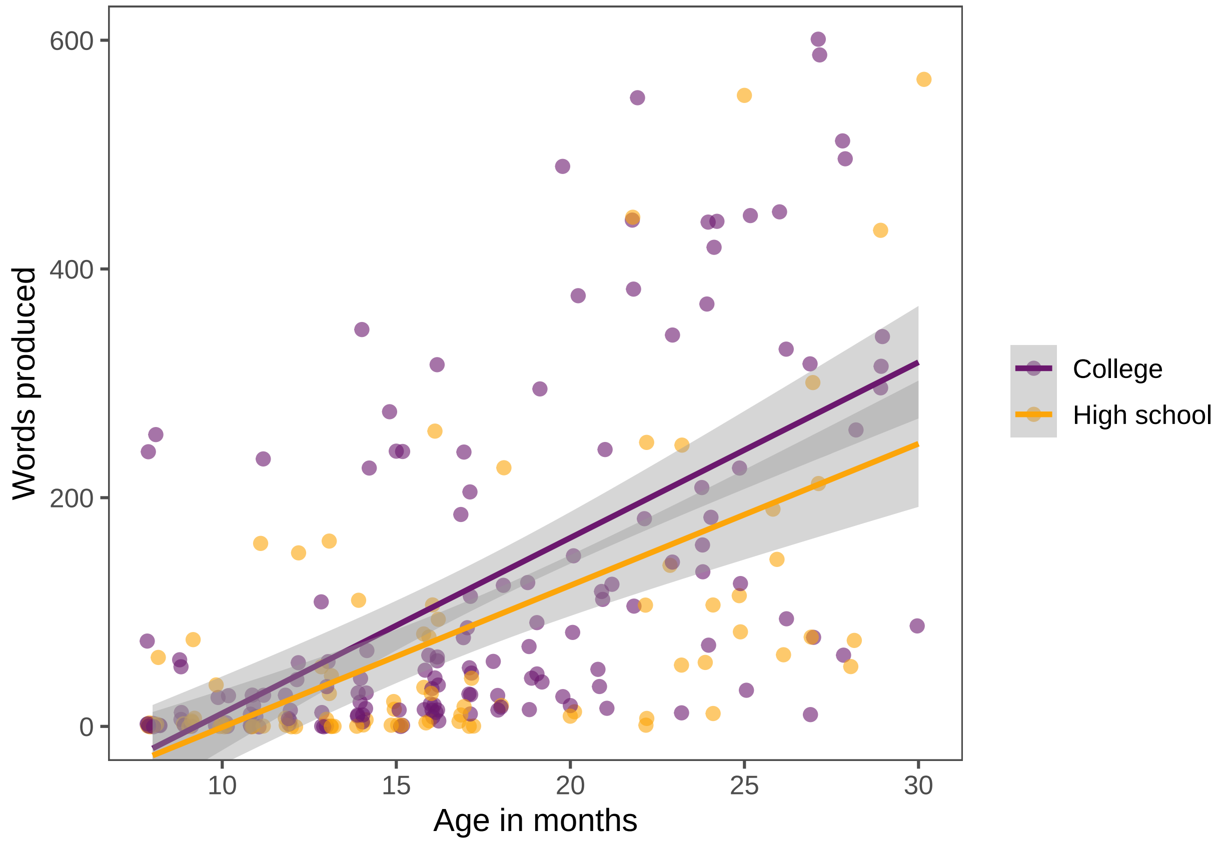

Figure 12. Individual children's vocabulary production scores from Dataset 2 (recent data collection efforts) plotted by age and level of primary caregiver education, binned into those with a high school diploma or less education and those with some college education or a college diploma (N = 243). Lines show best linear fits and associated 95% confidence intervals.

our data show similar general patterns to other CDI datasets with other populations (Frank, Braginsky, Yurovsky, & Marchman, 2021).

Importantly, Dataset 2 showed a substantial improvement in reaching non-white or less highly-educated participants. After exclusions, Dataset 2 has a higher proportion of non-white participants than Dataset 1 (the overall Web-CDI sample) and the norms established by Fenson et al. (2007) (Figure 13). Black participants in particular showed a marked increase in representation, from 10.5% in the 2007 norms to 30.7% in Dataset 2,

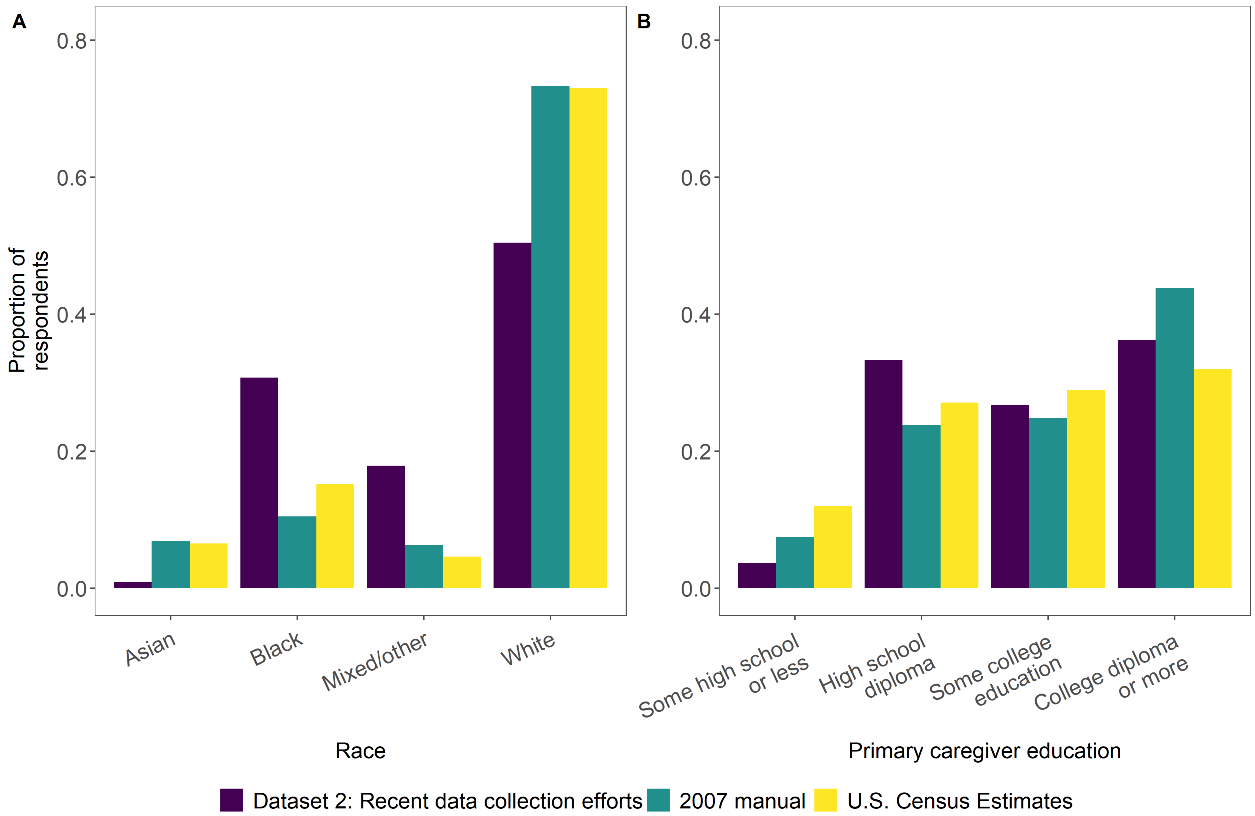

Figure 13. Proportion of respondents plotted by child race (A) and educational level of primary caregiver (B) from Dataset 2, recent data collection efforts aimed towards oversampling non-white, less highly-educated families (N = 243), compared with norming sample demographics from Fenson (2007). Latinx participants can be of any race and are thus not represented as a separate category here.

while the proportion of white participants decreased from 73.3% in the 2007 norms to 50.5% in Dataset 2. Representation on the basis of families' reported primary caregiver education also improved (Figure 13). Participants with only a high school diploma accounted for 33.3% of Dataset 2 as compared to 23.8% in the 2007 norms, and representation of those with a college diploma or more education decreased from 43.8% in the 2007 norms to 36.2% in Dataset 2. Notably, the distribution of Dataset 2 with regards to primary caregiver education level is quite similar to Kristoffersen et al. (2013), who collected a large, nationally-representative sample of CDI responses in Norway and obtained a sample with 30%, 42%, and 24% for participants reporting 12, 14-16, and 16+ years of education, respectively.

Discussion: Dataset 2. The results from Dataset 2 indicate that Web-CDI could is a promising platform to collect vocabulary development data in non-white populations and communities with lower levels of education attainment when paired with online recruitment methods that yield legitimate, representative participant samples. At the same time, however, these data convey clear limitations of our approach. Perhaps most conspicuously, more than half of completed administrations in this sample had to be excluded, in many cases because the information provided by participants appeared rushed or incomplete: over 40% of administrations were completed in a shorter amount of time than that allowed by our cutoff criteria (Tables A4 and A5), and of these quick completions, well over 90% were missing demographic information that is rarely missing in other administrations of the form. Determining the precise reasons for the high exclusion rate, and how (if at all) this (self-)selection may bias data reflecting demographic trends in vocabulary development, requires a more thorough assessment of who is submitting hastily-completed forms. Such an assessment is beyond the scope of the current study. However, all respondents who got to the end of the form were compensated regardless of how thoroughly they completed it, creating the possibility that some participants who clicked the anonymous link may not have been members of the population of interest, but rather were other individuals motivated by compensation. To the extent that participants moved through the form quickly because they found the length burdensome, a transition to short forms, including computer adaptive ones (e.g., Chai, Lo, & Mayor, 2020; Kachergis et al., 2021; Makransky, Dale, Havmose, & Bleses, 2016; Mayor & Mani, 2019), would potentially increase data quality and completion rates substantially.

Additionally, the exclusion rates described previously provide information only on those participants who did, at some point, submit a completed form, but many individuals clicked the advertisement link and did not subsequently continue on to complete the form. Without an in-depth exploration of who is clicking the link and why they might choose not to continue, we cannot draw conclusions about the representativeness of the sample in Dataset 2 with regards to the communities we would like to include in our research. As such, a more thorough understanding of how users from different communities respond to various recruitment and sampling methods is needed in future work in order to draw conclusions about demographic trends above and beyond those already established in the literature.

Participants in Dataset 2 were recruited through a targeted post on social media, a technique that is considerably more anonymous than recruitment strategies which entail face-to-face or extended contact between researchers and community members. Online recruitment methods may not be suitable for all communities, especially when researchers ask participants to report potentially sensitive information about the health, developmental progress, ethnicity and geographic location of their children (even when such information is stored anonymously). Our goal here was to assess whether general trends in past literature could be recovered using such an online strategy, but future research should take into account that other more personal methods of recruitment, such as direct community outreach or liaison contacts, may improve participants' experiences and their willingness to engage with the study.

Finally, a significant limitation of the data collection process in Dataset 2 is that many people in the population of interest - particularly lower-income families - do not have reliable internet access. Having participants complete the Web-CDI on a mobile device may alleviate some of the issues caused by differential access to Wi-Fi, since the vast majority of American adults own a smartphone (Pew Research Center, 2019). Accordingly, improving Web-CDI's user experience on mobile platforms will be an important step towards ensuring that caregivers across the socioeconomic spectrum can easily complete the survey. For smartphone users on pay-as-you-go plans, who may be reluctant to use phone data to complete a study, a possible solution could be compensating participants for the amount of "internet time" they incurred completing the form.

General Discussion and Conclusions

In this paper, we have presented Web-CDI, a comprehensive online interface for researchers to measure children's vocabulary by administering the MacArthur-Bates Communicative Development Inventories family of parent-report instruments. Web-CDI provides a convenient researcher management interface, built-in data privacy protections, and a variety of features designed to make both longitudinal and social-media sampling easy. To date, over 3,500 valid administrations of the WG and WS forms have been collected on Web-CDI from more than a dozen researchers in the United States after applying strict exclusion criteria derived from previous norming studies (Fenson et al., 2007, 1994). Our analysis of Dataset 1 shows that demographic trends from previous work using the paper-and-pencil CDI form are replicated in data gleaned from Web-CDI, suggesting that the Web-CDI is a valid alternative to the paper form and captures similar results.

Many research laboratories, not only in the United States but around the world, collect vocabulary development data using the MacArthur-Bates CDI in its original or adapted form. With traditional paper-based forms, combining insights from various research groups can prove challenging, as each group may have slightly different ways of formatting and managing data from CDI forms. By contrast, if all of these groups' data come to be stored in a single repository with a consistent database structure, data from disparate sources can easily be collated and analyzed in a uniform fashion. As such, a centralized repository such as Web-CDI provides a streamlined data-aggregation pipeline that facilitates cross-lab collaborations, multisite research projects and the curation of large datasets that provide more power to characterize the vast individual differences present in children's vocabulary development.

Beyond the goal of simply getting more data, we hope that Web-CDI can advance efforts to expand the reach of vocabulary research past convenience samples into diverse communities. A key question in the field of vocabulary development concerns the mechanisms through which sociodemographic variables, such as race, ethnicity, income and education are linked to group differences in vocabulary outcomes. Large, population-representative samples of vocabulary development data are needed to understand these mechanisms, but research to date (including the full sample of Web-CDI administrations) has often oversampled non-Hispanic white participants and those with advanced levels of education.

We explored the use of Web-CDI as part of a potential strategy to collect data from non-white and less highly-educated communities in two phases (Dataset 2). Several overall patterns emerged which we expected: vocabulary scores grew with age, providing a basic validity check of the Web-CDI measure; females held a slight advantage in word learning over males; and children of caregivers with a college education showed slightly higher vocabulary scores. Nonetheless, the insights from these data, while aligned with past norming studies, are necessarily constrained by several features of our method.

Limitations of our method notwithstanding, a transition to web-based data collection streamlines the process by which historically underrepresented populations can be reached in child language research. In particular, recruitment methods involving community partners, such as parenting groups, childcare centers and early education providers, are simplified substantially if leaders in these organizations can distribute a web survey to their members that is easy to fill out, as compared with paper forms, which typically present logistical hurdles for distribution and collection. Additionally, we hope that Web-CDI can serve as an accessible, free, and easy to use resource for researchers already doing extensive work with underrepresented groups.

Web-based data collection can capture useful information about vocabulary development from diverse communities, but future research will need to examine which sampling methods can yield accurate, population-representative data that can advance our understanding of the link between sociodemographic variation and variation in language outcomes.

Acknowledgements

We thank Larry Fenson, Philip Dale, and Donna Thal for their assistance and helpful feedback preparing this manuscript.

Ethics statement

Data collected in the United States for this project are anonymized according to guidelines set forth by the United States Department of Health and Human Services. Data collection at Stanford University was approved by the Stanford Institutional Review Board (IRB), protocol 20398.

Data, code and materials availability statement

- Open data: All data analyzed in this work are available on the Open Science Framework at https://osf.io/nmdq4/.
- Code: All code for this work is avaiable on the Open Science Framework at https://osf.io/nmdq4/.
- Materials: All code and materials for the Web-CDI are openly available at https://github.com/langcog/web-cdi. If readers wish to view the Web-CDI interface in full from the participants' or researchers' perspectives, they are encouraged to contact `webcdi-contact@stanford.edu`.

Author contributions

- Conceptualization: Benjamin deMayo, Danielle Kellier, Mika Braginsky, Christina Bergmann, Caroline Rowland, Michael Frank and Virginia Marchman.
- Data Curation: Benjamin deMayo, Danielle Kellier and Virginia Marchman.
- Formal Analysis: Benjamin deMayo.
- Funding Acquisition: Caroline Rowland and Michael Frank.
- Investigation: Benjamin deMayo, Danielle Kellier and Virginia Marchman.
- Methodology: Benjamin deMayo, Danielle Kellier, Michael Frank and Virginia Marchman.
- Project Administration: Caroline Rowland, Michael Frank and Virginia Marchman.
- Software: Danielle Kellier, Mika Braginsky, Christina Bergmann and Cielke Hendriks.
- Supervision: Christina Bergmann, Caroline Rowland, Michael Frank and Virginia Marchman.
- Visualization: Benjamin deMayo.
- Writing - Original Draft Preparation: Benjamin deMayo, Michael Frank and Virginia Marchman.
- Writing - Review & Editing: Benjamin deMayo, Danielle Kellier, Mika Braginsky, Christina Bergmann, Cielke Hendriks, Caroline Rowland, Michael Frank and Virginia Marchman.

Software used

R [Version 4.0.3; R Core Team (2020)] and the R-packages *broman* [Version 0.71.6; Broman (2020)], *cowplot* [Version 1.1.0; Wilke (2020)], *dplyr* [Version 1.0.2; Wickham, François, Henry, and Müller (2020)], *estimatr* [Version 0.26.0; Blair, Cooper, Coppock, Humphreys, and Sonnet (2020)], *forcats* [Version 0.5.0; Wickham (2020a)], *fs* [Version 1.5.0; Hester and Wickham (2020)], *ggplot2* [Version 3.3.2; Wickham (2016)], *here* [Version 0.1; Müller (2017)], *kableExtra* [Version 1.3.4; Zhu (2020)], *papaja* [Version 0.1.0.9997; Aust and

Barth (2020)], *purrr* [Version 0.3.4; Henry and Wickham (2020)], *readr* [Version 1.4.0; Wickham and Hester (2020)], *scales* [Version 1.1.1; Wickham and Seidel (2020)], *stringr* [Version 1.4.0; Wickham (2019)], *tibble* [Version 3.0.4; Müller and Wickham (2020)], *tidyr* [Version 1.1.2; Wickham (2020b)], *tidyverse* [Version 1.3.0; Wickham et al. (2019)], *wordbankr* [Version 0.3.1; Braginsky (2020)], and *xtable* [Version 1.8.4; Dahl, Scott, Roosen, Magnusson, and Swinton (2019)]

References

Alcock, K., Meints, K., & Rowland, C. (2020). *The UK communicative development inventories: Words and gestures.* J&R Press.

Astivia, O. L. O., & Zumbo, B. D. (2019). Heteroskedasticity in multiple regression analysis: What it is, how to detect it and how to solve it with applications in r and SPSS. *Practical Assessment, Research, and Evaluation*, *24*(1), 1.

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Bates, E., & Goodman, J. C. (2001). On the inseparability of grammar and the lexicon: Evidence from acquisition.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., … Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *J Child Lang*, *21*(01), 85–123.

Blair, G., Cooper, J., Coppock, A., Humphreys, M., & Sonnet, L. (2020). *Estimatr: Fast estimators for design-based inference. Journal of Open Source Software* (Vol. 4, p. 1686). Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Bleses, D., Makransky, G., Dale, P. S., Højen, A., & Ari, B. A. (2016). Early productive vocabulary predicts academic achievement 10 years later. *Applied Psycholinguistics*, *37*(6), 1461–1476.

Bornstein, M. H., & Putnick, D. L. (2012). Stability of language in childhood: A multiage, multidomain, multimeasure, and multisource study. *Developmental Psychology*, *48*(2), 477.

Braginsky, M. (2020). *Wordbankr: Accessing the wordbank database.* Retrieved from https://CRAN.R-project.org/package=wordbankr

Broman, K. W. (2020). *Broman: Karl broman's r code*. *Journal of Open Source Software* (Vol. 4, p. 1686). Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Caselli, N. K., Lieberman, A. M., & Pyers, J. E. (2020). The ASL-CDI 2.0: An updated, normed adaptation of the MacArthur bates communicative development inventory for american sign language. *Behavior Research Methods*, 1–14.

Chai, J. H., Lo, C. H., & Mayor, J. (2020). A bayesian-inspired item response theory–based framework to produce very short versions of MacArthur–bates communicative development inventories. *Journal of Speech, Language, and Hearing Research*, *63*(10), 3488–3500.

Dahl, D. B., Scott, D., Roosen, C., Magnusson, A., & Swinton, J. (2019). *Xtable: Export tables to LaTeX or HTML*. Retrieved from https://CRAN.R-project.org/package=xtable

Dale, P. S. (2015). Adaptations, Not Translations! Retrieved from http://mb-cdi.stanford.edu/Translations2015.pdf

De Houwer, A. (2019). Equitable evaluation of bilingual children's language knowledge using the CDI: It really matters who you ask. *Journal of Monolingual and Bilingual Speech*, *1*(1), 32–54.

Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur communicative development inventories at ages one and two years. *Child Development*, *71*(2), 310–322.

Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates Communicative Development Inventories*. Brookes Publishing Company.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., … Stiles, J. (1994). Variability in early communicative development. *Monogr Soc Res Child Dev*, *59*(5).

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, *21*(1), 95–116.

Floccia, C., Sambrook, T. D., Delle Luche, C., Kwok, R., Goslin, J., White, L., … others. (2018). Vocabulary of 2-year-olds learning english and an additional language: Norms and effects of linguistic distance. *Monographs of the Society for Research in Child Development*, *83*(1), 1–135.

Frank, M. C., Bergelson, E., Bergmann, C., Cristia, A., Floccia, C., Gervain, J., … others. (2017). A collaborative approach to infant research: Promoting reproducibility, best practices, and theory-building. *Infancy*, *22*(4), 421–435.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2021). *Variability and consistency in early language learning: The wordbank project*. MIT Press.

Henry, L., & Wickham, H. (2020). *Purrr: Functional programming tools. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Hester, J., & Wickham, H. (2020). *Fs: Cross-platform file system operations based on 'libuv'. Journal of Open Source Software* (Vol. 4, p. 1686). Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Kachergis, G., Marchman, V., Dale, P., Mehta, H., Mankewitz, J., & Frank, M. (2021). *An online computerized adaptive test (CAT) of children's vocabulary development in english and mexican spanish*.

Kapalková, S., & Slanèová, D. (2007). Adaptation of CDI to the slovak language. In *Proceedings from the first european network meeting on the communicative* *development inventories: May 24-28 2006 dubrovnik croatia*. University of Gävle.

Kartushina, N., Mani, N., AKTAN-ERCIYES, A., Alaslani, K., Aldrich, N. J., Almohammadi, A., … al., et. (2021). COVID-19 first lockdown as a unique window into language acquisition: What you do (with your child) matters. PsyArXiv. https://doi.org/10.31234/osf.io/5ejwu

Kristoffersen, K. E., Simonsen, H. G., Bleses, D., Wehberg, S., Jørgensen, R. N., Eiesland, E. A., & Henriksen, L. Y. (2013). The use of the internet in collecting CDI data–an example from norway. *Journal of Child Language*, *40*(03), 567–585.

Makransky, G., Dale, P. S., Havmose, P., & Bleses, D. (2016). An item response theory–based, computerized adaptive testing version of the MacArthur–bates communicative development inventory: Words & sentences (CDI: WS). *Journal of Speech, Language, and Hearing Research*, *59*(2), 281–289.

Mayor, J., & Mani, N. (2019). A short version of the MacArthur–bates communicative development inventories with high validity. *Behavior Research Methods*, *51*(5), 2248–2255.

Müller, K. (2017). *Here: A simpler way to find your files. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Müller, K., & Wickham, H. (2020). *Tibble: Simple data frames. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Percentage of persons 18 to 24 years old and age 25 and over, by educational attainment, race/ethnicity, and selected racial/ethnic subgroups: 2010 and 2017. (2019). https://nces.ed.gov/programs/digest/d18/tables/dt18_104.40.asp?referer=raceindica.asp.

Pew research center mobile fact sheet. (2019). Retrieved from https://www.pewresearch.org/internet/fact-sheet/mobile/

R Core Team. (2020). *R: A language and environment for statistical computing. Journal of Open Source Software* (Vol. 4, p. 1686). Vienna, Austria: R Foundation for Statistical Computing; Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Snyder, T. D., De Brey, C., & Dillow, S. A. (2019). Digest of education statistics 2017, NCES 2018-070. *National Center for Education Statistics.*

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis. Journal of Open Source Software* (Vol. 4, p. 1686). Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Wickham, H. (2019). *Stringr: Simple, consistent wrappers for common string operations. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Wickham, H. (2020a). *Forcats: Tools for working with categorical variables (factors). Journal of Open Source Software* (Vol. 4, p. 1686). Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Wickham, H. (2020b). *Tidyr: Tidy messy data. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., … Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43), 1686. https://doi.org/10.21105/joss.01686

Wickham, H., François, R., Henry, L., & Müller, K. (2020). *Dplyr: A grammar of data manipulation. Journal of Open Source Software* (Vol. 4, p. 1686). Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Wickham, H., & Hester, J. (2020). *Readr: Read rectangular text data. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Wickham, H., & Seidel, D. (2020). *Scales: Scale functions for visualization. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Wilke, C. O. (2020). *Cowplot: Streamlined plot theme and plot annotations for 'ggplot2'. Journal of Open Source Software* (Vol. 4, p. 1686). Springer-Verlag New York. https://doi.org/10.21105/joss.01686

Zhu, H. (2020). *kableExtra: Construct complex table with 'kable' and pipe syntax. Journal of Open Source Software* (Vol. 4, p. 1686). https://doi.org/10.21105/joss.01686

Appendix

Table A1

Settings customizable by researchers when creating new studies to be run on the Web-CDI platform.

Study setting	**Default value**	**Notes**
Study name	none	–
Instrument	none	–
Age range for study	none	Defaults based on instrument selected.
Number of days before study expiration	14	Must be between 1 and 28 days.
Measurement units for birth weight	Pounds and ounces	Weight can also be measured in kilograms (kg).
Minimum time (minutes) a parent must take to complete the study	6	–
Waiver of documentation	blank	Can be filled in by researchers to include a Waiver of Documentation for the participant to approve before proceeding to the experiment.
Pre-fill data for longitudinal participants?	No, do not populate any part of the form	Researchers can choose to pre-fill the background information and the vocabulary checklist.

Table A1

Settings customizable by researchers when creating new studies to be run on the Web-CDI platform. (continued)

Study setting	**Default value**	**Notes**
Would you like to pay subjects in the form of Amazon gift cards?	No	If checked, researchers can enter gift codes to distribute to participants once they have completed the survey.
Do you plan on collecting only anonymous data in this study? (e.g., posting ads on social media, mass emails, etc)	No	If checked, researchers can set a limit for the maximum number of participants, as well as select an option that asks participants to verify that the information entered is accurate.
Would you like to show participants graphs of their data after completion?	Yes	–
Would you like participants to be able to share their Web-CDI results via Facebook?	No	–
Would you like participants to answer the confirmation questions?	No	Asks redundant demographic questions to serve as attention checks.

Table A1

Settings customizable by researchers when creating new studies to be run on the Web-CDI platform. (continued)

Study setting	**Default value**	**Notes**
Provide redirect button at completion of study?	No	Used to redirect users to external site after form completion.
Capture the Prolific Id for the participant?	No	For integration with Prolific.
Allow participant to print their responses at end of Study?	No	–
End message	Standard end-of-study message	Can be changed to customize end-of-study message.

Table A2

Regression output for WG comprehension measure.

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df
Intercept	122.275	2.427	50.381	0.000	117.515	127.035	1610
Age	20.050	0.767	26.127	0.000	18.545	21.556	1610
Caregiver education: Some college	17.445	8.179	2.133	0.033	1.403	33.487	1610
Caregiver education: High school or less	21.862	10.935	1.999	0.046	0.413	43.311	1610
Age * Caregiver education: Some college	-1.991	2.261	-0.881	0.379	-6.425	2.443	1610
Age * Caregiver education: High school or less	-6.604	3.159	-2.091	0.037	-12.800	-0.408	1610

Table A3

Regression output for WG production measure.

term	estimate	std.error	statistic	p.value	conf.low	conf.high	df
Intercept	29.771	1.332	22.358	0.000	27.159	32.382	1610
Age	7.599	0.498	15.264	0.000	6.622	8.575	1610
Caregiver education: Some college	5.640	4.919	1.147	0.252	-4.009	15.289	1610
Caregiver education: High school or less	20.455	7.693	2.659	0.008	5.366	35.545	1610
Age * Caregiver education: Some college	-1.357	1.327	-1.022	0.307	-3.960	1.247	1610
Age * Caregiver education: High school or less	-0.121	2.095	-0.058	0.954	-4.229	3.988	1610

Table A4

Minimum times to completion, WG measure

Age in months	Minimum time to completion (minutes)
8	3.496
9	4.057
10	4.619
11	5.181
12	5.743
13	6.305
14	6.867
15	7.429
16	7.991
17	8.553
18	9.115

Table A5

Minimum times to completion, WG measure

Age in months	Minimum time to completion (minutes)
16	8.129
17	8.613
18	9.097
19	9.581
20	10.065
21	10.55
22	11.034
23	11.518
24	12.002
25	12.486
26	12.97
27	13.455
28	13.939
29	14.423
30	14.907