CREDI

Caregiver-Reported Early Development Instruments

Data Management & Scoring Manual

June 8th, 2018

Dana Charles McCoy, PhD
Harvard Graduate School of Education

Günther Fink, PhD
Swiss Tropical and Public Health Institute

Marcus Waldman, MS
Harvard Graduate School of Education

INTRODUCTION

The **Caregiver Reported Early Development Instrument (CREDI)** was designed to serve as a population-level measure of early childhood development (ECD) for children from birth to age three. As the name suggests, the CREDI exclusively relies on caregiver reports, and thus primarily focuses on milestones and behaviors that are easy for caregivers to understand, observe, and describe.

This **Data Management & Scoring Manual** was developed to help users to calculate overall CREDI scores based on individual item responses. We recommend that users review this document in full to understand how to prepare data for score calculations, how to derive scores, and how to interpret scores.

As we detail in the *Interpretation* sections below, there are several considerations to keep in mind when using the CREDI scores:

1. The **CREDI Short Form** produces only one overall score that summarizes overall developmental status based on skills from multiple domains. The Short Form does not include domain-specific scores.

2. The **CREDI Long Form**, on the other hand, includes both an overall developmental score, as well as scores for each of the four domains: motor, cognitive, language, and social-emotional. (A scoring system for mental health items is pending.) The Long Form also produces both **norm-referenced standardized scores** (whose units can be interpreted as z-scores, but which are less appropriate for hypothesis testing) and **raw scaled scores** (whose units are not easily interpretable, but which are more appropriate for hypothesis testing).

3. Because of the way that the scoring system was developed, CREDI raw scaled scores should *not* be **compared** against one another either (A) across the Short and Long Form, or (B) across domain scores within the Long Form. In other words, if a child scores a 57 in the motor domain and a 55 in the cognitive domain, it does not necessarily mean that the child is more developed in the motor than cognitive skills. If you want to compare differences across domains within the Long Form, you can use norm-referenced standardized scores – they tell you how each domain-specific score compares to the CREDI reference population for a particular age.

Sincerely,

The CREDI Team

SHORT FORM SCORING INSTRUCTIONS

To score the CREDI Short Form, please see the *CREDI Short Form Scoring Table* provided on the CREDI website. This document provides instructions on how to clean CREDI data, calculate total scores, and determine age-specific standardized scores.

The CREDI Short Form currently produces only raw scaled scores. These scores can be interpreted in the same way as CREDI Long Form's raw scaled scores. Details on interpreting raw scaled scores are provided in the section below called *Interpreting CREDI Long Form Scores*.

LONG FORM SCORING INSTRUCTIONS

As noted above, the CREDI Long Form produces an overall developmental score, as well as scores for each developmental domain: motor, cognitive, language, and social-emotional. (A scoring system for mental health is pending.)

In this section, we show you how to score the CREDI Long Form in the software program R via the *credi* package. We have set up the R scoring software such that prior knowledge of either R or R studio is *not* necessary. In order to score your collected data, you will need to complete three basic steps:

1) **Clean your data** and save it as a .csv file
2) **Install** the *latest version* of R software (which is free and can be downloaded as described above)
3) **Run** the *credi* package in R

STEP 1: CLEAN YOUR DATA

The software package developed to score the CREDI requires your data to be organized and cleaned prior to scoring. You can clean your data in whatever software program you prefer (e.g., Excel, Stata, SPSS, R, etc.).

In order to be readable for the package, your clean data file must meet the following **five criteria**:

1) There must be an identifying variable labeled **ID**. The ID variable must be unique for each observation/row. There cannot be missing ID values. If you have data with duplicate IDs, you will need to either recode observations, or drop duplicates. Observations without identifiers should be dropped.
2) There must be an **AGE** variable with only numeric values that indicates the child's age in months. Observations with missing AGE values will not necessarily result in an error; however, scores *will not* be calculated for these children. Please note that the *credi* package will only create scores for children under the age of 3 years (36 months).
3) The remaining variables must be responses to CREDI items that are appropriately named. The naming convention for the CREDI has evolved over the years, so the *credi* program is flexible enough to allow you to specify items using item names from Pilot 4, the Long Form version 20 Apr 2017, or the Long Form version 29 January 2018. Appendix A of this document shows the conversion between the three different versions. Variables provided to the program with names *other than* ID, AGE, and one of the three sets of item names listed in Appendix A *will result in an error*.

4) The *credi* program assumes that a "yes" response on an item is represented by a numeric value of 1, while a "no" response is represented by a numeric value of 0. Numeric values other than 0 or 1 are treated as missing values. Please note that the software does not allow for character or string values for any of the questions. If you've coded "yes"/"no" values for the item responses, or you specified characters codes to differentiate different reasons for missingness (e.g., ".r", ".d", etc.), you'll need to convert all values to 0 or 1, and the missing data codes to some other numeric value (e.g., -9, -99, etc.) This can be done quickly using the "Find→Replace" feature in Excel, or through simple "replace" commands in other software packages.

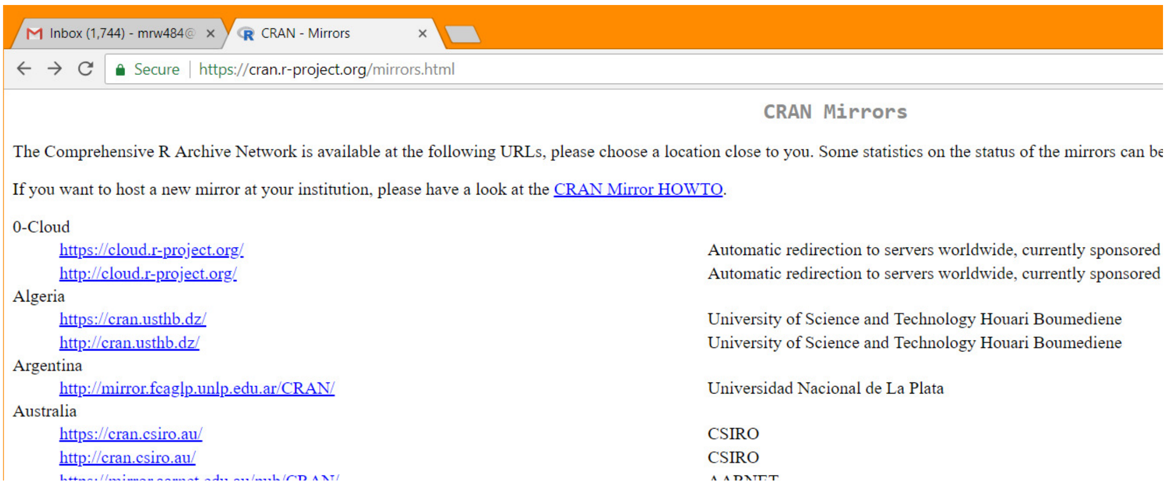
5) The cleaned data meeting requirements (1)-(4) must be saved as a ".csv" file. This can be done by using the "Save As" command in Excel, or using export functions in Stata, SAS, or SPSS.

STEP 2: INSTALL R

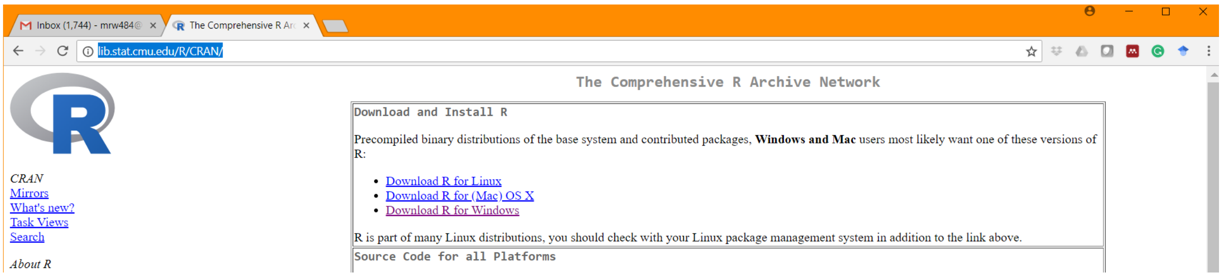
The *credi* package requires the latest version of R (R 3.5.0). R is a free software package that can be installed easily on both Windows and Mac computers, as described in further detail below.

Installing R in Windows

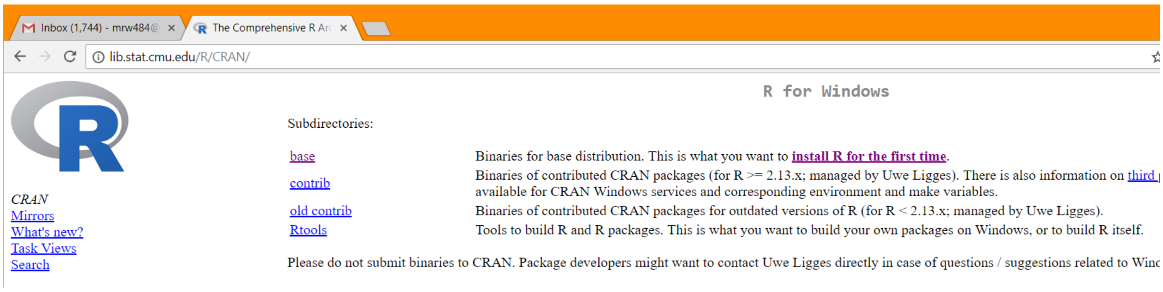
To install R in Windows, go to https://cran.r-project.org/mirrors.html. This will bring you to a website that looks like this:



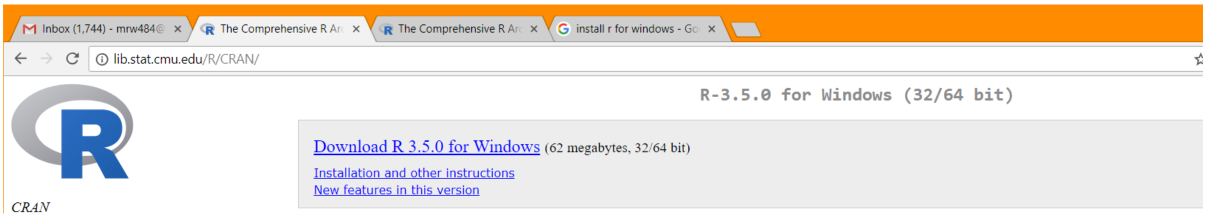
Click on one of the links. We suggest choosing a link that is in your country and closest to you geographically to minimize the time it takes to download R. For the example below, we used the "mirror" provided Carnegie Mellon University, which took us to a page that looked like this:



Click on the link that says "Download R for Windows." This will take you to a website that looks like this:



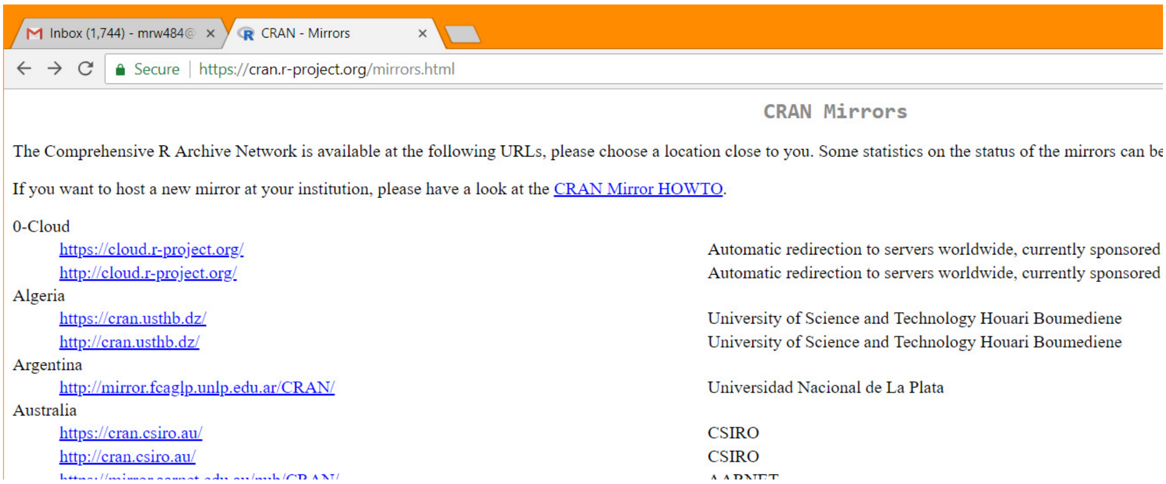
Next, click "install R for the first time," which will bring you (finally) to a link that will allow you to download R 3.5.0:



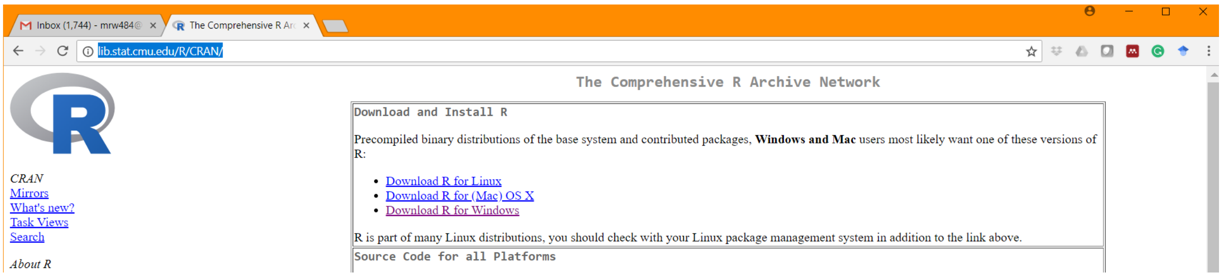
Click on the link to download the .exe file. It's approximately 80MB, so it shouldn't take too long to download. Once it is finished downloading, go to your downloads folder and run the executable ("R-3.5.0-win.exe"). Follow the instructions to install R.

Installing R on a Mac

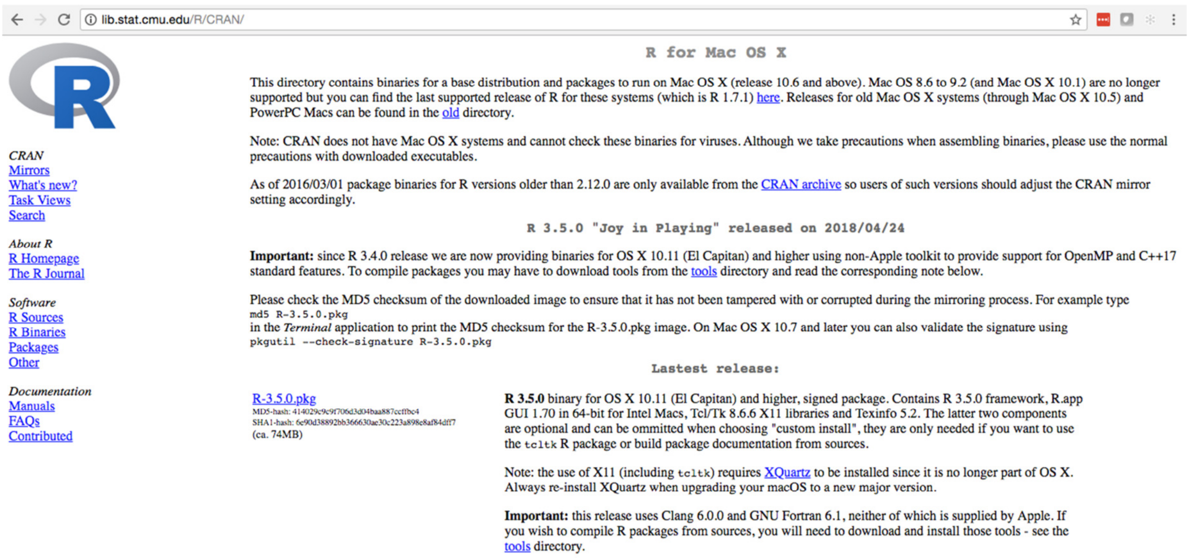
Go to https://cran.r-project.org/mirrors.html. This will bring you to a website that looks like this:



Next, click on one of the links. We suggest choosing a link that is in your country and closest to you geographically in order to minimize the time it takes to download R. For this example, we used the [“mirror” provided Carnegie Mellon University] and that took us to a page that looked like this:



Next, click on the link to “Download R for (Mac) OS X,” which will take you to this page:



Click on the link to download the latest release, which in this case is R-3.5.0.pkg. The package is less than 80MB, so it shouldn't take too long to download. Once it is finished downloading, run the executable ("R-3.5.0.pkg"). Follow the instructions to install R.

Note: Mac users with an older operating system will not be able to download the latest version of R. Please update your operating system before downloading R-3.5.0.

STEP 3: RUN THE *CREDI* PAGAGE IN R

Find the button for "R x64 3.5.0" (if running a 64 bit system) or "R i386 3.5.0" (if running a 32 bit system) in your applications folder to open the R program. If you don't know whether you are running a 64 bit or 32 bit system, you are most likely running a 64 bit system. You should see the R console (a pop-up window that will ask you to enter a command) once R is open.

Type or copy-paste the following code in the console and hit enter. (You likely will then be asked to select a CRAN mirror. As you did when you installed R, you should select one that is close by you geographically):

```
install.packages("devtools")
```

This next command will install the devtools package on your computer. After the installation is complete, load the devtools package by typing the following code and hitting enter:

```
library("devtools")
```

With the devtools package loaded, you can then install the *credi* package to conduct the scoring. Enter the following command:

```
install_github("marcus-waldman/credi")
```

After the installation is complete, load the *credi* package:

```
library("credi")
```

Finally, you can score the .csv file by running one of the two lines below:

```
X=credi::score(reverse_code = FALSE)
```

-or-

```
X=credi::score(reverse_code = TRUE)
```

Note: You should select reverse_code = FALSE only if you do *not* want the *credi* program to automatically reverse code the negatively worded items indicated by the last column of Appendix A. If you have not yet reverse coded relevant items from Appendix A (i.e., all items are reported exactly as they were collected), you should select reverse_code = TRUE.

You will then be asked to tell the software package where the .csv file with your data is saved. The *credi* package will then review your data file.

<u>If there are errors in your data</u>, you will be asked if you would like to save a log file of warning and error messages. We recommend that you say yes (by typing Y and saving the log file to a convenient folder), as this will help you to identify the source of your problems.

<u>If there are no errors in your data</u>, scoring will begin. You will be asked to save the scores as a .csv file once the scoring is finished. You should then be able to open the .csv in Excel or a similar program to view your scores for each ID number.

INTERPRETING CREDI LONG FORM SCORES

The .csv file with CREDI Long Form scores includes the following variables:

ID	Your original ID variable
z_MOT	A norm-referenced standardized score (z-score) for the motor domain
z_COG	A norm-referenced standardized score (z-score) for the cognitive domain
z_LANG	A norm-referenced standardized score (z-score) for the language domain
z_SEM	A norm-referenced standardized score (z-score) for the soc-emot. domain
z_OVERALL	A norm-referenced standardized score (z-score) for overall development
MOT	A raw scaled (factor) score for the motor domain
COG	A raw scaled (factor) score for the cognitive domain
LANG	A raw scaled (factor) score for the language domain
SEM	A raw scaled (factor) score for the social-emotional domain
OVERALL	A raw scaled (factor) score for overall development
MOT_SE	The standard error of measurement of the score for the motor domain
COG_SE	The standard error of measurement of the score for the cognitive domain
LANG_SE	The standard error of measurement of the score for the language domain
SEM_SE	The standard error of measurement of the score for the soc.-emot. domain
OVERALL_SE	The standard error of measurement of the score for overall development

Norm-referenced standardized scores are marked with **z_**. These scores were constructed by comparing the raw score in each domain to the average raw score in our CREDI reference population of a particular age[1]. Specifically, we subtract the average raw score of children of the same age from the observed raw score and then divide the difference by the age-specific standard deviation. A z-score of 0 thus means that the child has exactly the same score on that particular domain as the average same-age child in the CREDI reference population. A score of "-1" means that the child's raw score is 1 standard deviation below the same-age average of the reference population. These scores are useful for generating effect size metrics, as each "unit" can be interpreted as a standard deviation.

Preliminary analysis suggests that the norm-referenced scores are generally normally distributed. Thus, a norm-referenced standardized score between +/-2 roughly translates to a region where 95% of the scores *from the same-age reference group* are located. However, norm-referenced scores for children under 6 months old do not appear to be normally distributed; we observe that substantially *more than* 5% of the reference group scores are located outside of the +/-2 interval. In summary, the norm-reference scores behave as typical "z-scores" for children older than 6

[1] The CREDI reference population comprises all children in the original CREDI data base with an "ideal" home environment. Ideal home environments were defined through maternal educational attainment (college or higher), as well as through the number of activities done by adults with the child in the last three days (at least 4 out of the 6 MICS home stimulation activities).

months. We recommend interpreting the norm-referenced scores for children under 6 months with caution.

Raw scaled scores (also called [MAP] factor scores, ability estimates, or person location estimates) do not include a prefix or suffix in their variable name. The units for these scores are specific to the CREDI, and do not correspond to any other known metric (e.g., standard deviations, IQ points, etc.). Unlike the norm-referenced standardized scores, average raw scaled scores increase with age, reflecting developmental progressions.

It is almost always more appropriate to use the (raw) scaled scores for hypothesis testing (i.e., regression analysis, ANOVA, etc.) as compared to the z-scores. This is because hypothesis tests require that variables exhibit interval properties of measurement. An interval scale is one where the difference between values is meaningful. A practical example of an interval scale is temperature—the difference between 5 and 8 degrees Celsius is the same as the difference between 17 and 20 degrees Celsius. Although we are confident that the raw scaled scores demonstrate an interval scale, the transformation required to generate the z-scores described above almost certainly results in a non-interval scale. These z-scores are therefore less trustworthy for hypothesis testing.

Standard error of measurement (or conditional standard error of measurement) are marked with _SE. These variables refer to the precision of the raw scaled score estimates described above and are useful to apply weighting when conducting statistical tests comparing scores across groups (see example below). The smaller the standard error of measurement, the more precise the estimate of the child's score. In item response theory, "extreme" scores at the tail of the distribution of scores (i.e., a score below 40 or above 55) are generally less precise than scores near the mean of the distribution of scores (i.e., a score around 45-55). Missing responses to age-appropriate items can also decrease precision.

Example of Analyzing and Interpreting CREDI Long Form Scores

As a pedagogical exercise, consider the following example research question:

[RQ1] Are there mean differences in CREDI overall scores between girls and boys aged 12-17 months in Brazil?

To answer this research question, we will use a sample of CREDI responses from caregivers in Brazil ($N = 692$), with females being slightly over-represented. All children are between 12-17 months at the time that the CREDI was administered.

Sex of child	Freq.	Percent	Cum.
Male	322	46.53	46.53
Female	370	53.47	100.00
Total	692	100.00	

Scores could not be computed for *n* = 2 of the children using the *credi* R package. The log file indicates that this occurred because these children did not answer at least 5 item responses (text in box below). The *credi* package did not compute scores for these children because the scores would be too imprecise to be useful. We choose to employ listwise deletion and discard these observations to keep this example simple, resulting in a total sample size of N = 690. Although we used listwise deletion here, we encourage users to adopt more sophisticated missing data strategies, as appropriate.

```
* Warning:  The following 2 observation(s) contain less than 5 non-missing item responses and will not be scored:

  ID = 7221, 7232

* Warning:  A total of 2 (0.3%) observation(s) cannot be scored for the following reason(s):

Reason                        Number      Percent

1 Less than 5 item responses  2           0.3%
```

To investigate mean differences in the overall CREDI Long Form developmental score by gender, we select the raw scaled OVERALL score instead of the norm-referenced standardized Z_OVERALL score. We do so based on the above suggestion to use raw scaled scores for hypothesis testing, given that these raw scaled scores are on an interval scale. We conduct all analyses in Stata.

The *credi* scoring program supplies not only an OVERALL score for each child, but also the conditional standard error of measurement (OVERALL_SE). Point estimates are known to be more precise if observations are weighted by the inverse of the squared standard error of measurement. Intuitively this makes sense: scores that are less precise should not inform the results as much as scores that are more precise. The code we used to create a weighting variable is:

```
gen wgt = 1/OVERALL_SE^2
```

To answer our research question, we are interested in calculating descriptive statistics regarding the OVERALL scores in the sample, by gender. We obtain the descriptives using the tabstat function in Stata. Note that inverse weighting is accomplished by using the [aweight = wgt] option.

```
tabstat OVERALL [aweight = wgt], by(SEX) statistics(mean SD N) column(statistics)
```

SEX	mean	sd	N
Male	49.14518	1.309539	320
Female	49.44811	1.260939	370
Total	49.30463	1.292144	690

According to the table above, caregivers with female children report more positive OVERALL scores, as there is a 0.30 (= 49.45 – 49.15) point difference in the raw scaled scores. Although this finding is difficult to interpret without additional information, *hypothesis tests should be conducted with the raw scores.*

We used one-way ANOVA to test whether the 0.30 point difference can be explained by sampling error. We find that the F-test is statistically significant (p = .002), providing evidence that the difference in OVERALL scores is likely not the result of sampling error.

```
anova OVERALL SEX [aweight = wgt]
```

```
Number of obs =        690     R-squared     =  0.0137
Root MSE      =    1.28418     Adj R-squared =  0.0123
```

Source	Partial SS	df	MS	F	Prob>F
Model	15.786528	1	15.786528	9.57	0.0021
SEX	15.786528	1	15.786528	9.57	0.0021
Residual	1134.5927	688	1.6491173		
Total	1150.3793	689	1.6696361		

Although the norm-referenced standardized scores are not useful for statistical tests, they are useful for calculating effect sizes to interpret the difference. This is because the z_OVERALL score is in standard deviation units (specifically, the standard deviation of the reference population), making this a Cohen's □ statistic. Once again, we use tabstat to identify differences in mean z_OVERALL scores by gender.

```
tabstat z_OVERALL [aweight = wgt], by(SEX) statistics(mean) column(statistics)
```

SEX	mean
Male	.2135692
Female	.5284169
Total	.3792885

We find that there is a Cohen's □= 0.32 SD (= 0.53-0.21) difference in score between females and males.

Overall, these results can be summarized as: The data suggests that females exhibit more positive overall scaled scores compared to males (Est. = 0.30 pts, F(1,688) = 9.57, p = 0.002), with the difference being moderate-to-strong in size (Cohen's d = 0.32).

FREQUENTLY ASKED QUESTIONS

Q: I get the following statement. What should I do?

```
[1] "*Error: Processing the provided response data resulted in errors. See log for more details."

Would you like to save a log file of warning and error messages? [Y/N]:
```

A: This occurred because one of the 5 requirements of the data from Step 1 was not met. Type in Y and hit enter, as this will prompt you to save a .txt file (i.e., a log), which you can then open and read to identify the cause of the error.

Q: Can I compare raw scores across domains?

A: It is *not possible* to directly compare raw scores across domains. The four domains in CREDI have different age-specific means and standard deviations, which means that a score of 57 in the motor domain and a score of 55 in the cognitive domain does not necessarily mean that the child is more developed in the motor domain than in the cognitive domain. If you want to compare differences across domains, you can use z-scores – they tell you how each domain-specific score compares to the reference population.

Q: Are the overall scores the same as the average of the domain scores?

A: No. The overall score represents a single developmental score, which captures the overall development of the child. This overall score will be highly correlated with all four domains, but is computed based on a model combining all items, and cannot be directly computed from the domain-specific scores.

FOR MORE INFORMATION

A. ACADEMIC PAPERS

McCoy, D. C., Waldman, M., CREDI Field Team, & Fink, G. (in press). Measuring early childhood development at a global scale: Evidence from the Caregiver-Reported Early Development Instruments. *Early Childhood Research Quarterly.*

McCoy, D. C., Sudfeld, C., Bellinger, D. C., Muhihi, A., Ashery, G., Weary, T. E., Fawzi, W., & Fink, G. (2017). Development and validation of an early childhood development scale for use in low-resourced settings, *Population Health Metrics, 15*(3).

McCoy, D. C., Black, M., Daelmans, B., & Dua, T. (2016). *Measuring population-level development in 0-3. Early childhood matters.* The Hague, Netherlands: Bernard van Leer Foundation.

B. WEBSITE

sites.sph.harvard.edu/credi/

A. CONTACT INFORMATION

Dana Charles McCoy, PhD
Assistant Professor
Harvard Graduate School of Education
dana_mccoy@gse.harvard.edu

Günther Fink, PhD
Head of Research Unit
Swiss Tropical and Public Health Institute
Associate Professor
University of Basel
guenther.fink@swisstph.ch

Marcus Waldman, MS
Ed.D. Candidate
Harvard Graduate School of Education
marcus_waldman@gse.harvard.edu

APPENDIX A

Table 1: CREDI naming conventions allowed in the *credi* R package

Pilot 4	Long Form 20 Apr 2017	Long Form 29 Jan 2018	Reverse Coded Item (i.e. negatively worded)
QM08	CREDI_LM01	LF1	FALSE
QM07	CREDI_LM02	LF2	FALSE
QC06	CREDI_LC01	LF3	FALSE
QS01	CREDI_LS02	LF4	FALSE
QS04	CREDI_LS01	LF5	FALSE
QM06	CREDI_LM03	LF6	FALSE
QM03	CREDI_LM04	LF7	FALSE
QC14	CREDI_LC02	LF8	FALSE
QM01	CREDI_LM05	LF9	TRUE
QC04	CREDI_LC03	LF10	FALSE
QM10	CREDI_LM06	LF11	FALSE
QC05	CREDI_LC04	LF12	FALSE
QS17	CREDI_LS03	LF13	FALSE
QM12	CREDI_LM08	LF14	FALSE
QC08	CREDI_LC06	LF15	FALSE
QM09	CREDI_LM09	LF16	FALSE
QC09	CREDI_LC05	LF17	FALSE
QM05	CREDI_LM07	LF18	FALSE
QC07	CREDI_LC08	LF19	FALSE
QM11	CREDI_LM10	LF20	FALSE
QC11	CREDI_LC07	LF21	FALSE
QC10	CREDI_LC11	LF22	FALSE
QC13	CREDI_LC10	LF23	FALSE
QC12	CREDI_LC09	LF24	FALSE
QM14	CREDI_LM12	LF25	FALSE
QM16	CREDI_LM11	LF26	FALSE
QM20	CREDI_LM14	LF27	FALSE
QM13	CREDI_LM13	LF28	FALSE
QM15	CREDI_LM15	LF29	FALSE
QC15	CREDI_LC12	LF30	FALSE
QM19	CREDI_LM16	LF31	FALSE
QM22	CREDI_LM17	LF32	FALSE
QM23	CREDI_LM18	LF33	FALSE
QM18	CREDI_LM19	LF34	FALSE
QC21	CREDI_LC13	LF35	FALSE
QM17	CREDI_LM20	LF36	FALSE
QC20	CREDI_LC14	LF37	FALSE
QS07	CREDI_LS04	LF38	FALSE

QS20	CREDI_LS06	LF39	FALSE
QM21	CREDI_LM21	LF40	FALSE
QM24	CREDI_LM22	LF41	FALSE
QS06	CREDI_LS05	LF42	FALSE
QC18	CREDI_LC15	LF43	FALSE
QM26	CREDI_LM23	LF44	FALSE
QC22	CREDI_LC16	LF45	FALSE
QS14	CREDI_LS07	LF46	FALSE
QM25	CREDI_LM24	LF47	FALSE
QC17	CREDI_LC17	LF48	FALSE
QM27	CREDI_LM25	LF49	FALSE
QC19	CREDI_LC18	LF50	FALSE
QM30	CREDI_LM27	LF51	FALSE
QM28	CREDI_LM26	LF52	FALSE
QC16	CREDI_LC19	LF53	FALSE
QM29	CREDI_LM28	LF54	FALSE
QS25	CREDI_LS08	LF55	FALSE
QC26	CREDI_LC20	LF57	FALSE
QC23	CREDI_LC21	LF58	FALSE
QM31	CREDI_LM29	LF60	FALSE
QS21	CREDI_LS12	LF61	FALSE
QC25	CREDI_LC23	LF62	FALSE
QC28	CREDI_LC24	LF64	FALSE
QM37	CREDI_LM30	LF67	FALSE
QS24	CREDI_LS11	LF68	FALSE
QM32	CREDI_LM31	LF69	FALSE
QC34	CREDI_LC26	LF72	FALSE
QC24	CREDI_LC28	LF73	FALSE
QC33	CREDI_LC25	LF74	FALSE
QS19	CREDI_LS18	LF75	FALSE
QC27	CREDI_LC29	LF76	FALSE
QM35	CREDI_LM32	LF77	FALSE
QM34	CREDI_LM33	LF82	FALSE
QC29	CREDI_LC32	LF83	FALSE
QM33	CREDI_LM34	LF86	FALSE
QC31	CREDI_LC34	LF88	FALSE
QC35	CREDI_LC37	LF92	FALSE
QM36	CREDI_LM35	LF93	FALSE
QM38	CREDI_LM36	LF98	FALSE
QM39	CREDI_LM37	LF105	FALSE
QS09	CREDI_LMH01	LFMH1	TRUE
QS23	CREDI_LMH02	LFMH2	TRUE
QS16	CREDI_LMH03	LFMH3	TRUE
QS11	CREDI_LMH04	LFMH4	TRUE

QS18	CREDI_LMH07	LFMH7	TRUE
QC37	-	LF81	FALSE
QC50	-	LF95	FALSE
QS43	-	LF56	FALSE
QC36	-	LF84	FALSE
QS46	-	LF85	FALSE
QS45	-	LF59	FALSE
QC42	-	LF87	FALSE
QS50	-	LF101	FALSE
QC46	-	LF89	FALSE
QC41	-	LF90	FALSE
QS58	-	LF91	FALSE
QC40	-	LF65	FALSE
QS44	-	LF66	FALSE
QC44	-	LF94	FALSE
QC47	-	LF108	FALSE
QC43	-	LF100	FALSE
QC39	-	LF97	FALSE
QC38	-	LF71	FALSE
QS30	-	LF63	FALSE
QS51	-	LF104	FALSE
QS29	-	LFMH6	FALSE
QS53	-	LF102	TRUE
QC45	-	LF107	FALSE
QS55	-	LFMH8	TRUE
QS41	-	LF78	FALSE
QS38	-	LF79	FALSE
QC49	-	LF96	FALSE
QS35	-	LF80	FALSE
QS33	-	LFMH5	TRUE
QS36	-	LF70	FALSE
QS57	-	LFMH9	TRUE
QC51	-	LF99	FALSE
QC48	-	LF106	FALSE
QS52	-	LF103	FALSE