Using EdSurvey to Analyze ECLS-K:2011 Data

Michael Lee, Ting Zhang, and Paul Bailey*†

October 09, 2020

Overview of the EdSurvey Package

Data from large-scale educational assessment programs, such as the Early Childhood Longitudinal Study, Kindergarten Class of 2010-11 (ECLS-K:2011), require special statistical methods to analyze. Because of their scope and complexity, the `EdSurvey` package gives users functions to perform analyses that account for complex sample survey designs.

The `EdSurvey` package also seamlessly takes advantage of the `LaF` package to read in data only when required for an analysis. Users with computers that have insufficient memory to read in entire datasets can still do analyses without having to write special code to read in just the appropriate variables. This is all addressed directly in the `EdSurvey` package—behind the scenes and without any special tuning by the user.

Vignette Outline

This vignette will describe the basics of using the `EdSurvey` package for analyzing Early Childhood Longitudinal Study data using ECLS-K:2011 data as an example. The vignette sections are organized as follows:

- Notes
  - Vignette notation
  - Software requirements
- Setting up the environment for data analysis
  - Installing and loading `EdSurvey`
  - Philosophy of conducting analyses using the `EdSurvey` package
  - Downloading data
  - Reading in data
  - Getting to know the data format
    - Retrieving survey weights
    - Retrieving stratum and PSU variables
  - Recoding data
  - Removing special values
- Explore variable distributions with `summary2`
- Retrieving data for further manipulation with `getData`

*This publication was prepared for NCES under Contract No. ED-IES-12-D-0002 with the American Institutes for Research. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

†The authors would like to thank Howard Huo, Cameron McPhee, Shannon Russell, and Yuan Zhang for reviewing this document.

Notes

Vignette Notation

This vignette displays examples using notation for R console input and output. R console input will be displayed in a gray box:

```
inputCode <- c(2,"neat")
```

R console output will be displayed next to a double hash mark (`##`). Here is an example where the user types `inputCode` into the console and the code output R gives after the double hash marks:

```
inputCode
```

```
## [1] "2"    "neat"
```

Software Requirements

Unless you already have R version 3.5.0 or later, install the latest R version—which is available online at https://cran.r-project.org/. Users also may want to install RStudio desktop, which has an interface that many find easier to follow. RStudio is available online at https://www.rstudio.com/products/rstudio/download/.

Analyses of the ECLS-K:2011 Kindergarten–**Fourth** Grade data files requires `EdSurvey` version 2.3.2, while Kindergarten–**Fifth** Grade data requires `EdSurvey` version 2.6.0.

Setting Up the Environment for Data Analysis

Installing and Loading EdSurvey

Inside R, run the following command to install `EdSurvey` as well as its package dependencies:

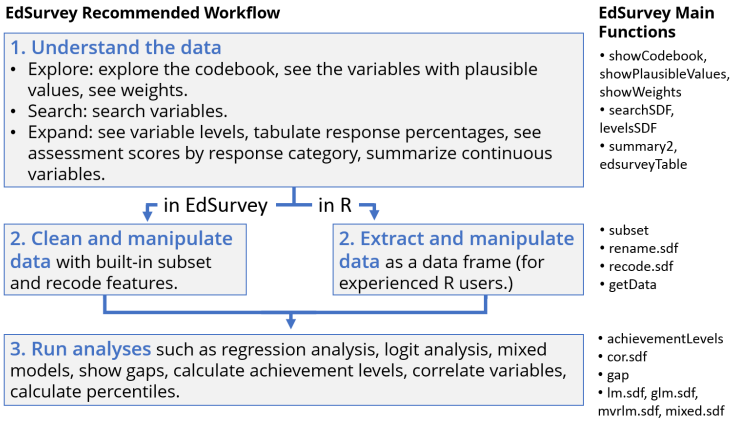
```
install.packages("EdSurvey")
```

Once the package is successfully installed, `EdSurvey` can be loaded with the following command:

```
library(EdSurvey)
```

Philosophy of Conducting Analyses Using the EdSurvey Package

Recognizing that researchers usjng R statistical software come with varying levels of experience, the `EdSurvey` package has provided multiple workflows to aid in this process of conducting survey analysis. The following graphic details the two recommended workflows:



The workflow has three phases:

1. Understanding the data
2. Preparing the data for analysis
3. Running the analysis

The first phase of the workflow directs users to better understand survey data by exploring, searching, and summarizing variables of interest. The second phase diverges into two approaches depending on user preference. While the `EdSurvey` package provides functions for users to clean and manipulate their data experienced R programmers might prefer to extract and manipulate their data using other R methods or supplementary packages to do so; each method is supported for use with `EdSurvey` analytical functions in phase three.

Downloading Data

If you do not have data ready in your computer, you can use `EdSurvey`'s download functions to acquire the survey data of interest, including the following:

- ECLS-K: Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 and Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (`downloadECLS_K`)
- ELS: Education Longitudinal Study (`downloadELS`)
- HSLS: High School Longitudinal Study (`downloadHSLS`)
- TIMSS: Trends in International Mathematics and Science Study and TIMSS Advanced (`downloadTIMSS`, `downloadTIMSSAdv`)
- PIRLS: Progress in International Reading Literacy Study (`downloadPIRLS`)
- ePIRLS: Electronic Progress in International Reading Literacy Study (`download_ePIRLS`)
- CIVED: The Civic Education Study 1999 and International Civic and Citizenship Study (`downloadCivEDICCS`)
- ICILS: International Computer and Information Literacy Study (`downloadICILS`)
- PISA: The Programme for International Student Assessment (`downloadPISA`)
- PIAAC: Programme for the International Assessment of Adult Competencies (`downloadPIAAC`)
- TALIS: Teaching and Learning International Survey (`downloadTALIS`)

For ECLS-K:2011, use the `downloadECLS_K` function to download data to a directory that the user specifies; for example, `"C:/Data"`. Please note that the `downloadECLS_K` function works for both 1998 and 2011 data and defaults to the most current public-use data file. One also can manually download desirable ECLS data (such as previous public-use data files) from the ECLS page of the NCES Data Products site.

The following example shows how to download the ECLS-K:2011 Kindergarten-Fifth Grade public-use data file:

```
downloadECLS_K(years = 2011, root = "C:/", cache=FALSE)
```

Given the size of the survey data it often takes quite a long time to download. Setting `cache = TRUE` will cache the text version of files and save time for future use of the data, but the process can take several hours, so we leave it at the default level of `FALSE`.

The data will be automatically stored in a folder in the directory that you specified. For example, the ECLS-K:2011 data will be saved in the "ECLS_K/2011" folder in the C drive. The R program assigns the folder name, but you can manually change it.

On the Mac OS, the user will need to set the root to a folder that exists. Here we use the user's home directory:

```
# for Mac OS
downloadECLS_K(years = 2011, root = "~\")
```

Then, on the Mac OS, future calls to `readECLS_K2011` should have `"C:/ECLS_K/2011"` replaced with `"~/ECLS_K/2011"`.

Reading in Data

The next step to running an analysis is reading in the data. For each study, there is a `read` function to assist in reading and processing its data. Once the data have been downloaded for your system, the read family of functions will open a connection to the specified data file to conduct your analysis. The read functions are as follows:

- ECLS-K: Early Childhood Longitudinal Study, Kindergarten Class of 1998–99 (`readECLS_K1998`)
- ECLS-K: Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (`readECLS_K2011`)
- ECLS-B: Early Childhood Longitudinal Study, Birth Cohort (`readECLS_B`)
- ELS: Education Longitudinal Study (`readELS`)

- HSLS: High School Longitudinal Study (`readHSLS`)
- TIMSS: Trends in International Mathematics and Science Study and TIMSS Advanced (`readTIMSS`, `readTIMSSAdv`)
- PIRLS: Progress in International Reading Literacy Study (`readPIRLS`)
- ePIRLS: Electronic Progress in International Reading Literacy Study (`read_ePIRLS`)
- CIVED: The Civic Education Study 1999 and International Civic and Citizenship Study (`readCivEDICCS`)
- ICILS: International Computer and Information Literacy Study (`readICILS`)
- PISA: The Programme for International Student Assessment (`readPISA`)
- PIAAC: Programme for the International Assessment of Adult Competencies (`readPIAAC`)
- TALIS: Teaching and Learning International Survey (`readTALIS`)
- NAEP: National Assessment of Educational Progress (`readNAEP`)

This section shows how to read in ECLS-K:2011 data using `EdSurvey`'s `readECLS_K2011` function. The `readECLS_K2011` function creates an `edsurvey.data.frame` that stores information about the specific data files processed, including the location on disk and the file format/layout of those files, as well as information about the weights, achievement levels, omitted levels, and so forth for the user. An ECLS-K:2011 `edsurvey.data.frame` includes information for all variables at multiple data levels: student, teacher, school, and/or home.

To load the ECLS-K:2011 U.S. data for the fifth graders and create an `edsurvey.data.frame`, select the pathway to the ECLS-K:2011 data folder and assign it the name `eclsk11` with this call:

```
eclsk11 <- readECLS_K2011("C:/ECLS_K/2011")
```

The function may take several minutes to run the first time; subsequent calls to `readECLS_K2011` are stored on the user's drive for easy access and near instant retrieval. Once read in, data from the ECLS-K:2011 data set can be analyzed and merged after loading the data into the R working environment.

Getting to Know the Data Format

Information about an `edsurvey.data.frame` can be obtained in several ways. To get general data information, simply call `print` by typing the name of the `data.frame` object (i.e., `eclsk11`) in the console.

```
eclsk11
```

Some of the basic functions that work on a `data.frame`, such as `dim`, `nrow`, and `ncol`, also work on an `edsurvey.data.frame`.[1] They help check the dimensions of `eclsk11`.

```
dim(eclsk11)
```

```
## [1] 18174 26060
```

```
nrow(eclsk11)
```

```
## [1] 18174
```

[1] Use `?function` in the R console to view documentation on base R and `EdSurvey` package functions (e.g., `?readECLS_K2011` or `?lm.sdf`).

```
ncol(eclsk11)
```

```
## [1] 26060
```

The `colnames` function can be used to list all variable names in the data:

```
colnames(eclsk11)
```

To conduct a more powerful search of `eclsk11` data variables, use the `searchSDF` function, which returns variable names and labels from an `edsurvey.data.frame` based on a character string such as `"book"`.

```
searchSDF("audio books", eclsk11)
```

```
##   variableName                                    Labels
## 1    a4useaubk A4 C1K READ USING READ-ALONG/AUDIO BOOKS
## 2    a6useaubk A6 C1K READ USING READ-ALONG/AUDIO BOOKS
```

The levels and labels for each variable searched via `searchSDF()` also can be returned by setting `levels = TRUE`. Let's use one of book variables `"a4useaubk"` as an example:

```
searchSDF("a4useaubk", eclsk11, levels = TRUE)
```

```
## Variable: a4useaubk
## Label: A4 C1K READ USING READ-ALONG/AUDIO BOOKS
## Levels (Lowest level first):
##      1. 1: NEVER OR HARDLY EVER
##      2. 2: ONCE OR TWICE A MONTH
##      3. 3: ONCE OR TWICE A WEEK
##      4. 4: ALMOST EVERY DAY
##      -9. -9: NOT ASCERTAINED
```

We also can use `levelsSDF()`to return the levels and labels for a particular variable:

```
levelsSDF(varnames = "x_raceth_r", data = eclsk11)
```

```
## Levels for Variable 'x_raceth_r' (Lowest level first):
##     1. 1: WHITE, NON-HISPANIC (n=8488)
##     2. 2: BLACK/AFRICAN AMERICAN, NON-HISPANIC (n=2396)
##     3. 3: HISPANIC, RACE SPECIFIED (n=4207)
##     4. 4: HISPANIC, NO RACE SPECIFIED (n=385)
##     5. 5: ASIAN, NON-HISPANIC (n=1543)
##     6. 6: NATIVE HAWAIIAN/PACIFIC ISLANDER, NON-HISPANIC (n=117)
##     7. 7: AMERICAN INDIAN/ALASKA NATIVE, NON-HISPANIC (n=168)
##     8. 8: TWO OR MORE RACES, NON-HISPANIC (n=827)
##     -9. -9: NOT ASCERTAINED* (n=43)
##     NOTE: * indicates an omitted level.
```

Access a full codebook using `showCodebook()`, retrieving the variable names, variable labels, and value labels of a survey. This function pairs well with the `View()` function to more easily explore a dataset:

```
View(showCodebook(eclsk11))
```

Retrieving Survey Weights

The variables associated with survey weights can be seen from the `showWeights` functions, respectively, when the `verbose` argument is set to `TRUE`.

```
showWeights(data = eclsk11, verbose = TRUE)
```

Here we hide the (lengthy) results, but the user can easily see them by running the same code. Selection of survey weights is especially important for ECLS-K and ECLS-K:2011. Once selected, the survey weight is specified using the `weightVar` argument in `EdSurvey` analytical functions.

Please see the full names of the weights for each data release in Appendix A of this vignette. To learn more about the selection of sample weights for analyses using ECLS:K-2011 data consult the **Calculation and Use of Sample Weights** section of the *Public-Use Data File User's Manuals* for the respective public-use file of interest. The *ECLS-K:2011 Kindergarten-Fifth Grade User's Manual, Public Version* is relevant for the K5 data used in this vignette. Alternative releases can be found on the NCES site here: *ECLS-K:2011 Public-Use Data File User's Manuals*.

Retrieving Stratum and PSU Variables

The functions `getStratumVar` and `getPSUVar` return the default stratum variable name or a PSU variable associated with a weight variable. As ECLS-K:2011 doesn't have default weights users need to specify a weight to return it's associated psu/stratum variables. For example, the total student weight from the 9th round `weightVar = "w9c29p_9a0"` returns the following:

```
EdSurvey:::getStratumVar(data = eclsk11, weightVar = "w9c29p_9a0")
```

```
## [1] "w9c29p_9astr"
```

```
EdSurvey:::getPSUVar(data = eclsk11, weightVar = "w9c29p_9a0")
```

```
## [1] "w9c29p_9apsu"
```

These are particularly useful for accessing the variables associated with the weights in longitudinal surveys.

Recoding Data

Data recoding is of particular importance when performing analyses with ECLS:K-2011 data. By default, the `EdSurvey` package omits special values, such as multiple entries, skipped values, or `NA`s. Typically, this setting helps users by dropping the levels of factors that are not typically included in regressions, tables, correlations, and other analyses. For ECLS:K-2011, this default setting requires careful consideration. There are many instances in which we'd like to keep special values for our analyses; in these cases it's advised to recode your data.

Special codes are used to indicate item nonresponse, legitimate skips, and unit nonresponse in ECLS:K-2011.

Table 1: SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K: 2011), kindergarten-fifth grade (K-5) restricted-use data file.

Value	Description
-1	Not applicable, including legitimate skips
-2	Data suppressed (public-use data file only)
-4	Data suppressed due to administration error
-5	Item not asked in School Administrator Questionnaire form B
-7	Refused (a type of item nonresponse)
-8	Don't know (a type of item nonresponse)
-9	Not ascertained (a type of item nonresponse)
(blank)	System missing, including unit nonresponse

We'll detail the method for recoding these values later in this vignette in the section *Recoding Variables in a Dataset* section of the **Retrieving Data for Further Manipulation With getData** portion of this vignette as follows.

Removing Special Values

The `EdSurvey` package uses listwise deletion to remove special values in all analyses by default, such as those detailed in Table 1. To use a different method, set `omittedLevels = FALSE` when running your analysis. You can then remove levels that you want to remove with a call to `subset`, which is discussed in the "Subsetting the Data" section of this vignette.

Explore Variable Distributions With `summary2`

The `summary2` function produces weighted and unweighted descriptive statistics for a variable. This functionality is particularly useful for gathering response information for survey variables when conducting data exploration. By default estimates are not weighted. For example, the variable `x9povty_i` ("Imputed Poverty Level") returns the following output:

```
summary2(eclsk11, "x9povty_i")
```

```
## Estimates are not weighted.
##                                                                    x9povty_i    N
## 1                                                                  (Missing) 7954
## 2                                                1: BELOW POVERTY THRESHOLD 2185
## 3 2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD 2226
## 4                          3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD 5809
##    Percent
## 1 43.76582
## 2 12.02267
## 3 12.24827
## 4 31.96324
```

Note that by default, the `summary2` function includes omitted levels; to remove those, set `omittedLevels = TRUE`:

```
summary2(eclsk11, "x9povty_i", omittedLevels = TRUE)
```

```
## Estimates are not weighted.
##                                                                          x9povty_i    N
## 1                                                  1: BELOW POVERTY THRESHOLD 2185
## 2 2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD 2226
## 3                          3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD 5809
##    Percent
## 1 21.37965
## 2 21.78082
## 3 56.83953
```

The `summary2` function returns the weighted number of cases, the weighted percent, and the weighted standard error for a categorical variable when specified in the argument `weightVar`, here using the total student weight from the 9th round `weightVar = "w9c29p_9a0"`:

```
summary2(eclsk11, "x9povty_i", weightVar = "w9c29p_9a0")
```

```
## Warning in calcEdsurveyTable(formula, data, weightVar, jrrIMax,
## pctAggregationLevel, : Removing rows with 0 weight from analysis.
```

```
## Estimates are weighted using weight variable 'w9c29p_9a0'
##                                                                          x9povty_i    N
## 1                                                                        (Missing)    0
## 2                                                  1: BELOW POVERTY THRESHOLD 1720
## 3 2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD 1823
## 4                          3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD 4999
##   Weighted N Weighted Percent Weighted Percent SE
## 1          0          0.00000                 NA
## 2     887797         22.28798          0.9054683
## 3     936566         23.51232          0.6704569
## 4    2158936         54.19970          1.0531340
```

Retrieving Data for Further Manipulation With getData

Retrieving a Set of Variables in a Dataset

Although the `EdSurvey` package allows for rudimentary data manipulation and analysis directly on a `edsurvey.data.frame` connection, the function `getData()` can extract a dataset of variables for manipulation and analyses in the same manner as other `data.frame` objects. This object—referred to as a `light.edsurvey.data.frame` —can then be used with packaged `EdSurvey` analytical functions.

Variables are extracted from an `edsurvey.data.frame` and returned as a `light.edsurvey.data.frame` by specifying a set of variable names in `varnames` or by entering a formula in `formula`.[2]

To access and manipulate data the `x_chsex_r` ("Sex of Students"), the weight variable `w5cf5pf_50`, `p5sumsch` ("Child attended summer school"), and `p5nhrprm` ("Hours per day child attended summer school") variables in `eclsk11`, call `getData`.

[2]Use `?getData` for details on default `getData` arguments.

```
gddat <- getData(data = eclsk11, varnames = c("x_chsex_r", "w5cf5pf_50", "x12sesl",
                                              "p5sumsch", "p5nhrprm"),
                                 omittedLevels = FALSE, addAttributes = TRUE)
```

By default, setting `omittedLevels` to `TRUE` removes special values, such as multiple entries or `NA`s. `getData` tries to help by dropping the levels of factors for regression, tables, and correlations that are not typically included in analyses. Here we set `omittedLevels` to `FALSE` in order to recode special values in an example to follow.

The argument `addAttributes = TRUE` ensures that the analysis functions shown so far can continue to be used with the resulting dataset, `gddat`.

Retrieving All Variables in a Dataset

To extract all the data in an `edsurvey.data.frame`, define the `varnames` argument as `names(eclsk11)`, which will query all variables. Setting the argument `omittedLevels = FALSE` ensures that values that would normally be removed are included:

```
lsdf0 <- getData(data = eclsk11, varnames = colnames(eclsk11), addAttributes = TRUE,
                 omittedLevels = FALSE)
dim(lsdf0)
dim(eclsk11)
```

Additional details on the features of the `getData` function appear in the vignette titled *Using the `getData` Function in EdSurvey.*

Recoding Variables in a Dataset

As mentioned earlier, data recoding is of particular importance when performing analyses with ECLS:K-2011 data given the complexity of it's survey design on the data set. The `EdSurvey` package offers methods of recoding data to fit these needs.

Lets suppose you were interested in exploring student performance in Mathematics based on the number of hours/day a parent reported their child attended summer school (`p5nhrprm`). We'd first need to recode that variable so that students who didn't attend summer school (where `p5sumsch` coded as a `2: NO`) are included in the analytic subset with zero minutes.

The `table` function is a simple method of ascertaining the number of values for each level of a variable in a data set. Using the table function for the `p5nhrprm` variable indicates that parents reported their child attending summer school anywhere from 2 to 7 hours per day:

```
table(gddat$p5nhrprm,useNA = "ifany")
```

```
##
##     2     3     4     5     6     7  <NA>
##    55    72   126    43    87    62 17729
```

In order to include children who attended summer school for zero hours per day - those that were skipped by the design of the survey - recode `p5nhrprm` values to zero where `p5sumsch == "2: NO"`:

```
gddat$p5nhrprm <- ifelse(gddat$p5sumsch == "2: NO", 0, gddat$p5nhrprm)
table(gddat$p5nhrprm,useNA = "ifany")
```

```
##
##     0     2     3     4     5     6     7  <NA>
##  3913    55    72   126    43    87    62 13816
```

Alternatively, for demonstration purposes, a researcher may also choose to recode the `-1` values for the `p5nhrprm` variable directly:

```
gddat$p5nhrprm <- ifelse(gddat$p5nhrprm == "-1: NOT APPLICABLE*", 0, gddat$p5nhrprm)
table(gddat$p5nhrprm,useNA = "ifany")
```

```
##
##     0     2     3     4     5     6     7  <NA>
##  3913    55    72   126    43    87    62 13816
```

A second example of recoding a variable in response to a skip pattern pertains to the frequency that a) a child does homework and b) a parent/someone else helps; variables `p9hmwork` and `p9hlphwk` respectively. The `levelsSDF` function is useful to show a variable's levels and their unweighted n-sizes.

```
levelsSDF("p9hmwork",eclsk11)
```

```
## Levels for Variable 'p9hmwork' (Lowest level first):
##     1. 1: NEVER (n=294)
##     2. 2: LESS THAN ONCE A WEEK (n=381)
##     3. 3: 1 TO 2 TIMES A WEEK (n=1406)
##     4. 4: 3 TO 4 TIMES A WEEK (n=4236)
##     5. 5: 5 OR MORE TIMES A WEEK (n=3838)
##     -1. -1: NOT APPLICABLE* (n=51)
##     -7. -7: REFUSED* (n=1)
##     -8. -8: DON'T KNOW* (n=13)
##     -9. -9: NOT ASCERTAINED* (n=0)
##     NOTE: * indicates an omitted level.
```

```
levelsSDF("p9hlphwk",eclsk11)
```

```
## Levels for Variable 'p9hlphwk' (Lowest level first):
##     1. 1: NEVER (n=643)
##     2. 2: LESS THAN ONCE A WEEK (n=1786)
##     3. 3: 1 TO 2 TIMES A WEEK (n=3774)
##     4. 4: 3 TO 4 TIMES A WEEK (n=2558)
##     5. 5: 5 OR MORE TIMES A WEEK (n=1094)
##     -1. -1: NOT APPLICABLE* (n=359)
##     -7. -7: REFUSED* (n=2)
##     -8. -8: DON'T KNOW* (n=4)
##     -9. -9: NOT ASCERTAINED* (n=0)
##     NOTE: * indicates an omitted level.
```

The skip pattern for this sequence of survey questions is as follows: If `p9hmwork == "1: NEVER"` then `p9hlphwk` is skipped and coded `"-1: NOT APPLICABLE"`. To include this subset of data in the analysis, the variable `p9hlphwk` can be recoded to `0`. First, retrieve the data via `getData` (along with a few other variables for a subsequent example) to recode using `ifelse`:

```
mvData <- getData(data = eclsk11, varnames = c("p9hmwork", "p9hlphwk", "x_chsex_r",
                                               "x9rscalk5", "x9mscalk5", "w9c29p_9t90"),
                                  omittedLevels = FALSE, addAttributes = TRUE)
mvData$p9hlphwk <- ifelse(mvData$p9hmwork == "1: NEVER" &
                          mvData$p9hlphwk == "-1: NOT APPLICABLE", 0,
                          mvData$p9hlphwk)
```

Then use `table` to view the counts of each level.

```
table(mvData$p9hlphwk,useNA = "ifany")
```

```
##
##    0    1    2    3    4    5    6    7    8 <NA>
##  294  643 1786 3774 2558 1094   65    2    4 7954
```

Note that the 294 cases of the variable `mvData$p9hmwork == "1: NEVER"` are now included as a level in our recoded `mvData$p9hlphwk`.

With a few recoding steps the appropriate value levels can be included in the data set in preparation for analysis with `EdSurvey`. To find more information about special values specific to ECLS-K:2011, consult the *Missing Values* section of the *ECLS-K:2011 Public-Use Data File User's Manual*.

Applying `rebindAttributes` to Use `EdSurvey` Functions With Manipulated Data Frames

A helper function that pairs well with `getData` is `rebindAttributes`. This function allows users to reassign the attributes from a survey dataset to a data frame that might have had its attributes stripped during the manipulation process. Once attributes have been rebinded, all variables—including those outside the original dataset—are available for use in `EdSurvey` analytical functions.

The `p9hlphwk` variable from the second example in the *Recoding Variables in a Dataset* section is used in the *Multivariate Regression With `mvrlm.sdf`* section of this vignette, therefore the following example will display how to apply survey attributes to this object for analysis.

Using the `mvData` object created earlier apply `rebindAttributes` from the attribute data `eclsk11` to the manipulated data frame `mvData`. The new variables are now available for use in `EdSurvey` analytical functions:

```
mvData <- rebindAttributes(mvData, eclsk11)
lm2 <- lm.sdf(formula = x9rscalk5 ~ x_chsex_r + p9hlphwk, data = mvData,
              weightVar = "w9c29p_9t90")
```

```
## Warning in calc.lm.sdf(formula = formula, data = data, weightVar = weightVar, :
## Removing 1422 rows with nonpositive weight from analysis.
```

```
summary(lm2)
```

```
##
## Formula: x9rscalk5 ~ x_chsex_r + p9hlphwk
##
## Weight variable: 'w9c29p_9t90'
## Variance method: jackknife
## JK replicates: 80
## full data n: 18174
## n used: 7906
##
## Coefficients:
##                          coef        se        t    dof  Pr(>|t|)
## (Intercept)        142.46410   0.74660 190.8168 67.852 < 2.2e-16 ***
## x_chsex_r2: FEMALE   1.44536   0.38854   3.7200 46.223 0.0005384 ***
## p9hlphwk            -2.11562   0.21672  -9.7619 54.904 1.338e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.0284
```

Correlating variables with cor.sdf

The `EdSurvey` package features multiple correlation methods for data exploration and analysis that fully accounts for the complex sample design in ECLS data by using the `cor.sdf` function.[3] This includes the following correlation procedures:

- Pearson product-moment correlations for continuous variables
- Spearman rank correlation for ranked variables
- Polyserial correlations for one categorical and one continuous variable
- Polychoric correlations for two categorical variables

Weighted Correlations

In the following example, `x9mscalk5` ("Mathematics IRT scale"), `x9povty_i` ("Imputed Poverty Level"), and the total student weight (`w9c29p_9a0`) are read in to calculate the correlation using the Pearson method. Similar to other `EdSurvey` functions, the data are removed automatically from memory after the correlation is run.

```
cor_pearson <- cor.sdf(x = "x9mscalk5", y = "x9povty_i", data = eclsk11,
                       method = "Pearson", weightVar = "w9c29p_9a0")
```

```
## Warning in cor.sdf(x = "x9mscalk5", y = "x9povty_i", data = eclsk11, method =
## "Pearson", : Removing rows with 0 weight from the analysis.
```

Note the error message `"Removing rows with 0 weight from analysis."` indicates that this call has excluded rows with 0 weights from the calculation. It is important to take note of the order of levels to ensure that the correlations are functioning as intended. Printing a correlation object will provide a condensed summary of the correlation details and the order of levels for each variable:

[3] Use `?cor.sdf` for details on default `cor.sdf` arguments.

```
cor_pearson
```

```
## Method: Pearson
## full data n: 18174
## n used: 8487
##
## Correlation: 0.369858
## Standard Error: 0.0162821
## Confidence Interval: [0.3369019, 0.4019089]
##
## Correlation Levels:
##   Levels for Variable 'x9povty_i' (Lowest level first):
##     1. 1: BELOW POVERTY THRESHOLD
##     2. 2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD
##     3. 3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD
```

Variables in `cor.sdf` can be recoded and reordered. Variable levels and values can be recoded given the desired specifications. Recoding can be useful when a level is very thinly populated (so that it might merit combination with another level) or when changing the value label to something more appropriate for a particular analysis. In the example that follows, `x9mscalk5` and `x9povty_i` are correlated using the Pearson method, with the levels `"2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD"` and `"3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD"` of the variable `x9povty_i` being recoded to `"AT OR ABOVE POVERTY THRESHOLD"` within a list of lists in the `recode` argument:

```
cor_recode <- cor.sdf(
  x = "x9mscalk5", y = "x9povty_i", data = eclsk11,
  method = "Pearson", weightVar = "w9c29p_9a0",
  recode = list(x9povty_i =
                  list(from = c("2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOL
                                "3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD"),
                       to = c("AT OR ABOVE POVERTY THRESHOLD"))))
cor_recode
```

```
## Method: Pearson
## full data n: 18174
## n used: 8487
##
## Correlation: 0.3130048
## Standard Error: 0.01818094
## Confidence Interval: [0.2758924, 0.3491853]
##
## Correlation Levels:
##   Levels for Variable 'x9povty_i' (Lowest level first):
##     1. 1: BELOW POVERTY THRESHOLD
##     2. AT OR ABOVE POVERTY THRESHOLD
```

Changing the order of levels might be useful to modify a variable that is out of order or when reversing the orientation of a series. Note that reordering a level will not change the magnitude of the correlation coefficient but only the direction—from negative to positive in this example. The `reorder` argument also is suitable when implemented in conjunction with recoded levels. The variables `x9mscalk5` and `x9povty_i` are correlated using the Pearson method in the following example, with the variable `x9povty_i`'s values `"1: BELOW POVERTY THRESHOLD", "2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT`

OF POVERTY THRESHOLD", "3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD", being reordered to "3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD", "2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD", "1: BELOW POVERTY THRESHOLD" within a list:

```
cor_reorder <- cor.sdf(
  x = "x9mscalk5", y = "x9povty_i", data = eclsk11,
  method = "Pearson", weightVar = "w9c29p_9a0",
  reorder = list(x9povty_i = c("3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD",
                               "2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD",
                               "1: BELOW POVERTY THRESHOLD")))

cor_reorder
```

```
## Method: Pearson
## full data n: 18174
## n used: 8487
##
## Correlation: -0.369858
## Standard Error: 0.0162821
## Confidence Interval: [-0.4019089, -0.3369019]
##
## Correlation Levels:
##   Levels for Variable 'x9povty_i' (Lowest level first):
##     1. 3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD
##     2. 2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD
##     3. 1: BELOW POVERTY THRESHOLD
```

Note: As an alternative, recoding also can be completed within `getData`, a function detailed earlier in the vignette. To see additional examples of recoding and reordering, use `?cor.sdf` in the R console.

The `cor.sdf` function features multiple methods for data exploration and analysis using correlations. The following example shows the differences in correlation coefficients among the Pearson, Spearman, and polychoric methods using a `subset`[4] of the `edsurvey.data.frame` data, where `x_chsex_r %in% 1` (saved as the `sdf_dnf` object), `x9mscalk5`, `x9povty_i`, and the full sample weight (`w9c29p_9a0`):

```
sdf_dnf <- subset(eclsk11, x_chsex_r %in% 1)
cor_pearson <- cor.sdf(x = "x9mscalk5", y = "x9povty_i", data = sdf_dnf,
                       method = "Pearson", weightVar = "w9c29p_9a0")
cor_spearman <- cor.sdf(x = "x9mscalk5", y = "x9povty_i", data = sdf_dnf,
                       method = "Spearman", weightVar = "w9c29p_9a0")
cor_polychoric <- cor.sdf(x = "x9mscalk5", y = "x9povty_i", data = sdf_dnf,
                      method = "Polychoric", weightVar = "w9c29p_9a0")
```

```
##            Correlation
## Pearson      0.3825147
## Spearman     0.3860714
## Polychoric   0.4406290
```

Unweighted Correlations

The `cor.sdf` function also features the ability to perform correlations without accounting for weights by setting `weightVar = NULL`. The following example shows the correlation coefficients of the Pearson and Spearman methods of the variables `x9mscalk5` and `x9povty_i` while excluding weights:

[4] `subset` will be further detailed in this vignette; use `?subset` to access function documentation.

```
cor_pearson_unweighted <- cor.sdf(x = "x9mscalk5", y = "x9povty_i", data = eclsk11,
                                  method = "Pearson", weightVar = NULL)
cor_pearson_unweighted
```

```
## Method: Pearson
## full data n: 18174
## n used: 9329
##
## Correlation: 0.3840037
## Standard Error: NA
## Confidence Interval: [NA]
##
## Correlation Levels:
##   Levels for Variable 'x9povty_i' (Lowest level first):
##     1. 1: BELOW POVERTY THRESHOLD
##     2. 2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD
##     3. 3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD
```

```
cor_spearman_unweighted <- cor.sdf(x = "x9mscalk5", y = "x9povty_i", data = eclsk11,
                                   method = "Spearman", weightVar = NULL)
cor_spearman_unweighted
```

```
## Method: Spearman
## full data n: 18174
## n used: 9329
##
## Correlation: 0.3854115
## Standard Error: NA
## Confidence Interval: [NA]
##
## Correlation Levels:
##   Levels for Variable 'x9povty_i' (Lowest level first):
##     1. 1: BELOW POVERTY THRESHOLD
##     2. 2: AT OR ABOVE POVERTY THRESHOLD, BELOW 200 PERCENT OF POVERTY THRESHOLD
##     3. 3: AT OR ABOVE 200 PERCENT OF POVERTY THRESHOLD
```

Note the difference in `full data n` and `n used` in each correlation result - listwise deletion removes special values by default.

Making a Table with `edsurveyTable`

Summary tables can be created in the `EdSurvey` package using the `edsurveyTable` function. A call to `edsurveyTable`[5] with two variables, `x_chsex_r` ("Sex of Students") and `p9curmar` ("current marital status"), creates a table that shows the number and percentage of students by gender and their parent's current marital status. Percentages add up to 100 within each gender.

```
es1 <- edsurveyTable(formula = ~ x_chsex_r + p9curmar, data = eclsk11,
                     weightVar = "w9c29p_9t90",
                     varMethod = "jackknife")
```

[5] Use `?edsurveyTable` for details on default `edsurveyTable` arguments.

```
## Warning in calcEdsurveyTable(formula, data, weightVar, jrrIMax,
## pctAggregationLevel, : Removing rows with 0 weight from analysis.
```

This `edsurveyTable` is saved as the object `es1`, and the resulting table can be displayed by printing the object:

Table 2: Weighted and unweighted sample size, percentage distribution, and standard error of percentage distribution of children by students' gender and their parents' marital status

x_chsex_r	p9curmar	N	WTD_N	PCT	SE(PCT)
1: MALE	1: MARRIED (1)	2938	1367616.83	67.608642	1.1756039
1: MALE	2: SEPARATED (2)	151	86412.02	4.271810	0.3944507
1: MALE	3: DIVORCED OR WIDOWED (3, 4)	442	250607.34	12.388866	0.7625198
1: MALE	4: NEVER MARRIED (5)	425	273190.02	13.505249	0.9075561
1: MALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	81	45017.01	2.225432	0.3368478
2: FEMALE	1: MARRIED (1)	2870	1319848.64	69.131210	1.0257652
2: FEMALE	2: SEPARATED (2)	143	80672.81	4.225491	0.4357400
2: FEMALE	3: DIVORCED OR WIDOWED (3, 4)	428	224738.15	11.771365	0.6138104
2: FEMALE	4: NEVER MARRIED (5)	385	237346.10	12.431746	0.7270084
2: FEMALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	82	46587.90	2.440187	0.2406026

Given that the previous analysis uses parent data from round 9, the weight variable `"w9c29p_9a0"` might also be appropriate. Both weights, `"w9c29p_9t90"` and `"w9c29p_9a0"`, could be used for this analysis although both include nonresponse adjustments for additional data components or rounds of data collection than those of interest in the current analysis. Therefore analysts need to determine which weight they prefer to use given there is not a weight that adjusts for nonresponse for only the sources used in this analysis. Successive analyses in this vignette that mix round 9 child and parent variables might also substitute the selected weight chosen. Note the slight differences in `n used` and results.

Consult the *4.3.1 Types of Sample Weights* section of the *ECLS-K:2011 Kindergarten-Fifth Grade User's Manual, Public Version* for additional guidance on choosing the most appropriate sample weight for an analysis.

```
es1p <- edsurveyTable(formula = ~ x_chsex_r + p9curmar, data = eclsk11,
                      weightVar = "w9c29p_9a0",
                      varMethod = "jackknife")
```

```
## Warning in calcEdsurveyTable(formula, data, weightVar, jrrIMax,
## pctAggregationLevel, : Removing rows with 0 weight from analysis.
```

Table 3: Weighted and unweighted sample size, percentage distribution, and standard error of percentage distribution of children by students' gender and their parents' marital status - using parent weights

x_chsex_r	p9curmar	N	WTD_N	PCT	SE(PCT)
1: MALE	1: MARRIED (1)	3160	1384646.17	67.662995	1.2201761
1: MALE	2: SEPARATED (2)	165	87807.35	4.290850	0.4019977
1: MALE	3: DIVORCED OR WIDOWED (3, 4)	473	257917.26	12.603548	0.7437032
1: MALE	4: NEVER MARRIED (5)	465	273250.63	13.352838	0.8934741
1: MALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	85	42764.76	2.089770	0.3269484
2: FEMALE	1: MARRIED (1)	3072	1340918.74	69.585339	0.9959942
2: FEMALE	2: SEPARATED (2)	147	78575.20	4.077564	0.4313110
2: FEMALE	3: DIVORCED OR WIDOWED (3, 4)	449	223633.62	11.605193	0.5995680
2: FEMALE	4: NEVER MARRIED (5)	417	234899.85	12.189841	0.7083241
2: FEMALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	90	48985.90	2.542063	0.2864988

The function also features variance estimation by setting the `varMethod` argument.[6] As shown in the previous example, the default `varMethod = "jackknife"` indicates that the call used the jackknife method for variance estimation. By setting `varMethod = "Taylor"`, the same `edsurveyTable` call used in the previous example can return results using Taylor series variance estimation:

```
es1t <- edsurveyTable(formula = ~ x_chsex_r + p9curmar, data = eclsk11,
                      weightVar = "w9c29p_9t90",
                      varMethod = "Taylor")
```

Table 4: Weighted and unweighted sample size, percentage distribution, and standard error of percentage distribution of children by students' gender and their parents' marital status - Taylor Series

x_chsex_r	p9curmar	N	WTD_N	PCT	SE(PCT)
1: MALE	1: MARRIED (1)	2938	1367616.83	67.608642	1.3241743
1: MALE	2: SEPARATED (2)	151	86412.02	4.271810	0.4490012
1: MALE	3: DIVORCED OR WIDOWED (3, 4)	442	250607.34	12.388866	0.7819802
1: MALE	4: NEVER MARRIED (5)	425	273190.02	13.505249	1.1456202
1: MALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	81	45017.01	2.225432	0.3459628
2: FEMALE	1: MARRIED (1)	2870	1319848.64	69.131210	1.1836404
2: FEMALE	2: SEPARATED (2)	143	80672.81	4.225491	0.4377188
2: FEMALE	3: DIVORCED OR WIDOWED (3, 4)	428	224738.15	11.771365	0.6806314
2: FEMALE	4: NEVER MARRIED (5)	385	237346.10	12.431746	0.8638412

[6] Use `?lm.sdf` for details on default `lm.sdf` arguments.

x_chsex_r	p9curmar	N	WTD_N	PCT	SE(PCT)
2: FEMALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	82	46587.90	2.440187	0.3446689

If the percentages do not add up to 100 at the desired level, an adjustment can be made in the `pctAggregationLevel` argument to change the aggregation level. By default, `pctAggregationLevel = 1`, indicating that the formula will be aggregated by each level the first variable in the call; in our previous example this is `x_chsex_r`. Setting `pctAggregationLevel = 0` aggregates by each level of each variable in the call.

The calculation of means and standard errors requires computation time that the user may not want to wait for. If you wish to simply see a table of the levels and the N sizes, you can set the `returnMeans` and `returnSepct` arguments to `FALSE` to omit those columns as follows:

```
es1b <- edsurveyTable(formula = ~ x_chsex_r + p9curmar, data = eclsk11,
                      weightVar = "w9c29p_9t90", jrrIMax = Inf,
                      returnMeans = FALSE, returnSepct = FALSE)
```

In this `edsurveyTable`, the resulting table can be displayed by printing the object:

Table 5: Weighted and unweighted sample size and percentage distribution of children by students' gender and their parents' marital status

x_chsex_r	p9curmar	N	WTD_N	PCT
1: MALE	1: MARRIED (1)	3718	1367616.83	67.608642
1: MALE	2: SEPARATED (2)	210	86412.02	4.271810
1: MALE	3: DIVORCED OR WIDOWED (3, 4)	570	250607.34	12.388866
1: MALE	4: NEVER MARRIED (5)	599	273190.02	13.505249
1: MALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	103	45017.01	2.225432
2: FEMALE	1: MARRIED (1)	3585	1319848.64	69.131210
2: FEMALE	2: SEPARATED (2)	194	80672.81	4.225491
2: FEMALE	3: DIVORCED OR WIDOWED (3, 4)	545	224738.15	11.771365
2: FEMALE	4: NEVER MARRIED (5)	558	237346.10	12.431746
2: FEMALE	5: CIVIL UNION/DOMESTIC PARTNERSHIP (6)	114	46587.90	2.440187

For more details on the arguments in the `edsurveyTable` function, look at the examples using `?edsurveyTable`.

Subsetting the Data

A subset of a dataset can be used with `EdSurvey` package functions. In this example, a summary table is created with `edsurveyTable` after filtering the sample to include only female students (`x_chsex_r == "FEMALE"`) and those who are `1: WHITE, NON-HISPANIC` or `2: BLACK/AFRICAN AMERICAN, NON-HISPANIC`(noted by `(x_raceth_r == 1 | x_raceth_r == 2)`). The weight variable `w9c29p_9t90` is also used in the analysis.

```
eclsk11F <- subset(eclsk11, x_chsex_r %in% "2: FEMALE" & (x_raceth_r %in% 1 | x_raceth_r %in% 2))
es2 <- edsurveyTable(formula = ~ x_raceth_r, data = eclsk11F, weightVar = "w9c29p_9t90")
```

```
## Warning in calcEdsurveyTable(formula, data, weightVar, jrrIMax,
## pctAggregationLevel, : Removing rows with 0 weight from analysis.
```

```
es2
```

Table 6: Weighted and unweighted sample size, percentage distribution, and standard error of percentage distribution of white and black female students

x_raceth_r	N	WTD_N	PCT	SE(PCT)
1: WHITE, NON-HISPANIC	2067	982255.1	79.32355	1.779226
2: BLACK/AFRICAN AMERICAN, NON-HISPANIC	326	256034.2	20.67645	1.779226

Regression Analysis with lm.sdf

After the data are read in with the `EdSurvey` package, a linear model can be fit to fully account for the complex sample design used for the ECLS data by using `lm.sdf`.

```
lm1 <- lm.sdf(formula = x9mscalk5 ~ x12sesl + x_chsex_r,
              data = eclsk11, weightVar = "w9c29p_9t90")
summary(lm1)
```

```
##
## Formula: x9mscalk5 ~ x12sesl + x_chsex_r
##
## Weight variable: 'w9c29p_9t90'
## Variance method: jackknife
## JK replicates: 80
## full data n: 18174
## n used: 7877
##
## Coefficients:
##                          coef        se        t    dof  Pr(>|t|)
## (Intercept)        121.69721   0.36292 335.3305 47.291 < 2.2e-16 ***
## x12sesl              9.14174   0.28348  32.2488 52.998 < 2.2e-16 ***
## x_chsex_r2: FEMALE  -2.17136   0.41993  -5.1707 38.322 7.619e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.1711
```

After regression is run, the data are automatically removed from memory. The data are read in and analyzed by the `lm.sdf` function—in this case, `x_chsex_r`, `x12sesl`, and the Mathematics IRT scale `x9mscalk5`. The `w9c29p_9t90` weight variable and its associated replicate weights are selected for the analysis. Users can define the weight variable through the argument `weightVar`.

The option `jrrIMax` is omitted in the following example; therefore, the default jackknife variance estimator is used.[7] Since ECLS-K:2011 doesn't have a default weight an explicit weight variable must be set for the `lm.sdf` function.

[7] Use `?lm.sdf` for details on default `lm.sdf` arguments.

By default, `lm.sdf` uses "treatment contrasts," where one level is dropped from the regression for categorical variables. This cannot be changed, but the omitted and comparison group can be changed with the `relevels` argument. In the following example, `"1: MALE"` is omitted from the analysis for the variable `x_chsex_r`:

```
lm1m <- lm.sdf(formula = x9mscalk5 ~ x12sesl + x_chsex_r, data = eclsk11,
               weightVar = "w9c29p_9t90",
               relevels = list(x_chsex_r = "1: MALE"))
summary(lm1m)
```

```
##
## Formula: x9mscalk5 ~ x12sesl + x_chsex_r
##
## Weight variable: 'w9c29p_9t90'
## Variance method: jackknife
## JK replicates: 80
## full data n: 18174
## n used: 7877
##
## Coefficients:
##                         coef        se        t    dof  Pr(>|t|)
## (Intercept)        121.69721   0.36292 335.3305 47.291 < 2.2e-16 ***
## x12sesl              9.14174   0.28348  32.2488 52.998 < 2.2e-16 ***
## x_chsex_r2: FEMALE  -2.17136   0.41993  -5.1707 38.322 7.619e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.1711
```

Note that the coefficient on `x_chsex_r` changed from positive in the previous run to negative of the exact same magnitude, whereas none of the other coefficients (aside from the intercept) changed—the expected result. The change results from the switch of the reference gender from "FEMALE" in the first regression model to "MALE" in the second regression model.

The standardized regression coefficient also can be returned by adding `src = TRUE` into the summary call to your regression model object:

```
summary(lm1m, src=TRUE)
```

```
##
## Formula: x9mscalk5 ~ x12sesl + x_chsex_r
##
## Weight variable: 'w9c29p_9t90'
## Variance method: jackknife
## JK replicates: 80
## full data n: 18174
## n used: 7877
##
## Coefficients:
##                         coef         se        t    dof   Pr(>|t|) stdCoef   stdSE
## (Intercept)        121.697214   0.362917 335.3305 47.291 0.0000e+00      NA      NA
## x12sesl              9.141740   0.283475  32.2488 52.998 0.0000e+00  0.4093 0.01269
## x_chsex_r2: FEMALE  -2.171360   0.419933  -5.1707 38.322 7.6194e-06 -0.0622 0.01203
##
```

```
## (Intercept)
## x12sesl            *
## x_chsex_r2: FEMALE *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Multiple R-squared: 0.1711
```

By default, the standardized coefficients are calculated using standard deviations of the variables themselves. When `standardizeWithSamplingVar` is set to `TRUE`, the variance of the standardized coefficient is calculated similar to a regression coefficient and therefore includes the sampling variance in the variance estimate of the outcome variable.

Multivariate Regression With `mvrlm.sdf`

A multivariate regression model can be fit to fully account for the complex sample design used for NCES data by using `mvrlm.sdf`. This function implements an estimator that correctly handles multiple dependent variables that are numeric, which allows for variance estimation using the jackknife replication method.

The vertical line symbol `|` separates dependent variables on the left-hand side of formula. In the following example, a multivariate regression is fit with two subject scales as the outcome variables (`x9mscalk5` and `x9rscalk5`) by two predictor variables signifying gender and frequency that parent helps with homework (`x_chsex_r` and `p9hlphwk`). In addition, we'll use the `mvData` data set we prepared in the *Recoding Variables in a Dataset* section, recoding specific `-1` values as `0` for the `p9hlphwk` variable to account for survey skip patterns.

```
mvrlm1 <- mvrlm.sdf(x9mscalk5 | x9rscalk5 ~ x_chsex_r + p9hlphwk, data = mvData,
                    weightVar = "w9c29p_9t90")
summary(mvrlm1)
```

```
##
## Formula: x9mscalk5 | x9rscalk5 ~ x_chsex_r + p9hlphwk
##
## Weight variable: 'w9c29p_9t90'
## Variance method:
## JK replicates: 80
## full data n: 18174
## n used: 7902
##
## Coefficients:
##
## x9mscalk5
##                         coef        se         t    dof  Pr(>|t|)
## (Intercept)        129.48360   0.91514 141.49092 43.911 < 2.2e-16 ***
## x_chsex_r2: FEMALE  -2.13178   0.43307  -4.92250 38.010 1.690e-05 ***
## p9hlphwk            -2.77123   0.28759  -9.63617 55.219 1.987e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## x9rscalk5
##                         coef        se         t    dof  Pr(>|t|)
## (Intercept)        142.51179   0.74123 192.26519 67.227 < 2.2e-16 ***
```

```
## x_chsex_r2: FEMALE   1.43687   0.38947   3.68934 46.336 0.0005897 ***
## p9hlphwk            -2.12750   0.21480  -9.90446 54.691 8.359e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual correlation matrix:
##
##           x9mscalk5 x9rscalk5
## x9mscalk5     1.000     0.751
## x9rscalk5     0.751     1.000
##
## Multiple R-squared by dependent variable:
##
## x9mscalk5 x9rscalk5
##    0.0384    0.0289
```

The `mvrlm.sdf` documentation provides examples to compare the regression outputs. See `?mvrlm.sdf` for an overview of additional details that can be accessed through components of the returned object. In addition, the vignette titled *Statistical Methods Used in EdSurvey* goes into further detail by describing estimation of the reported statistics.

Logistic Regression Analysis With `glm.sdf`, `logit.sdf`, and `probit.sdf`

A logistic regression model can be fit to fully account for the complex sample design used for NCES data by using `glm.sdf`, `logit.sdf`, and `probit.sdf`. These functions predict *binary* outcomes from a set of predictor variables factoring in appropriate weights and variance estimates. `glm.sdf` is an umbrella function that currently fits logit and probit models. Alternatively, users can choose `logit.sdf` or `probit.sdf` functions for binomial outcomes. The following shows an example for the logit function, but most components also apply to the probit model. For more information about the `glm.sdf` and `probit.sdf` models, please refer to their R help files.

Although some variables might already be binary, the function `I()` can be used to specify the desired outcome level for a nonbinary variable. The following example demonstrates how to predict whether a student participates in organized performing arts programs (`p9perfrm`) based on their sex (`x_chsex_r`) using the weight `w9c29p_9t90`. The dependent variable `p9perfrm` specified using the outcome level of the regression with `I(p9perfrm == 1)`:

```
logit1 <- logit.sdf(I(p9perfrm == 1) ~ x_chsex_r,
                    weightVar = "w9c29p_9t90",
                    data = eclsk11)
summary(logit1)
```

```
##
## Formula: p9perfrm ~ x_chsex_r
## Family: binomial (logit)
##
## Weight variable: 'w9c29p_9t90'
## Variance method: jackknife
## JK replicates: 80
## full data n: 18174
## n used: 7552
```

```
##
## Coefficients:
##                           coef         se          t    dof  Pr(>|t|)
## (Intercept)         -1.921820   0.067824 -28.335533 49.136 < 2.2e-16 ***
## x_chsex_r2: FEMALE   1.278549   0.078918  16.201046 57.840 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The log odds of of performing in an arts program increases by `1.278549` for female students compared with male students.

Logistic regression results can be further interpreted with the assistance of the `oddsRatio` and `waldTest` functions. Note that these coefficients are in unmodified logistic units and are not linearized.

oddsRatio

The `oddsRatio` helper function allows for the conversion of coefficients from an `EdSurvey` logit regression model (i.e., `logit.sdf`) to odds ratios. Odds ratios are useful for understanding the real likelihood of an event occurring based on a transformation to the log odds returned in a logistic model.

In `EdSurvey`, odds ratios can be returned by specifying the logistic model object (`logit1`)

```
oddsRatio(logit1)
```

```
##                            OR      2.5%     97.5%
## (Intercept)         0.1463404 0.1276955 0.1677077
## x_chsex_r2: FEMALE 3.5914247 3.0665979 4.2060719
```

The odds of performing in an arts program increases by `3.5914247` for female students compared with male students.

waldTest

The `waldTest` function allows the user to test composite hypotheses with multiple coefficients involved. Because there is no likelihood test for residuals, the Wald test fills the role of the likelihood ratio test, ANOVA, and F-test.

Wald tests can be run by specifying the model and coefficients. The 2nd coefficient in our `logit1` model object (`x12sesl`) is tested in the following example:

```
waldTest(model = logit1, coefficients = 2)
```

```
## Wald test:
## ----------
## H0:
## x_chsex_r2: FEMALE = 0
##
## Chi-square test:
## X2 = 262.5, df = 1, P(> X2) = 0.0
##
## F test:
## W = 262.5, df1 = 1, df2 = 70, P(> W) = 0
```

To learn more about conducting Wald tests, consult the vignette titled *Methods and Overview of Using EdSurvey for Running Wald Tests* at the AIR website.

Quantile Regression Analysis with `rq.sdf`

The `rq.sdf` function computes an estimate on the tau-th conditional quantile function of the response, given the covariates, as specified by the formula argument. Similar to `lm.sdf`, the function presumes a linear specification for the quantile regression model (i.e., the formula defines a model that is linear in parameters). Note that Jackknife is the only applicable variance estimation method used by the function.

To conduct quantile regression at a given tau value (by default, tau is set as 0.5), specify using the `tau` argument (in this example `tau = 0.8`); all other arguments are otherwise consistent with `lm.sdf`, except for `returnVarEstInputs`, `returnNumberOfPSU`, and `standardizeWithSamplingVar`, which are not available.

```
rq1 <- rq.sdf(x9mscalk5 ~ x12sesl + x_chsex_r, data = eclsk11,
                    weightVar = "w9c29p_9t90", tau = 0.8)
summary(rq1)
```

```
##
## Formula: x9mscalk5 ~ x12sesl + x_chsex_r
##
## tau: 0.8
## Weight variable: 'w9c29p_9t90'
## Variance method: jackknife
## JK replicates: 80
## full data n: 18174
## n used: 10344
##
## Coefficients:
##                          coef        se         t    dof  Pr(>|t|)
## (Intercept)        134.83310   0.21299 633.03856 34.729 < 2.2e-16 ***
## x12sesl              6.43454   0.19969  32.22254 30.165 < 2.2e-16 ***
## x_chsex_r2: FEMALE  -2.40032   0.48376  -4.96183 49.922 8.505e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

For further details on quantile regression models and how they are implemented in R, see the vignette from the `quantreg` package (accessible by `vignette("rq", package="quantreg")`), on which the `rq.sdf` function is built.

Endnotes

Additional Resources

Supplementary vignettes are available to assist in analyzing NCES data. Note that some of them are written with NAEP Primer data as examples, whereas others are relevant to international assessment or longitudinal data.[8]

- *Using EdSurvey to Analyze NCES Data: An Illustration of Analyzing NAEP Primer* is an introduction to the basics of using the `EdSurvey` package for analyzing NCES data, using the NAEP Primer as an example. The vignette covers topics such as preparing the R environment for processing, creating summary tables, running linear regression models, and correlating variables.

[8] Refer to the vignette titled *Using EdSurvey to Analyze NCES Data: An Illustration of Analyzing NAEP Primer* for an introduction to the NAEP Primer dataset.

- *Exploratory Data Analysis on NCES Data* provides examples of conducting exploratory data analysis on NAEP data.
- *Calculating Adjusted* p*-Values From EdSurvey Results* describes the basics of adjusting p-values to account for multiple comparisons.
- *Using the* `getData` *Function in EdSurvey* describes the use of the `EdSurvey` package when extensive data manipulation is required before analysis.
- *Using EdSurvey to Analyze NAEP Data With and Without Accommodations* provides an overview of the use of NAEP data with accommodations and describes methods used to analyze these data.
- *Using EdSurvey to Analyze TIMSS Data* is an introduction to the methods used in the analysis of large-scale educational assessment programs such as TIMSS using the `EdSurvey` package. The vignette covers topics such as preparing the R environment for processing, creating summary tables, running linear regression models, and correlating variables.
- *Using EdSurvey for Trend Analysis* describes the methods used in the `EdSurvey` package to conduct analyses of statistics that change across time in large-scale educational studies.
- *Producing LaTeX Tables From edsurveyTable Results With edsurveyTable2pdf* details the creation of pdf summary tables from summary results using the `edsurveyTable2pdf` function.

Methodology Resources

Documents that describe the statistical methodology used in the `EdSurvey` package include the following:

- *Statistical Methods Used in EdSurvey* details the estimation of the statistics in the `lm.sdf`, `achievementLevel`, and `edsurveyTable` functions.
- *Analyses Using Achievement Levels Based on Plausible Values* describes the methodological approaches for analyses using NAEP achievement levels.
- *Methods Used for Gap Analysis in EdSurvey* covers the methods comparing the gap analysis results of the `EdSurvey` package to the NAEP Data Explorer.
- *Methods Used for Estimating Percentiles in EdSurvey* describes the methods used to estimate percentiles.
- *Methods Used for Estimating Mixed-Effects Models in EdSurvey* describes the methods used to estimate mixed-effects models with plausible values and survey weights and how to fit different types of mixed-effects models using the `EdSurvey` package.
- *Methods and Overview of Using EdSurvey for Multivariate Regression* details the estimation of multivariate regression models using `mvrlm.sdf`.
- *Methods and Overview of Using EdSurvey for Running Wald Tests* describes the use of the Wald test to jointly test regression coefficients estimated using `lm.sdf` and `glm.sdf`.

Memory Usage

Because many NCES databases have hundreds of columns and hundreds of thousands of rows, the `EdSurvey` package allows users to subset data and run regressions without storing it in the global environment. Alternatively, the `getData` function retrieves `light.edsurvey.data.frames` into the global environment, which can be costly to memory usage.

This package uses the `LaF` package to read in only the necessary data when needed for an analysis. Instead of storing all the data in memory, only some "header" information is stored as well as a link to the file in question. When the user calls a function, only the data needed for that function are read in. It works seamlessly and reduces the memory requirements for a user's machine.

Factors and Factor Analysis

R uses the concept of *factors* for data storage. This is a separate concept from factor analysis. In the case of the R storage method, it is simply a way of enforcing that only valid data labels are used.

Summary

This vignette covered the basics of the `EdSurvey` package, such as setting up the R environment for analysis, creating summary tables with `edsurveyTable`, running linear regression models with `lm.sdf`, correlating variables with `cor.sdf`, and retrieving data for manipulation with the `getData` function. Other aspects of the package relating to memory usage also were considered.

For a full list of `EdSurvey` functions and documentation, use the R help viewer:

```
help(package = "EdSurvey")
```

Appendix A.

Weights developed for use with the ECLS-K:2011 base-year data

Exhibit 4-2. Weights developed for use with the ECLS-K:2011 base-year data, by components for which nonresponse adjustments are made: School year 2010–11

	Fall kindergarten				Spring kindergarten						
Weight	Child assessment	Parent	Teacher, teacher-level	Teacher, child-level	Child assessment	Parent	Teacher, teacher-level	Teacher, supplemental	Teacher, child-level	BASC	School administrator
School-level weight											
W2SCH0	†	†	†	†	†	†	†	†	†	†	Yes
Child-level weight											
W1C0	Yes	†	†	†	†	†	†	†	†	†	†
W1A0[1]	†	†	*Yes*	†	†	†	†	*Yes*	†	†	†
W1T0	†	†	†	Yes	†	†	†	†	†	†	†
W1P0	†	Yes	†	†	†	†	†	†	†	†	†
W2P0	†	†	†	†	†	Yes	†	†	†	†	†
W12P0	†	Yes	†	†	†	Yes	†	†	†	†	†
W1_2P0[2]	†	*Yes*	†	†	†	*Yes*	†	†	†	†	†
W12T0	†	†	†	Yes	†	†	†	†	Yes	†	†
W12AC0	Yes	†	†	†	†	†	Yes	†	†	†	†
W1PZ0	†	Yes	†	†	†	†	†	†	†	Yes	†
W12PZ0	†	Yes	†	†	†	Yes	†	†	†	Yes	†

† Not applicable.
[1] The italicized *Yes* indicates an "or" condition. If teachers did not complete a fall kindergarten teacher-level questionnaire, in the spring of kindergarten they were asked to complete a Supplemental Teacher Questionnaire, which was a survey that collected important information about the teacher's characteristics, such as demographics, education, and teaching experience.
[2] The italicized *Yes* indicates an "or" condition. A case had to either have a fall kindergarten parent interview or a spring kindergarten parent interview to have a valid W1_2P0 weight.
NOTE: "Yes" indicates that the weight includes nonresponse adjustments for that component. BASC= Before- and After-School Care surveys.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), fall 2010 and spring 2011.

Weights developed for use with the ECLS-K:2011 first-grade data

Exhibit 4-2. Weights developed for use with the ECLS-K:2011 first-grade data, by components for which nonresponse adjustments were made: School year 2011–12

	Fall kindergarten		Spring kindergarten			Fall first grade			Spring first grade		
	Child assessment	Parent	Child assessment	Parent	BASC	Child assessment	Parent	Teacher	Child assessment	Parent	Teacher[1]
Weight	C1	P1	C2	P2	Z2	C3	P3	T3	C4	P4	T4
W3CF3P_30	†	*Yes*	Yes	*Yes*	†	Yes	Yes	†	†	†	†
W3CF3P3T0	†	Yes	Yes	Yes	†	Yes	Yes	Yes	†	†	†
W4CF4P_20	Yes	*Yes*	Yes	*Yes*	†	Yes	†	†	Yes	†	†
W4CF4P20	Yes	Yes	Yes	Yes	†	Yes	†	†	Yes	†	†
W4PF40	†	Yes	†	Yes	†	†	Yes	†	†	Yes	†
W4CF4P40	†	†	†	†	†	Yes	†	†	Yes	Yes	†
W4CF4P4T0	†	†	†	†	†	Yes	†	†	Yes	Yes	Yes
W4C4P_20	Yes	*Yes*	Yes	*Yes*	†	†	†	†	Yes	†	†
W4C4P_40	Yes	*Yes*	Yes	*Yes*	†	†	†	†	Yes	Yes	†
W4C4P_2T0	Yes	*Yes*	Yes	*Yes*	†	†	†	†	Yes	†	Yes
W4C4P_4T0	Yes	*Yes*	Yes	*Yes*	†	†	†	†	Yes	Yes	Yes
W4CS4P_20	†	*Yes*	Yes	*Yes*	†	†	†	†	Yes	†	†
W4CS4P_40	†	*Yes*	Yes	*Yes*	†	†	†	†	Yes	Yes	†
W4CS4P_2T0	†	*Yes*	Yes	*Yes*	†	†	†	†	Yes	†	Yes
W4CS4P_4T0	†	*Yes*	Yes	*Yes*	†	†	†	†	Yes	Yes	Yes
W4C4P4TZ0[2]	Yes	Yes	Yes	†	Yes	†	†	†	Yes	Yes	Yes
W4C_4P_4TZ0[2,3]	*Yes*	*Yes*	*Yes*	*Yes*	Yes	†	†	†	Yes	Yes	Yes

† Not applicable.
[1] A case had to have either teacher/classroom or child-level teacher data from a first-grade or a kindergarten teacher questionnaire in the spring first-grade data collection to have a valid weight.
[2] The weights involving BASC are for the sample of children who have child and/or parent and/or teacher data as defined in this table. These children do not have to have BASC data, but adjustments for BASC unknown eligibility and BASC nonresponse were included in the computation of the weights.
[3] This weight also includes the presence of either teacher/classroom or child-level teacher data from one of the base year teacher questionnaires.
NOTE: "Yes" indicates that the weight includes nonresponse adjustments for that component. An italicized *Yes* indicates an "or" condition. BASC = before- and after-school care surveys. The weight designations (C1, C2, etc.) use the same prefixes that are used for other variables in the kindergarten–first grade data file. The prefixes are listed in exhibit 7-1.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), fall 2011 and spring 2012.

Weights developed for use with the ECLS-K:2011 second-grade data

Exhibit 4-2. Weights developed for use with the ECLS-K:2011 second-grade data, by components for which nonresponse adjustments were made

	Fall kindergarten			Spring kindergarten				Fall first grade		Spring first grade			Fall second grade		Spring second grade		
Weight	C1	P1	T1[1]	C2	P2	T2[2]	Z2	C3	P3	C4	P4	T4[3]	C5	P5	C6	P6	T6[4]
W6CF6P_2A0	†	*Yes*	†	†	*Yes*	†	†	†	†	†	†	†	Yes	†	Yes	†	†
W6C6P_60	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	Yes	Yes	†
W6C6P_20	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	Yes	†	†
W6C6P60	Yes	Yes	†	Yes	Yes	†	†	†	†	Yes	Yes	†	†	†	Yes	Yes	†
W6CF6PF60	Yes	Yes	†	Yes	Yes	†	†	Yes	Yes	Yes	Yes	†	Yes	Yes	Yes	Yes	†
W6CF6P_2T0	Yes	*Yes*	†	Yes	*Yes*	Yes	†	Yes	†	Yes	†	Yes	Yes	†	Yes	†	Yes
W6CF6P_2B0	Yes	*Yes*	†	Yes	*Yes*	†	†	Yes	†	Yes	†	†	Yes	†	Yes	†	†
W6CS6P_20	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	Yes	†	†
W6CS6P_2T0	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	Yes	†	Yes
W6CS6P_6A0	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	Yes	Yes	†
W6CS6P_6TA0	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	Yes	Yes	Yes
W6CS6P_6TB0	†	*Yes*	†	Yes	*Yes*	Yes	†	†	†	Yes	†	Yes	†	†	Yes	Yes	Yes
W6CS6P_6B0	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	Yes	†	†	†	Yes	Yes	
W6C6P_6T0	Yes	*Yes*	Yes	Yes	*Yes*	Yes	†	†	†	Yes	Yes	Yes	†	†	Yes	Yes	Yes
W5CF5PF_50	†	*Yes*	†	Yes	*Yes*	†	†	Yes	Yes	Yes	†	†	Yes	Yes	†	†	†
W6C_6P_6TZ0[5,6]	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	Yes	†	†	†	†	†	†	†	Yes	Yes	Yes

† Not applicable.

[1] A case had to have either teacher/classroom- or child-level teacher data from the fall kindergarten data collection to have a valid weight.

[2] A case had to have either teacher/classroom- or child-level teacher data from a teacher questionnaire or supplemental teacher questionnaire from the spring kindergarten data collection to have a valid weight.

[3] A case had to have either teacher/classroom- or child-level teacher data from a first-grade or a kindergarten teacher questionnaire in the spring first-grade data collection to have a valid weight.

[4] A case had to have either teacher/classroom- or child-level teacher data from the spring second-grade data collection to have a valid weight.

[5] The weights involving BASC are for the sample of children who have child and/or parent and/or teacher data as defined in this table. These children do not have to have BASC data, but adjustments for BASC unknown eligibility and BASC nonresponse were included in the computation of the weights.

[6] This weight includes the presence of assessment data (C1 or C2), parent interview data (P1 or P2), and teacher data (T1 or T2).

NOTE: C indicates child assessment data. P indicates parent interview data. T indicates teacher data. Z indicates BASC data. BASC= before- and after-school care surveys. The weight designations (C1, C2, etc.) use the same prefixes that are used for other variables in the kindergarten–second grade data file. The prefixes are listed in exhibit 7-1. "Yes" indicates that the weight includes nonresponse adjustments for that component. An italicized *Yes* indicates an "or" condition.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), kindergarten-second grade (K-2) data file.

Weights developed for use with the ECLS-K:2011 third-grade data

Exhibit 4-2. Weights developed for use with the ECLS-K:2011 third-grade data, by components for which nonresponse adjustments were made

	Fall kindergarten			Spring kindergarten			Fall first grade			Spring first grade			Fall second grade			Spring second grade			Spring third grade		
Weight	C1	P1	T1[1]	C2	P2	T2[2]	C3	P3	T3[3]	C4	P4	T4[4]	C5	P5	T5[5]	C6	P6	T6[6]	C7	P7	T7[7]
W7C7P_2	†	*Yes*	†	†	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	†	†
W7C17P_2	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	†	†
W7C17P_7T7	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W7C17P_7	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	Yes	†	†	†	†	†	Yes	†	Yes	Yes	†
W7C17P_2T27	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	Yes	†	†	†	†	†	Yes	Yes	†	Yes
W7C17P_7T27A	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	Yes	†	†	†	†	†	Yes	Yes	Yes	Yes
W7C17P_7T27B	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	Yes	Yes	†	†	†	†	Yes	Yes	Yes	Yes	Yes
W7C17P_7T17	Yes	*Yes*	Yes	Yes	*Yes*	Yes	†	†	†	Yes	Yes	Yes	†	†	†	Yes	Yes	Yes	Yes	Yes	Yes
W7C27P_7T7	†	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W7C27P_7A	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	Yes	†
W7C27P_2T7	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	†	Yes
W7C27P_7B	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	Yes	†	†	†	†	Yes	Yes	†	Yes	Yes	†
W7C27P_2T27	†	*Yes*	†	Yes	*Yes*	Yes	†	†	†	Yes	†	Yes	†	†	†	Yes	†	Yes	Yes	†	Yes
W7C27P_7T27	†	*Yes*	†	Yes	*Yes*	Yes	†	†	†	Yes	Yes	Yes	†	†	†	Yes	Yes	Yes	Yes	Yes	Yes
W7CF7P_7	Yes	*Yes*	†	Yes	*Yes*	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†
W7CF7P_2T17	Yes	*Yes*	Yes	Yes	*Yes*	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes

† Not applicable.

[1] A case had to have either teacher/classroom- or child-level teacher data from the fall kindergarten data collection to have a valid weight.

[2] A case had to have either teacher/classroom- or child-level teacher data from a teacher questionnaire or supplemental teacher questionnaire from the spring kindergarten data collection to have a valid weight.

[3] A case had to have child-level teacher data from the fall first-grade data collection to have a valid weight.

[4] A case had to have either teacher/classroom- or child-level teacher data from a first-grade or a kindergarten teacher questionnaire in the spring first-grade data collection to have a valid weight.

[5] A case had to have child-level teacher data from the fall second-grade data collection to have a valid weight.

[6] A case had to have either teacher/classroom- or child-level teacher data from the spring second-grade data collection to have a valid weight.

[7] A case had to have either teacher/classroom- or child-level teacher data from the third-grade data collection to have a valid weight.

NOTE: C indicates child assessment/child questionnaire data. P indicates parent interview data. T indicates teacher data. The weight designations (C1, C2, etc.) use the same prefixes that are used for other variables in the kindergarten–third grade data file. The prefixes are listed in exhibit 7-1. "Yes" indicates that the weight includes nonresponse adjustments for that component. An italicized *Yes* indicates an "or" condition.

SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), kindergarten–third grade (K-3) data file.

Weights developed for use with the ECLS-K:2011 fourth-grade data

Exhibit 4-3. Weights developed for use with the ECLS-K:2011 fourth-grade data, by components for which nonresponse adjustments were made

	Fall kindergarten			Spring kindergarten			Fall first grade			Spring first grade			Fall second grade			Spring second grade			Spring third grade			Spring fourth grade		
Weight	C1	P1	T1[1]	C2	P2	T2[2]	C3	P3	T3[3]	C4	P4	T4[4]	C5	P5	T5[5]	C6	P6	T6[6]	C7	P7	T7[7]	C8	P8	T8[8]
W8C8P_2	†	*Yes*	†	†	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	†	†
W8C18P_2	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	†	†
W8C18P_8	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	†	Yes	Yes	†
W8C28P_8A	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	†	†	Yes	Yes	†
W8C28P_8B	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	Yes	†	†	†	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†
W8CF8P_8	Yes	*Yes*	†	Yes	*Yes*	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†
W8C18P_2T28	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	Yes	†	Yes
W8C18P_8T28A	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	Yes	Yes	Yes
W8C18P_8T28B	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	Yes	Yes	†	†	†	†	Yes	Yes	†	Yes	Yes	Yes	Yes	Yes
W8C18P_8T18	Yes	*Yes*	Yes	Yes	*Yes*	Yes	†	†	†	Yes	Yes	Yes	†	†	†	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W8C28P_2T28	†	*Yes*	†	Yes	*Yes*	Yes	†	†	†	Yes	†	Yes	†	†	†	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes
W8CF8P_2T18	Yes	*Yes*	Yes	Yes	*Yes*	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes
W8C18P_8T8	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W8C18P_8T8Z[9]	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W8C18P_8T28C	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W8C18P_8T28Z[9]	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes

See notes at end of exhibit.

Exhibit 4-3. Weights developed for use with the ECLS-K:2011 fourth-grade data, by components for which nonresponse adjustments were made—Continued

	Fall kindergarten			Spring kindergarten			Fall first grade			Spring first grade			Fall second grade			Spring second grade			Spring third grade			Spring fourth grade		
Weight	C1	P1	T1[1]	C2	P2	T2[2]	C3	P3	T3[3]	C4	P4	T4[4]	C5	P5	T5[5]	C6	P6	T6[6]	C7	P7	T7[7]	C8	P8	T8[8]
W8C28P_8T8	†	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W8C28P_8T8Z[9]	†	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W8C28P_2T8	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	†	†	Yes	†	Yes
W8C28P_2T8Z[9]	†	*Yes*	†	Yes	*Yes*	†	†	†	†	Yes	†	†	†	†	†	Yes	†	†	Yes	†	†	Yes	†	Yes
W8C28P_8T28	†	*Yes*	†	Yes	*Yes*	Yes	†	†	†	Yes	Yes	Yes	†	†	†	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W8C28P_8T28Z[9]	†	*Yes*	†	Yes	*Yes*	Yes	†	†	†	Yes	Yes	Yes	†	†	†	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

† Not applicable.
[1] A case had to have either teacher/classroom- or child-level teacher data from the fall kindergarten data collection to have a valid weight.
[2] A case had to have either teacher/classroom- or child-level teacher data from a teacher questionnaire or supplemental teacher questionnaire from the spring kindergarten data collection to have a valid weight.
[3] A case had to have child-level teacher data from the fall first-grade data collection to have a valid weight.
[4] A case had to have either teacher/classroom- or child-level teacher data from a first-grade or a kindergarten teacher questionnaire in the spring first-grade data collection to have a valid weight.
[5] A case had to have child-level teacher data from the fall second-grade data collection to have a valid weight.
[6] A case had to have either teacher/classroom- or child-level teacher data from the spring second-grade data collection to have a valid weight.
[7] A case had to have either teacher/classroom- or child-level teacher data from the third-grade data collection to have a valid weight.
[8] A case had to have either teacher/classroom- or child-level teacher data from the fourth-grade data collection to have a valid weight.
[9] This weight is for the analysis of data that include the mathematics/science teacher/classroom or child-level mathematics/science teacher data from the fourth grade
NOTE: C indicates child assessment/child questionnaire data. P indicates parent interview data. T indicates teacher data. The weight designations (C1, C2, etc.) use the same prefixes that are used for other variables in the kindergarten–fourth grade data file. The prefixes are listed in exhibit 7-1. "Yes" indicates that the weight includes nonresponse adjustments for that component. An italicized *Yes* indicates an "or" condition.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), kindergarten–fourth grade (K-4) restricted-use data file.

Weights developed for use with the ECLS-K:2011 fifth-grade data

Exhibit 4-3. Weights developed for use with the ECLS-K:2011 fifth-grade data, by components for which nonresponse adjustments were made: Spring 2016

	Fall kindergarten			Spring kindergarten			Spring first grade			Spring second grade			Spring third grade			Spring fourth grade			Spring fifth grade		
Weight	C1	P1	T1[1]	C2	P2	T2[2]	C4	P4	T4[3]	C6	P6	T6[4]	C7	P7	T7[5]	C8	P8	T8[6]	C9	P9	T9[7]
W9C9P_20	†	*Yes*	†	†	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	†	†
W9C19P_20	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	†	†
W9C19P_90	Yes	*Yes*	†	Yes	*Yes*	†	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	Yes	Yes	†
W9C29P_9A0	†	*Yes*	†	Yes	*Yes*	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	Yes	†
W9C29P_9B0	†	*Yes*	†	Yes	*Yes*	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†
W9C19P_2T290	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	Yes	†	Yes
W9C19P_9T29A0	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	Yes	Yes	Yes
W9C19P_9T29B0	Yes	*Yes*	†	Yes	*Yes*	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	Yes	Yes	Yes
W9C29P_2T290	†	*Yes*	†	Yes	*Yes*	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes	Yes	†	Yes
W9C19P_9T90	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W9C19P_9T9Z0[8]	Yes	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W9C19P_9T29C0	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W9C19P_9T29Z0[8]	Yes	*Yes*	†	Yes	*Yes*	Yes	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W9C29P_9T90	†	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W9C29P_9T9Z0[8]	†	*Yes*	†	Yes	*Yes*	†	†	†	†	†	†	†	†	†	†	†	†	†	Yes	Yes	Yes
W9C29P_2T90	†	*Yes*	†	Yes	*Yes*	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	Yes
W9C29P_2T9Z0[8]	†	*Yes*	†	Yes	*Yes*	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	†	Yes	†	Yes
W9C29P_9T290	†	*Yes*	†	Yes	*Yes*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W9C29P_9T29Z0[8]	†	*Yes*	†	Yes	*Yes*	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
W9C790	†	†	†	†	†	†	†	†	†	†	†	†	Yes	†	†	Yes	†	†	Yes	†	†
W9C79P_9T790	†	*Yes*	††	†	*Yes*	††	†	†	††	†	†	†	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

† Not applicable.
[1] A case had to have either teacher/classroom- or child-level teacher data from the fall kindergarten data collection to have a valid weight.
[2] A case had to have either teacher/classroom- or child-level teacher data from a teacher questionnaire or supplemental teacher questionnaire from the spring kindergarten data collection to have a valid weight.
[3] A case had to have either teacher/classroom- or child-level teacher data from a first-grade or a kindergarten teacher questionnaire in the spring first-grade data collection to have a valid weight.
[4] A case had to have either teacher/classroom- or child-level teacher data from the spring second-grade data collection to have a valid weight.
[5] A case had to have either teacher/classroom- or child-level teacher data from the third-grade data collection to have a valid weight.
[6] A case had to have either teacher/classroom- or child-level teacher data from the fourth-grade data collection to have a valid weight.
[7] A case had to have either teacher/classroom- or child-level teacher data from the fifth-grade data collection to have a valid weight.
[8] This weight is for the analysis of data that include the mathematics/science teacher/classroom or child-level mathematics/science teacher data from the fifth grade.
NOTE: C indicates child assessment/child questionnaire data. P indicates parent interview data. T indicates teacher data. "Yes" indicates that the weight includes nonresponse adjustments for that component. An italicized *Yes* indicates an "or" condition.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Early Childhood Longitudinal Study, Kindergarten Class of 2010–11 (ECLS-K:2011), kindergarten–fifth grade (K-5) restricted-use data file.