

Created variables for CCSSE 2017–Present

Table of Contents

Documentation for all colleges	1
Additional documentation for variables created for oversamples	4

Introduction

This document provides detailed descriptions of how all calculated variables in the *CCSSE* data set are created. These variables include, among others, developmental status, first-generation status, traditional-age/nontraditional-age, race/ethnicity, and weights.

This document includes two major sections. The first section describes created variables that apply to all colleges. The second section (beginning on page 4) describes additional created variables that apply only to oversample records.

Documentation for all colleges

Developmental Status [DEVED]

Item 8: Which of the following have you done, or are you currently doing at this college?

8c. Developmental/remedial reading course (also referred to as Basic Skills, College Prep, etc.)	DONEDEVRD	0 = No	1 = Yes
8d. Developmental/remedial writing course (also referred to as Basic Skills, College Prep, etc.)	DONEDEVWR	0 = No	1 = Yes
8e. Developmental/remedial math course (also referred to as Basic Skills, College Prep, etc.)	DONEDEVMT	0 = No	1 = Yes

If respondents indicate that they have taken or are currently taking any one or more of these three types of courses, they are classified as developmental (1); if respondents indicate that they have not nor are currently taking any of these three types of courses, they are classified as non-developmental (0). Additionally, to be classified as developmental or non-developmental, a student must have responded to all three items.

First-Generation Status [FIRSTGEN]

Item 47: Who in your family has attended at least some college? (Mark all that apply).

First-Generation status is based on responses for mother and father's college education experiences only. This maintains the logic from the original *CCSSE* instrument and aligns with item 38 on the *SENSE* instrument.

Relation	Variable	Value if not marked	Value if marked
47a. Mother	SOMECOLMO	0	1
47b. Father	SOMECOLFA	0	1
47c. Brother/Sister	SOMECOLSIB	0	1
47d. Child	SOMECOLCHLD	0	1
47e. Spouse/Partner	SOMECOLSP	0	1
47f. Legal Guardian	SOMECOLGUAR	0	1
47g. No one	SOMECOLNONE	0	1

If a student marked neither 47a [SOMECOLMO] nor 47b [SOMECOLFA], the student is coded as first-generation (1) regardless of whether or not the student marked other subitems.

If a student marked either 47a [SOMECOLMO] or 47b [SOMECOLFA] or both, the student is coded as not first-generation (0) regardless of whether the student marked other subitems.

Traditional/Non-Traditional Age [TRADAGE]

Item 38: Mark your age group

- 1 = Under 18
- 2 = 18–19
- 3 = 20–21
- 4 = 22–24
- 5 = 25–29
- 6 = 30–39
- 7 = 40–49
- 8 = 50–64
- 9 = 65+

Students under the age of 18 are excluded from all data sets.

If a student marked 2 (18–19), or 3 (20–21), or 4 (22–24) on item 38, the student is coded as traditional-age (1).

If a student marked 5 (25–29), or 6 (30–39), or 7 (40–49), or 8 (50–64), or 9 (65+) on item 38, the student is coded as non-traditional age (0).

Race/Ethnicity [RACE_ETH]

Item 45: What is your racial or ethnic identification? (Mark all that apply)

- 45a. American Indian or Alaska Native
- 45b. Asian
- 45c. Black or African American
- 45d. Hispanic or Latino
- 45e. Native Hawaiian

- 45f. Pacific Islander (non-Native Hawaiian)
- 45g. White
- 45h. Other
- 45i. I prefer not to respond.

If a student marked only one of the above population subgroups, the student is classified as belonging to that subgroup. If a student marked two or more of the above responses, the student is reported as “Two or More Races” (9). Finally, if the student marked “I prefer not to respond” (10), regardless of any other subgroups that may have been marked, the student’s race/ethnicity classification is set to missing.

Hours Completed [CREDIT]

This calculated variable reflects the number of credit hours a student completed **prior** to the academic term the student completed *CCSSE*.

Item 33: How many total credit hours have you earned at this college, not counting the courses you are currently taking this academic term? [TOTCREARND]

- 0 = None
- 1 = 1–14 credits
- 2 = 15–29 credits
- 3 = 30–44 credits
- 4 = 45–60 credits
- 5 = Over 60 credits

If a student marked 0 (None), or 1 (1–14 credits), or 2 (15–29 credits) on item 33, the student is coded as earning 0-29 credits (1).

If a student marked 3 (30–44 credits), or 4 (45–60 credits), or 5 (Over 60 credits) on item 33, the student is coded as earning 30 or more credits (2).

Credential-Seeking [CREDENTIAL]

Item 26: Indicate which of the following are your reasons/goals for attending this college?

(Please respond to each item). The response options are “Yes” and “No.”

- a. Complete a certificate program [GOALCERT]
- b. Obtain an associate degree [GOALAA]
- c. Transfer to a 4-year college or university [GOALTR4YR]
- d. Obtain or update job-related skills [GOALJOB SKILL]
- e. Change careers [GOALCHGCAR]
- f. Self-improvement/personal enjoyment

If a student marked “No” for subitems a, b, and c, (these are not goals), the student is coded as a non-credential-seeking student (0).

If a student responded “Yes” for any sub-item a, b, or c, (these are goals), AND did not leave any of these blank, the student is coded as a credential-seeking student (1).

Institutional Weight [IWEIGHT]

This variable is the weight used to correct for the known sampling bias resulting from sampling at the classroom level. Since full-time students have a greater likelihood of being sampled simply by the fact that they enroll in more classes than part-time students, full-time students are over-represented in the data set. To correct for this bias, the center creates the variable IWEIGHT using the following formula.

Part-time weight:

$$PT_{WEIGHT} = \frac{(\% \text{ PT population at college as reported to IPEDS})}{(\% \text{ PT respondents from the college's surveys})}$$

Full-time weight:

$$FT_{WEIGHT} = \frac{(\% \text{ FT population at college as reported to IPEDS})}{(\% \text{ FT respondents from the college's surveys})}$$

If a student marked a response of being enrolled full-time on item 3 (ENRLMENT), then IWEIGHT is set to the full-time weight value above. If the student marked a response of being enrolled less-than full-time, their IWEIGHT is set to the part-time weight value above. If a student does not mark either response, IWEIGHT will be missing since the information necessary to assign the proper weight is missing.

Additional documentation for variables created for oversamples

Pseudo-Weights [IWEIGHT_P]

Oversample observations do not have an IWEIGHT value because they were not included in the calculation of the weights used for CCSSE reporting. However, to facilitate analyses of the raw data files, the Center includes a “pseudo-weight” [IWEIGHT_P], which is the same as the primary sample IWEIGHT.

Responses are also excluded from online reporting if they meet one of the following criteria: no response on item 2 (ENRLMENT) or straight-lined responses to all subitems in item 4.

If the variable PSAMPLE is 1, the respondent is in the primary sample, and if it is 0 (zero), the respondent is in the oversample.

Weights for oversample respondents (IWEIGHT_P) are set to match the weights for primary sample respondents. Therefore, the formula shown above applies to oversample respondents as well and is not repeated here.

Raw Benchmark Scores for Oversamples

The raw benchmark score is simply the average of the rescaled items composing the respective benchmark. As with weights, this is the same process for both primary and oversample respondents. For complete documentation of how raw benchmarks are created, see "[How Benchmarks Are Calculated.](#)"

Standardized Benchmark Scores (pseudo-benchmark scores)

Because oversample respondents are not included in the creation of the standardized benchmark scores, a different process is followed to create pseudo-benchmark scores. This involves four steps:

- For each benchmark score, subtract the mean of the raw benchmark scores (based only on the primary sample) from each student's score;
- Divide the difference above by the standard deviation of the mean of the raw benchmark scores (based only on the primary sample);
- Multiply this by 25 (standard deviation for standardized scores across the cohort);
- Add 50 (mean for the standardized scores across cohort).

Following the steps above, the formulae for calculation of pseudo-benchmark scores are as follows:

Active and Collaborative Learning (7 items: 4a, 4b, 4f, 4g, 4h, 4i, and 4q)

$$\text{ACTCOLL_STD_P} = (((\text{actcoll} - \text{actcoll_m}) / \text{actcoll_s}) * 25) + 50$$

Student Effort (8 items: 4c, 4d, 4e, 6b, 10a, 12d1, 12e1, and 12h1)

$$\text{STUEFF_STD_P} = (((\text{stueff} - \text{stueff_m}) / \text{stueff_s}) * 25) + 50$$

Academic Challenge (10 items: 5b, 5c, 5d, 5e, 5f, 6a, 6c, 7, 9a, and 4o)

$$\text{ACCHALL_STD_P} = (((\text{acchall} - \text{acchall_m}) / \text{acchall_s}) * 25) + 50$$

Student-Faculty Interaction (6 items: 4j, 4k, 4l, 4m, 4n, and 4p)

$$\text{STUFAC_STD_P} = (((\text{stufac} - \text{stufac_m}) / \text{stufac_s}) * 25) + 50$$

Support for Learners (7 items: 9b, 9c, 9d, 9e, 9f, 12a1, and 12b1)

$$\text{SUPPORT_STD_P} = (((\text{support} - \text{support_m}) / \text{support_s}) * 25) + 50$$

In the formulae above, the [benchmark]_m variables represent the mean of the raw benchmark scores for the primary sample; oversample respondents are not included in the calculation of the benchmark mean and standard deviation.

Please note, this process cannot be replicated with an individual college's data because the full cohort of primary respondents is required to obtain the appropriate mean and standard deviation.