# How Benchmarks Are Calculated

For prior versions of this document, please visit www.cccse.org/reportdocs

The six <u>SENSE</u> benchmarks are groups of conceptually related survey items that focus on institutional practices and behaviors that promote engagement among entering students. Benchmarks are used to compare each institution's performance to that of similar institutions and to the <u>SENSE</u> Cohort. As part of each data release, colleges that participate in <u>SENSE</u> receive two types of benchmark scores, raw and standardized. Both types are useful, but for different purposes. See "<u>When to Use Standardized and Raw Benchmark Scores</u>" for a brief discussion of when it is appropriate to use each of these types of benchmark scores.

## **Creating Benchmark Scores**

Creating both types of benchmark scores involves several steps, including reverse coding items as necessary and converting all responses to the same scale. After these initial steps are taken, raw benchmark scores are computed for each respondent. Based on the raw benchmark scores, *SENSE* then computes standardized benchmark scores for each respondent. Please note that individual colleges cannot compute standardized benchmark scores as this process can only be completed using the complete 3-year cohort data set. However, pre-calculated standardized benchmark and raw benchmark scores for each student record are included in the college raw data file. Once the benchmark scores are created (either raw or standardized), college-, campus-, and group-level benchmarks can be calculated. The steps used to create these benchmark scores are explained in detail below.

#### 1. Reverse code items where necessary.

The first step is to determine which items, if any, need to be reverse coded so that a low score on the item represents a desirable behavior. For example, Item 19s, which asks how often the student skipped class, is originally coded such that 1 = never and 4 = four or more times. In this case, never skipping class should have a higher positive impact on the benchmark score than skipping class four or more times. The easiest way to reverse code this item (and others that have these same four response options) is to use the following formula:

ReverseScore = 5 - OriginalScore

For 19s, where the original variable name is SKIPCL, the reverse codes would be calculated as follows:

SKIPCL Rev = 5 - SKIPCL

4 = 5 - 1 (4 becomes the value for "Never")

3 = 5 - 2 (3 becomes the value for "Once")

2 = 5 - 3 (2 becomes the value for "Two or three times")

1 = 5 - 4 (1 becomes the value for "Four or more times")

Several yes/no response items are used in calculating benchmarks. These are normally coded as 1 = yes and 2 = no. The same logic presented above is used to reverse code these scores so that 0 = no and 1 = yes (ReverseScore = 2 - OriginalScore).

SENSE has eight reverse-coded items: 12a ,12b, 14, 19c, 19d, 19f, 19s, and 23.



#### 2. Convert all items to a common scale with a range of 0 to 1.

After reverse coding the items, the next step is to convert all items to a common 0 (zero) to 1 scale. The following formula is used to accomplish this conversion:

```
RescaleScore = (OriginalScore - 1) / (MaxResponseValue - 1)
```

Using 19s [SKIPCL] as an example again, the following formula should be used:

```
SKIPCL_RevRaw = (SKIPCL_Rev - 1) / 3

0.00 = 1 - 1 / 3

0.33 = 2 - 1 / 3

0.66 = 3 - 1 / 3

1.00 = 4 - 1 / 3
```

Remember when working with reverse-coded items to use the reverse-coded variables.

#### 3. Create raw benchmark scores.

Calculation of the raw benchmark scores uses the newly-created rescaled (0-1) variables. These scores are created by calculating the average score of the items that make up the benchmark. Using the Early Connections benchmark [EARLYCON] as an example and the item numbers from the survey as variable names, the following is the formula for computing the raw benchmark score:

#### 4. Compute standardized benchmark scores.

Before explaining how this step is carried out, it is important to note that standardized benchmark scores cannot be computed without having the entire cohort data set (all respondents included in the 3-year cohort). As such, this step is only briefly explained.

SENSE uses the STANDARD procedure in SAS to create the standardized benchmark scores across the 3-year cohort so the average benchmark is 50 with a standard deviation of 25 at the student record level. To account for the inherent sampling bias, this calculation includes weights, the utilization of which is explained in the next step.

#### 5. Compute group-level benchmark scores.

The process explained above creates benchmark scores (raw and standardized) for every respondent in the primary sample. The process for creating group-level (e.g., men and women) benchmark scores is the same for both raw and standardized benchmarks. In most circumstances, the group-level benchmarks are created by calculating the weighted average of a benchmark variable for the members of the group. Colleges can choose to administer *CCSSE* and *SENSE* on paper in the classroom or online. For the colleges administering on paper in the classroom, the survey is administered to a sample of students. Sampling is conducted at the class level and, as such, full-time students are more likely to be included in the sample than part-time students because full-time students take more classes. Sampling is not conducted for the colleges that administer the survey online; all eligible

students are invited to participate. However, online respondents tend to be disproportionately full-time and women. To account for these known biases in the respondent populations for both types of survey administration, computation of group-level benchmark scores must incorporate weights so the results are more representative of the actual distribution of students at a given college, even when analyzing benchmarks when either enrollment status and/or gender are used as break-out variables. See "When to Use Weights" for a more detailed discussion of using weights in analyzing SENSE data. (NOTE: Standardized benchmark scores are not created for oversample respondents.)

## Computing the Six SENSE Benchmark Scores

Raw data sets returned to participating colleges include all core *SENSE* survey items as well as special-focus item data, if applicable. Two sets of benchmark scores are included in each college's data set: raw benchmark scores and corresponding standardized benchmark scores for each respondent. These scores can be used to calculate sub-group benchmark scores.

The process SENSE uses to calculate raw benchmark scores and how to create sub-group benchmarks using the raw and standardized benchmark scores are described in this section. See "When to Use Standardized and Raw Benchmark Scores" for a brief discussion of when it is appropriate to use each of these types of benchmark scores. Please note that individual colleges cannot compute individual respondent-level standardized benchmark scores as this process can only be completed with the complete 3-year cohort data set.

The standard process for calculating individual respondent-level benchmark scores involves the following:

- 1. Reverse coding items (where applicable)
- 2. Converting scores on benchmark items to a common scale with a range of 0-1 (zero to one)
- 3. Creating the benchmark scores
- 4. Computing group-level benchmark scores
  - a. Raw benchmark scores
  - Standardized benchmark scores

#### **Early Connections**

Items 18a, 18i, 18j, 18p, and 23

Item 23 is the one variable in the Early Connections benchmark [EARLYCON] that requires reverse coding (Note: The following code also converts the ASNPER variable to a 0-1 scale.)

```
Item 23 (2-point item): ASNPERS Rev = (2 - ASNPER)
```

The process for converting the original scale for the remaining items to a 0-1 scale is the same, varying only by the number of response options for any given item. The math for converting each item is presented below. (Note: The lowest numeric response value for all items in the original scale is 1 and the highest value is 5 for all items except Item 23, a yes/no response item, which is 2.)

```
Item 18a (5-point scale): WELCOME_Raw = (WELCOME - 1) / 4
Item 18i (5-point scale): FAINFO_Raw = (FAINFO - 1) / 4
Item 18j (5-point scale): QUALFA_Raw = (QUALFA - 1) / 4
Item 18p (5-point scale): CSTAFNAM_Raw = (CSTAFNAM - 1) / 4
Item 23 (2-point scale): ASNPERS RevRaw = ASNPERS Rev
```



The new rescaled variables can now be used to calculate the raw individual-level benchmark scores. This is simply a matter of computing the average of the five rescaled items:

```
EARLYCON = (WELCOME_Raw + FAINFO_Raw + QUALFA_Raw + CSTAFNAM_Raw + ASNPERS_RevRaw) / 5
```

The final step is creating the raw benchmark score for a given population subgroup. This is accomplished by computing the weighted average of the raw benchmark score [EARLYCON] for all respondents in the subgroups of interest. (IMPORTANT NOTE: If your population subgroup is based on enrollment status (full-time and part-time) or gender identity (man and woman), then the weight should not be used; instead, the unweighted average should be calculated. See "When To Use Weights" for a more detailed discussion of using weights in analyzing SENSE data.

The raw benchmark variable for the Early Connections benchmark is EARLYCON and the standardized benchmark variable is EARLYCON\_STD. Computation of a population subgroup standardized benchmark score follows the same procedure as just described for the raw subgroup population benchmark score substituting EARLYCON\_STD for EARLYCON.

### **High Expectations and Aspirations**

Items 18b, 18t, 18u, 19c, 19d, 19f, and 19s

The High Expectations and Aspirations benchmark [HIEXPECT] contains four items that require reverse coding. These items are reverse coded using the following process:

```
Item 19c (4-point scale): LATETRN_Rev = (5 - LATETURN)
Item 19d (4-point scale): NOTTURN_Rev = (5 - NOTTURN)
Item 19f (4-point scale): NOTCOMPL_Rev = (5 - NOTCOMPL)
Item 19s (4-point scale): SKIPCL Rev = (5 - SKIPCL)
```

The process for converting the original scale for each item to a 0-1 scale is the same, varying only by the number of response options for any given item. The math for converting each item is presented below. (Note: The lowest numeric response value for all items in the original scale is 1 and the highest value is 5 for Item 18 and 4 for Item 19.)

```
Item 18b (5-point scale): WNTSCCD_Raw = (WNTSCCD - 1) / 4
Item 18t (5-point scale): ITTAKES_Raw = (ITTAKES - 1) / 4
Item 18u (5-point scale): ACPRPRD_Raw = (ACPRPRD - 1) / 4
Item 19c (4-point scale): LATETRN_ RevRaw = (LATETRN_Rev - 1) / 3
Item 19d (4-point scale): NOTTURN_ RevRaw = (NOTTURN_Rev - 1) / 3
Item 19f (4-point scale): NOTCOMPL_ RevRaw = (NOTCOMP_Rev - 1) / 3
Item 19s (4-point scale): SKIPCL_ RevRaw = (SKIPCL_Rev - 1) / 3
```

The new recoded variables can now be used to calculate the raw individual-level benchmark scores. This is simply a matter of computing the average of the seven recoded items in this scale:

```
HIEXPECT = (WNTSCCD_Raw + ITTAKES_Raw + ACPRPRD_Raw + LATETRN_ RevRaw + NOTTURN RevRaw + NOTCOMPL RevRaw + SKIPCL RevRaw) / 7
```



The final step is creating the raw benchmark score for a given population subgroup. This is accomplished by computing the weighted average of the raw benchmark score [HIEXPECT] for all respondents in the subgroups of interest. (IMPORTANT NOTE: If your population subgroup is based on enrollment status (full-time and part-time) or gender identity (man and woman), then the weight should not be used; instead, the unweighted average should be calculated. See "When To Use Weights" for a more detailed discussion of using weights in analyzing SENSE data.)

The raw benchmark variable for the High Expectations and Aspirations benchmark is HIEXPECT, and the standardized benchmark variable is HIEXPECT\_STD. Computation of a population subgroup standardized benchmark score follows the same procedure as just described for the raw subgroup population benchmark score substituting HIEXPECT\_STD for HIEXPECT.

### **Clear Academic Plan and Pathway**

Items 18d, 18e, 18f, 18g, and 18h

The Clear Academic Plan and Pathway benchmark [ACADPLAN] does not include any items that require reverse coding.

The process for converting the original scale for each item to a 0-1 scale is the same, varying only by the number of response options for any given item. The math for converting each item is presented below. (Note: The lowest numeric response value for all items in the original scale is 1 and the highest value is 5 for all items.)

```
Item 18d (5-point scale): AACONTIM_Raw = (AACONTIM - 1) / 4
Item 18e (5-point scale): AASELMAJ_Raw = (AASELMAJ - 1) / 4
Item 18f (5-point scale): ACADGOAL_Raw = (ACADGOAL - 1) / 4
Item 18g (5-point scale): CRSADV_Raw = (CRSADV - 1) / 4
Item 18h (5-point scale): OSCOMM_Raw = (OSCOMM - 1) / 4
```

The new recoded variables can now be used to calculate the raw individual-level benchmark scores. This is simply a matter of computing the average of the five recoded items in this scale:

```
ACADPLAN = (AACONTIM_Raw + AASELMAJ_Raw + ACADGOAL_Raw + CRSADV_Raw + OSCOMM_Raw) / 5
```

The final step is creating the raw benchmark score for a given population subgroup. This is accomplished by computing the weighted average of the raw benchmark score [ACADPLAN] for all respondents in the subgroups of interest. (IMPORTANT NOTE: If your population subgroup is based on enrollment status (full-time and part-time) or gender identity (man and woman), then the weight should not be used; instead the unweighted average should be calculated. See "When To Use Weights" for a more detailed discussion of using weights in analyzing SENSE data.)

The raw benchmark variable for the Clear Academic Plan and Pathway benchmark is ACADPLAN and the standardized benchmark variable is ACADPLAN\_STD. Computation of a population subgroup standardized benchmark score follows the same procedure as just described for the raw subgroup population benchmark score substituting ACADPLAN\_STD for ACADPLAN.

### **Effective Track to College Readiness**

Items 12a, 12b, 14, 21a, 21b, and 21c

Three items in the Effective Track to College Readiness benchmark [COLLREAD] require reverse coding (Note: The following code also converts the these three variables to a 0-1 scale.)

```
Item 12a (2-point scale): REQPTEST_Rev = (2 - REQPTEST)
Item 12b (2-point scale): TKPTEST_Rev = (2 - TKPTEST)
Item 14 (2-point scale): REQCLASS_Rev = (2 - REQCLASS)
```

The process for converting the original scale for each item to a 0-1 scale is the same, varying only by the number of response options for any given item. The math for converting each item is presented below. (Note: The lowest numeric response value for all items in the original scale is 1 and the highest value is 2 for Items 12 and 14 and 5 for the Item 21.)

```
Item 12a (2-point scale): REQPTEST_RevRaw = REQPTEST_Rev
Item 12b (2-point scale): TKPTEST_RevRaw = TKPTEST_Rev
Item 14 (2-point scale): REQCLASS_RevRaw = REQCLASS_Rev
Item 21a (5-point scale): LNDSTUDY_Raw = (LNDSTUDY - 1) / 4
Item 21b (5-point scale): LNDACAWK_Raw = (LNDACAWK - 1) / 4
Item 21c (5-point scale): LNDSKLLS_Raw = (LNDSKLLS - 1) / 4
```

The new recoded variables can now be used to calculate the raw individual-level benchmark scores. This is simply a matter of computing the average of the six recoded items in this scale:

```
COLLREAD = (REQPTEST_RevRaw + TKPTEST_RevRaw + REQCLASS_RevRaw + LNDSTUDY Raw + LNDACAWK Raw + LNDSKLLS Raw) / 6
```

The final step is creating the raw benchmark score for a given population subgroup. This is accomplished by computing the weighted average of the raw benchmark score [COLLREAD] for all respondents in the subgroups of interest. (IMPORTANT NOTE: If your population subgroup is based on enrollment status (full-time and part-time) or gender identity (man and woman), then the weight should not be used; instead the unweighted average should be calculated. See "When To Use Weights" for a more detailed discussion of using weights in analyzing SENSE data.)

The raw benchmark variable for the Effective Track to College Readiness benchmark is COLLREAD and the standardized benchmark variable is COLLREAD\_STD. Computation of a population subgroup standardized benchmark score follows the same procedure as just described for the raw subgroup population benchmark score substituting COLLREAD\_STD for COLLREAD.

#### **Engaged Learning**

Items 19a, 19b, 19e, 19g, 19h, 19l, 19j, 19k, 19L, 19m, 19n, 19o, 19q, 20d2, 20f2, and 20h2

The Engaged Learning benchmark [ENGAGLRN] does not include any items that require reverse coding.

The process for converting the original scale for each item to a 0-1 scale is the same, varying only by the number of response options for any given item. The math for converting each item is presented below. (Note: The lowest numeric response value for all items in the original scale is 1 and the highest value is 4 for all items.)

```
Item 19a
              (4-point scale): ASKQUES_Raw = (ASKQUES - 1) / 3
Item 19b
              (4-point scale): PREPDRFT Raw = (PREPDRFT - 1) / 3
              (4-point scale): SUPINSTR Raw = (SUPINSTR - 1) / 3
Item 19e
              (4-point scale): PINCLASS_Raw = (PINCLASS - 1) / 3
Item 19g
Item 19h
              (4-point scale): PREPOUTC_Raw = (PREPOUTC - 1) / 3
Item 19i
              (4-point scale): GRPSTUDY Raw = (GRPSTUDY - 1) / 3
              (4-point scale): NRGSTUDY Raw = (NRGSTUDY - 1) / 3
Item 19i
              (4-point scale): USEINTMG Raw = (USEINTMG - 1) / 3
Item 19
              (4-point scale): MAILFAC_Raw = (MAILFAC - 1) / 3
Item 19
Item 19m
              (4-point scale): FACASSN Raw = (FACASSN - 1) / 3
Item 19n
              (4-point scale): CLASSREL_Raw = (CLASSREL - 1) / 3
Item 19o
              (4-point scale): FEEDBACK Raw = (FEEDBACK - 1) / 3
              (4-point scale): FACIDOC Raw = (FACIDOC - 1) / 3
Item 19q
Item 20d2
              (4-point scale): FFTUSE_Raw = (FFTUSE - 1) / 3
              (4-point scale): SKLABUSE Raw = (SKLABUSE - 1) / 3
Item 20f2
Item 20h2
              (4-point scale): COMLBUSE_Raw = (COMLBUSE - 1) / 3
```

The new recoded variables can now be used to calculate the raw individual-level benchmark scores. This is simply a matter of computing the average of the sixteen recoded items in this scale:

```
ENGAGLRN = (ASKQUES_Raw + PREPDRFT_Raw + SUPINSTR_Raw + PINCLASS_Raw + PREPOUTC_Raw + GRPSTUDY_Raw + NRGSTUDY_Raw + USEINTMG_Raw + MAILFAC_Raw + FACASSN_Raw + CLASSREL_Raw + FEEDBACK_Raw + FACIDOC_Raw + FFTUSE_Raw + SKLABUSE_Raw + COMLBUSE_Raw) / 16
```

The final step is creating the raw benchmark score for a given population subgroup. This is accomplished by computing the weighted average of the raw benchmark score [ENGAGLRN] for all respondents in the subgroups of interest. (IMPORTANT NOTE: If your population subgroup is based on enrollment status (full-time and part-time) or gender identity (man and woman), then the weight should not be used; instead, the unweighted average should be calculated. See "When To Use Weights" for a more detailed discussion of using weights in analyzing SENSE data.)

The raw benchmark variable for the Engaged Learning benchmark is ENGAGLRN and the standardized benchmark variable is ENGAGLRN\_STD. Computation of a population subgroup standardized benchmark score follows the same procedure as just described for the raw subgroup population benchmark score substituting ENGAGLRN \_STD for ENGAGLRN.

#### **Academic and Social Support Network**

Items 18I, 18m, 18n, 18o, 18q, 18r, and 18s

The Academic and Social Support Network benchmark [ACSOCSUP] does not include any items that require reverse coding.

The process for converting the original scale for each item to a 0-1 scale is the same, varying only by the number of response options for any given item. The math for converting each item is presented below. (Note: The lowest numeric response value for all items in the original scale is 1 and the highest value if 4.)



```
Item 18I (5-point scale): RESOURCE_Raw = (RESOURCE - 1) / 4
Item 18m (5-point scale): GRADEPOL_Raw = (GRADEPOL - 1) / 4
Item 18n (5-point scale): SYLLABI_Raw = (SYLLABI - 1) / 4
Item 18o (5-point scale): FACMEET_Raw = (FACMEET - 1) / 4
Item 18q (5-point scale): OSTUDNAM_Raw = (OSTUDNAM - 1) / 4
Item 18r (5-point scale): FACNAM_Raw = (FACNAM - 1) / 4
Item 18s (5-point scale): STUNAM_Raw = (STUNAM - 1) / 4
```

The new recoded variables can now be used to calculate the raw individual-level benchmark scores. This is simply a matter of computing the average of the seven recoded items in this scale:

```
ACSOCSUP= (RESOURCE_Raw + GRADEPOL_Raw + SYLLABI_Raw + FACMEET_Raw + OSTUDNAM Raw + FACNAM Raw + STUNAM Raw) / 7
```

The final step is creating the raw benchmark score for a given population subgroup. This is accomplished by computing the weighted average of the raw benchmark score [ACSOCSUP] for all respondents in the subgroups of interest. (IMPORTANT NOTE: If your population subgroup is based on enrollment status (full-time and part-time) or gender identity (man and woman), then the weight should not be used; instead the unweighted average should be calculated. See "When To Use Weights" for a more detailed discussion of using weights in analyzing SENSE data.)

The raw benchmark variable for the Academic and Social Support Network benchmark is ACSOCSUP, and the standardized benchmark variable is ACSOCSUP \_STD. Computation of a population subgroup standardized benchmark score follows the same procedure as just described for the raw subgroup population benchmark score substituting ACSOCSUP STD for ACSOCSUP.

#### When to Use Standardized or Raw Benchmark Scores

CCCSE creates two types of benchmark scores based on member colleges' data: raw and standardized. Both types of benchmarks are useful, but for different purposes. The standardized *SENSE* benchmark scores are useful for comparing any given college to the cohort at one point in time. Standardized benchmark scores can also be used to compare subgroups within each college: a subgroup (e.g., full-time students) with a standardized benchmark score of 52 will be more engaged on that benchmark than a subgroup (e.g., part-time students) with a standardized score of 47.

For colleges wishing to conduct longitudinal analyses of trends at their campus, the raw benchmark scores are the appropriate measures to use. The standardized benchmark scores are not appropriate for longitudinal analysis as they are recalculated every year and are based on the distribution of responses for each annual 3-year cohort. The raw benchmark scores, on the other hand, are always on a 0-1 scale and are not affected by fluctuations in the distribution of national responses from year to year.

## When to Use Weights

Beginning with SENSE 2021, colleges can choose to administer SENSE on paper in the classroom or online. Consistent with prior years, the in-class survey is administered to a sample of all eligible classes. Since full-time students, by definition, are enrolled in more classes than part-time students, the former have a higher probability of being sampled. This introduces a sampling bias resulting in disproportionately more full-time students receiving the survey. Statistical weighting, discussed below, is used to adjust for this bias.



For the online survey, sampling is not employed; the survey invitation is sent to all eligible students. Research shows that full-time students and women are more likely to complete online surveys than part-time students and men. As a result, full-time students and women are overrepresented in the final sample. Statistical weighting, discussed below, is used to adjust for this bias. Analysis of the *SENSE* respondents nationally reflect this bias among online respondents.

The disproportionality of respondents compared to institutional populations is addressed using weights based on the most recently released IPEDS enrollment status and gender counts (FT women, FT men, PT women, PT men). An additional change from prior years is the IPEDS data are limited to first-time at the college and transfer in degree-/certificate-seeking fall enrollment rather than the total fall enrollment to better align with the *SENSE* population. These new weights are applied to both paper-and-pencil and online survey results. College data sets include a variable called IWEIGHT, which contains the appropriate weight for each respondent. This variable is also used in the *SENSE* online reporting tool.

Because weights are based on enrollment status and gender, analysis of *SENSE* data in which enrollment status and/or gender are used as a breakout variable should remain unweighted; this is automatically handled in the online analysis tool. For example, when comparing subgroups of part-time first-generation students to full-time first-generation students, weights should not be used; the same applies if the comparison is between first-generation men and first-generation women. On the other hand, when analyzing first-generation students compared to not first-generation students and both full-time and part-time students and women and men are included in both groups, for example, weights should be applied. When analyzing simple demographics (e.g., number of men and women, number of respondents by race), weights should not be used to obtain the actual number of respondents in each demographic group.

As noted above, weights are determined using the most recent publicly available IPEDS data. As IPEDS data are approximately two years old, they may not accurately represent a college's current student population. For example, in the case that a college has experienced a significant change in enrollment characteristics during the two years prior to administering *SENSE*, the college's institutional research department may want to consider whether the weights based on IPEDS numbers are completely appropriate.

A final example of when to consider not using weights is when the vast majority of students at a college are either full-time or part-time and either men or women. For example, if a college has 92% full-time students the college may want to look at unweighted results; the same would apply if the institution population was comprised of 92% women. However, if the institutional distribution of either enrollment status or gender is not dominated by one group, then the decision to not use weights should be carefully considered.

CCCSE encourages each member college to carefully compare the student characteristics of its *SENSE* respondents with the characteristics of its student population from which the sample was drawn to evaluate the effect of a possible sampling or respondent bias.