Effect of Changes to Benchmark Items on *CCSSE* 2017

With the introduction of the 2017 version of the Community College Survey of Student Engagement (*CCSSE*), the Center made modifications to items in three of the *CCSSE* benchmarks: Student Effort, Academic Challenge, and Support for Learners.

To test the effect of these changes on the benchmark scores, administration data were reviewed for the 2017 participating colleges that had administered *CCSSE* three or more times prior to 2017. This limited the comparison data set to 254 out of 297 institutions. Pre-2017 data were limited to the most recent three administrations for each institution so that no institution had an inordinate influence of the analysis. For some of the colleges, the most recent three years of data were 2014, 2015, and 2016; for others, the three most recent administrations stretched back to 2005 with two additional administrations between 2006 and 2016. These data were merged with the response data from 2017 for a total of four time points.

The change to the Academic Challenge benchmark was the replacement of the "synthesizing" item (5c) because too many students who participated in cognitive interview testing of the refreshed survey did not know what the word "synthesize" meant. Because the item was completely changed, there is no way to directly compare the new item with the old item. Effects of this change on the benchmark, however, can be evaluated. When examining the trends across the four administrations, the 2017 benchmark scores appear to be in line with the pre-2017 score trends. Therefore, we can conclude that replacing this one item has had no appreciable impact on the Academic Challenge benchmark score.

The Student Effort and Support for Learners benchmarks include five items that ask students about their use of support services (Part 1 of Item 12). The wording of the items was not changed, but the response scale was changed. The five items, organized by benchmark are:

Student Effort

- 12.1d Peer or other tutoring
- 12.1e Skill labs (writing, math, etc.)
- 12.1f Computer lab

Support for Learners

- 12.1a Academic advising/planning
- 12.1b Career counseling

Table 1 shows how the response scale was changed.

Table 1. Response scale changes for Part 1 of item 12.

Old responses	New Responses
0 = Don't know / N.A.	0 = Never
1 = Rarely / Never	1 = 1 time
2 = Sometimes	2 = 2–4 times
3 = Often	3 = 5 or more times

This change was made to eliminate ambiguity in the old response scale ("Don't know/N.A." and "Rarely/Never"). In both cases, the two options in the old response categories are qualitatively and/or quantitatively different; not knowing is different from not applicable and "Never" is different from "Rarely." With these combined response categories, it is impossible to know which of the options the student was selecting. Additionally, "Rarely" may not mean the same for use of academic advising as it does for use of tutoring services. Given these ambiguities, we have replaced the old response options with values that we believe to be more concrete and actionable for colleges.

Because the items themselves did not change, there were two options for calculating the Support for Learners and Student Effort benchmark scores for the 2017 administration:

1. Collapse the "Never" and "1 time" responses into a single response in order to mimic the "Rarely/Never" response option on the old survey and use this to calculate the benchmarks. *(Since this version is attempting to mimic the old benchmark items, this will be referred to as the "old benchmark.")*
2. Calculate the benchmarks using the full 4-level response set. *(Since this version uses the full response scale for the refresh, this will be referred to as the "new benchmark.")*

Variables were created to represent both of these scenarios and to create two sets of raw benchmark scores for Student Effort and Support for Learners. These scores were combined with the pre-2017 data, and trends were examined for both sets of 2017 benchmark scores.

Table 2 presents the distribution of responses: Raw benchmark scores were calculated for each college, and then the college-level benchmark scores were analyzed. Across the 254 colleges, the average differences between raw benchmark scores using the two different methods of calculation are very small (even considering the range of scores is zero to one).

Table 2. Distribution Statistics for Benchmark Changes Between Administrations

Benchmark Comparisons[1]	Old / New Benchmark	Mean	Standard Deviation	Skewness	Kurtosis
STUEFF 1-2	Old	0.0037	0.0201	0.7630	3.1770
STUEFF 2-3	Old	-0.0012	0.0218	-2.3068	16.8921
STUEFF 3-7	Old	0.0282	0.2419	-0.1400	3.4563
	New	-0.0237	0.0203	0.2475	1.5102
SUPPORT 1-2	Old	0.0161	0.2855	0.4294	2.6944
SUPPORT 2-3	Old	0.0088	0.0301	-0.7201	4.2516
SUPPORT 3-7	Old	0.0434	0.0288	0.2247	0.7473
	New	0.0025	0.0269	0.2080	0.9844

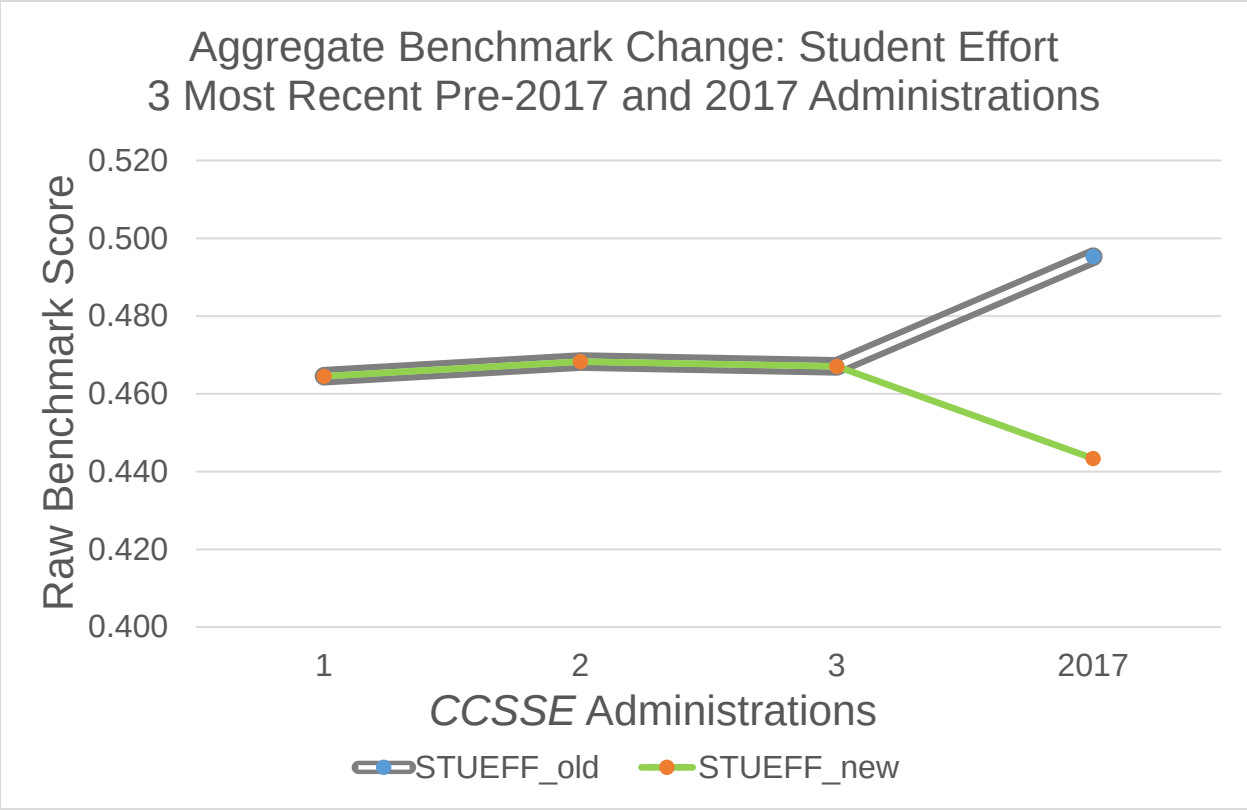
Source: ccsse_ref103._bmrkUniv.rtf.

Note: [1]The "Benchmark Comparisons" column lists the benchmark abbreviation and the change between administrations. For example, STUEFF 1-2 represents the change between the earliest and second-earliest administration and STUEFF 3-7 represents the change between the most recent pre-2017 administration and the 2017 administration.

The average benchmark mean changes between these two calculation methods are about the same size in magnitude, but the direction of change is opposite (with the old version being a positive change and the new version being a negative change). Choosing the better calculation

method was not clear due to the small number of pre-2017 data points and the opposite change in the mean trend line for STUEFF 3-7. However, taking into account the other statistics presented in Table 2 (standard deviation, skewness, and kurtosis) when using the full response scale generates a more normally distributed benchmark score. Therefore, the calculation of the STUEFF benchmark using the new computational approach appears to provide the better estimate. For a different perspective on the trends and the effect of the different computation methods, Figure 1 shows the aggregate trend lines for all colleges included in this analysis across the four administrations. Because the magnitude of change between the two calculation methods is almost exactly the same, the new method was selected as the more conservative alternative. Center research staff will continue to monitor this over the next couple of years as the full three-year cohort is established.

Figure 1. Trends for Three Most Recent Pre-2017 Student Effort Raw Benchmark Scores and the 2017 Score Using Different Computation Methods

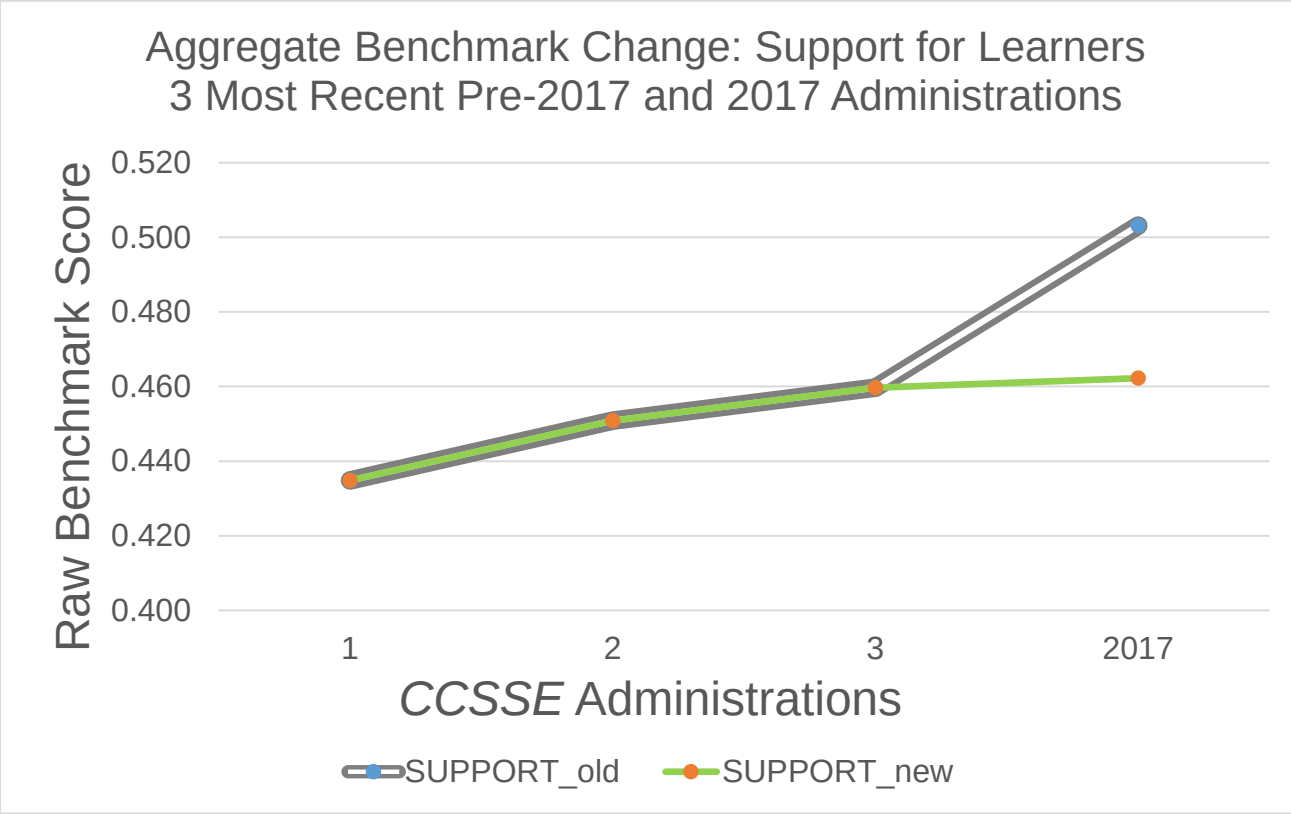


Source: ccsse_ref104._bmrkmns.xlsx.

Turning to the Support for Learners (SUPPORT) benchmark, the trend of inter-administration changes is positive throughout for both methods of computation. The calculations for changes between the most recent pre-2017 administration and the 2017 administration, however, diverge, depending on how the benchmark is calculated. Using the combined "Never" and "1 time" ("Old") in the benchmark formula shows a marked increase in the benchmark score (see the hollow grey line in the Figure 2). Yet the resulting trend based on using the full response scale to calculate the benchmarks is very consistent with the pre-2017 trend line (see solid

green line in Figure 2). Considering the results of the two approaches to calculating the 2017 SUPPORT benchmark, other than the difference in the mean value, there is little difference between each of the other three statistics reported in Table 2. It is entirely possible, that the trend for SUPPORT would have shown the change represented by the grey line in Figure 2 had no change to the survey been made; however, the green line could also be a possible viable outcome. Visually, the green line appears to be more consistent with the pre-2017 trend. The statistical data in Table 2 and the graphical presentation in Figure 2 helped confirm the decision for the Student Effort benchmark to use the full response scale to compute the benchmark scores.

Figure 2. Trends for Three Most Recent Pre-2017 Support for Learners Raw Benchmark Scores and the 2017 Score Using Different Computation Methods



Source: ccsse_ref104._bmrkmns.xlsx.

In conclusion, the analyses presented here suggest that the better option for computing benchmark scores is to use the full response scale for the new response option set for the student services items included in Item 12 on the refreshed *CCSSE*. For a detailed description of the benchmarks for the refreshed *CCSSE*, please see How Benchmarks are Created: *CCSSE* 2017 - Present.

A longer-term analysis of the benchmark scores could be more informative by the inclusion of a longer baseline for the intra-administration changes. However, further restricting the colleges to those with four or more pre-2017 administrations would result in the exclusion of 30% of the colleges in the 2017 sample. Exclusion of so many colleges from the analysis could bias these

results as this would likely increase the proportion of high-performing colleges in the analysis. The analysis limited to those colleges with three or more pre-2017 administrations of *CCSSE* only excluded 17.5% (n=52) of the 2017 sample.