Overview of the *CCSSE* Instrument and Psychometric Properties

C. Nathan Marti
The Community College Survey of Student Engagement

The Community College Survey of Student Engagement (*CCSSE*) was developed to measure and assess the extent to which students are engaged in good educational practices at community and technical colleges. Data obtained from the *CCSSE* instrument, the *Community College Student Report* (*CCSR*), are intended to be used to understand student engagement and to serve as a tool for improving teaching and learning. These data allow colleges both to look inward by making comparisons across various areas of student engagement and between constituents within a college and to look outward, benchmarking their performance in relation to student engagement at other two-year colleges.

The *CCSR* was adapted from the National Survey of Student Engagement (NSSE), with permission from Indiana University. The NSSE instrument was developed in 1999 for use in four-year colleges and universities. There is a high degree of intentional overlap between the NSSE and *CCSSE* instruments. Of the 79 items measuring student engagement on the NSSE instrument, 56 of those items appear on the 2003 version of the *CCSR*, representing a 71% overlap between the two instruments. Psychometric properties of the NSSE instrument have been explored extensively and have demonstrated that the instrument is reliable and valid (Kuh, Hayek, Carini, Ouimet, Gonyea, & Kennedy, 2001; Kuh, 2002).

An emphasis on quality educational practices is important at all levels of community colleges. Engaging activities enhance students' educational experiences, which means that they learn more and learn more effectively ─ meaning that they learn more and learn more effectively. The amount of time and energy that students spend engaged in educationally meaningful activities has consistently been shown to be related to desired learning and personal development outcomes (Pace, 1980). For educators, understanding how their students are engaged provides new possibilities for enhancing their teaching through using techniques that get students involved. For administrators, having the holistic view of student engagement on their campus provides an opportunity to consider campus-wide initiatives to enhance the learning experiences at their institutions.

One of the primary uses of the *CCSR* is to benchmark a college's performance in key areas of student engagement. The purpose of benchmarking is to look at others that are engaged in similar practices and use that information to set and achieve goals for improvement. This is fundamental for understanding an institution's performance as it situates its results in a relevant context. One limitation to self-studies is that without external comparisons, there is no benchmark to establish what is typical or exemplary in a larger context; and, as a result, self-studies beg the question of how other similar institutions are performing.

The desirability of making internal and external comparisons requires that a survey instrument be administered to a sample that is representative of the population from which it was drawn and that the instrument itself be both reliable and valid. The administration of the *CCSR* uses a sampling methodology that is consistent across all participating institutions, thus making the results a viable tool for benchmarking student engagement. Benchmarking with data collected through *CCSSE* provides colleges an opportunity to make use of the national picture to set goals for themselves.

Overview of the *CCSR*

The *CCSR* is designed to measure student engagement, and most items on the survey pertain to time spent on activities that previous research has shown to be related to desired outcomes of a college education. There are five series of items that directly measure educational engagement. The College Activities section uses twenty items to measure the frequency with which students engage with instructors, other students, and in classroom activities. For example, items measure the frequency with which students ask questions in class, work with other students inside and outside of class, and speak with instructors about grades or career plans. The Mental Activities section has six questions on the extent to which course work emphasizes activities such as analyzing the basic elements of an idea, synthesizing ideas, and making judgments about information and arguments. Academic Preparation items measure the number of textbooks assigned, the number of non-assigned books read, and the number of papers written. Opinions about Your College is a set of seven items that measure the extent to which a college emphasizes providing social support, exposure to diverse backgrounds, and financial support. Student Services items measure the frequency, satisfaction, and importance of eleven services, such as academic advising, tutoring, and financial aid advising.

In addition to the five series of items pertaining to engagement, there are several other related item sets. A series of fifteen Educational and Personal Growth items asks students to rate how much their college has contributed to their development in areas such as writing, speaking, solving numerical problems, understanding themselves, and acquiring work-related knowledge and skills. Student Goals are measured with a series of items that ask students if their educational goals are career-related, to transfer to a four-year institution, to obtain a certificate or degree, and so on. Student involvement in other educational experiences is measured with eleven items that ask students if they have taken developmental/remedial, ESL, or honors courses. There are also several questions about the students' demographics, goals, financial arrangements, prior education, and support. These items allow comparisons in engagement and other areas between important subgroups served by the two-year sector.

Sampling and Administration

Ensuring that a sample is representative of a larger population is an important requirement for generalizing responses to that population. A standard sampling methodology is used across all colleges that participate in *CCSSE*, so that participating colleges can have confidence that the results from the survey generalize to their population of students and are comparable to results from other institutions.

A stratified random cluster sample scheme is used for each participating institution. Samples are pulled from a list of all credit courses at an institution. Each class represents a cluster, as it contains multiple students. The stratification is conducted at three levels based upon the time of day in which the class begins: (1) 11:59 a.m. and earlier, (2) 12:00 p.m. to 4:59 p.m., and (3) 5:00 p.m. to 12:00 a.m.

Examination of self-reported demographic variables obtained from the *CCSR* compared with institution-level data reported to the Integrated Postsecondary Education Data System (IPEDS) revealed that there was a very close match between the proportions of race, sex, and age in the 2003 *CCSSE* sample and the numbers reported to IPEDS for the institutions. (See Table 1 for a comparison of key demographic variables.) However, the proportion of part-time to full-time students was notably different for the sample and the population. This is undoubtedly due to the sampling methodology ─ because full-time students by definition are enrolled in more classes, they have a greater probability of being sampled. To correct for this effect, statistics are weighted by part-time and full-time status to more accurately estimate campus-wide statistics.

Table 1: Comparison of the 2003 *CCSSE* Sample to Underlying Population

	CCSSE 2003 Sample	*CCSSE* 2003 Population
Gender		
Female	59.6%	57.5%
Male	40.4%	42.5%
Race/Ethnicity		
White	67.1%	64.6%
Asian	5.2%	5.5%
Hispanic	10.3%	13.4%
Black	10.3%	11.3%
Native American	3.5%	0.9%
Other	3.3%	4.2%
Enrollment Status		
Full-time	67.4%	35.6%
Part-time	32.6%	64.4%
Student Age		
18 to 29	73.3%	67.6%
30 to 39	14.3%	16.2%
40 to 49	8.8%	10.3%
50 and over	3.6%	5.8%

The method of survey administration ensures that there will be a high response rate and that the response rate will not be systematically biased. Survey administration takes place in the classroom during regularly scheduled class meeting times and is not announced to the students in advance. In addition to producing a higher response rate than purely voluntary surveys, classroom administration avoids a non-respondent bias. Such a bias exists when portions of the population are not sampled and thus bias the population estimate. For example, in a survey administered via the Internet, non-response can be correlated with an extraneous variable, such as access to the Internet, which will mean that non-respondents are systematically different from respondents. Classroom administration also ensures that the survey will obtain similar response rates across administration years and is therefore not as sensitive to historical fluctuations in survey response rates. For example, over-surveying during a period of increased use of the Internet as a mode of survey administration has resulted in decreases in survey responses (Sax, Lee, Hagedorn, Gilmartin, & Gale, 2003). The consistency across time and equal accessibility to all students is perhaps the principle benefit of the classroom administration as there is little reason to expect that the Internet per se would bias responses; examination of Internet and paper survey respondents in the four-year sector show slightly more favorable responses from Internet respondents, but results ultimately differ little in terms of effect size (Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003). Because the *CCSR* is administered in the classroom, survey administrators have a captive audience, one less able to decline participation or simply fail to find the time to participate.

Exploration of Constructs Underlying the *CCSR*

One of the primary intended uses of the *CCSR* was to establish meaningful groups of items that could be used as benchmarks of effective educational practices. Prior to constructing benchmarks of effective educational practice, extensive analyses were conducted on the items to understand how they empirically related to each other. To achieve this, the first phase of analysis for the 2003 survey data was exploratory and intended to enhance understanding of the empirical relationships between items through the use of correlation matrices and exploratory factor analysis (EFA). The results of exploratory analyses were solely intended to inform factor structures to be tested in confirmatory factor analyses (CFA). The results of the CFA were then used as the empirical basis of the *CCSSE* benchmarks.

There were forty-nine items originally considered for use in the benchmarks. These included all items from the College Activities, Mental Activities, Academic Preparation, Opinions about Your College, and the frequency of use items in the Student Services section. In addition, items on the extent to which exams are challenging and amount of time spent preparing for class were included in the factor analysis.

Exploratory Factor Analysis

Prior to construction of the *CCSSE* benchmarks, exploratory techniques such as correlation matrices and exploratory factor analyses (EFA) were used to assess the strength of relationships between variables. In the first phase of analysis, EFA was used to examine relationships between variables, to identify the number of factors present in the data, and to identify items that either did not load on a factor or loaded on more than one factor. A variety of rotations (e.g., varimax, direct oblimin) were used to determine the number of factors, to identify items that consistently loaded together, and to identify items whose loadings were sensitive to rotation methods. Results for the EFA are not reported as this phase was purely exploratory and intended only to inform decisions about the factor structure for the subsequent CFA models.

Confirmatory Factor Analysis

Confirmatory factor analysis was used to establish a model with the closest fit to the data. Confirmatory factor analysis is an application of structural equation modeling in which items are associated *a priori* with factors, and the adequacy of a model is tested through fit indices that measure the degree to which the factor model reproduces the empirical covariance matrix. All CFA models were constructed using the Mplus application (Muthén & Muthén, 2000). The data from 2003 represented approximately 65,300 surveys from students enrolled in 93 community and technical colleges in 31 states. After excluding respondents that did not indicate whether they were part-time or full-time; did not indicate that it was the first time they had taken the survey; did not take the survey in the class to which that survey had been assigned; or were not part of an over-sample that did not conform to the prescribed sampling frame, there were 53,358 respondents used in the CFA. Models developed from participants in the 2003 administration of the *CCSSE* survey were tested on data from the 2002 *CCSSE* field test. There were 31,545 participants in the field test[1].

It should be noted that the instrument was designed to measure engagement holistically and not with an intended underlying factor structure. Thus, while the empirical work is important for understanding relationships between items, the goal is not necessarily to confirm or deny a particular factor structure. Despite this, CFA offers a great deal of value through goodness-of-fit tests comparing sub-groups and year-to-year comparisons as the technique assesses the degree to which the specified structure models the observed data and provides information about the appropriateness of using models across various subpopulations represented in the two-year sector.

After examination of exploratory analyses, a model with nine factors and 39 items was specified. The Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Residual (SRMR) fit indices were used following the two-index presentation strategy recommended by Hu and Bentler (1999). The two-index presentation strategy reflects the fact that several fit indices have been shown to be highly correlated and thus provide somewhat redundant information. To compensate for this, two indices that are not highly correlated should be used. RMSEA and SRMR were chosen because these indices have been shown to be dissimilar under various sample sizes, distributional violations, and model misspecifications. Using the combinatorial cutoff of RMSEA < .06 and SRMR < .09 minimizes Type I and Type II error rates and was thus selected for evaluating all models. Using this rule, the nine factor structure suggested by the earlier EFA was confirmed as having good model fit (RMSEA = .053, SRMR = .060). Coefficients and the factor structure are shown in Table 2.

[1] Because all EFA and CFA procedures require complete data within respondents, a Markov Chain Monte Carlo imputation was used to replace missing values.

Table 2: Coefficients, Standard Errors, and Alphas for nine factor CFA

Item	Parameter Estimate	Standard Error	Estimate/ S. E.	Standardized Coefficient
Faculty Interactions (α = .73)				
BFACGRAD	1	0	0	0.521
BFACPLAN	1.094	0.009	116.551	0.571
BFACIDEA	1.018	0.009	116.69	0.531
BFACOTH	0.666	0.007	93.735	0.347
BFACFEED	0.740	0.008	87.711	0.386
BCLQUEST	0.655	0.008	80.584	0.342
Class Assignments (α = .64)				
BREWROPA	1	0	0	0.638
BINTEGRA	1.123	0.009	118.492	0.716
BCLPRES	0.748	0.008	96.335	0.477
Exposure to Diversity (α = .74)				
BDIVRSTU	1	0	0	0.865
BDIFFSTU	0.994	0.006	159.245	0.860
BOOCIDEA	0.465	0.005	95.958	0.402
Collaborative Learning (α = .61)				
BCLASSGR	1	0	0	0.457
BOCCGRP	1.279	0.014	90.309	0.585
BCOMMPRO	0.630	0.008	75.289	0.288
BTUTOR	0.750	0.010	77.637	0.343
Information Technology (α = .54)				
BITACADE	1	0	0	0.619
BEMAIL	1.025	0.013	77.190	0.635
Mental Activities (α = .83)				
BANALYZE	1	0	0	0.581
BSYNTHSZ	1.139	0.007	158.438	0.662
BEVALUAT	1.126	0.007	151.488	0.654
BAPPLY	1.174	0.007	157.859	0.682
BPERFORM	0.982	0.007	132.966	0.570
BWORKHRD	0.587	0.007	85.204	0.341
School Opinions (α = .78)				
BENVSPRT	1	0	0	0.511
BENVDVRS	1.296	0.011	120.551	0.662
BENVNACD	1.370	0.011	129.676	0.699
BENVSCL	1.413	0.011	133.592	0.721
BFINSUPP	0.987	0.011	88.443	0.504
BENVSCHL	0.715	0.008	86.599	0.365
Student Services (α = .67)				
USEACAD	1	0	0	0.415
USECACOU	0.917	0.009	96.757	0.380
USETUTOR	0.811	0.009	86.786	0.336
USELAB	0.995	0.011	90.104	0.413
USECOMLB	0.791	0.011	72.403	0.328
Academic Preparation (α = .56)				
BREADASG	1	0	0	0.419
BWRITE	1.285	0.015	85.081	0.538
BEXAMS	0.472	0.008	55.575	0.198
BACADPR0	0.676	0.009	72.520	0.283

Multiple Group Analyses

Having established that there is a latent variable model underlying the *CCSR* with close fit to the data, the next step of analysis modeled data across multiple groups to ensure that the factor structure was consistent across major subgroups. Three multiple group analyses were undertaken to test goodness-of-fit across subgroups within the sample: (1) respondents from the 2002 field test were compared with 2003 respondents; (2) males were compared with females; and (3) part-time students were compared with full-time students. Multiple group analysis tests the adequacy of model fit by constraining certain parameters to be equal across the groups that are being compared. All of the multiple group models constrained factor loadings and factor variances to be equal across both groups.

The multiple group model containing data from the *CCSR* 2002 pilot test and the 2003 data resulted in a good model fit (RMSEA = .051, SRMR = .055). Both values were below their respective cutoffs, indicating that the model fit well with the constraints imposing equality of factor loadings and variances across data from the two years. As with the multiple group analysis comparing 2002 and 2003 data, model fit was very similar across male and female students (RMSEA = .051 and SRMR = .056) and part- and full-time students produced good model fit as well (RMSEA = .050 and SRMR = .056).

The results of these three models show that the *CCSR* is measuring the same latent constructs across subpopulations in the sample. Perhaps the most critical of the three multiple group analyses is the year to year comparison, as it demonstrates that the instrument is measuring the same constructs in different years and can thus be used to track changes across time. Other analyses have shown that there are differences in overall levels of engagement between male and female students. Finding that a model that constrains factor loadings to be the same across two groups maintains the close fit that was observed in the single group model and suggests that differences across groups are in levels of engagement, rather than structural differences. Demonstrating the assumption of identical factor structures across years and subpopulations within community colleges is critical for research efforts using factors derived from the instrument, given that many of the most interesting analyses are comparisons between various subgroups within the larger population.

Interpretation of the Factors

Having constructed a model that closely reproduced the empirical covariance matrix, factors in each major section of the survey were examined to determine their face validity and to understand how items grouped together.

<u>College Activities</u>

The first twenty items on the *CCSR* are intended to measure the frequency with which college students engage in practices which previous research shows are related to higher levels of student learning and persistence (Kuh, 2002). With few exceptions, items in this portion of the survey load on readily interpretable latent constructs. One item, *Come to class without competing readings or assignments*, did not correlate well with other items on the survey and was therefore not included in subsequent factor analyses. Another item, *Worked harder than you thought you could to meet an instructor's standards or expectations*, was included with items from the Mental Activities scale. Eighteen of the items were used to form five factors that were readily interpretable. These five factors are interpreted as representing the following latent constructs: Faculty Interactions, Class Assignments, Collaborative Learning, Information Technology, and Exposure to Diversity

There are few surprises in the manner that the variables group together in the College Activities portion of the survey. While there are disparate numbers of items loading on the factors, this is likely a reflection of the survey instrument's greater focus on some areas in contrast to others. Items on the Faculty Interactions factor generally pertain to verbal interactions between students and faculty. Asking questions in class is not as obviously related, but it is not surprising that students who are more likely to talk with faculty are also more likely to engage in the classroom by asking questions. Faculty feedback tends to have weaker correlations than other correlations within the scale; likely as a result of feedback being controlled by the faculty rather than the student. It is nevertheless related, likely indicating that students are more likely to engage with faculty who do provide prompt feedback. The items related to Class Assignments measure effort in preparing and presenting class-related materials and include items on writing more than one draft of a paper, integrating information in a paper, and classroom presentations. Collaborative Learning includes items on working in groups with other students in and out of class as well as tutoring. The relatively poor alpha for this factor is likely a reflection of the fact that activities such as these are some of the least common activities and thus the extent to which students engage in them varies greatly. The Information Technology factor was comprised of the only two items dealing with information technology; it also had a low alpha, which is undoubtedly a result of the scale only having two items ─ despite a strong correlation between these items ─ because alphas are penalized for paucity of items. Exposure to Diversity measures interactions with students of different backgrounds and interactions with people outside of the classroom.

Mental Activities

Items on the Mental Activities scale were all highly correlated with each other and formed an easily interpretable factor dealing with academic challenge. One item, *Memorizing facts, ideas, or methods from your courses and readings so you can repeat them in pretty much the same form*, was not included in the analysis because the goal of the analysis was to establish benchmarks of good educational practice ─ and memorization is not widely considered such. In addition to the Mental Activities items, the aforementioned item, *working hard to meet an instructor's expectations*, was included on this scale.

College Opinions

The items on Supportiveness of College Environment all deal with the extent to which colleges emphasize supportiveness in areas that foster a quality learning environment. Items address encouragement to study, contact with diversity, coping with non-academic responsibilities, social support, financial support, and using computers. All of the items on this scale were well correlated with the exception of *using computers in academic work*. It is not obvious why this item differed, though there is a large range in the importance of computers across academic programs and institutions, whereas Supportiveness of College Environment items may be of relatively even importance across programs of study and two-year institutions.

Academic Preparation

The Academic Preparation items measure the amount of reading and writing that students do in response to course requirements and the amount of reading they do on their own. The three items on this scale formed a factor that measured the volume of effort directed toward such endeavors. In addition, an item measuring *number of hours spent preparing for class* was combined with the three Academic Preparation items as exploratory analysis indicated that it was empirically related.

Academic Support

There were eleven Student Services items, though only the five that were directly related to academic engagement were used in the factor analysis. These included academic advising, career counseling, skill labs, computer labs, and tutoring. These represented five of the seven most frequently used services; with the other two being financial aid advising and transfer credit assistance.

Limitations of Factor Analysis with the *CCSR*

The usefulness of factor analysis alone in creation of the *CCSSE* benchmarks is limited by the fact that the survey was not designed to measure a set of latent constructs defined *a priori.* Instruments specifically designed to represent a set of underlying factors construct each item with the intent of that item contributing to a latent construct that is derived from each of the items that contribute to it; for example, an IQ test with a latent construct of "spatial ability" will have several items specifically designed to measure spatial ability. Also, there is typically an even balance of items for each construct. When applying factor analysis to an instrument in which each item has been designed to load on a particular latent construct, factor indeterminacy is not as difficult of a problem for interpreting results, as the analysis largely serves to determine whether there is empirical evidence that items designed to load together actually do load together. Without an *a priori* number of factors, it is difficult to determine the best number of factors underlying the set of items, especially given that engaged students often tended to be engaged across several latent constructs. In other words, the confirmatory factor analysis models used assume orthogonality, and it is evident that the dimensions of student engagement are non-orthogonal.

Another way in which the *CCSR* differs from classical applications of factor analysis is that the practices being assessed are not necessarily under the control of a single agent. IQ tests, for example, are measures of various dimensions of intelligence for a particular individual. In contrast, *CCSR* questions measure aspects of engagement that are influenced by students' willingness or ability to engage in the educational activities being measured, but students are also influenced by an instructor's implementation of good educational practices and the institution's development of programs, services, and allocation of resources to support them. For example, a student who would take the initiative to rewrite papers may not be taking classes from faculty that assign papers and thus the item that measures students' engagement in writing multiple drafts of a paper is not entirely under the student's control. While all engagement items measured on the *CCSR* can be influenced at the institutional level, there are multiple forces that influence student engagement, and it is not reasonable to assume that conceptually related items will be empirically related.

For these reasons, the final construction of a smaller number of benchmarks was sought through the advice of a panel of experts in educational research. The benchmarks correspond well with the major latent variables in the CFA, but important items from minor factors were assigned for compelling conceptual and empirical reasons to a condensed set of constructs.

Creation of the *CCSSE* Benchmarks

Factor analytic solutions were informative with regard to student engagement behavior; however, for several reasons noted above, the-nine factor CFA solution was not represented verbatim in the *CCSSE* benchmarks. To establish the final benchmarks, a group of survey research experts (*CCSSE's* Technical Advisory Panel) reviewed the CFA results and then assigned items to benchmarks, taking into account the results of factor analysis and reliability tests ─ and also applying expert judgment based on both the conceptual framework and empirical evidence related to student engagement in undergraduate learning. The objective was to create benchmarks that are reliable, useful, and intuitively compelling to community college educators.

Reliability of the *CCSSE* Benchmarks

Reliability of an instrument is fundamentally defined as the instrument's ability to provide consistent results. Results should be consistent across individuals and over time. Latent constructs, such as the *CCSSE* benchmarks, need to demonstrate that items comprising each construct are reliably measuring the same underlying latent factor. Also, the constructs should measure the same phenomena over time.

Reliability of the latent constructs was measured with Cronbach's alpha. While Cronbach's alpha is widely used to assess psychometric properties of a battery of items, it may not be equally appropriate for each of the *CCSSE* benchmarks. As has been previously noted, the instrument was not designed to measure a set of latent constructs hypothesized *a priori,* and therefore, questions that are conceptually or empirically related may not be measured on an equivalent scale. Furthermore, Cronbach's alpha is designed for one-dimensional concepts and is therefore problematic for scales that have both high and low frequency items that, when they are treated numerically, may not appear to be measuring the same underlying concept. For example, asking questions in class and tutoring are both measures of active and collaborative learning, but asking questions in class occurs far more frequently than tutoring, which from an empirical standpoint would ordinarily suggest that they are not measuring the same concept. Despite the limitations of Cronbach's alpha with the *CCSR*, the benchmark scales had reasonable reliability measures.

The five-construct solution reproduces the empirical covariance matrix reasonably well (RMSEA = .066, SRMR = .066). The RMSEA falls into a range considered adequate fit and the SRMR also was in the range of good fit. Table 3 displays alphas, coefficients and standard errors for the *CCSSE* benchmarks.

Distributional Analysis of *CCSSE* Benchmarks

While showing that distributional properties closely approximate a normal distribution is not necessary for demonstrating the psychometric properties of the *CCSR*, they were examined for each composite benchmark score because the final presentation of the benchmarks was in a standardized format. This requires that the raw values be normally distributed in order to justify the assumption that a value that is *x* less than the mean is equal distance from the mean as a value that is *x* greater than the mean. In order for the standardized scores to be interpreted relative to the mean, it is important that there are approximately equal numbers of respondents that fall above and below the mean for a given benchmark so that there are approximately equal numbers of respondents above and below the mean. For example, a score that is half a standard deviation above the mean should have approximately the same number of individuals between the mean and that score as there are between the mean and a value half a standard deviation below the mean.

Table 3: Coefficients, Standard Errors, and Alphas *CCSSE* Benchmarks CFA

Item	Parameter Estimate	Standard Error	Estimate/ S. E.	Standardized Coefficient
Active and Collaborative Learning (α = .67)				
BOCCGRP	1	0	0	0.498
BCOMMPRO	0.530	0.007	78.247	0.264
BTUTOR	0.602	0.008	78.240	0.300
BCLASSGR	0.858	0.010	86.581	0.427
BCLQUEST	0.712	0.009	78.199	0.354
BCLPRES	0.898	0.010	87.480	0.447
BOOCIDEA	0.861	0.010	83.885	0.429
Student Effort (α = .56)				
BREWROPA	1	0	0	0.614
BINTEGRA	1	0.01	103.286	0.613
BACADPR0	0.363	0.005	66.623	0.223
BCLUNPRE	0.031	0.006	4.957	0.019
BREADOWN	0.249	0.006	39.214	0.153
BUSETUTR	0.511	0.009	59.738	0.314
BUSESLAB	0.620	0.010	62.102	0.380
BUSECLAB	0.651	0.011	61.443	0.400
Academic Challenge (α = .80)				
BANALYZE	1	0	0	0.577
BSYNTHSZ	1.134	0.007	156.14	0.654
BEVALUAT	1.110	0.008	147.86	0.640
BAPPLY	1.152	0.008	153.588	0.664
BPERFORM	0.983	0.007	131.717	0.567
BREADASG	0.335	0.006	55.034	0.193
BWRITE	0.398	0.007	57.919	0.229
BEXAMS	0.359	0.005	71.885	0.207
BENVSCHL	0.591	0.007	87.550	0.341
BWORKHRD	0.642	0.007	92.061	0.370
Student-Faculty Interaction (α = .72)				
BFACPLAN	1	0	0	0.562
BFACIDEA	0.950	0.008	119.615	0.534
BFACOTH	0.629	0.007	96.365	0.353
BFACGRAD	0.925	0.008	114.667	0.520
BEMAIL	0.736	0.009	82.579	0.413
BFACFEED	0.690	0.008	89.131	0.388
Support for Learners (α = .76)				
BENVSPRT	1	0	0	0.507
BENVDVRS	1.287	0.011	118.113	0.653
BENVNACD	1.388	0.011	128.408	0.704
BENVSCL	1.409	0.011	131.015	0.715
BENVFAC	0.494	0.006	77.860	0.251
BUSEACAD	0.680	0.010	68.817	0.345
BUSECACO	0.603	0.009	65.981	0.306

It should be noted that while the composite benchmarks scores should be normally distributed, the individual items that comprise the benchmarks do not need to be normally distributed in order to justify the standardization; therefore, distributional properties of individual items were not examined. In fact, non-normally distributed items can increase the normality of a composite score as highly-engaged students are more likely to reflect engagement on some of the lower frequency items. This property is similar to tests designed to approximate a normal distribution, such as the SAT test, in which only high-performing students answer the most challenging questions correctly; thus, the most challenging questions that are the most likely to be non-normally distributed are the questions that are crucial for producing a normally shaped distribution of exam scores.

Examination of the *CCSSE* benchmark scores in Figures 1 through 5 shows that each of the benchmarks is approximately normally distributed and there are certainly no major divergences from normality[2]. As a whole, there is a very tight fit between the density distribution and the empirical data from which it is derived. Results in Table 4 illustrate that skewness and kurtosis statistics are acceptably close to 0. The means and standard deviations vary greatly across composite scores, illustrating why standardization was used in the final presentation of the data. For example, .186 points above the mean on Student-Faculty interaction represents a distance of one *z* score above the mean, whereas .186 points on the Student Effort benchmark is 1.185 *z* scores above the mean. Standardization across scales ensures that a given number of points above or below the mean represent equal distances from the mean in the standardized format regardless of the benchmark.

---

[2] Items that contributed to the benchmarks were rescaled so that the low and high bounds of each item were equal across all scales. This was done by converting all scores to proportions of their totals so that the low end of the scale was always zero and the high end was always one. For example, a four on a seven-point scale and a three on a five-point scale both equal .5, as they are the midpoint of their respective scales. Don't Know/Not Applicable responses on items measuring frequency of use were not included in the computation of benchmark scores. For each participant, a benchmark score was computed by averaging the items comprising each benchmark.

Figure 1: Distribution of Active and Collaborative Learning Benchmark

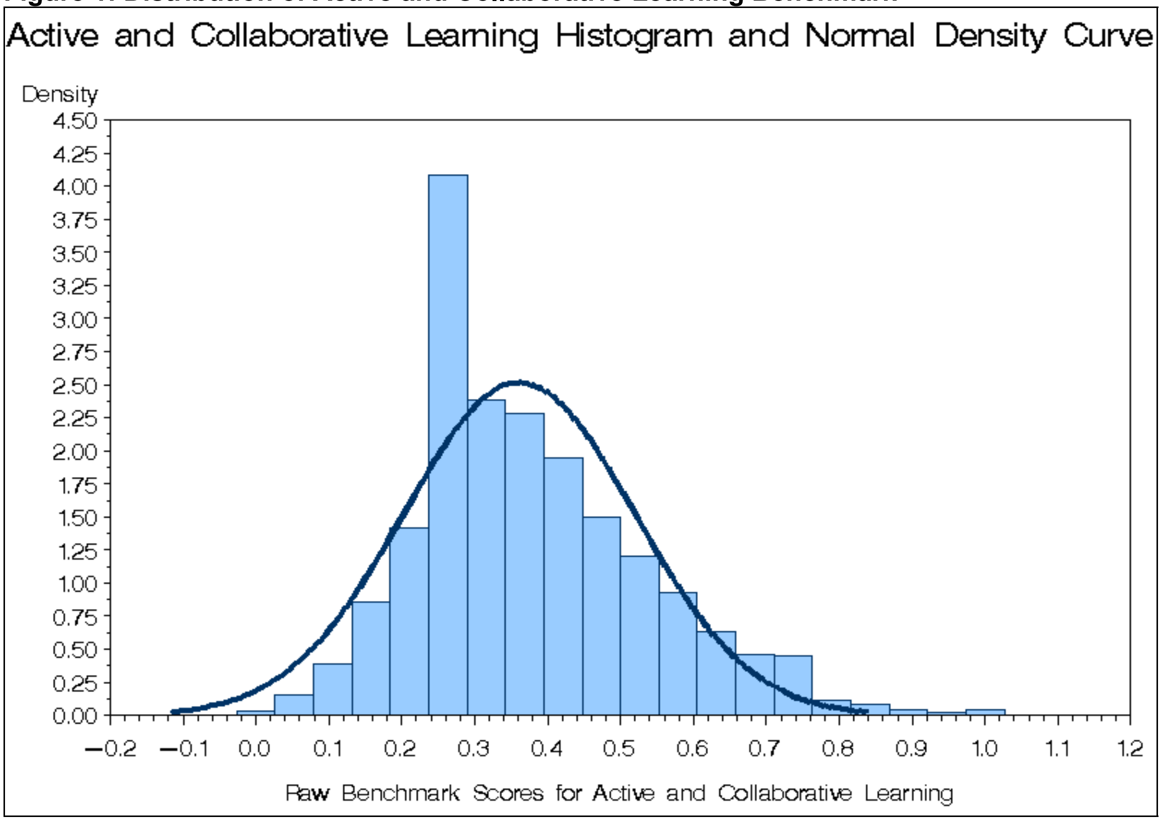


Figure2: Distribution of Student Effort Benchmark

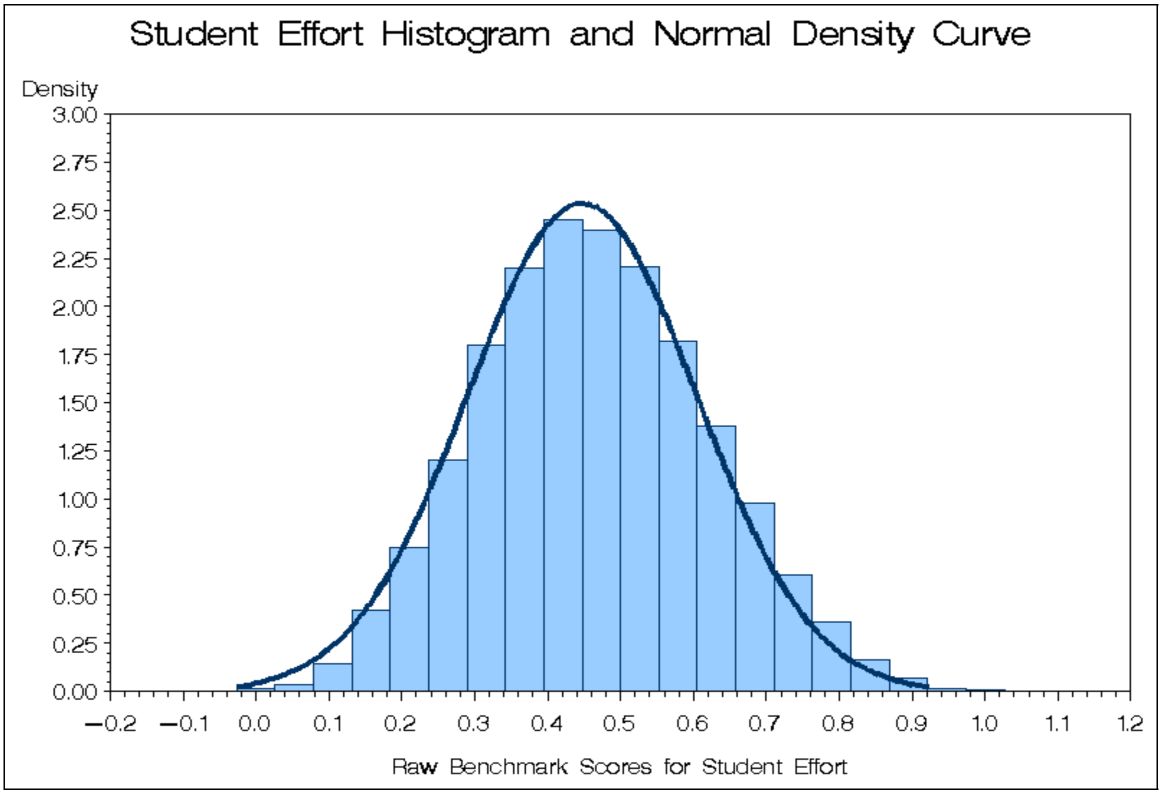


Figure 3: Distribution of Academic Challenge Benchmark

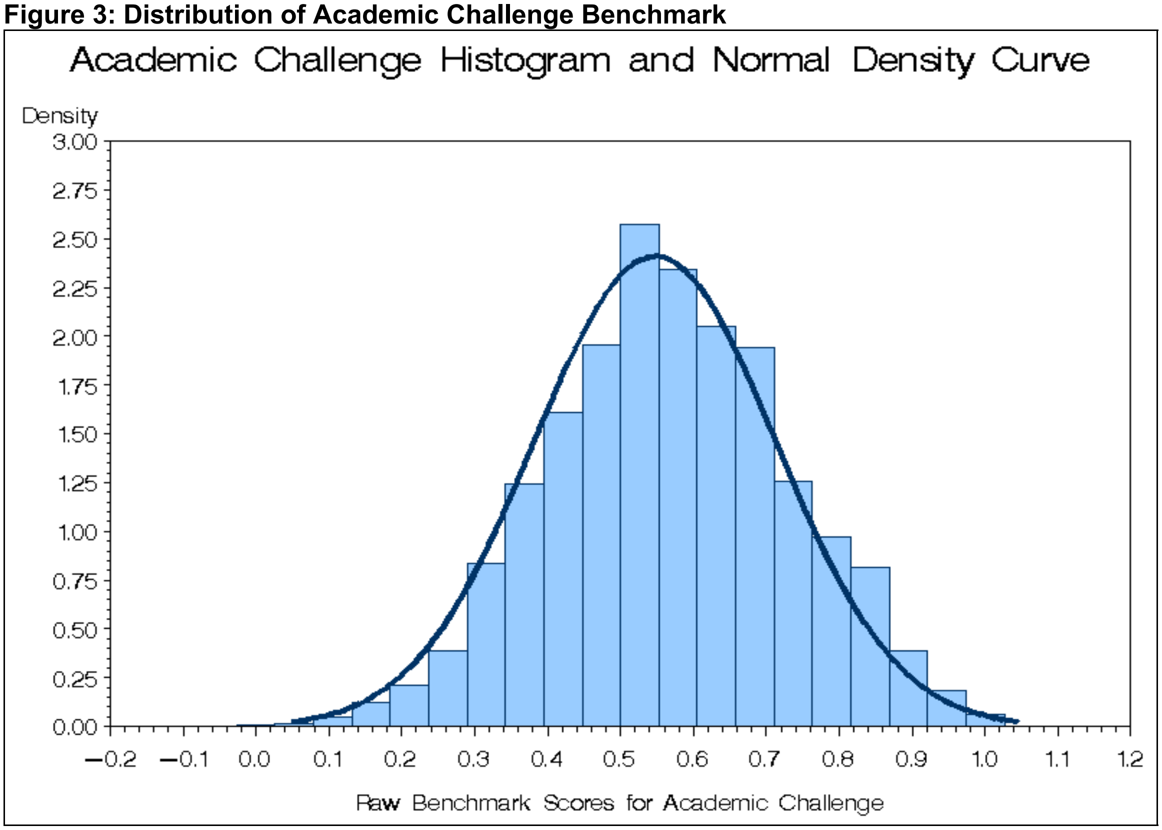


Figure 4: Distribution of Student-Faculty Interaction Benchmark

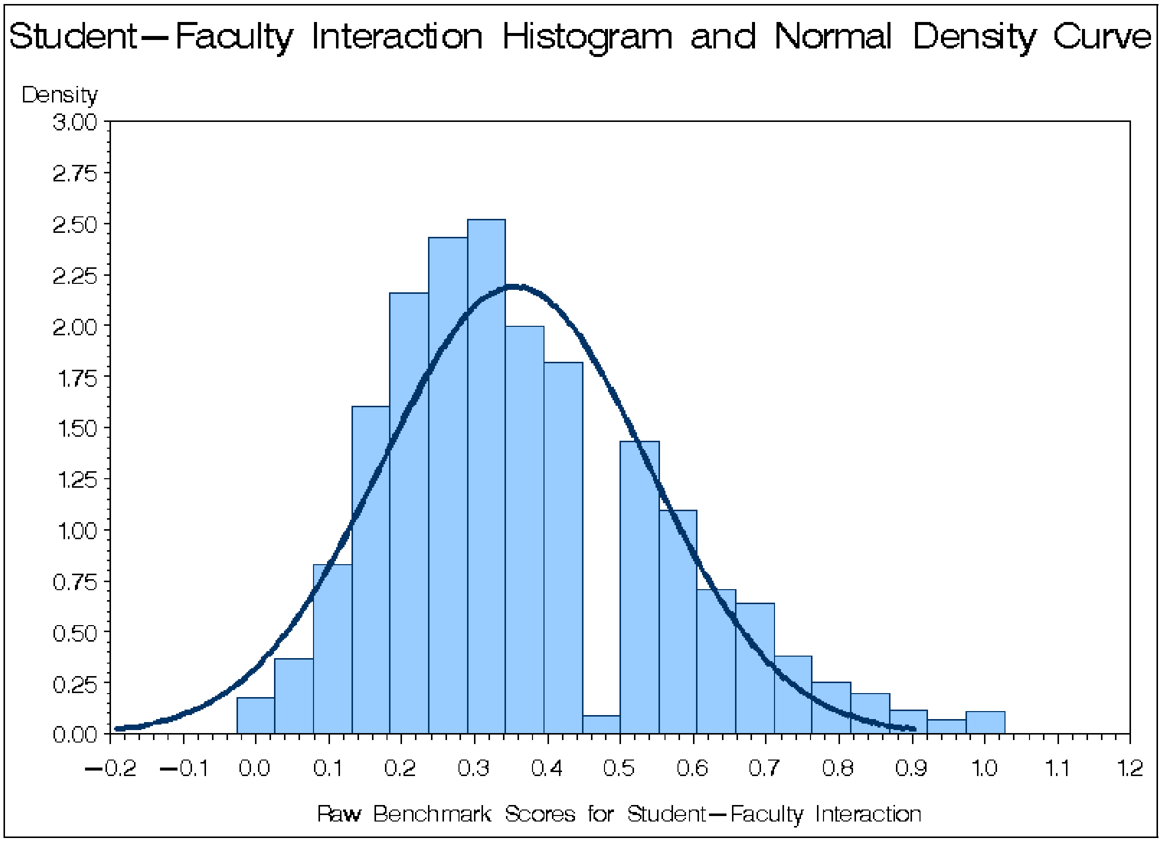


Figure 5: Distribution of Support for Learners Benchmark

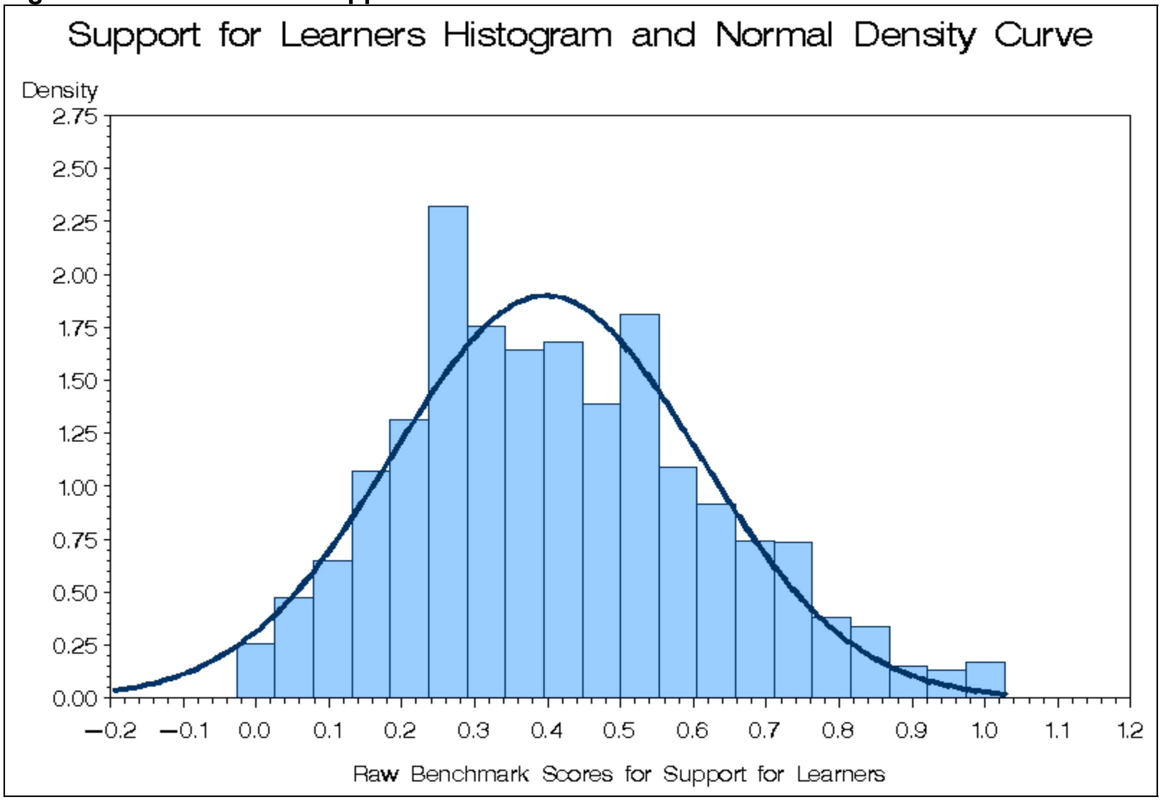


Table 4: Distributional Statistics for *CCSSE* Benchmarks

	Mean	Standard Deviation	Skewness	Kurtosis
Active and Collaborative Learning	.378	.161	.595	.466
Student Effort	.463	.157	.146	-.251
Academic Challenge	.567	.164	-.002	-.226
Student-Faculty Interaction	.373	.186	.688	.430
Support for Learners	.414	.211	.405	-.235

Validity of Benchmark Scores

In order for *CCSSE* benchmarks to be useful to colleges, the constructs must show reasonable construct validity. The *CCSR* itself has several items independent of the benchmarks that should be related to each of the benchmarks, and showing a relationship between these variables and benchmark scores is a powerful demonstration that the benchmarks are related to educational outcomes. Validation is perhaps most powerfully demonstrated by showing that an instrument's outcomes can be shown to be appropriately related to external measures. While there are several important external measures, such as retention, graduation, workforce outcomes, and direct assessments of student learning that have not yet been examined in relation to data obtained from the *CCSR*, grade point average (GPA) is one variable measured on the *CCSR* that can reasonably be considered an external measure of student performance.

Grade point average plays an important role in validating the benchmark constructs as it is nearly universally used as a measure of academic success. It is the single item on the *CCSR* survey that is an external measure of students' academic achievement, making it a valuable measure of how well students are doing in their college education. GPA is not without controversy, as grades are assigned differently across instructors and institutions. Nevertheless, as a widely used measure in nearly all institutions of higher education, it can reasonably be assumed to be a gross measure of how much students are learning in their academic endeavors.

To examine the relationship between self-reported GPA and the *CCSSE* benchmarks, GPA was regressed on each of the benchmarks in a random slopes and intercepts model[3]. Hierarchical linear models were used as participants were nested within colleges, making the data a typical multilevel structure in which respondents within an institution are expected to have correlated responses as a result of attending the same institution that are accounted for through random effects (Raudenbush & Bryk, 2002). The SAS Mixed procedure was used to construct the random slopes and intercepts models. It was anticipated that there would be differences in GPA across institutions, which is accounted for by random intercepts, as well as differences in the strength of the relationship between benchmarks and GPA across colleges, which is accounted for by random slopes (Singer, 1998; Littell, Milliken, Stroup, & Wolfinger, 1998). Not surprisingly, significance tests revealed that there were statistically significant amounts of variance accounted for by the random parameters in each model.

Outcomes from these models showed that there was a positive relationship between GPA and four of the five benchmarks. Active and Collaborative Learning, t (1, 52,705) = 18.90, p

[3] Individuals' scores were standardized to have a mean of 50 and standard deviation of 25 around a mean weighted by full- and part-time status so that scores represent distance from the weighted mean of all survey participants. To compute benchmarks for colleges, the students' standardized scores are computed as an average weighted on full- and part-time enrollment status. Benchmark scores thus contain information about strengths and weaknesses in various areas of student engagement as well as providing information about how an institution is doing in relation to other institutions.

< .001, Student Effort, *t* (1, 52,724) = 10.65, *p* < .001, Academic Challenge, *t* (1, 52,713) = 13.75, *p* < .0001, and Student-Faculty Interaction, *t* (1, 52,650) = 12.72, *p* < .001, were all positively related to GPA. Support for Learners was the one benchmark that did not demonstrate a clearly positive relationship with GPA, *t* (1, 52,685) = .78, *p* = .44. The results for Student-Faculty interaction model are shown in Figure 6 and closely resemble results for the other benchmarks with the exception of the Support for Learners benchmark.

The benchmark on Support for Learners is comprised of items specifically designed to reflect institutional practices that are important in student retention. Thus, it is perhaps unsurprising that this benchmark appears not to be as strongly related to GPA as it is not directly related to learning. The relationship between each of the benchmarks and retention data for participating institutions is a topic for future research using results from the *CCSR*.

The relationship between benchmarks of student engagement and other self-reported measures is also important for demonstrating construct validity. To examine the relationship between benchmarks and other measures, a series of items that asked students to rate the extent to which experiences at their college have contributed to their knowledge, skills and personal development were analyzed in relation to the benchmarks. Prior analyses suggest that there are three primary factors among these items (Marti & Ouimet, 2003), representing gains in traditional academic skills, gains in personal development, and gains in career-related knowledge and skills. Academic skills items measure gains in general education, writing, speaking, critical thinking, and numerical problem-solving. Personal development items measure gains in working with others, learning on your own, understanding yourself, understanding people with different backgrounds, and developing personal values. Career items include acquiring work-related knowledge, developing career goals, and leaning about career opportunities.

Random slope and intercept models were conducted on each of the factors among the gain items and each of the benchmarks. In each case there was a statistically significant relationship between the benchmark scores and the gain factors. Results were extremely consistent across models and were very similar to those shown in Figure 7. The consistency is an impressive display that latent constructs on different parts of the scale are related to each other in

Figure 6: GPA and Student-Faculty Interaction Model

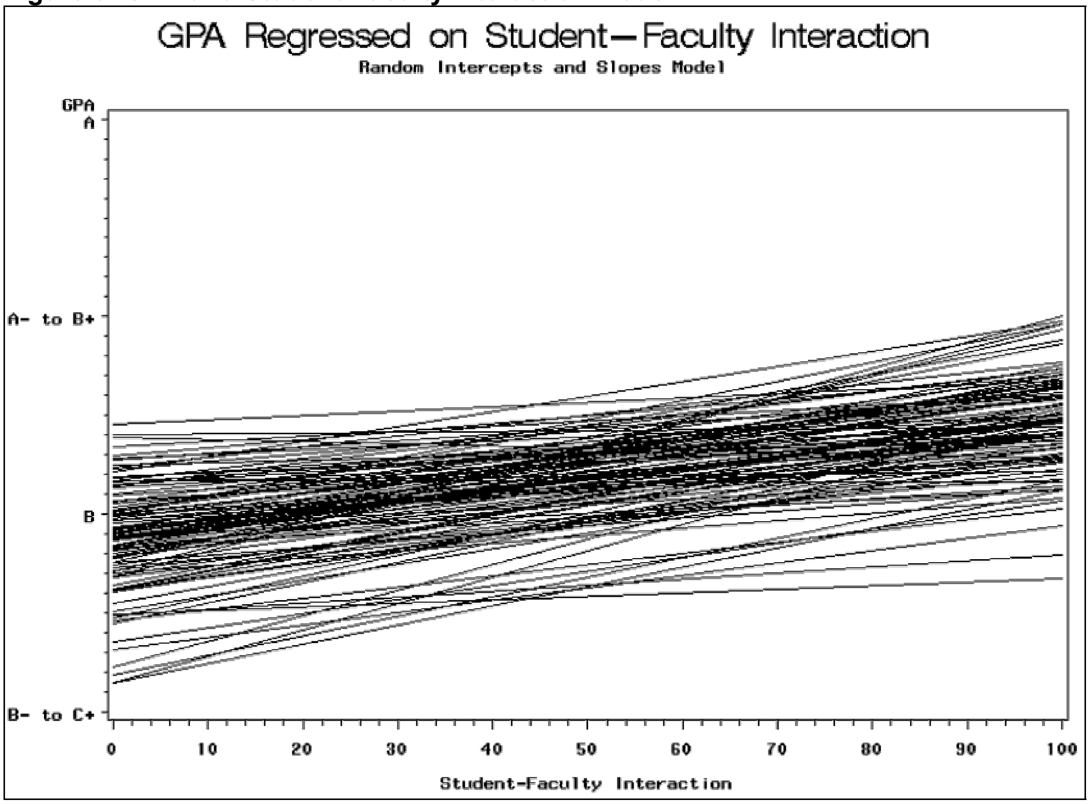
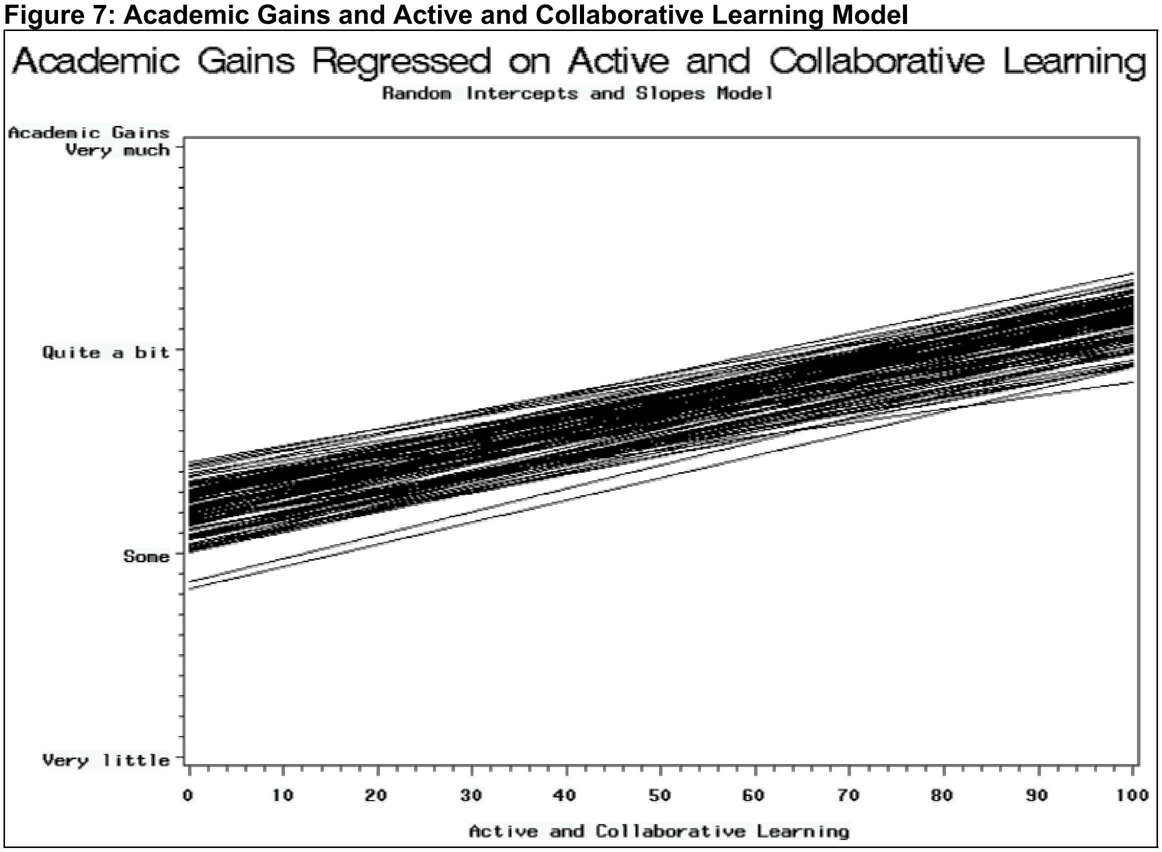


Figure 7: Academic Gains and Active and Collaborative Learning Model



a manner that would be predicted; specifically, students that are more engaged are reporting higher levels of gains in their academic skills, personal development, and career-related items.

Conclusions

Analyses of the psychometric properties of the *CCSR* and the benchmarks of effective educational practice that are derived from the instrument indicate that the data are both reliable and valid. CFA shows that modeling the data can closely reproduce the empirical results and that the instrument is stable from year to year. Validation analyses examined the ability of the benchmarks to predict other outcomes, such as GPA, and showed that engagement benchmarks are indeed predictably related to the outcomes to which they should be related.

With the firm psychometric basis demonstrated here, there is considerable potential for examining the relationship between the *CCSSE* benchmarks and other outcomes. While student engagement is certainly not the only factor contributing to student success, future work should examine how it relates to direct assessments of student learning outcomes in community colleges and to other important outcomes including retention, transfer, graduation, and career successes.

References

Carini, R. M., Hayek, J. H., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to Web and paper surveys: Does mode matter? *Research in Higher Education*, *44*, 1-19.

Hu, L., Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6(1)*, 1-55.

Kim, J., Mueller, C. W. (1978). Introduction to factor analysis: What it is and how to do it. Beverly Hills, CA: Sage.

Kuh, G. D., (2002). *The National Survey of Student Engagement: Conceptual Framework and Overview of Psychometric Properties*. Retrieved November 15, 2003 from the National Survey of Student Engagement Web site: http://www.iub.edu/~nsse/acrobat/psychometric_framework_2002.pdf.

Kuh, G. D., Hayek, J. C., Carini, R. M., Ouimet, J. A., Gonyea, R. M., and Kennedy, J. (2001). *NSSE Technical and Norms Report*, Indiana University Center for Postsecondary Research and Planning, Bloomington, IN.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). SAS system for mixed models. Cary, N.C.: BBU Press.

Marti, C. N., & Ouimet, J. H. (April, 2003). *Are students' educational gains related to their educational goals?* Paper presented at the Council for the Study of Community Colleges, Dallas, TX.

Muthén, L.K., & Muthén, B.O. (1998). Mplus User's Guide. Los Angeles: Muthén & Muthén.

Pace, C. R. (1980) Measuring the quality of student effort. *Current Issues in Higher Education*, *2*, 10-16.

Raudenbush, S. W., Bryk, A. (2002). Hierarchical linear models: Applications and data analysis methods, 2nd Ed. Beverly Hills, CA: Sage.

Sax, L. J., Lee, J. J., Hagedorn, L. S., Gilmartin, S. K., & Gale, A. (May, 2003). Combining Web and Paper Survey Responses: Can We? Should We? Paper presented at the Association of Institutional Research, Tampa, FL.

Singer, J. D., (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Education and Behavioral Statistics*, *24 (4)*, 323-355.