Updating the National Survey of Student Engagement:
Analyses of the NSSE 2.0 Pilots

Allison BrckaLorenz
Robert M. Gonyea
Angie Miller
Indiana University Center for Postsecondary Research

The authors would like to thank the following individuals for their assistance in the analysis and written sections of this technical report: Jim Cole, Yiran Dong, Kevin Fosnacht, Kevin Guidry, Heather Haeger, Amber D. Lambert, Thomas Nelson Laird, Wen Qi, Amy Ribera, Louis Rocconi, Shimon Sarraf, Rick Shoup, and Malika Tukibayeva.

Presented at the Association for Institutional Research Annual Forum in New Orleans, LA, June 2012
Updating the National Survey of Student Engagement

**Updating the National Survey of Student Engagement: Analyses of the NSSE 2.0 Pilots**

Updating a large-scale survey is not a simple task. This paper gives an overview of how the National Survey of Student Engagement (NSSE) used quantitative analyses to update their survey over the course of several years and two survey pilot administrations. A description of the pilot instruments and pilot data will be followed with an overview of the quantitative methods used to analyze both individual items and groupings of items for the improvement of the survey. A brief overview of the results from these analyses gives a picture of how the 2013 NSSE evolved.

**Goals and Purposes of the NSSE Update**

Completing a major update of the NSSE survey came out of a decision in 2004 to resist annual amendments to the core survey items and benchmarks, which made year-to-year comparisons increasingly difficult for institutions. After a number of significant changes were made to the survey in the first several years, we moved to take a longer view and compile desired updates for a new version sometime in the future. This allowed us to engage in a more scholarly, deliberate, and participative process to updating the survey. The process of updating NSSE began in earnest in 2009, and continues with the following goals:

- To continue in our core purpose of assessing student engagement in effective educational practices to inform improvement efforts;
- To apply what we have learned over the past decade about student engagement and institutional improvement;
- To stay current with movements and trends in higher education such as technology for learning, distance learning programs, faculty development, infusion of high-impact practices, diverse and global curriculum development;
- To improve the clarity, consistency, and applicability of the survey language, and to eliminate the use of outdated terminology, primarily related to technology;
- To improve the measurement properties of existing measures (e.g., benchmarks and other measures); and
- To incorporate new measures relevant to effective teaching and learning

With an ultimate goal of administering the updated NSSE in 2013, two pilot administrations were planned for 2011 and 2012.
Pilot Instruments

The focus of revisions for the first pilot in 2011 included enhancing item clarity, improving the properties of aggregate measures, eliminating outdated content, and including new content related to student development and learning. This pilot included new items about quantitative reasoning and effective teaching practices, expanded on existing items about collaborative learning, and significantly revised the language of the higher-order learning items (which were based on Bloom’s taxonomy, 1956). Additional item sets about technology usage, global awareness, diverse perspectives, learning strategies, and behaviors associated with enhanced reading comprehension were also tested.

The 2012 pilot extended many of the changes made in the 2011 pilot, with adjustments based on quantitative and qualitative analysis. This pilot also introduced a few new content areas for testing in the core survey such as learning strategies and more interactions with faculty. The 2012 pilot was also NSSE’s first opportunity to test the administration of modules—brief item sets that focus on singular content areas not addressed on the core survey and provide institutions an opportunity to collect additional information about a particular area of engagement. Each pilot institution had the opportunity to append two of five available modules to the 2012 pilot instrument. Modules involved questions about advising, civic engagement, diverse activities, transferable skills, and technology.

Compared to the standard NSSE, the overall survey length of the second pilot (118 survey items) is greater by about twelve items, nine being core engagement items and the other three demographics. From the original NSSE instrument, 24 items were deleted and 36 were new. Of the original 82 items that appear on the standard NSSE instrument, 24 did not change (29%), 31 received minor modifications (38%), and 27 received major modifications (33%).

Pilot Administrations and Samples

Pilot institutions were selected to cover a range of Carnegie classifications and of institutional differences by size, selectivity, minority-serving status, religious affiliation, urban status, geographic region, and online instruction. In the spring of 2011 the first pilot instrument was administered at 19 institutions, gathering data from over 20,000 first-year students and seniors. The average institutional response rate for the 2011 pilot was 35%, compared to 37% for the standard NSSE 2010 administration, and institutional response rates ranged from 18% to 74% with three-quarters achieving a response rate of 30% or higher. In the spring of 2012, the second pilot instrument was administered at 55 institutions and collected responses from over 50,000 students. Institutional response rates in 2012 ranged from 9% to 63%, with an average of 28%, and nearly half of institutions achieving a response rate of 30% or higher.
Both pilots included all first-years and seniors at participating institutions, using a web-only administration. For both pilots, women made up about two-thirds of respondents and the majority was under 24 years old. About half reported earning mostly ‘A’ grades while fewer than 10% reported mostly ‘C’ grades or lower. Respondents were mostly White (around two-thirds) and proportions of minority students were similar on both pilots as well. Respondents were more likely to be enrolled full-time (90% of first-years and 80% of seniors), and about half of respondents were first-generation students (i.e., neither parent had earned a bachelor’s degree). Proportionally more men pursued studies in business and engineering, while majors in education, the social sciences, and other professions (e.g., nursing) were more popular among women. One of the few differences between the two pilot samples were that the 2011 pilot sample included more senior transfer students (57%) than the 2012 pilot (45%).

Methods

Pilot data were analyzed quantitatively with two purposes – to examine the quality of the new and modified items individually, and to explore new groupings of items as potential reportable aggregate measures of engagement. Although not a part of this paper, qualitative information was collected in 2011 and 2012 through cognitive interviews of 120 students and 10 focus groups with a total of 79 students at 12 different campuses, phone interviews with students on specific questions, write-in responses from students taking the pilot, and feedback from outside sources such as from our institutional users and board members. (For more information on the qualitative methods and results, see Haeger et al., 2012).

In the following description of methods, analyses were conducted similarly for both pilots unless otherwise indicated. Individual item properties were evaluated quantitatively in a variety of ways. Item descriptives included frequencies, means, standard deviations, standard errors, skewness, kurtosis, and percent missing were calculated by class level, as well as by gender and major. Effect size differences were calculated to compare by gender and enrollment status as well. Because some items remained unchanged or only slightly modified from the standard NSSE survey, comparisons were made between the pilot descriptives and standard NSSE descriptives when available. Items were compared longitudinally in this way both at the aggregate and individual institution level for several institutions. These individual item analyses resulted in recommendations on rewording or removing of items. Both the pilot and standard NSSE survey were co-administered at seven institutions in 2012 so that items could be compared on the two different surveys from students in the same time period and same institution.

Demographic items were analyzed separately from the individual engagement items. Demographics were analyzed descriptively with frequencies and means (when appropriate) as well as item-nonresponse.
Updating the National Survey of Student Engagement

Demographics were also cross-checked with each other as a form of known-groups validity looking at such things as whether or not younger students spent less time caring for dependents or students living on campus spent fewer hours per week commuting. Demographic proportions were also compared to several external data sources such as the National Center for Education Statistics and the Integrated Postsecondary Education Data System. To test temporal stability reliability in 2011, demographic questions were compared longitudinally by matching students that took the pilot as seniors to their participation in standard NSSE as first-year students. The overall survey was also analyzed by looking at individual-item missing data as well as response drop off rates. Straight-lining (selecting the same response option for a set of items) was examined for sections of items, and duration was examined to provide additional information about survey length.

Exploratory factor analysis was then used to create a preliminary list of aggregate measures for further analysis. In 2011, second exploratory factor analysis was done for the online institution so that the appropriateness of aggregate measures could be investigated for students taking their courses online. In 2012, a second exploratory factor analysis was done for all students who reported taking their courses entirely online regardless of whether or not they were at an online institution. Items administered in the pilot modules were included in the factor analyses as well. Aggregate measure descriptives included means, standard deviations, standard errors, skewness, and kurtosis. In 2012, the percent of variation between institutions was examined for each measure as well. Confirmatory factor analyses were then conducted using the preliminary aggregate measures. If multiple measures in a single content area were suggested, these measures were included in the same model and allowed to be correlated. Because the exploratory factor analysis resulted in several different options for measures within a single content area in 2011, each suggestion for the arrangement of measures within a single content area were tested as individual models. Only one measure was evaluated with two different options in 2012. Once good models were fit and correlations between measures within content areas were determined to be acceptable, further tests of validity and reliability were conducted on the refined set of measures.

Additional evidence for validity was collected by looking at differences between various groups in 2011: gender, transfer status, enrollment status, Greek status, campus living situation, distance education, athletic status, age, first-generation status, international status, veteran status, disability, sexual orientation, ethnicity, academic disciplinary field, and institutional control. Evidence for concurrent validity was gathered by correlating pilot measures to individual items from the core pilot survey and three experimental item sets about collaborative learning, interest in diversity, and teaching clarity. Predictive validity was examined using OLS regression models to investigate relationships between pilot
measures and students’ institution-reported GPA. These analyses will be completed again for the 2012 data, but will not be completed in time for the presentation of this paper.

Internal consistency reliability was measured with Cronbach’s alphas for each proposed measure, Cronbach’s alpha for a measure if a single item is removed, correlations between an item and the remaining items in the measure, average inter-item correlations, and the individual inter-item correlations of the measure. Equivalence reliability was tested for one set of revised items by comparing the standard survey’s Higher Order Learning items (appended to the pilot survey as an extra item set) to the modified pilot items in 2011. Reliability and factor analysis results were strengthened using item response theory analysis to check for unidimensionality as well as determine which items within a measure were providing more or less information. In 2012, generalizability theory was used to examine how generalizable measures were for groups of students.

Results

Informed by both years of the pilot data analyses, the original five NSSE Benchmarks of Effective Educational Practice were reshaped into smaller group of items with the intention that each measure consist of a singular or unidimensional concept. Thus, the five general themes of the Benchmarks now consist of a variety of content areas which contain one or more “engagement indicators.” These themes intend to provide institutional users a more useful, actionable, and diagnostic collection of results for benchmarking purposes. The results of both pilot analyses for the newly formed indicators appear below.

Academic Challenge

The Academic Challenge content area was reformulated from the standard NSSE Benchmark entitled Level of Academic Challenge. Although the new content area covers basically the same content as the Benchmark, the separate dimensions included in this topic are no longer combined into a single score. The topics covered in the Academic Challenge content area include measures of the quality and quantity of students’ writing and reading; how often they analyze, evaluate, and apply numerical information; and how often students engage in various learning strategies such as reviewing their notes after class or identifying key information from reading assignments. Individual items in this content area asked students such things as how much their courses challenged them to do their best work, how much time they spend preparing for class, and how much their institution emphasizes spending time on studying and academic work.
Updating the National Survey of Student Engagement

Rather than combine these separate areas of challenging academic work into a single score, separate engagement indicators were computed for two sets of items called Quantitative Reasoning and Learning Strategies. Individual items in this content area will stand as individual measures of quality although additional indices of ‘challenge’ and ‘writing’ will be explored for the future. In 2011, items within the Quantitative Reasoning indicator were very closely interrelated so some edits were made to items in the 2012 pilot in the hopes of making these items more distinct for students. In 2012, the Quantitative Reasoning and Learning Strategies indicators performed well with an overall good confirmatory factor analysis model fit, meeting even the most conservative indices of model fit. Results for online students did not differ noticeably from the more traditional students implying that these indicators are appropriate for online students as well. Reliability analyses concluded that both of these indicators were internally consistent and both models met all IRT assumptions. A large group of students (greater than 100) would be needed in order to generalize both of these indicators as constructs. An understandably high variability between students on Quantitative Reasoning (from disciplinary differences) and a low variability between institutions on Learning Strategies leads to these restrictions on generalizability.

Deep Approaches to Learning

Deep Approaches to Learning has emerged as an important measure on the NSSE survey (Nelson Laird et al., 2008), although it previously had not been reported as one of the Benchmarks of Effective Educational Practice. Several of the component items were included in different Benchmarks (e.g., Level of Academic Challenge and Active and Collaborative Learning), but with the NSSE update, this important area for learning and student development was considered its own content area for pilot analysis. The topics in this content area include how often students integrate diverse perspectives or ideas from different courses in their studies, how often they reflect on their understanding of concepts, and how much their courses include higher-order learning tasks such as application or evaluation. Although a single item about how much their coursework emphasizes memorizing course material is included in this content area, it is not included in either of the two indicators of this content area.

The two indicators for this content area were Higher Order Learning and Integrative and Reflective Learning. The changes made to the wording of the Higher Order Learning items on the standard survey were very positive, resulting in items that were much easier for students to understand. In 2012, this indicator had a good overall confirmatory factor analysis model fit, and reliability analyses showed it to be internally consistent. Although descriptives and IRT models looked slightly different for traditional students and online students (online students appear to do these things slightly less often), there was no reason to believe that this indicator was inappropriate for either population of students. Generalizability
Updating the National Survey of Student Engagement

results suggest that only a very large sample of students (greater than 200) should be used in order to
generalize Higher Order Learning as a higher order construct.

Although reflective learning and integrative learning were generally discussed as separate concepts with
standard NSSE data, high intercorrelations on the pilot surveys suggested that a single indicator would be
more appropriate. Several items were modified and added to these content areas in 2011 in order to help
distinguish the two different concepts for students. In 2012, the Reflective and Integrative Learning
indicator performed well with an overall good confirmatory factor analysis model fit, with all model fit
indices meeting criteria. Reliability analyses suggested this was an internally consistent grouping of
items, and they met all IRT assumptions with a good amount of information throughout the majority of
the distribution. A smaller sample of 25 students would be all that is required to generalize this indicator
to represent a higher order construct. Analyses of both traditional and online students were similar,
implying that this indicator would be appropriate for both populations of students.

Collaborative Learning

The Collaborative Learning content area came out of the standard NSSE Benchmark Active and
Collaborative Learning. A goal of the first pilot was to focus this content area on the collaborative
learning activities such as group presentations, co-writing, tutoring, in-class and out-of-class projects, and
study groups as opposed to the combined ideas of active and collaborative learning. Analysis of these
items found that a single scale was recommended after several items were removed from the survey due
to redundancy in content coverage. One of the more interesting 2011 findings regarding Collaborative
Learning surfaced when examining these items for students at an online institution. It was clear that
working collaboratively was very different for these students compared to those in a more traditional
classroom setting when looking at item distributions and factor analysis results. When investigating the
two preliminary subscales for collaborative learning, a scale involving collaboration such as preparing for
exams with other students or receiving feedback from other students on assignments still worked well for
online students, but a subscale about writing with or presenting with other students did not work so well
for online students. This subscale, however, was not preserved on the 2012 pilot for a variety of reasons,
including it not being applicable to all students and redundancy in content coverage.

Given the removal and rewording of many items in the Collaborative Learning content area, in 2012 only
four items were further tested on the second pilot. In 2012, these four items created the Collaborative
Learning indicator and performed well in all analyses. There was an overall very good confirmatory
factor analysis model fit, with the model meeting even the most conservative indices of model fit.
Reliability analyses suggested the indicator is internally consistent for both online and traditional students.
Updating the National Survey of Student Engagement

although descriptives showed that online students are doing these collaborative activities less often than their traditional peers. IRT results also indicated that online students answered these questions in different ways from traditional students, but there was no indication that these items were inappropriate for online students. It makes sense that online students interact differently with peers than their traditional counterparts and may do things such as preparing for exams with classmates or asking another student for help less often. Online institutions or institutions with large proportions of online students will have to be careful with how they compare their Collaborative Learning scores to other institutions.

Experiences with Faculty

The content area Experiences with Faculty borrowed from the standard NSSE Benchmark Student-Faculty Interaction, but expanded coverage of the topic. In addition to a variety of individual items, two different areas of focus were tested: instructors’ use of clear teaching behaviors such as using examples to explain difficult points, and faculty practices related to mentoring such as getting to know students and creating an atmosphere conducive to learning. Individual items included getting feedback from instructors, working with faculty on activities other than coursework, and talking about career plans with faculty or advisors. After analyses were completed, many of the items in this content area were rewritten, with some getting new response options and stems. Additional items were explored to cover topics such as mentoring and in-class interactions.

In 2012, two indicators and a variety of individual items were tested in the Experiences with Faculty content area. The individual items were about instructors providing feedback on a work in progress, discussing ideas for coursework with faculty, and asking faculty for guidance. Two indicators were considered for further testing. The Student-Faculty Interaction indicator included items about in-class and out-of-class interactions with faculty such as talking about students’ academic performance or discussing course concepts outside of class. There was an overall very good confirmatory factor analysis model fit, with the model meeting the most conservative model indices. Descriptives showed that online students have lower Student-Faculty Interaction scores than their traditional peers, and IRT results did show some evidence that these two populations answered these questions differently. Reliability analyses showed that this indicator was internally consistent for both online and traditional students, and there was no other evidence that this indicator was inappropriate for online students even though the two populations answered the questions differently. It makes sense that online students might have less out-of-course contact with their faculty. Generalizability results show that this indicator can be generalized to a higher order construct with a sample of at least 25 students.
Updating the National Survey of Student Engagement

The indicator Good Teaching Practices included the 2011 clear teaching behavior items in addition to an item about instructors giving prompt and detailed feedback. This indicator also performed well with an overall very good model fit that met the most conservative indices of model fit. As with the previous indicator, descriptives showed that online students have lower Good Teaching Practice scores than their traditional peers. Reliability analyses showed that this indicator was internally consistent, but this indicator did fail to meet the IRT assumption of unidimensionality. Generalizability analyses also showed that this indicator could only be generalized for very large samples of students (more than 200 students). It appears that these problems are due to issues raised by part-time and full-time students answering the same question about “in how many of your courses” their instructors engage in certain behaviors. Because, for example, the responses of “some” or “all” mean very different things for part- and full-time enrolled students, this set of items will be reframed in the 2013 survey. Instead, the questions will ask to what extent students’ instructors have engaged in the various teaching practices. This change will likely clean up the generalizability and IRT issues seen in 2012.

Diverse Interactions

Student interactions with diversity, originally part of the NSSE Benchmark Enriching Educational Experiences, were considered to be an important issue to expand upon in both the core survey and a future module. In 2011, a set of experimental items about diverse activities and interactions was tested with the pilot survey. These items were analyzed with the other pilot items in order to determine which items would be kept either on the main survey or in a future module on diversity-related activities. One group of items, Diverse Interactions, was chosen to replace the two diversity items on the standard pilot survey. These items focused on how often students had serious discussions with people who were different from them in various ways. The remaining two groups of items, Diverse Events and Diverse Conversations, and an individual item about attending campus events were chosen to be further tested in a future module on diversity. These items asked students how often they had conversations about various topics of diversity and attended various campus events about diversity.

In 2012 during cognitive interviews (see Haeger et al., 2012) some students indicated that the items in this content area could be interpreted in a variety of ways so further wording revisions were considered and additional qualitative information was gathered. Although the indicator Diverse Interactions performed well (good overall confirmatory factor analysis model fit, internally consistent, and generalizable for small groups of students), the items continued to have issues in cognitive interviews. Students would misinterpret the question as having to be more about the topic of discussing diversity than about having discussions with people who were different from them. The items were reworded again for the 2013
survey and were found to be much more understandable and correctly interpreted by students. Future analyses will further test the quality of this indicator in 2013.

High Impact Practices

Due to the independent and discrete nature of the items, much of the standard NSSE Benchmark Enriching Educational Experiences will be reported individually as High-Impact Practices, namely, participating in a learning community, internship, study abroad, research with faculty, culminating senior experience, and service learning. In 2011, items under this content area were not tested as a single group of items, but they did function well as individual items with sensible distributions and only minor trouble in cognitive interviews (see Haeger et al., 2012). The new question about formal leadership experiences appears to work well with students, but an additional item about first-year seminars was found to be too difficult for students to understand. These items continued to perform well (sensibly distributed, low skip-rate, etc.) in the 2012 pilot, and all of these high-impact practices will continue on the 2013 pilot with only a minor word tweak to make the learning community item more understandable.

Campus Environment

The Campus Environment content area comes out of the standard NSSE benchmark Supportive Campus Environment and largely covers the same content as before. This content area included students’ perceptions of their quality of interactions with various people on their campus as well as their perceptions of different ways that their institution supports their success and encourages them to participate in beneficial activities. In 2011, two groups of items were created in this content area: Quality of Interactions and Supportive Environment. The Quality of Interactions measure in 2011 included a wide variety of people with whom students might interact, including library staff and learning support staff. Because students had difficulty answering some of the questions in this item set, many changes were made including the removal of items, rewording of items, and the inclusion of a ‘not applicable’ option. Small changes were made to items in the Supportive Environment set such as adding an item about using learning support services and other wording changes to make items more clear.

In 2012, the Quality of Interactions indicator had an overall adequate confirmatory factor analysis model fit, and reliability analyses showed this indicator to be internally consistent. Although IRT analyses indicated that online students answer these questions differently than traditional students, other results for online students and traditional students were similar implying that this indicator is appropriate for both populations. Generalizability results showed that this indicator could be generalized to a higher order construct for samples of 80 students or more. The inclusion of a ‘not applicable’ option for students to
choose when considering the quality of their interactions with various people was considered to be a success. Most of the ‘not applicable’ responses were chosen for students’ interactions with student services staff, particularly for online students. Sensibly, few students chose ‘not applicable’ for interactions with students and faculty.

The Supportive Environment indicator also had an overall adequate confirmatory factor analysis model fit, and reliability analyses showed this to be an internally consistent indicator. Although descriptives were slightly lower for online students, this difference was hardly noticeable. Only a small number of students, at least 25, would be needed to generalize this indicator to a higher order construct. An individual item in this content area about the number of times a student has seen an academic advisor was moved to a module on advising as it was thought to be more appropriate when considered with other items about advising.

Self-Reported Gains

Self-reported gains are appropriately understood to be students’ general perceptions of their learning in specified areas. Although the self-reported gains have always been a part of NSSE, they have not been reported as benchmarks. Because of their potential utility among users, NSSE plans to provide indicator scores for self-reported gains in the future. In 2011, it was found that either one or two measures would work equally well from a quantitative perspective. Conceptually, however, two measures, one for students’ perceived academic gains and one for applied gains, were sensible and may be more useful for institutions. Several items ended up being recommended for deletion from this item set as they did not cover critical content area.

In 2012, analyses of the self-reported gains items showed that with the removal of items from the 2011 pilot, one indicator for gains was more appropriate. The single indicator of gains had overall good confirmatory factor analysis model fit, and reliability analyses showed the indicator to be internally consistent. Results were similar for online students and traditional students, but IRT and generalizability analyses indicated that this grouping of items should probably not be interpreted as a unidimensional construct. It makes sense that these items may be well correlated, but do not represent a construct as the items cover gains on such varied concepts as writing clearly and effectively, acquiring job- or work-related knowledge and skills, and developing or clarifying a personal code of values and ethics. An item that was previously removed from the 2011 pilot about becoming an active and informed citizen was added back on to the survey for 2013 in order to add to the survey’s overall coverage of the topic of civic engagement.
Updating the National Survey of Student Engagement

Demographic Items

Demographic items appeared to function as one would expect with no notable issues with non-response. All cross-checks of demographic items suggested a high degree of evidence for validity—in all cases, the data supported the expected result. Although there were several limitations to comparing the 2011 NSSE pilot results to external sources (the external sources collect, define, and aggregate information differently), most comparisons were favorable, showing little difference proportionally. Comparing demographics results for seniors on the 2011 pilot to first-years on standard NSSE revealed greater than 90% agreement for most demographic items. Agreement on race/ethnicity was at 89%, but increased to 96% after accounting for students who chose “I prefer not to respond” on one of the administrations. The variable with the lowest percent agreement (77%) was parental education. This low rate of matching is likely impacted by a number of factors including the substantial rewording of the item on the pilot, a parent’s obtainment of a degree, or students’ knowledge of their parents’ educational attainment. In 2012, demographic items continued to be stable and had no notable issues. The item asking students about race/ethnicity was changed slightly to become more consistent with IPEDS reporting standards.

Looking Ahead

In sum, institutions that have participated with NSSE over the years will experiences several changes with the NSSE update in 2013. We expect most of these changes to be welcome improvements, but some may present challenges or require adjustments for institutional users. Therefore, we want to be sensitive to the experiences and positions of our users so we can better serve them through this transition. NSSE 2013 will involve the following changes because of the analyses of the two pilot surveys: updated survey content with both new and modified items, new items within optional modules, and new groupings of items to serve as indicators of engagement. As part of NSSE’s commitment to transparency, giving insight into the quantitative methods used to analyze the pilot administrations of NSSE can help users understand the process of change and the many considerations accounted for in the updated NSSE 2013 survey.

References


Updating the National Survey of Student Engagement