The Dependability of the Updated NSSE: A Generalizability Study

Kevin Fosnacht
Robert M. Gonyea

Center for Postsecondary Research
Indiana University School of Education

Corresponding author:
Kevin Fosnacht, Ph.D.
Indiana University Center for Postsecondary Research
1900 East Tenth Street, Suite 419
Bloomington, IN 47406-7512
812-856-1961
kfosnach@indiana.edu
The Dependability of the Updated NSSE: A Generalizability Study

Abstract

NSSE updated its widely used undergraduate survey and created the Engagement Indicators (EI) to succeed the Benchmarks. This study examined the situations where the EI group means at the institutional level are dependable utilizing data from the 2013 NSSE administration. Using Generalizability Theory, it found that the EI produce dependable group means that can be generalized to the institution level from samples as small as 25-50 students. The results also suggest that the EIs should be used as indexes at the group level, rather than higher-order constructs.

Keywords: National Survey of Student Engagement, generalizability theory, group means, university students
The Dependability of the Updated NSSE: A Generalizability Study

Despite decades of dialogue, higher education still struggles with assessing the quality of undergraduate education and no longer enjoys respectful deference from governments, media, and the public who are collectively anxious about cost and quality. Such anxieties have stimulated considerable pressure for assessment and accountability. The dominant paradigm focusing on resources and reputation – most visible in the *U.S. News and World Report* rankings – has been roundly criticized for its neglect of students’ educational experiences (Carey 2006; McGuire 1995). In response, higher education leaders, researchers, and assessment professionals have explored many ways for higher education to improve – through reforming the curriculum, faculty development, and improved assessment (Association of American Colleges and Universities n.d.; Barr and Tagg 1995; Gaston 2010; Lumina Foundation 2011). In recent years, the measurement of student engagement has emerged as a viable alternative for institutional assessment, accountability, and improvement efforts. Student engagement represents collegiate quality in two important ways. The first is the amount of time and effort students put into their coursework and other learning activities, and the second is how the institution allocates resources, develops the curriculum, and promotes enriching educational activities that decades of research studies show promote student learning. (Kuh 2003b, 2009; Kuh et al. 2001; McCormick, Kinzie, and Gonyea 2013).

The National Survey of Student Engagement (NSSE) collects information at hundreds of bachelor’s-granting universities to estimate how students spend their time and how their educational experiences are shaped. Institutions use NSSE primarily in two ways. The first is to compare, or benchmark, their students’ responses with those of students at other institutions. Such an approach provides the institution with diagnostic information about how their students
are learning, and which aspects of the undergraduate experience have been effective and which are in need of improvement. The second way institutions use NSSE is to assess subgroups of their own students to determine how student engagement varies within the institution and to uncover areas for institutional improvement for groups such as first-generation students, part-time students, adult and commuter students, students enrolled in different majors, transfer students, and so on. Both of these approaches utilize NSSE scores by comparing the aggregate score of one group with that of another group, whether they be different institutions or different types of students within the same institution.

Thus, the NSSE instrument depends foremost on its reliability at the group level, and upon its ability to accurately generalize an outcome to the aggregated group. The examination of the reliability of a group mean score requires methodological techniques that can account for and identify multiple sources of error. Consequently, this paper explores the notion that generalizability theory (GT) may provide the proper methodological framework to assess the dependability of benchmarking instruments such as NSSE, and uses GT to investigate the number of students needed to produce a reliable group mean for the NSSE Engagement Indicators. Finally, the appropriate uses of NSSE data are discussed in light of the study’s findings.

**Updating NSSE**

Since NSSE’s initial launch in 2000, higher education scholars have learned more about collegiate activities and practices that positively influence student outcomes (Kuh et al. 2006; McClenney and Marti 2006; Pascarella, Seifert, and Blaich 2010; Pascarella and Terenzini 2005). Many areas of higher education are seeing growth, innovation, and rapid adoption of new ideas such as distance learning and other technological advances. To meet these challenges and
improve the utility and actionability of its instrument, NSSE introduced an updated version in 2013, which both refines its existing measures and incorporates new measures related to emerging practices in higher education (NSSE 2018a). The new content includes items investigating quantitative reasoning, interactions among diverse populations, learning strategies, and teaching practices. Additionally, the update provides the opportunity to improve the clarity and consistency of the survey’s language and to improve the properties of the measures derived from the survey. Despite these changes, the updated instrument is consistent with the purpose and design of the original version of NSSE (Kuh et al. 2001), as it continues to focus on whether an institution’s environment emphasizes participation in effective educational practices, and is administered to samples of first-year students and seniors at various types of baccalaureate-granting institutions.

Validity of NSSE

With the updated survey, NSSE continues its core purpose of providing institutions with valid and reliable assessment information for the improvement of the educational experience such as helping faculty and senior academic leaders to shape faculty development programs, revise curricula, or develop student support programs. Studies that link student engagement to university outcomes such as critical thinking, moral development, and leadership capacity, or to other indicators of success such as grades, persistence, and graduation, give credence to NSSE’s validity and support such valid uses of the data.

For example, research have found positive associations with persistence (Kuh et al. 2008; Kuh 2008; Hughes and Pace 2003; McClennen and Marti 2006), critical thinking (Loes, Pascarella, and Umbach 2012), GRE scores (Carini, Kuh, and Klein 2006), moral reasoning (Mayhew et al. 2012), and need for cognition (Padgett et al. 2010). Using institution-level data,
NSSE benchmarks had at least one significant positive association with institution-level outcome scores (effective reasoning and problem-solving, moral character, inclination to inquire and lifelong learning, intercultural effectiveness, and personal well-being) for first-year students after controlling for pre-test outcome scores (Pascarella, Seifert, and Blaich 2010).

Prior research has supported the use of self-reported data on university students (see Pace 1985 and Pike 2011), although some (e.g., Porter 2011) have raised questions about the validity of university student surveys. Cited concerns included a lack of a sufficient theoretical basis for survey content, difficulties in the response process, the lack of a factor structure and adequate reliability for NSSE’s benchmarks, and poor relationships between measures of student engagement and direct observations of the same behavior. In response, NSSE researchers explain that while the student engagement survey items are supported in the literature, the survey was created for institutional assessment, not for theory building or testing of a narrow theoretical construct. In addition, students’ ability to respond to the survey items has been established by extensive testing with hundreds of students at dozens of institutions using focus groups and cognitive interviews. Finally, the notion that a survey response on a question would match an objective source of the same information represents both a narrow conception of validity and a fundamental misunderstanding of how NSSE data are typically used: to make relative comparisons among groups of students. What matters is not a point estimate but that certain groups may be more or less engaged than their peers. For a more comprehensive discussion of NSSE’s validity, see McCormick, Kinzie, and Gonyea (2013) and NSSE’s (2018b) psychometric portfolio.
Generalizability Theory

GT, first detailed in a monograph by Cronbach, Gleser, Nanda, and Rajaratnam (1972), is a conceptual framework useful in determining the reliability and dependability of measurements. While classical test theory (CTT) focuses on determining the error of a measurement, GT recognizes that multiple sources of error may exist and examines their magnitude. These potential sources of error (e.g., individuals, raters, items, and occasions) are referred to as facets. The theory assumes that any observation or data point is drawn from a universe of possible observations. For example, an item on a survey is assumed to be sampled from a universe of comparable items, just as individuals are sampled from a larger population. Consequently, the notion of reliability in CTT is replaced by the question of the “accuracy of generalization or generalizability” to a larger universe (Cronbach et al. 1972, 15).

As a methodological theory, GT is intimately associated with its methods. The theory utilizes ANOVA to estimate the magnitude of variance components that are associated with the types of error identified by the researcher. The variance components are then used to calculate the generalizability coefficient. The coefficient is an intraclass correlation coefficient, however, the “universe score variance replac[es] the true score variance of CTT” (Kane and Brennan 1977, 271).

GT also distinguishes between a generalizability (G) study and a decision (D) study. The G-study estimates the variance components used to calculate the generalizability coefficient. However, the components can also be used in a D-study to estimate the generalizability coefficient in different contexts. This allows a researcher to efficiently optimize a study or to determine the conditions under which a score is generalizable.
Due to the focus on groups in educational assessment, GT makes important contributions to determining the validity of surveys, such as NSSE. The flexibility of GT allows for researchers to determine the conditions under which group means will be accurate and dependable. This is in contrast to the methods based on CTT that look at the internal consistency of a set of items, but fail to identify the conditions under which a measure is accurate. This weakness of CTT approaches may lead well-intentioned researchers to use a measure under conditions where its validity is questionable. Despite the benefits of GT, it has been underutilized in higher education research even after Pike’s (1994) work that introduced GT and its methods to the field.

**Research Questions**

Guided by GT, the study answered the following questions:

1. How dependable are the NSSE Engagement Indicators?
2. How many students are needed to produce dependable group means for the NSSE Engagement Indicators?

**Methods**

**Data**

The study utilized data from the 2013 NSSE administration. The survey was administered to 334,808 first-year students and seniors at 568 baccalaureate-granting institutions in the U.S. in the winter and spring of 2013. Table 1 presents the characteristics of the institutions, which roughly mirror the US landscape although public institutions and larger master’s colleges and universities were overrepresented. Additionally, bachelor’s-granting specialized institutions that did not receive one of the primary Carnegie Classifications were underrepresented in the NSSE dataset. Table 2 shows the gender, race/ethnicity, and enrollment status of the respondents by
class status. Approximately, two out of three respondents were female, the same proportion was White, and the vast majority enrolled as full-time students.

Measures

The measures used in the study were the survey items that comprised the NSSE Engagement Indicators (EI), groups of related items designed by NSSE researchers to measure the extent to which an institution’s environment promotes effective educational practices. The ten indicators are Higher-Order Learning (HOL; 4 items), Reflective & Integrative Learning (RIL; 7 items), Learning Strategies (LS; 3 items), Quantitative Reasoning (QR; 3 items), Collaborative Learning (CL; 4 items), Discussions with Diverse Others (DDO; 4 items), Student-Faculty Interaction (SFI; 4 items), Effective Teaching Practices (ETP; 5 items), Quality of Interactions (QI; 5 items), and Supportive Environment (SE; 8 items). The full list of the items and their coding can be found in Appendix A. The QI items had a “not applicable” option which was recoded to missing for this analysis. All items within each indicator shared the same response set and were not recoded.

Analyses

Guided by GT, the study examined the group mean generalizability of the NSSE Engagement Indicators at the institution level. We identified two facets, students and items, as potential sources of error for the indicators. We performed the following procedures to assess the generalizability of each EI by class. First, we conducted a G-study using a split-plot, random effects ANOVA design, where students were nested within institutions and crossed with survey items. In this design, each institution has a different set of students, but all students answered the same items. The design was also balanced, with 50 students randomly selected from each institution. The value of 50 was selected to maximize the number of students and institutions
included in the study after the exclusion of cases with missing data. The mathematical model for the design is:

\[ X_{usi} = \mu + \alpha_u + \pi_{s(u)} + \beta_i + \alpha\beta_{ui} + \beta\pi_{is(u)} + e_{usi} \]  

(1)

Where,

\( X_{usi} \) = Response by student \( s \) in institution \( u \) on item \( i \)

\( \mu \) = grand mean

\( \alpha_u \) = effect for institution \( u \)

\( \pi_{s(u)} \) = effect for student \( s \) nested within institution \( u \)

\( \beta_i \) = effect for item \( i \)

\( \alpha_u\beta_i \) = institution by item interaction

\( \beta_i\pi_{s(u)} \) = item by student, nested within institution, interaction, and

\( e_{usi} \) = error term.

Apart from the grand mean, each of the parameter estimates varies by institution, student, and/or item. This variation allows for the estimation of the variance components which decompose the total model variation into portions attributable to each effect. We used the G1 program for SPSS to analyze the data and estimate the variance components (Mushquash and O’Connor 2006).

After calculating the variance components in the G-study, we performed D-studies for each EI by class. A D-study allows a researcher to estimate how a generalizability coefficient would change if the study parameters changed. We estimated the generalizability coefficients over sample sizes of 25, 50, 75, and 100 students within an institution. By varying the number of students in the D-studies, the results allow us to investigate the dependability of the EIs and describe situations where the use of a group mean is and is not appropriate.
In the D-studies, we calculated two generalizability coefficients using formulas outlined by Kane, Gillmore, and Crooks (1976). The first coefficient generalized over both facets—students and items—and can be interpreted as the expected correlation of the group means derived from two samples of students at the same institution, who answered separate, but comparable items. This version of the generalizability coefficient should be used if a set of items is believed to represent a higher-order construct. This correlation could also be produced by developing a large number of survey items, giving half of the items to half of the students at each institution and correlating the mean to the mean of the other half of items given to the remaining students.

The second coefficient generalized only over students by treating the survey items as fixed, rather than random, effects. This coefficient can be interpreted as the expected correlation of the aggregated means of two samples of students who answered the same items. This formula should be used when a conclusion is to be drawn about a set of items, but not a higher-order construct. An analogous method to produce this correlation is to correlate the group means of two samples of students at each institution answering the same items.

**Limitations**

The primary limitation of the study is that it used the institution as the object of measurement. As there is more variability within than between institutions (NSSE 2008), the results may exhibit a non-trivial difference if the object of measurement utilized was major field or a demographic characteristic. For example, the dependability of an indicator like quantitative reasoning may be higher when the object of measurement is the group mean of a major field as this measure varies more between majors than it does between institutions (Rocconi, Lambert, McCormick, and Sarraf 2013). Additionally, missing data may have impacted the results. Due to
The balanced design, students who did not answer all of the items in an EI were dropped from that EI’s analysis. The failure to answer all of the items within an EI may indicate that a student engaged less in that activity than their peers, which would potentially bias the results. A less obvious source of missing data is that many students did not respond to the survey. The average institutional response rate was 27% for first-year students, and 33% for seniors (NSSE 2013). Data from these students could possibly alter the study’s conclusions. However, research based on the previous version of NSSE indicates that non-responders have relatively similar engagement patterns to responders and NSSE produces accurate group means at lower response rates (Fosnacht, Sarraf, Howe, and Peck 2017; Kuh 2003a; Sarraf 2005). Finally, it should be noted that generalizability or the related concept of reliability in classical test theory does not alone indicate that a measure is valid. Validity is a multifaceted topic and practitioners and researchers should examine other aspects of validity prior to concluding that an indicator is an accurate measure of student engagement (Messick 1989, 1995).

Results

Table 3 contains the generalizability coefficients when generalizing over students and items by class and the four sample sizes investigated in the D-studies. For first-year students, the CL, DDO, SE, and SFI EIs met accepted standards for dependability ($E_{p^2} \geq .70$) when a group mean was derived from sample of 25 to 75 students. The other EIs required substantially larger samples to meet the same threshold. For seniors, the same EIs had generalizability coefficients in excess of .70 using sample sizes of 25 to 50 students. However, the LS and QI EIs were dependable when a group mean contained at least 75 students. Coefficients for the remaining EIs were below the .70 standard using sample sizes of less than 100 when generalizing over students and items.
The generalizability coefficients when generalizing only over students are located in Table 4. In contrast to the coefficients over both students and items, nearly all of the EIs met standards for dependability using samples as low as 25 students. The exceptions were QR for both classes and LS for first-year students. All the generalizability coefficients were greater than .80 when group means contained 50 seniors, and all were in excess of .80 when group means contained 75 first-year students.

**Discussion**

The study found that the means of the NSSE Engagement Indicators can be reliably generalized to a larger population from small samples of students at postsecondary institutions. Therefore, the EIs appear to be dependable measurements of undergraduates’ engagement in beneficial activities at an institution during university. Eight of the ten indicators had generalizability coefficients above .70 for both first-year students and seniors, when an institution’s mean was derived from just 25 students. All EIs had generalizability coefficients in excess of .70 when the sample size increased to 50 students. Thus, the NSSE EIs can efficiently discriminate institutional environments that promote engagement in effective educational practices. In other words, using a relatively small sample of respondents, the EIs can identify institutions with high and low levels of engagement.

However, the results revealed that only some of the indicators could be dependably generalized to a higher-order construct. The CL, DD SFI, and SE EIs appear to be dependable group-level measures when generalizing over students and items and using sample sizes of 25 to 75 students. LS and QI also appear to be dependable for seniors using a sample size of at least 75 students. However, the remaining EIs do not appear to produce dependable group means representing a higher-order construct, except when the sample contains hundreds of students.
Therefore, when the object of measurement is an institution, the indicators with lower levels of dependability when generalizing over students and items would be most reliably treated as indexes (groups of items that, when combined, indicate a more general characteristic) rather than higher-order constructs. The lower level of dependability in these indicators, when generalizing over students and items, is generally caused by the small amount of variability accounted for by the institutional effects, which limits the ability to discriminate between institutional means.

It is not surprising that some of the indicators have poor dependability as higher-order constructs. NSSE was designed to estimate to undergraduates’ engagement in effective educational practices “known to be related to important [university] outcomes” (Kuh et al. 2001, 3). Therefore, the original NSSE Benchmarks and the newer Engagement Indicators do not contain items randomly selected from a domain of all possible questions related to a higher-order construct, but rather function as an index or snapshot of the level of engagement in specific beneficial activities known to improve university outcomes. While this study examined the generalizability of the Engagement Indicators over students and items, the purpose of and methods used to construct the survey suggest that this is not the appropriate criterion to assess the dependability of the NSSE Engagement Indicators. Instead, the more appropriate measure is the generalizability coefficient when generalizing over students, but not items.

NSSE is a rather blunt survey instrument that briefly examines multiple forms of student engagement to increase its utility for institutions and to ensure a reasonable survey length for respondents. The downside of this approach is that NSSE is unable to ask a detailed set of questions about each type of student engagement. As the accuracy of a student’s score is a function of a measurement’s reliability (Wainer and Thissen 1996) and the reliability is related to the number of items in a measurement (Brown 1910; Spearman 1910), the relatively small
number of items in each Engagement Indicator suggests that an individual’s score is associated with a nontrivial amount of error. However, NSSE overcomes this limitation by shifting the object of measurement from an individual student to the group level and aggregating the EIs into group means. Aggregation naturally increases the number of items in a measurement, which results in a higher degree of reliability and thus its accuracy. The same principle can be observed in this study’s results, as the generalizability coefficient increases along with the sample size.

Due to the survey’s design and the measurement error associated with an individual’s score, the main utility of NSSE data is in its ability to compare the experiences of different groups of students, where the grouping may be by institution, major, demographic group, program participation, etc. Measurement error is the most probable reason why some researchers have not observed strong correlations between the NSSE benchmarks and outcomes at the student level (Campbell and Cabrera 2011; DiRamio and Shannon 2010; Gordon, Ludlum, and Hoey 2008), while others have found clear associations between institutional aggregates of the NSSE benchmarks and their persistence and graduation rates (Pike 2013). Furthermore, these disparate results highlight the need for researchers to better understand the intended purposes of survey instruments like NSSE. As stated above, NSSE was designed to measure the extent to which an institution adopts programs and practices that have been demonstrated to be effective in promoting undergraduate learning and development. However, the survey was not designed to predict an individual student’s GPA or their probability of persistence; therefore, NSSE data most likely provide limited utility to these types of research endeavors.

The vast majority of variation in NSSE data occurs within institutions (NSSE 2008). In other words, students vary considerably more than institutions. Institutional researchers can exploit this variation to examine how a program or academic unit with a high graduation rate
impacts students. For example, by comparing the NSSE Engagement Indicators between participants and non-participants in a learning community with a high graduation rate, an institutional researcher may discover that the participants have more academic interactions with their peers and perceive a more supportive campus environment. Administrators may use this finding to justify expanding the program or to implement a portion of the program for all students. Similarly, enrollment in a major may be low because the faculty have poor pedagogical practices that can be improved upon through workshops or another type of intervention. These hypothetical examples illustrate how NSSE data can be used by institutions to identify areas of strength and weakness. After identifying these areas, institutions can intervene to improve areas of weakness and encourage other programs or academic units to adopt the practices of successful programs.

In summary, the mean of the NSSE Engagement Indicators can be dependably and accurately generalized to a broader population of students when derived from a relatively small sample of undergraduates. The number of students required to produce a dependable group mean varies by Engagement Indicator; however, a sample of 25 or 50 students is typically sufficient. Due to the relatively small number of students needed to produce a dependable group mean, the NSSE Engagement Indicators provide the opportunity for assessment professionals to investigate the level of student engagement in a variety of subpopulations. Finally, researchers should keep in mind that NSSE is intended to be used as a group-level instrument and was not designed to predict the outcome of an individual student.
References


Kuh, George D., Jillian Kinzie, Ty Cruce, Rick Shoup, and Robert M. Gonyea. 2006. *Connecting The Dots: Multi-Faceted Analyses Of The Relationships Between Student Engagement Results From The NSSE, And The Institutional Practices And Conditions That Foster Student Success: Final Report Prepared For Lumina Foundation For Education.* Bloomington, IN: Indiana University, Center for Postsecondary Research.


The Dependability of the Updated NSSE


NSSE (National Survey of Student Engagement). 2018a. *Information About the 2013 Update.*
http://nsse.indiana.edu/nsse-update/

NSSE (National Survey of Student Engagement). 2018b. *NSSE’s Commitment to Data Quality.*
http://nsse.indiana.edu/html/psychometric_portfolio.cfm


Table 1. Characteristics of Institutions Participating in NSSE 2013

<table>
<thead>
<tr>
<th>Institutional Characteristics</th>
<th>2013 NSSE¹</th>
<th>U.S.²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td><strong>Basic Carnegie Classification</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research Universities (very high research activity)</td>
<td>22</td>
<td>4</td>
</tr>
<tr>
<td>Research Universities (high research activity)</td>
<td>49</td>
<td>9</td>
</tr>
<tr>
<td>Doctoral/Research Universities</td>
<td>38</td>
<td>7</td>
</tr>
<tr>
<td>Master's Colleges and Universities (larger programs)</td>
<td>159</td>
<td>28</td>
</tr>
<tr>
<td>Master's Colleges and Universities (medium programs)</td>
<td>56</td>
<td>10</td>
</tr>
<tr>
<td>Master's Colleges and Universities (smaller programs)</td>
<td>29</td>
<td>5</td>
</tr>
<tr>
<td>Baccalaureate Colleges—Arts &amp; Sciences</td>
<td>81</td>
<td>14</td>
</tr>
<tr>
<td>Baccalaureate Colleges—Diverse Fields</td>
<td>101</td>
<td>18</td>
</tr>
<tr>
<td>Other</td>
<td>33</td>
<td>6</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Private</td>
<td>353</td>
<td>62</td>
</tr>
<tr>
<td>Public</td>
<td>215</td>
<td>38</td>
</tr>
<tr>
<td><strong>Region</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Far West</td>
<td>46</td>
<td>8</td>
</tr>
<tr>
<td>Great Lakes</td>
<td>93</td>
<td>16</td>
</tr>
<tr>
<td>Mid East</td>
<td>91</td>
<td>16</td>
</tr>
<tr>
<td>New England</td>
<td>43</td>
<td>8</td>
</tr>
<tr>
<td>Plains</td>
<td>69</td>
<td>12</td>
</tr>
<tr>
<td>Rocky Mountains</td>
<td>20</td>
<td>4</td>
</tr>
<tr>
<td>Southeast</td>
<td>143</td>
<td>25</td>
</tr>
<tr>
<td>Southwest</td>
<td>60</td>
<td>11</td>
</tr>
<tr>
<td>Outlying areas</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>U.S. Service Schools</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>568</td>
<td>100</td>
</tr>
</tbody>
</table>

¹ Excludes non-U.S. institutions and institutions with non-standard administrations

² U.S. percentages are based on data from the 2012-13 IPEDS Institutional Characteristics File
Table 2. Characteristics of NSSE 2013 Respondents

<table>
<thead>
<tr>
<th>Respondent Characteristics</th>
<th>First-Year (%)</th>
<th>Senior (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>35</td>
<td>37</td>
</tr>
<tr>
<td>Female</td>
<td>65</td>
<td>63</td>
</tr>
<tr>
<td><strong>Enrollment Status</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-time</td>
<td>6</td>
<td>19</td>
</tr>
<tr>
<td>Full-time</td>
<td>94</td>
<td>81</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Asian</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Black or African American</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>&lt; 1</td>
<td>&lt; 1</td>
</tr>
<tr>
<td>White</td>
<td>64</td>
<td>67</td>
</tr>
<tr>
<td>Other</td>
<td>&lt; 1</td>
<td>1</td>
</tr>
<tr>
<td>Foreign or Nonresident alien</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Two or more races/ethnicities</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Unknown</td>
<td>5</td>
<td>7</td>
</tr>
</tbody>
</table>

Notes: Characteristics were institution reported. Excludes non-randomly sampled students and students attending institutions outside of the U.S. or with non-standard administrations.
Table 3. D-study generalizability coefficients over students and items by class and sample size

<table>
<thead>
<tr>
<th></th>
<th>First-Year</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25 50 75 100</td>
<td>25 50 75 100</td>
</tr>
<tr>
<td><strong>Academic Challenge</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher-Order Learning</td>
<td>.48 .56 .59 .61</td>
<td>.46 .54 .57 .58</td>
</tr>
<tr>
<td>Reflective &amp; Integrative Learning</td>
<td>.56 .62 .64 .65</td>
<td>.60 .65 .67 .68</td>
</tr>
<tr>
<td>Quantitative Reasoning</td>
<td>.39 .48 .52 .55</td>
<td>.46 .55 .59 .61</td>
</tr>
<tr>
<td>Learning Strategies</td>
<td>.47 .58 .63 .65</td>
<td>.60 .70 .74 .76</td>
</tr>
<tr>
<td><strong>Learning with Peers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborative Learning</td>
<td>.75 .81 .83 .84</td>
<td>.75 .81 .83 .84</td>
</tr>
<tr>
<td>Discussions with Diverse Others</td>
<td>.62 .69 .72 .73</td>
<td>.62 .69 .71 .72</td>
</tr>
<tr>
<td><strong>Experiences with Faculty</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student-Faculty Interaction</td>
<td>.65 .72 .75 .76</td>
<td>.78 .83 .85 .86</td>
</tr>
<tr>
<td>Effective Teaching Practices</td>
<td>.47 .55 .58 .59</td>
<td>.50 .57 .60 .61</td>
</tr>
<tr>
<td><strong>Campus Environment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of Interactions</td>
<td>.52 .60 .63 .65</td>
<td>.58 .67 .70 .72</td>
</tr>
<tr>
<td>Supportive Environment</td>
<td>.68 .72 .74 .75</td>
<td>.71 .76 .77 .78</td>
</tr>
</tbody>
</table>
Table 4. D-study generalizability coefficients over students, but not items by class and sample size

<table>
<thead>
<tr>
<th></th>
<th>First-Year</th>
<th>Senior</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25 50 75 100</td>
<td>25 50 75 100</td>
</tr>
<tr>
<td><strong>Academic Challenge</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher-Order Learning</td>
<td>.72 .84 .89 .91</td>
<td>.72 .84 .89 .91</td>
</tr>
<tr>
<td>Reflective &amp; Integrative Learning</td>
<td>.81 .90 .93 .95</td>
<td>.85 .92 .94 .96</td>
</tr>
<tr>
<td>Quantitative Reasoning</td>
<td>.63 .77 .84 .87</td>
<td>.67 .80 .86 .89</td>
</tr>
<tr>
<td>Learning Strategies</td>
<td>.63 .77 .84 .87</td>
<td>.71 .83 .88 .91</td>
</tr>
<tr>
<td><strong>Learning with Peers</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborative Learning</td>
<td>.85 .92 .95 .96</td>
<td>.85 .92 .94 .96</td>
</tr>
<tr>
<td>Discussions with Diverse Others</td>
<td>.81 .89 .93 .94</td>
<td>.81 .90 .93 .95</td>
</tr>
<tr>
<td><strong>Experiences with Faculty</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student-Faculty Interaction</td>
<td>.80 .89 .92 .94</td>
<td>.88 .93 .96 .97</td>
</tr>
<tr>
<td>Effective Teaching Practices</td>
<td>.73 .84 .89 .92</td>
<td>.75 .86 .90 .92</td>
</tr>
<tr>
<td><strong>Campus Environment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quality of Interactions</td>
<td>.74 .85 .90 .92</td>
<td>.74 .85 .90 .92</td>
</tr>
<tr>
<td>Supportive Environment</td>
<td>.87 .93 .95 .96</td>
<td>.89 .94 .96 .97</td>
</tr>
</tbody>
</table>
Appendix A. Items Comprising the NSSE Engagement Indicators

Higher-Order Learning

During the current school year, how much has your coursework emphasized the following [Response options: Very little, Some, Quite a bit, Very much]:

- Applying facts, theories, or methods to practical problems or new situations
- Analyzing an idea, experience, or line of reasoning in depth by examining its parts
- Evaluating a point of view, decision, or information source
- Forming a new idea or understanding from various pieces of information

Reflective & Integrative Learning

During the current school year, how often have you [Response options: Never, Sometimes, Often, Very Often]:

- Combined ideas from different courses when completing assignments
- Connected your learning to societal problems or issues
- Included diverse perspectives (political, religious, racial/ethnic, gender, etc.) in course discussions or assignments
- Examined the strengths and weaknesses of your own views on a topic or issue
- Tried to better understand someone else's views by imagining how an issue looks from his or her perspective
- Learned something that changed the way you understand an issue or concept
- Connected ideas from your courses to your prior experiences and knowledge

Learning Strategies

During the current school year, how often have you [Response options: Never, Sometimes, Often, Very Often]:
• Identified key information from reading assignments
• Reviewed your notes after class
• Summarized what you learned in class or from course materials

Quantitative Reasoning
During the current school year, how often have you [Response options: Never, Sometimes, Often, Very Often]:
  • Reached conclusions based on your own analysis of numerical information (numbers, graphs, statistics, etc.)
  • Used numerical information to examine a real-world problem or issue (unemployment, climate change, public health, etc.)
  • Evaluated what others have concluded from numerical information

Collaborative Learning
During the current school year, how often have you [Response options: Never, Sometimes, Often, Very Often]:
  • Asked another student to help you understand course material
  • Explained course material to one or more students
  • Prepared for exams by discussing or working through course material with other students
  • Worked with other students on course projects or assignments

Discussions with Diverse Others
During the current school year, how often have you had discussions with people from the following groups [Response options: Never, Sometimes, Often, Very Often]:
  • People from a race or ethnicity other than your own
  • People from an economic background other than your own
• People with religious beliefs other than your own
• People with political views other than your own

**Student-Faculty Interaction**

During the current school year, how often have you [Response options: Never, Sometimes, Often, Very Often]:

• Talked about career plans with a faculty member
• Worked with a faculty member on activities other than coursework (committees, student groups, etc.)
• Discussed course topics, ideas, or concepts with a faculty member outside of class
• Discussed your academic performance with a faculty member

**Effective Teaching Practices**

During the current school year, to what extent have your instructors done the following [Response options: Very little, Some, Quite a bit, Very much]:

• Clearly explained course goals and requirements
• Taught course sessions in an organized way
• Used examples or illustrations to explain difficult points
• Provided feedback on a draft or work in progress
• Provided prompt and detailed feedback on tests or completed assignments

**Quality of Interactions**

Indicate the quality of your interactions with the following people at your institution [Response options: 1 = Poor, 2, 3, 4, 5, 6, 7 = Excellent, Not applicable]:

• Students
• Academic advisors
• Faculty
• Student services staff (career services, student activities, housing, etc.)
• Other administrative staff and offices (registrar, financial aid, etc.)

Supportive Environment

How much does your institution emphasize the following [Response options: Very little, Some, Quite a bit, Very much]:

• Providing support to help students succeed academically
• Using learning support services (tutoring services, writing center, etc.)
• Encouraging contact among students from different backgrounds (social, racial/ethnic, religious, etc.)
• Providing opportunities to be involved socially
• Providing support for your overall well-being (recreation, health care, counseling, etc.)
• Helping you manage your non-academic responsibilities (work, family, etc.)
• Attending campus activities and events (performing arts, athletic events, etc.)
• Attending events that address important social, economic, or political issues