UNIVERSITY OF HOUSTON
LAW CENTER
INSTITUTE FOR HIGHER EDUCATION LAW & GOVERNANCE
100 LAW CENTER
HOUSTON, TEXAS 77204-6060
713.743.2075 713.743.2085 FAX
WWW.LAW.UH.EDU/LAWCENTER/PROGRAMS/IHELG

MICHAEL A. OLIVAS
William B. Bates
Distinguished Chair in Law
Director, IHELG
molivas@uh.edu
713.743.2078

DEBORAH Y. JONES
Program Manager
dyjones@uh.edu

How Sound is NSSE? Investigating the Psychometric Properties of NSSE at a Public, Research Extensive Institution

IHELG Monograph 10-09

Corbin M. Campbell
301-793-9616
corbin@umd.edu
Doctoral Student
Department of Education Leadership, Higher and International Education
University of Maryland College Park
and
Alberto F. Cabrera
Professor
301-405-2875
Cabrera@umd.edu
Department of Education Leadership, Higher and International Education
University of Maryland College Park

IHELG

University of Houston Law Center/Institute for Higher Education Law and Governance (IHELG)

The University of Houston Institute for Higher Education Law and Governance (IHELG) provides a unique service to colleges and universities worldwide. It has as its primary aim providing information and publications to colleges and universities related to the field of higher education law, and also has a broader mission to be a focal point for discussion and thoughtful analysis of higher education legal issues. IHELG provides information, research, and analysis for those involved in managing the higher education enterprise internationally through publications, conferences, and the maintenance of a database of individuals and institutions. IHELG is especially concerned with creating dialogue and cooperation among academic institutions in the United States, and also has interests in higher education in industrialized nations and those in the developing countries of the Third World.

The UHLC/IHELG works in a series of concentric circles. At the core of the enterprise is the analytic study of postsecondary institutions--with special emphasis on the legal issues that affect colleges and universities. The next ring of the circle is made up of affiliated scholars whose research is in law and higher education as a field of study. Many scholars from all over the world have either spent time in residence, or have participated in Institute activities. Finally, many others from governmental agencies and legislative staff concerned with higher education participate in the activities of the Center. All IHELG monographs are available to a wide audience, at low cost.

Programs and Resources

IHELG has as its purpose the stimulation of an international consciousness among higher education institutions concerning issues of higher education law and the provision of documentation and analysis relating to higher education development. The following activities form the core of the Institute's activities:

Higher Education Law Library

Houston Roundtable on Higher Education Law

Houston Roundtable on Higher Education Finance

Publication series

Study opportunities

Conferences

Bibliographical and document service

Networking and commentary

Research projects funded internally or externally

Running Head: How sound is NSSE?

How sound is NSSE?
Investigating the psychometric properties of NSSE at a public, research extensive institution

Corbin M. Campbell
301-793-9616
corbin@umd.edu
Doctoral Student
Department of Education Leadership, Higher and International Education
University of Maryland College Park

Alberto F. Cabrera, Ph.D.
Professor
301-405-2875
Cabrera@umd.edu
Department of Education Leadership, Higher and International Education
University of Maryland College Park

Paper presented before the annual meeting of the Association of the Study of Higher Education (ASHE), November 2010, Indianapolis, Indiana

Abstract

The National Survey of Student Engagement (NSSE) Benchmarks has emerged as competing paradigm for assessing institutional effectiveness vis-a-vis the News and World Report. While gaining popularity, Porter (2009) noticed that NSSE did not seem to meet validity and reliability standards, two necessary conditions to endorse NSSE benchmarks as sound universal yardsticks of institutional quality. This study investigated whether the NSSE five benchmarks had construct and predictive validity for non-transfer seniors at one, mid-Atlantic research extensive institution. Findings indicated that a five benchmark model does not hold, with high inter-correlations among the benchmarks, low item loadings, and low reliability scores. Regarding predictive validity, we found that the NSSE Benchmark model was not a valid predictor of cumulative GPA for this institution.

Appraising the quality of American higher education has been dominated by the US News and World Report rankings for decades. The popularity of the US News and World Report Rankings' annual publication (2.2 million copies sold per year) attest to the importance of the rankings to both institutions and the public (Pike, 2006; Ehrenberg, 2003).

The unprecedented focus on US News rankings has had unintended consequences. One of them has been appraising quality based on student characteristics rather than the value added colleges and universities bestow to their students. This criticism was articulated the best by Russell Edgerton of the Pew Foundation in a 1999 *Chronicle of Higher Education* article, when he found rankings to be problematic due to their emphasis on standardized test scores while neglecting what students learned. Edgerton's ominous warning still rings true after one decade:

> Unless we develop measures of quality where colleges can actually provide evidence of their contribution to student learning, then this whole system [of ranking colleges] turns on resources and reputation, and reinforces the elitism of higher education (Gose, 1999, A65).

In the midst of calls for changing the rankings, the National Survey of Student Engagement (NSSE) emerged as the alternative paradigm in institutional accountability: *measuring the extent to which the students are engaged in high-impact practices as a way to assess institutional effectiveness* (Kuh, 2003, 2009). A panel of experts, gathered by the Pew foundation, was tasked with developing a survey that would measure student engagement under the well-accepted theory that more engagement means more learning (NSSE, 2008; Chickering & Gamson, 1987; Kuh, et al., 1994). The panel developed the

NSSE instrument, appraising institutional quality into five domains, known as Benchmarks of Effective Educational Practice: Level of Academic Challenge (LAC), Active and Collaborative Learning (ACL), Enriching Educational Experiences (EEE), Student-Faculty Interaction (SFI), and Supportive Campus Environment (SCE). The benchmarks are measured on a score of 0-100 in order "to facilitate comparisons across time, as well as between individual institutions and types of institutions" (NSSE, 2009).

NSSE appeared in the most auspicious moment: the public and policy arena were clamoring for valid tools for assessing institutional quality (Gose, 1999). A major endorsement on behalf of NSSE came in 2006 when Charles Miller, Chair of the Commission on the Future of Higher Education, suggested using the NSSE benchmarks as a viable tool for assessing educational quality (Miller & Malandra, 2006). Organizations representing higher education have lent more credence to NSSE. Sponsored by the American Association of State Colleges and Universities (AASCU) and the Association of Public Land-grant Universities (APLU), the *Voluntary System of Accountability* reports NSSE benchmarks, aimed at assisting students and parents in their college choice process (www.collegeportraits.com). NSSE itself has played a key role in marketing the benchmarks by encouraging institutions to post their benchmark results in outlets such as USA Today. On the USA today site, students and parents can see the results of the five benchmarks and how the institution compares to colleges and universities with a similar Carnegie classification.

Following the public calls for accountability and the strong endorsement of NSSE as an assessment tool, institutions have bought into NSSE as a way to benchmark their progress and satisfy external constituents. NSSE participation is growing each year. In

1999, 56 colleges participated (NSSE, 2010a). In 2009, nearly 1400 different colleges and universities have joined. In 2009, 443 institutions agreed to post their NSSE results to USA Today (up from 257 in 2007). Furthermore, institutions have used the benchmark scores internally to make substantial changes in policies, to highlight effective practices, and to compare with peer institutions.

The NSSE benchmarks are built with two important assumptions. The first is that student engagement is as precursor of such important students outcomes as learning and persistence (Kuh, 2003). The second assumption is that the NSSE benchmarks underscore five well defined, different, though interrelated, latent constructs of undergraduate student engagement with the institution that are applicable to all types of colleges and universities irrespective of their mission, Carnegie classification, location and type of students served.

While the items making up NSSE have strong theoretical grounding, to date, little work has been done to investigate the internal validity of the five NSSE benchmarks and the extent to which they predict relevant student outcomes (Carle, Jaffe, Vaughan, & Eder, 2009; Gordon, Ludlum, & Hoey, 2008; LaNasa, Cabrera & Tangsrud, 2009; Pascarella, Seifert & Blaich, 2008; Porter, 2009). Researchers at NSSE have conducted only minimal analyses on the construct validity of the benchmarks including Cronbach's alpha, stability correlations across time, and correlations across institutional types and student groups. To date, NSSE has not released on its website Confirmatory Factor Analyses, analyses using Item Response Theory, or other forms of more sophisticated techniques that are the most accepted methods for determining construct validity (Porter, 2009).

Of particular concern is the lack of research on whether the NSSE benchmarks prove reliable and valid on an institutional level (LaNasa, et al., 2009; Gordon, et al, 2008). A few studies have investigated the reliability and validity of other facets of the NSSE unrelated (or minimally related) to the five benchmarks (see, for example, Pike, 2006; Carle, et al., 2009; Kuh, et al., 2008). Even fewer studies examine the psychometric properties of the benchmarks themselves at an institutional level. Such dearth of research is perplexing considering the vast use of the benchmarks by institutions in comparing across time and to other institutions/groups of institutions. On the other hand, the two studies that investigated the benchmarks at a single institution have not produced strong results on behalf of the internal and predictive validity of NSSE benchmarks (LaNasa, et al., 2009; Gordon, et al., 2008). Those studies found that 1) the construct validity of certain benchmarks was either marginal or poor, 2) the benchmarks did not appear to be strongly associated with important student outcomes, like GPA, and 3) the benchmarks were highly inter-correlated: they appear not to measure distinct domains of student engagement.

The paradigm shift away from USNWR rankings towards measuring effectiveness through engagement is no doubt an improvement. However, as more institutions buy-in to NSSE as a universal tool for assessing engagement and effectiveness, the need to validate the NSSE becomes more and more critical. This study seeks to contribute to the body of research on NSSE benchmarks by examining the construct validity and reliability of the NSSE five benchmarks at a large, public, research extensive institution in the Mid-Atlantic.

In this context, this study seeks to answer three research questions pertaining to the construct and predictive validity of the NSSE-five benchmark model. These questions are:

1) Are there five separate, stable benchmarks that appraise engagement?
2) Do they apply to a single, large, public, research institution?
3) Do they predict cumulative GPA?

Review of the Literature

The NSSE benchmarks were developed based on theory that links student engagement to key collegiate outcomes, such as learning and development (NSSE, 2008; Kuh, 2009; Pascarella, et al, 2008). In this respect, NSSE rests on a set of well known and respected seven practices in undergraduate education formulated by Chickering and Gamson (1987). These practices, while elegant, underscore a decisively simple principle: students learn and develop when they participate in educationally purposeful activities with the different components of an institution (Astin, 1993; Kuh, 2000a; Chickering & Gamson, 1987).

Substantial research supports the engagement practices as valid predictors of learning and development. Engagement has been positively linked to a wide range of student outcomes such as persistence (DeSousa & Kuh, 1996; Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008), leadership development (Pascarella et al., 2008; Posner, 2004), identity development (Harper, Carini, Bridges, & Hayek, 2004; Hu & Kuh, 2003), moral development (Pascarella et al 2008), academic performance or GPA (Carini, Kuh, & Klein, 2006; Gordon et al., 2008; Kuh et al., 2008), and critical thinking skills (Anaya, 1996; Pascarella et al., 2008; Pike, 2000). Furthermore, student engagement is

influenced by a number of factors that are shaped by the collegiate environment, such as students' involvement on campus (Astin, 1999; Kuh, Hu, & Vesper, 2000) and the frequency of interaction between students and faculty (Pascarella & Terenzini, 2005) and between students and staff (Flowers, 2003; Kuh, 2009). Pre-college experiences also seem to matter in engagement. For example, students' experiences with diverse peers prior to attending college has been found to influence collegiate engagement (Locks, Hurtado, Bowman, & Oseguera, 2008).

According to the NSSE conceptual framework, engagement equals success. In this respect, NSSE conceptualizes engagement as falling within five well delineated domains. These five domains or benchmarks are: 1. Level of Academic Challenge (LAC), 2. Active and Collaborative Learning (ACL), 3. Student-Faculty Interaction (SFI), 4. Enriching Educational Experiences (EEE), and 5. Supportive Campus Environments (SCE). These domains reflect the two sides of the engagement equation: what the student does to become involved and what the institutions does to create meaningful engagement experiences.

Predictive Validity of NSSE Benchmarks

The principle of predictive validity calls for evidence of the extent to which a test or scale predicts subsequent outcomes or behaviors that are conceptually linked to the predictor measures they are supposed to index (Krathwohl, 2004, Vogt, 2005). If the NSSE benchmarks are indeed a valid measure of student engagement, they should be predictive of student learning across a variety of institutional types and student populations. Such evidence gathered from single and multi-institutional samples would

support the claim that indeed NSSE benchmarks are a valid method to ascertain institutional effectiveness.

We identified three studies that examined the predictive validity of NSSE benchmarks. Two of them drew from multi-institutional samples (Pascarella, et al, 2008; Carini, et al, 2006) and one study from a single institution (Gordon, et al, 2008). Overall, research on the predictive validity of NSSE has yielded mixed results.

Pascarella and colleagues (2008) used data from the Wabash National Study of Liberal Arts Education to investigate the predictive validity of NSSE benchmarks on six liberal arts outcomes (e.g. critical thinking, effective reasoning and problem solving, moral character, leadership, personal well-being). The study is compelling for two reasons. First, it uses a pre-post design, with controls for pre-college abilities. Second, it explores a wide variety of college outcomes, including previously validated measures such as the Collegiate Assessment of Academic Proficiency (CAAP). The researchers conducted analyses at both the individual and institutional levels. We focus on results at the individual level since averages for only 19 institutions were used in the institutional level analyses. Pascarella and associates excluded the EEE benchmark due to its low reliability. They also estimated upper bound (without controls of pre-college characteristics or the other NSSE benchmarks) and lower bound estimates (with controls) of the associations between NSSE and the liberal arts outcomes. At least one of the four benchmarks was significantly and positively associated with all of the outcomes when considering the upper bound estimates. However, when controls were introduced to account for correlations among NSSE benchmarks and precollege characteristics, many associations became non-significant. For example, the Supportive Faculty Interaction

benchmark displayed a significant and negative lower bound estimate with only one outcome, Critical Thinking, out of the 15 outcomes under examination. Notably, the effect size of associations that remained significant in the lower bound estimates ranged from a negative and weak association (-.049) to small associations (.15); the median association was also rather small (.082). While the longitudinal nature of this study is promising, it is based on NSSE data from liberal arts institutions only, and further research on the predictive validity for research institutions is warranted.

Another multi-institutional study by Carini and associates (2006) examined the association between the NSSE benchmarks and a few different measures of academic performance. The NSSE researchers included 1,058 first year and senior students at 14 institutions. They reported weak, but positive associations with certain NSSE benchmarks and academic achievement outcomes. However, Carini and colleagues did not account for intercorrelations among the NSSE benchmarks, as the Pascarella and associates' (2008) study did.

By contrast, Gordon and colleagues (2008) focused on an individual research institution in assessing the predictive validity of NSSE benchmarks. Gordon and associates used multiple years of freshmen and senior data at Georgia Tech to investigate whether the NSSE benchmarks were associated with several outcomes, including freshmen retention, GPA, pursuit of graduate education, and employment outcome at time of graduation. They found that three benchmarks had minimal associations with GPA for freshmen (LAC, ACL, and EEE). Of the three benchmarks, Enriching Educational Experiences had a negative association. For seniors, only Supportive Campus Environment was significant. None of them had effect sizes larger than .3.

Additionally, only one benchmark (SCE) had an association with freshmen retention. None of the benchmarks was significantly related to employment, but Student-Faculty Interaction was related to further education. This study is helpful in uncovering the possible associations with multiple important outcomes. However, this study did not examine the predictive validity of each of the five benchmarks simultaneously, while accounting for the degree of inter-correlation among the five benchmarks, an approach for which Structural Equation Modeling is ideally suited (Byrne, 2006; Hancock & Muller, 2009).

Construct Validity of NSSE

The principle of construct validity asks for evidence of the extent to which items or scales measure the construct they purport to represent (Cronbach & Meehl, 1955; Byrne, 2006; Bollen, 1989; Kline, 2005). In the case of NSSE benchmarks, the principle of construct validity calls for evidence that, indeed, the five benchmarks measure five distinct, yet inter-correlated, domains of engagement.

Porter (2009) reviewed the literature investigating the validity and reliability of the NSSE benchmarks using several methods (e.g. response bias and internal structure analyses). He found that the NSSE failed to meet validity or reliability standards. For example, he reported that the presumed structure of the five dimensions of engagement (i.e. NSSE benchmarks) had not been replicated and reliability values consistently failed to meet basic standards (i.e. alpha levels at or above .7). Regarding the individual items within the benchmarks, he found that students were not self-reporting accurate responses and that there was ambiguous wording. He also noted that there was not sufficient

evidence to support the claim that the NSSE benchmarks were associated with student outcomes.

A few scholars have investigated the construct validity of NSSE. LaNasa and associates (2009) used confirmatory factor analysis to investigate the construct validity of NSSE at a single research institution. Their findings indicated that the model fit was reasonable, but there were several item-level measurement problems. Loadings of many items were very low, particularly in the Active and Collaborative Learning benchmark. Moreover, they noted a high degree of overlap between the Student-Faculty Interaction and the Active and Collaborative Learning benchmarks; these two benchmarks had a large correlation of .89. For two of the five benchmarks the reliability index was below the recommended .7 threshold. While the confirmatory approach in this study was useful in understanding the construct validity of NSSE, the data were from a single institution. Evidently, the study needs replication to determine whether the measurement problems were more wide-spread or isolated to this one institution.

Some research has departed from using the five benchmarks in documenting the construct and predictive validity of NSSE by reconstructing NSSE into sub-scales or scalelets (e.g., Pike, 2006, Gordon et al., 2008; LaNasa, Olson, & Alleman, 2007). Pike (2006) used 2001 NSSE senior data to create 12 scalelets that measured engagement and reported that some NSSE scalelets were positively associated with student self-reports of gains in general education and practical skills. Gordon and colleagues (2008) later on used Pike's scalelets and found a marginal improvement over the NSSE benchmarks in terms of predictive validity, but they displayed poor reliability measures (all scalelets had alpha values less than .7). LaNasa, Olson, and Alleman (2007) found that NSSE items

grouped together into eight dimensions of student engagement. Several dimensions were moderately associated with both student satisfaction and first-year grade point average. Carle, et al (2009) used freshmen and senior data from a single institution to investigate the construct validity of three alternative engagement constructs. Carle, et al's study was particularly useful because of the sophistication of modeling techniques (Item-Response Theory) and confirmatory approaches. Even George Kuh, founder of NSSE, has been using alternative engagement measures. Kuh, et al (2008) employed three separate measures from NSSE: time spent in curricular activities, time spent studying, and a "global measure of engagement" made up of 19 educational activities to determine whether engagement was associated with persistence and grades. While this line of research is informative of the applicability of NSSE to examine relevant outcomes, it is unclear why these studies would focus on alternative scales instead of researching the actual benchmarks. The NSSE five benchmarks are the scales most often used to appraise institutional effectiveness.

NSSE's Psychometric Portfolio

Outside of peer reviewed publications, NSSE created a "psychometric portfolio" on its website to showcase research it conducted to ensure that the survey adequately represents measures of engagement and that it is psychometrically sound (NSSE, 2010b). According to the psychometric portfolio on the NSSE website, "As part of NSSE's commitment to transparency as well as continuous improvement, we routinely assess the quality of our survey and resulting data, and we embrace our responsibility to share the results with the higher education community" (NSSE, 2010b). The psychometric

portfolio includes research briefs on studies of reliability, validity, and other quality indicators.

The validity section of the portfolio includes studies of seven types of validity: response process, content, construct, concurrent, predictive, known groups, and consequential. Of particular relevance to the present study are the NSSE portfolio briefs on construct and predictive validity. In the construct validity brief, NSSE reports the results of a second order confirmatory factor analysis for a construct they call "deep learning" that has three sub-scales: higher-order, integrative, and reflective, proving that the select NSSE items seem to represent well the deep learning framework. It is pertinent to note that there is no portfolio brief on the construct validity of the five NSSE benchmarks of effective educational practices, which are the primary scales used by institutions to report NSSE results. The brief on predictive validity includes a study that found a positive relationship between four of the five NSSE benchmarks (all but SFI) and persistence/credits earned.

The reliability section of the psychometric portfolio on the NSSE website offers research briefs on three aspects of reliability: internal consistency, temporal stability, and equivalence (NSSE, 2010b). For internal consistency, they report Cronbach's alpha levels each year for the five benchmarks, which ranged from .591 (EEE) to .787 (SCE) for freshmen in 2009. They also report the intercorrelations among the benchmarks (mainly moderate) and inter-item correlations within each benchmarks (varies by benchmark, but mainly low to moderate).

It is clear that researchers at NSSE have conducted a wide array of psychometric studies on the properties of NSSE, but there are two glaring omissions. First, they have

not reported construct validation of the five benchmarks of effective educational practices. Second, they cite no research examining how well the benchmarks hold true for individual institutions. In fact, Kuh (2009) acknowledges that "Institution-specific analysis sometimes produce factor structures different than the five benchmarks or clusters of effective educational practices that NSSE uses to report its findings" (p. 687).

Methods

Data Collection

The focus institution contracted with NSSE to survey the population of senior students in the spring of 2009. The NSSE survey included 42 items that comprised the five benchmarks, among other items related to student engagement and institutional activities. The survey was administered by NSSE via the web, yielding a response rate of 28%.

Data Sources

This study used 2009 NSSE data from a large, public, research extensive university. Only non-transfer seniors were utilized, yielding an analytical sample of 1,026. Non-transfer seniors were chosen because they had four years to experience and be engaged with the institution. Institutional variables such as college cumulative GPA, high school GPA, and SAT score were added to the dataset.

Procedures

First, descriptive analyses were obtained to determine assumptions related to normality of the data. Certain items were recoded to categorical or dichotomous due to low variability and non-normality. Other items were recoded to match the NSSE coding of items for creating benchmarks.

Following Kuh (2000a, b, 2004), we tested a confirmatory factor model assuming that five dimensions account for the inter-correlations underlying the 42 items comprising the NSSE benchmarks. The model also presumed that the five dimensions are inter-correlated while being defined by unique items. Following recommendations from the Structural Equation Modeling literature (e.g., Bollen, 1988; Byrne, 2006; Kline, 2005), we adopted a two-step strategy in answering our three research questions. First, we used Confirmatory Factor Analysis (CFA) to answer the first two research questions, regarding whether there are five stable and distinct benchmarks. We set the variance of each latent construct associated to each benchmark to one, allowing us to ascertain the extent to which the items indeed loaded in their corresponding conceptual latent factor (Kline, 2005).

Next, we employed Structural Equation Modeling to answer our third research question regarding whether the benchmarks predict college cumulative GPA. While some scholars have debated whether GPA is useful as a sole predictor of achievement or collegiate success (Baird, 1985; Porter, 2009), there is strong evidence that college GPA is a relevant criterion of success in college. For example, it is related to graduation which is an ultimate goal of college (Cabrera, Burkum, & LaNasa, 2005). College GPA has a demonstrated relationship to admission to graduate school, and certain forms of occupational success (Allen, et al, 2008; Baird, 1985; Carini, et al, 2006; Pascarella & Terenzini, 1991, 2005; Tinto, 1993). Additionally, studies that have investigated the relationship between student engagement and college success often use GPA (as a proxy for academic ability) and persistence as the outcome variables of interest (see, for example, Kuh, et al 2008; Gordon, et al, 2008; Klein, et al, 2005). Finally, GPA has been

found to be one of the best predictors of persistence in college (e.g., Pascarella & Terenzini, 2005; Cabrera, Nora & Castaneda, 1992).

The model testing the predictive validity of NSSE on college GPA included a measure of pre-college academic ability that was a composite factor of high school GPA and SAT Math score. High school GPA and SAT Math score have been consistently cited in the literature as predictors of college academic success and persistence (see, for example, Tinto, 1993; Pascarella & Terenzini, 1991, 2005; Allen, et al, 2008; Kuh, et al, 2008).

We relied on four robust measures of fit to judge the CFA model and the SEM models. These indices include: a) the Satorra-Bentler Maximum Likelihood estimate of chi-square ($S\text{-}B\chi^2$), b) the Comparative Fit Index (CFI), c) the Non-Normed Fit Index (NNFI), and d) the Root Mean Square Error of Approximation (RMSEA). We guided our selection of goodness of fit values based on recommendations from the SEM literature (Byrne, 2006; Hu & Bentler, 1999). Accordingly, we sought CFI and NNFI values of 0.95 or higher to signify an excellent fit; but; we also considered values greater than .90 to be appropriate. In terms of RMSEA, we judged values ranging from 0 to .05 excellent; but, we also considered RMSEA values less than .08 to be suitable. In addition, we estimated 90% confidence intervals to check that RMSEA values did not fall beyond the cut off value of .10, signifying the rejection of the model. In contrasting alternative models underscoring NSSE benchmarks, we relied on the Satorra-Bentler corrected for non-normality chi-square test (see Byrne, 2006; Hancock & Muller, 2009).

Limitations

There are a few limitations of this study that are noteworthy. First, this study is based on NSSE data from a single, large, research extensive institution. While this study contributes to our understanding of the validity of NSSE on an institutional level, results may not be generalizable to other institutions, particularly institutions of other institutional types. Second, this study only investigates GPA as a measure of student success. While there is evidence that GPA is one useful measure (Allen, et al, 2008; Baird, 1985; Carini, et al, 2006; Pascarella & Terenzini, 1991, 2005; Tinto, 1993), further studies should investigate the predictive validity of NSSE in relation to other measures of success (e.g. graduation, other measures of student learning). A third limitation is that this study investigates only non-transfer seniors. If other populations were included, results could be different. However, non-transfer seniors likely have the best ability to respond to a survey of student engagement that appraises institutional effectiveness due to their longevity of experiences with the institution. Lastly, the response rate was 28%, which is somewhat lower than the NSSE average response rate in 2009 of 36%. However, the NSSE average response rate includes all participating institutions from all institutional types. It is possible that research institutions have a slightly lower response rates.

Results

Analyses were run using EQS software. After running initial descriptive analyses, we determined that 10 variables should be recoded and treated as dichotomous or ordinal, the remaining 32 variables were treated as continuous[1]. The first EQS model revealed that the data departed from normality. The standardized Mardia value of 6.71 was higher than the threshold of

[1] The following NSSE items were treated as dichotomous or ordinal: WRITEMOR, COMMPROJ, RESRCH04, INTERN04, VOLNTR04, LRNCOM04, FORLNG04, STDABR04, INDSTD04, SNRX04RE

5.0. Consequently, we relied on stepwise least robust methods contained in EQS 6.1 version 2010 in our model estimation. This method adjusts estimates due to departure of normality and can handle ordinal and continuous variables as the ones present in this study (Bentler, 2005; Byrne, 2006).

Construct Validity of NSSE

In testing our first two research questions, we used confirmatory factor analysis (CFA). The first CFA model tested the hypothesis that NSSE captures five benchmarks or factors that are independent of one another. This was used as the null model, in comparison to an alternative model that tested the hypothesis that the 42 items in NSSE account for five correlated constructs or benchmarks. Table 1 reports the results of testing these two alternative representations of NSSE.

Insert Table 1 here

Our first model is non-tenable. The TLI/NNFI and CFI fall far below the .90 threshold. The RMSEA for the model is .085 with a tenuous 90% confidence interval (CI_{90}).

The second model approached NSSE as representing five interrelated benchmarks. As shown in Table 1, this model represents a significant improvement over the null model ($\Delta S\text{-}B\ \chi^2 = 89.6$, $p < .05$). While the five correlated benchmark model is better than the five uncorrelated factor model, it does not appear to be a good representation of the data. Both the NNFI (.795) and the CFI (.807) were not supportive of the model, while the RMSEA was acceptable (.056, CI90= .053, .058).

A close of examination of factor results reveals substantial correlations among the latent constructs representing NSSE five benchmarks (see Table 2). Notable is the .864 correlation between the Active & Collaborative Learning and Student-Faculty Interactions benchmarks, suggesting that these two constructs may be indexing only one domain of student engagement. Equally interesting is the strong correlation between Enriching Educational Experiences and Active and Collaborative Learning (.749). The lowest correlation between benchmarks was between Supportive Campus Environment and Level of Academic Challenge, which was still a moderate-sized correlation (.435).[2]

Insert Table 2 here

Table 3 presents the loadings corresponding to each of the five benchmarks. More than half of the NSSE items have poor loadings, well below the recommended 0.500 threshold (Kline, 2005). Eight items had loadings less than .3; six additional items had loadings less than .4; and eight additional items had loadings less than.5. This leaves only 20 out of 42 items with loadings greater than the recommended threshold. When looking at each benchmark separately, the Supportive Campus Environment and Student Faculty Interaction benchmarks have a majority of items with loadings greater than the recommended threshold. By contrast, the Level of Academic Challenge, Active and

[2] We tested an alternative model to NSSE. One in which ACL & SFI benchmarks were collapsed into a single factor. This alternative 4 factor model was not supported. The corrected chi-square indicated that the alternative model significantly worsen the fit ($\Delta S\text{-}B\ \chi^2 = 130$, $p < .05$). Moreover, all robust measures of fit for this model were below the recommended threshold with the exception of the RMSEA.

Collaborative Learning, and Enriching Educational Experiences include a majority of items with loadings less than the recommended threshold.

Insert Table 3 here

Additionally, many items appear to explain little variance in the model and have substantial error (see Table 3). For the Level of Academic Challenge benchmark, the average percent error across the eleven items was 79%. The variance explained for each item in the LAC benchmark ranged from 2% to 42%, with six of eleven accounting for less than 15% of variance. For the Active and Collaborative Learning benchmark, the average percent error across the seven items was 79%. Variance explained for each item in ACL ranged from 14% to 28%. The average percent error across the six items in the Student-Faculty Interactions benchmark was 64%, with variance explained ranging from 9% to 48%. For the Enriching Educational Experiences benchmark, the average percent error across the twelve items was 84%. Variance explained for each item ranged from 1% to 31%, with less than 10% of variance explained for five out of twelve items. Of the five benchmarks, the Supportive Campus Environment benchmark had the least error in its items, with an average of 63% across loadings, with variance explained ranging from 25% to 49%.

In total, the average percent of error across the item loadings in each benchmark ranged from 63-84%. Particularly noteworthy is the average percent error in items for the Level of Academic Challenge (79%), Active and Collaborative Learning (79%), and

Enriching Educational Experiences (84%) benchmarks. Additionally, certain items seem to explain almost no variance and are almost entirely error. For example, one item in the Level of Academic Challenge benchmark accounts for only 2% of variance with 98% remaining of error. One item in the Enriching Educational Experiences benchmark accounts for only 1% of variance, with 99% remaining of error.

Predictive Validity of NSSE

In answering the third research question, we relied on Structural Equation Modeling to estimate the predictive validity of the five NSSE benchmarks in relation to college cumulative GPA, when controlling for high school academic ability (GPA and SAT math score). Results showed that the model explained 56% of the variance in cumulative GPA and the RMSEA of .057 is acceptable (CI_{90} = .054, .059). However, the CFI=.78 and the NNFI=.77 are well below the .90 thresholds. As a result, when all the Goodness of Fit indices are considered, the model is not a strong representation of the data (see Table 4).

Insert Table 4 here

Moreover, the results demonstrate that there are only two individual paths that have a significant influence on college cumulative GPA ($p<.05$). The first is high school academic achievement (GPA and SAT Math score), which was set as a control in this study. Expectedly, the effect of high school academic achievement on college cumulative GPA is quite strong (.663). By contrast, out of the five benchmarks, only Enriching Educational Experiences has a significant effect on cumulative GPA ($p<.05$),

and this effect, is rather strong (.653 in standardized units). The resulting model is shown in Figure 1, displaying the correlations among the five benchmarks and the paths between each benchmark and GPA.

Insert Figure 1 here

We also ran a model that included only the benchmarks as a predictor of GPA, without high school academic achievement as a control. The model explained only 21% of the variance in cumulative GPA. In terms of goodness of fit indices, the model represents a poor fit of the data. The CFI=.79 and the NNFI=.774 are well below the .90 thresholds. However, the RMSEA of .057 is acceptable (CI_{90} = .055, .060). Similar to results of the model that included pre-college academic achievement, Enriching Educational Experiences was the only benchmark that had a significant effect on cumulative GPA ($p<.05$), and this effect, is rather strong (.686 in standardized units).

Discussion

In 2009, Porter found that NSSE did not meet validity and reliability standards. In particular, he raised questions as to the extent to which engagement items a) have content validity, b) prompt accurate responses from students, and c) group together into five distinct domains as postulated by Kuh (2004). If Porter's conclusions are correct, one would expect to find NSSE items displaying poor loadings in their corresponding conceptual latent factors, lack of support on behalf of the underlying five-dimensionality of NSSE, and poor predictive power as well.

All in all Porter's hypotheses seem to hold true for this study. Our Confirmatory Factor Analyses show that the five benchmarks model does not hold for the focus institution. Of particular note was the substantial overlap between Active & Collaborative Learning with Student-Faculty Interactions (with a structural correlation of .864) and Enriching Educational Experiences (with a structural correlation of .749). In other words, the amount of overlap between these two pairs of benchmarks ranged from 56% to 75%, which questions the extent to which these three benchmarks appraise distinct dimensions of student engagement.

Second, each benchmark does not appear to be well appraised by the individual items, as evidenced by poor alpha reliabilities and low factor loadings. Particularly troublesome is the amount of error found in the Enriching Educational Experiences benchmark, with 8 of 12 items having loadings lower than .5. If treated as a scale, the alpha reliability of this benchmark is.543, which lies well below the .70 recommended threshold (Porter, 2009). The only benchmarks that are moderately well appraised by the NSSE items are Supportive Campus Environment and Student-Faculty Interaction benchmarks with average percent of error across loadings of about 60% and alpha reliabilities at or slightly above the recommended .70 threshold (see Table 3).

In terms of validity, we found that the model linking the five benchmarks with GPA represented a poor fit of the data. Two out of three indicators of fit rejected this model. Under this questionable model, only one of the four benchmarks, Enriching Educational Experiences, is predictive of GPA. Paradoxically, this benchmark is the least reliable.

Our results seem to both contradict and replicate those of the extant literature. While our predictive validity results appear to be in sharp contrast with those based on multi-institutional samples (e.g., Carini, et al, 2006; Pascarella, et al., 2008), they are in alignment with those based on samples drawn from a single institution (e.g., Gordon, et al, 2008). For instance, Carini and associates reported significant, though relatively small, partial correlations between 4 of the NSSE benchmarks and GPA among 1,058 undergraduate students drawn from 14 four-year colleges and universities.

On the other hand, our results are consistent with those reported by Gordon and associates (2008) and LaNasa and associates (2009). Gordon and colleagues investigated the predictive validity of the benchmarks at a major Doctoral-Extensive institution. They reported small, if not negligible, validity coefficients for the association between Level of Academic Challenge and Active Collaborative Learning benchmarks and GPA among freshman. Only Supportive Campus Environment was found to have a significant, yet small, effect size on seniors' GPA. Additionally, LaNasa and associates relied on a sample of freshman students from a Midwestern university. Like the present study, they too noted that the NSSE five benchmark model was a poor representation of the data. They also observed that the NSSE benchmarks substantially overlapped among themselves.

As far as reliability is concerned, our results are quite consistent with the literature in singling out Enriching Educational Experiences to be the least reliable benchmark (LaNasa, et al., 2009; Gordon et al., 2008; Pascarella, et al, 2008). As a matter of fact, this benchmark was deemed so unreliable that Pascarella and associates excluded it in their validation study of NSSE as a predictor of outcomes linked to liberal arts education.

Conclusion

Recently, Pascarella and colleagues (2008) judged NSSE benchmarks to be good measures of precursors of liberal arts outcomes. This conclusion, based on a solid research design that is longitudinal with powerful controls, heightens the argument on behalf of NSSE as a viable substitute of USNWR rankings to appraise institutional quality at liberal arts institutions. Our study, however, raises questions about the validity of NSSE to appraise quality for research extensive institutions. The results of this study in combination with the LaNasa and associates' (2009) study suggest that, at least for two research-extensive institutions, the NSSE benchmarks did not hold. It is also important to note that, even for the liberal arts institutions, some questions about validity remain: the Pascarella et al study reported small effect sizes, and untenable reliability scores for the EEE benchmark. They also did not investigate the construct validity of the benchmarks.

Our findings question the extent to which NSSE benchmarks are a universal tool for appraising institutional quality and can predict student outcomes, such as GPA. We echo Gordon, et al's (2008) advice to institutional researchers and policy makers. They should carefully examine the extent to which the five NSSE benchmarks are reliable and valid for their own institutional contexts before committing themselves to major organizational changes. If each of the five benchmarks does not measure a distinct dimension of engagement and includes substantial error among its items, it is difficult to inform intervention strategies to improve undergraduates' educational experiences. For example, if it is unclear what the EEE benchmark actually measures, the interventions could be targeting the wrong precursor of learning.

The changing paradigm from US News rankings to measuring student engagement is both courageous and more reflective of institutional quality. Yet, if the NSSE benchmarks are not psychometrically sound, it has the potential to misinform institutional stakeholders, such as prospective students, administrators, and policy makers.

References

Allen, J., Robbins, S.B., Casillas, A., & Oh, In-Sue. (2008). Third-year College Retention and Transfer: Effects of Academic Performance, Motivation, and Social Connectedness. *Research in Higher Education, 49*(7), 647-664.

Anaya, G. (1996). College experiences and student learning: The influence of active learning, college environments, and cocurricular activities. *Journal of College Student Development, 37*, 611-622.

Astin, A. W. (1993). *What matters in college? Four critical years revisited*. San Francisco: Jossey-Bass.

Astin, A.W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development, 40*, 518-529.

Baird, L. L. (1985). Do grades and tests predict adult accomplishment? *Research in Higher Education, 23*(1), 3-85.

Bentler, P.M. (2005). *EQS 6 Structural Equations Program Manual*. Encini, CA.: Multivariate Software.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Byrne, B.M. (2006). *Structural Equation Modeling with EQS* (Second Edition). Newbury

Park, CA: Sage.

Cabrera, A.F., Nora, A., & Castaneda, M.B. (1992). The role of finances in the persistence process: A structural model. *Research in Higher Education, 33*(5), 571-593.

Cabrera, A. F., Burkum, K. R., & La Nasa, S. M. (2005). Determinants of transfer and degree completion. *College Student Retention: A Formula for Student Success.* Westport, CN, ACE/Praeger, 174-214.

Carini, R.M., Kuh, G.D., & Klein, S.P. (2006). Student engagement and student learning: Testing the linkages, *Research in Higher Education 47,* 1-32.

Carle, A.C., Jaffe, D., Vaughn, N.W., & Eder, D. (2009). Psychometric Properties of Three New National Survey of Student Engagement Based Engagement Scales: An Item Response Theory Analysis. *Research in Higher Education*, 50(8), 775 794.

Chickering, A., & Gamson, Z. (1987). Seven principles of good practice in undergraduate education. *AAHE Bulletin*, 39, 3–7.

Chickering, A. W., & Reisser, L. (1993). *Education and identity.*San Francisco: Jossey-Bass.

Cronbach, L. & Meehl, P. (1953). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Gose, B. (1999). A new survey of 'Good Practices' could be an alternative to rankings. *Chronicle of Higher Education, A65*. Retrieved from http://chronicle.com/article/A-New-Survey-of-Good/4747/

DeSousa, J. & Kuh, G. (1996). Does institutional racial composition make a difference in

what black students gain from college*? Journal of College Student Development, 37*(3), 257-267.

Ehrenberg, R. (2003). Reaching for the brass ring: The *U.S. News & World Report* rankings and competition. *Review of Higher Education,* 26(2), 145-162.

Flowers, L.A. (2003). Differences in self-reported intellectual and social gains between African American and While college students at predominately White institutions: implications for student affairs professionals. *NASPA Journal, 41*(1), 68-84.

Gordon, J., Ludlum, J., & Hoey, J. J. (2008). Validating NSSE against student outcomes: Are they related? *Research in Higher Education*, 49(1), 19–39.

Hancock, G., & Mueller, R. (2009). *Structural Equation Modeling: The second course.* Chicago, IL: Scientific Software International.

Harper, S.R., Carini, R., Bridges, B., & Hayek, J. (2004). Gender differences in student engagement among African American undergraduates at historically Black colleges and universities. *Journal of College Student Development, 45*(3), 271-284.

Hu, L.-T., & 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.

Hu, S., & Kuh, G. D. (2003). A learning productivity model for estimating student gains during college. *Journal of College Student Development*, *44*, 185-203.

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.

Krathwohl, D. R. (2004). *Methods of educational and social sciences research: An*

integrated approach (2nd edition). Long Grove, IL: Waveland Press.

Kuh, G. D., Douglas, K. B., Lund, J. P., & Ramin-Gyurnek, J. (1994). *Student learning outside the classroom: Transcending artificial boundaries.* Washington, DC: The George Washington University, Graduate School of Education and Human Development.

Kuh, G. D. (2000a). *The NSSE 2000 report: National benchmarks for effective educational practice*. Bloomington, IN: Center for Postsecondary Research and Planning.

Kuh, G. D. (2000b). *The national survey of student engagement: The college student report*. Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Kuh, G. D. (2003). What we're learning about student engagement from NSSE: Benchmarks for effective educational practices. *Change, 35*(2), 24–32.

Kuh, G. D. (2004). *The National Survey of Student Engagement: Conceptual Framework and Overview of Psychometric Properties*. Bloomington, IN: Indiana University Center for Postsecondary Research.

Kuh, G.D. (2009). What Student Affairs Professionals Need to Know About Student Engagement. *Journal of College Student Development, 50*(6), 683-706.

Kuh, G., Cruce, T.M., Shoup, T. Kinzie, J.& Gonyea, R. M. (2008). Unmasking the effects of student engagement on First-Year college grades and persistence. *Journal of Higher Education*, 79(5), 540-563.

Kuh, G. D., Hu, S., & Vesper, N. (2000). They shall be known by what they do: In activities-based topology of college students. *Journal of College Student*

Development, 41, 228-244.

LaNasa, S.,Cabrera, A. F., & Tangsrud, H. (2009). The construct validity of student engagement: A confirmatory factor analysis. *Research in Higher Education*, *50*(4), 313-352.

LaNasa, S., Olson, E., & Alleman, N. (2007). The Impact of On-Campus Student Growth on First-Year Student Engagement and Success. *Research in Higher Education, 48(8). 941-966.*

Locks, A.M., Hurtado, S., Bowman, N., & Oseguera, L. (2008). Extending notions of campus climate and diversity to students' transition to college. *The Review of Higher Education, 31*(3), 257-285.

Miller, C. & Malandra, G. (2006). A national dialogue: The secretary of Education's Commission on the Future of Higher Education. US Department of Education. Retrieved from: http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/miller-malandra.pdf

National Survey of Student Engagement (NSSE). (2008). Origins. Retrieved from: http://nsse.iub.edu/html/origins.cfm.

National Survey of Student Engagement (NSSE). (2009). Assessment for improvement: Tracking student engagement over time—Annual results 2009. Bloomington, IN: Indiana University Center for Postsecondary Research.

National Survey of Student Engagement (NSSE). (2010a). NSSE Timeline 1998-2009. Retrieved from: http://nsse.iub.edu/pdf/NSSE_Timeline.pdf

National Survey of Student Engagement (NSSE). (2010b). Psychometric Portfolio. Retrieved from: http://nsse.iub.edu/_/?cid=154

Pascarella, E.T., Seifert, T.A., & Blaich, C. (2009). Validation of the NSSE Benchmarks and deep approaches to learning against Liberal Arts outcomes. Paper presented before the Association for the Study of Higher Education.Vancouver, Canada.

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: A third decade of research.* San Francisco: Jossey-Bass.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research.* San Francisco: Jossey-Bass.

Pike, G. (2000). The influence of fraternity or sorority membership on students' college experiences and cognitive development. *Research in Higher Education, 41*(1), 117-139.

Pike, G. R. (2006). The convergent and discriminant validity of NSSE scalelet scores. *Journal of College Student Development*, 47(5), 550–563.

Porter, S. R. (2009). *Do college student surveys have any validity?* Paper presented before the Annual Meeting of the Association for the Study of Higher Education. Vancouver, Canada.

Posner, B.Z. (2004). A leadership instrument for students: Updated. *Journal of College Student Development, 45*(4), 443-456.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd edition). Chicago: University of Chicago Press.

Vogt, P. W. (2005). *Dictionary of statistics and methodology: A nontechnical guide for the social sciences* (3th edition). London: Sage.

Table 1. Goodness of fit indices for alternative models of NSSE Benchmarks

		Robust measures of goodness of fit							Changes in robust measures of fit		
Model	ML χ^2	***S-B*** χ^2	***df***	**TLI/ NNFI**	**CFI**	**RMSEA**	**RMSEA CI_{90}**		**Corrected** Δ ***S-B*** χ^2	Δ ***df***	***p-value***
1. Five independent constructs	6,117.3	8,107.2	819	.524	.547	.085	.083	.087	-	-	-
2. Five interdependent constructs	5,252.8	2,806.3	809	.795	.807	.056	.053	.058	-89.6	10	<.05

Table 2. Structural correlations among the NSSE five benchmarks					
Benchmark	1	2	3	4	5
1. Level of Academic Challenge	1				
2. Active & Collaborative Learning	.687	1			
3. Student-Faculty Interactions	.633	.864	1		
4. Enriching Educational Experiences	.593	.749	.663	1	
5. Supportive Campus Environment	.435	.490	.600	.516	1

Table 3. ***Loadings and variance accounted for in the Five NSSE Benchmark model***

Benchmark	Measure	Loading	Variance		Average % of error across loadings	Reliability of the Benchmark
			Explained	Error		
1. Level of Academic Challenge	ACADPR01	.317	.100	0.900	79%	.661
	WORKHARD	.542	.294	0.706		
	READASGN	.146	.021	0.979		
	WRITEMOR	.338	.114	0.886		
	WRITEMID	.267	.071	0.929		
	WRITESML	.205	.042	0.958		
	ANALYZE	.611	.374	0.626		
	SYNTHESZ	.650	.422	0.578		
	EVALUATE	.631	.398	0.602		
	APPLYING	.623	.389	0.611		
	ENVSCHOL	.324	.105	0.895		
2. Active & Collaborative Learning	CLQUEST	.435	.189	0.811	79%	.618
	CLPRESEN	.486	.236	0.764		
	CLASSGRP	.379	.144	0.856		
	OCCGRP	.440	.193	0.807		
	TUTOR	.425	.180	0.820		
	COMMPROJ	.532	.283	0.717		
	OCCIDEAS	.486	.237	0.763		
3. Student-Faculty Interactions	FACGRADE	.516	.266	0.734	64%	.708
	FACPLANS	.690	.476	0.524		
	FACIDEAS	.588	.346	0.654		
	FACFEED	.557	.310	0.690		
	FACOTHER	.579	.335	0.665		
	RESRCH04	.299	.089	0.911		
4. Enriching Educational Experiences	DIVRSTUD	.518	.268	0.732	84%	.543
	DIFFSTU2	.508	.258	0.742		
	ITACADEM	.448	.200	0.800		
	ENVDIVRS	.463	.214	0.786		
	INTERN04	.394	.155	0.845		
	VOLNTR04	.557	.310	0.690		
	LRNCOM04	.511	.262	0.738		
	FORLNG04	.106	.011	0.989		
	STABR04	.192	.037	0.963		
	INDSTD04	.282	.079	0.921		
	SNRX04RE	.268	.072	0.928		
	COCURR01	.305	.093	0.907		

Table 3. ***Loadings and variance accounted for in the Five NSSE Benchmark model***						
Benchmark	**Measure**	**Loading**	**Variance**		**Average % of error across loadings**	**Reliability of the Benchmark**
			Explained	Error		
5. Supportive Campus Environment	ENVSUPRT	.701	.492	0.508	63%	.758
	ENVNACAD	.639	.408	0.592		
	ENVSOCAL	.632	.399	0.601		
	ENVSTU	.498	.248	0.752		
	ENVFAC	.641	.411	0.589		
	ENVADM	.536	.287	0.713		

Table 4. Goodness of fit indices for model of NSSE Benchmarks' Predictive Validity with College Cumulative GPA

		Robust measures of goodness of fit						
Model	ML χ^2	***S-B*** χ^2	***df***	**TLI/ NNFI**	**CFI**	**RMSEA**	**RMSEA CI_{90}**	
NSSE Predictive Validity of GPA	9098.5	2693.4	930	.768	.783	.057	.054	.059

Figure 1. NSSE 5 BENCHMARKS AS PREDICTORS OF GPA

Standardized Solution