How Effective are the NSSE Benchmarks IN PREDICTING IMPORTANT EDUCATIONAL OUTCOMES?

By Ernest T. Pascarella, Tricia A. Seifert, and Charles Blaich

Ernest T. Pascarella is a professor and the Mary Louise Petersen Chair in Higher Education at the University of Iowa, where he also co-directs the Center for Research on Undergraduate Education. His work focuses on the impact of college on students. Tricia A. Seifert is an assistant professor in the higher education group at the Ontario Institute for Studies in Education at the University of Toronto. Her research interests include identifying postsecondary educational experiences and contexts that foster student success. Charles Blaich is the director of the Center of Inquiry in the Liberal Arts at Wabash College. He taught at Eastern Illinois University from 1987–1991 and moved to Wabash College in 1991, becoming director of the center in 2002.

This research was supported by a generous grant from the Center of Inquiry in the Liberal Arts at Wabash College to the Center for Research on Undergraduate Education at the University of Iowa.
Over the last three decades, colleges and universities increasingly have been expected to be accountable for the quality of undergraduate education. Although focused on K-12 education, *A Nation at Risk* launched this expectation in earnest in 1983. Numerous subsequent papers urging attention to student learning culminated in the 2006 report of the US Department of Education’s Commission on the Future of Higher Education (the “Spellings Commission”), where we see a palpable sense of urgency for colleges to demonstrate that they provide a high-quality undergraduate education by reporting on their students’ cognitive and personal development.

By the time the Spellings Commission did its work, the conversation that had previously been kept within a concerned group of higher education stakeholders had enlarged to engage not only higher education faculty and administrators but also parents, employers, and the general public. Beyond simply knowing the characteristics of entering students and graduation statistics, a growing segment of the population has become interested in knowing what students actually learn in college.
In response to this need for information, the American Association of State Colleges and Universities (AASCU) and the then-National Association of State Universities and Land Grant Colleges (now the Association for Public and Land-Grant Universities, or APLU) developed the Voluntary System of Accountability (VSA) for the purpose of helping postsecondary institutions demonstrate accountability by providing accessible, understandable, and comparable information about the effective educational practices they engage in and the learning outcomes they generate. Through the College Portrait, participating institutions provide information on student and campus characteristics, cost, and progression and success rates, as well as student educational experiences on campus and student learning outcomes. With 70 percent of four-year college students attending one of AASCU’s or APLU’s member institutions, the VSA has the potential to reach an extensive audience and create a market in which information about student experiences and learning inform students’ and parents’ college decisions.

Several college-experience surveys are acceptable for institutions’ completion of the “Student Experiences and Perceptions” portion of the College Portrait. None, however, reach the number of students as the National Survey of Student Engagement (NSSE), which is one of the most widely used annual surveys of undergraduates in the country. According to the NSSE 2007 annual report (Experiences That Matter: Enhancing Student Learning and Success, 2008), the NSSE survey has been completed by nearly 1.5 million students at nearly 1,200 colleges and universities in the last decade. In 2008 alone, 774 different colleges and universities participated in the annual spring administration of the 15–20 minute survey.

The NSSE is specifically designed to assess the extent to which college students are engaged in empirically vetted good practices in undergraduate education. Indeed, one of the major assumptions of the NSSE is that in measuring the extent to which students engage in such practices, one is indirectly measuring student cognitive and personal development during college. Other things being equal, the greater students’ engagement in or exposure to these educationally effective practices in college, the greater their development—or so the logic goes.

The individual items in the survey instrument are grouped into five NSSE Benchmarks of Effective Educational Practice: the Level of Academic Challenge, Active and Collaborative Learning, Student-Faculty Interaction, Enriching Educational Experiences, and Supportive Campus Environment. Given the NSSE’s broad-based national and even international use, it seems reasonable to ask whether the good practices in undergraduate education that it measures actually do predict important educational outcomes. With some narrowly focused exceptions (Carini, Kuh, & Klein, 2006; LaNasa, Olson, & Alleman, 2007) however, nearly all the predictive validity evidence in this regard is based on studies that link the various NSSE measures of good practices to student self-reported gains in intellectual and personal development that are assessed by a set of 16 items near the end of the NSSE instrument itself. (For a review of those studies, see Pascarella, Seifert, and Blaich [2008].)

Although such self-reported gains can be formed into psychometrically reliable scales, serious problems exist with the internal validity of any findings in which self-reported gains are taken to be a learning outcome of the educationally effective practices that the NSSE targets. Recent evidence reported by Bowman (in press) indicates that there is little or no overlap between self-reported gains and longitudinal (pre-test-post-test) gains made on standardized, more objective instruments.

Furthermore, students complete the self-reported gains part of the NSSE at the same time that they complete the benchmark items. But when researchers do not have a precollege measure of an individual student’s receptiveness to educational experiences, it is difficult—if not impossible—to distinguish how much of that student’s “gain” on some outcome is due to the added value of college from how much is simply due to his or her disproportionate openness and receptivity to the college experience. Two students having the same educational experience could report substantially different gains because they enter college differentially receptive to the effects of postsecondary education. Without a precollege measure of the students’ propensities to report learning gains (e.g., self-reports of such gains during high school), it is nearly impossible to take this differential receptiveness to educational experiences into account.

Thus, considering the NSSE self-reported gains in college to be an outcome of good practices risks confounding the effects of exposure to good practices with the individual characteristics of the students an institution attracts and admits (Astin & Lee, 2003; Pascarella, 2001).

• The Level of Academic Challenge is an eleven-item scale on which students report about the time they spend preparing for class, the amount of reading and writing they have done, and institutional expectations for academic performance.

• Active and Collaborative Learning is a seven-item scale on the extent of students’ class participation, the degree to which they have worked collaboratively with other students inside and outside of class, and the amount of tutoring and number of community-based projects in which they have been involved.

• The Student-Faculty Interaction scale consists of six items. Students report on the extent of their interaction with faculty members and advisors and their discussions of ideas with faculty members outside of class; they also report on the extent of prompt feedback on academic performance and work with faculty on research projects.

• Enriching Educational Experiences is a scale with twelve items probing the extent of students’ interaction with those of different racial or ethnic backgrounds or with different values or political opinions; their use of information technology; and their participation in activities such as internships, community service, study abroad, and co-curricular activities.

• Finally, Supportive Campus Environment is a six-item scale measuring the extent to which students feel that the campus helps them succeed academically and socially; assists them in coping with nonacademic responsibilities; and promotes supportive relations among students and their peers, faculty members, and administrative personnel and offices.
The bottom line is that we have, at present, very little internally valid evidence with respect to the predictive validity of the NSSE. This is a serious concern if participating post-secondary institutions are asked to consider the NSSE benchmark scales as a proxy for student growth in important areas. Consequently, pursuant to a subcontract from the Center of Inquiry in the Liberal Arts at Wabash College, the Center for Research on Undergraduate Education at the University of Iowa analyzed institution-level data from the first year of the Wabash National Study of Liberal Arts Education to estimate the validity of the NSSE benchmarks in predicting seven traits and skills thought to be the outcomes of a general liberal arts education. Our study measured those outcomes directly; it also addressed the limitations of past research on the NSSE by using a longitudinal pre-test-post-test approach. No other investigation of which we are aware provides such a comprehensive validation of the NSSE benchmark scales.

**The Wabash National Study of Liberal Arts Education**

The Wabash study is a longitudinal (using a pre-test-post-test design) investigation of the institutional experiences that enhance growth in important educational outcomes. It measures student development during the first year of college on a range of dimensions derived from the model of college outcomes historically associated with a liberal arts education that was developed by King, Kendall Brown, Lindsay, and VanHecke (2007). The five liberal arts outcomes addressed in our analyses were as follows:

- **Effective Reasoning and Problem Solving**
- **Moral Character**
- **Inclination to Inquire and Lifelong Learning**
- **Intercultural Effectiveness**
- **Personal Well-Being**

Nineteen institutions from eleven different states participated in the Wabash study. The group comprised a mix of liberal arts colleges, regional institutions, research universities, and community colleges. The analyses reported here are based on data from 1,426 first-year students at these 19 institutions who took the Critical Thinking Test, 1,446 different first-year students (including both previous samples) who completed all other measures (see sidebar).

Data were collected from these first-year students when they entered college in fall 2006 and again at the end of their first year of postsecondary education in early spring 2007. As the students entered in fall 2006, they completed the seven liberal arts outcome measures. In the follow-up data collection in spring 2007, the same students first completed the National Survey of Student Engagement and then once again completed the posttests of the seven liberal arts outcome measures.

**Analysis of the Wabash Study Data**

Since the NSSE benchmarks are designed to provide an institution-level assessment of exposure to good practices, institutions were our unit of analysis. Therefore, we aggregated the responses of the sample at each institution to obtain an average institution-level score on each of the seven liberal arts outcomes assessments and on each of the five NSSE benchmark scales.

With a sample of only 19 institutions, we were somewhat limited both in the statistical power to uncover significant findings and with respect to the sophistication of our analytic...
approach. However, the longitudinal nature of the Wabash study data did permit us to estimate the associations between the average NSSE benchmarks and the average of each liberal arts outcome in spring 2007, while taking into account what was arguably the most important confounding influence on the latter—the average institutional-level score of the fall 2006 entering students. We estimated the partial correlation between the average institution-level scores on each NSSE benchmark and each post-test liberal arts outcome measured in spring 2007, while statistically controlling for the pre-test score on the corresponding instrument measured in fall 2006.

**WHAT WE FOUND**

Our analyses completed, we concluded that institution-level NSSE benchmark scores had a significant overall positive association with the seven liberal arts outcomes at the end of the first year of college, independent of differences across the 19 institutions in the average score of their entering student population on each outcome. The mean value of all the partial correlations summarized in Table 1 was .34, which had a very low probability (.001) of being due to chance.

As Table 1 further shows, in the presence of controls for the average institutional precollege score, at least one of the NSSE benchmarks had a significant association with each of the end-of-first-year liberal arts outcomes measures except the Need for Cognition Scale. Across all liberal arts outcomes, the most influential NSSE benchmark appeared to be the Enriching Educational Experiences scale, which had significant associations with three of the seven outcomes: effective reasoning and problem-solving, moral character, and intercultural effectiveness. The Supportive Campus Environment benchmark also correlated with intercultural effectiveness, as well as with personal well-being. The Level of Academic Challenge benchmark had significant partial associations with critical thinking and the inclination-to-inquire and lifelong-learning goals, while the Active and Collaborative Learning benchmark had a significant partial correlation with intercultural effectiveness.

Only the Student/Faculty Interaction benchmark failed to have a significant partial correlation with at least one of the seven liberal arts outcomes. We posit that this lack of association could be the result of the variable circumstances under which students interact with faculty. For example, students who are in need of a lot of academic assistance may have as much interaction with faculty members as those students who excel academically.

Clearly our institution-level results are limited by the small sample (19 institutions). Although it contained a wide variety of institutional types, it certainly cannot be considered a statistically representative national sample of colleges and universities, and its attendant limiting effect on statistical power also reduced the sophistication of the analytic procedures we employed and led us to rely on rather straightforward partial correlations.

<table>
<thead>
<tr>
<th>Liberal Arts Outcome</th>
<th>NSSE Benchmark Scales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Level of Academic Challenge</td>
</tr>
<tr>
<td>Effective Reasoning and Problem Solving</td>
<td>.43*</td>
</tr>
<tr>
<td>CAAP Critical Thinking Test</td>
<td></td>
</tr>
<tr>
<td>Moral Character</td>
<td>.39</td>
</tr>
<tr>
<td>Defining Issues Test-N2 Score</td>
<td></td>
</tr>
<tr>
<td>Inclination to Inquire and Lifelong Learning</td>
<td>.27</td>
</tr>
<tr>
<td>Need for Cognition Scale</td>
<td></td>
</tr>
<tr>
<td>Positive Attitude Toward Literacy Scale</td>
<td>.51*</td>
</tr>
<tr>
<td>Intercultural Effectiveness</td>
<td></td>
</tr>
<tr>
<td>Miville-Guzman Universality-Diversity Scale</td>
<td>.33</td>
</tr>
<tr>
<td>Openness to Diversity/Challenge Scale</td>
<td>.30</td>
</tr>
<tr>
<td>Personal Well-Being</td>
<td></td>
</tr>
<tr>
<td>Ryff Scales of Psychological Well-Being (Total Score)</td>
<td>.31</td>
</tr>
</tbody>
</table>

*Statistically significant with controls for school-level precollege score.
However, the longitudinal nature of the Wabash study data permitted us to control for precollege scores on each first-year liberal arts outcome, yielding a more valid estimate of the “value added” by the college experience. Moreover, the Wabash study allowed us to look at first-year student development on a wide range of liberal arts outcomes that were measured with direct standardized instruments of vetted reliability and validity. We know of no other data that would permit such a comprehensive institutional-level assessment of the predictive validity of the NSSE benchmark scales.

**Implications for Policy**

One cannot make strict causal claims with correlational data, even in the best-case scenario when the study design is longitudinal. Although we controlled for the average institution precollege score on each outcome, this is certainly not the only possible source of confounding influence. So our results need to be considered with caution.

That said, our findings nevertheless support the claim that the NSSE results regarding educational practices and student experiences are good proxy measures for growth in important educational outcomes such as critical thinking, moral reasoning, intercultural effectiveness, personal well-being, and a positive orientation toward literacy activities. Even with controls for the average institutional-level precollege score, there were discernible differences among institutions in the average end-of-first-year educational outcomes that were significantly and positively linked to average institutional scores on the NSSE benchmarks. Our findings suggest that institutions using the NSSE can have reasonable confidence that the benchmark scales do, in fact, measure exposure to experiences that predict student progress on important educational outcomes, independent of the level on these outcomes at which an institution’s student body enters college.

Such findings have non-trivial implications for institutional assessment expenditures. In the present economic climate, the institutional costs incurred in gathering the information needed to complete the VSA College Portrait, particularly the Student Experiences and Perceptions and the Student Learning Outcomes sections, are daunting. Not all colleges may be able to absorb these costs. Our findings suggest that increases on institutional NSSE scores can be considered as reasonable proxies for student growth and learning across a range of important educational outcomes. Thus, if an institution can only afford to focus on the “process” of undergraduate education as measured by the NSSE benchmarks, this nevertheless seems likely to have implications for the “product.”

Of additional importance is the fact that these significant partial associations between the NSSE benchmarks and liberal arts outcomes were uncovered in a small sample with very low statistical power. Although institutional use of the NSSE is usually oriented toward a broader sample of students from all undergraduate classes, our findings suggest that the NSSE may also be used to focus on the effectiveness of the first year of college—a period of time during which, some multi-institutional evidence suggests, the greatest developmental impact of postsecondary education occurs (Flowers et al., 2001; Pascarella & Terenzini, 2005).

Moreover, our findings may understate the links between the NSSE results and our various outcome measures. One might reasonably expect good practices in undergraduate education to demonstrate somewhat stronger impacts on student development during the subsequent years of college, when such practices have had longer periods of time to exert their influence.

Our findings also provide fodder in the ongoing national debate over quality in undergraduate education. *U.S. News and World Report*’s annual ranking of postsecondary institutions has strongly shaped public understanding of what constitutes a high-quality undergraduate education in the US. But these rankings operationally define “quality” largely in terms of resources, reputation, and the academic selectivity of an institution’s undergraduate student body. Indeed, there is sound evidence to suggest that the *U.S. News* rankings can be essentially reproduced simply by knowing the average ACT/SAT score of each institution’s first-year class (Pascarella et al., 2006). This means that quality is measured by the characteristics students bring to college and not by the actual effectiveness of the academic and non-academic experiences students have after they arrive on campus.

The NSSE benchmark scales were designed specifically to provide another gauge of academic quality—students’ participation in academic and non-academic experiences that lead to learning—and there is little evidence that such experiences are substantially linked to the academic selectivity of the college one attends (Pascarella et al., 2006). Since our findings suggest the dimensions of the undergraduate experience measured by NSSE benchmarks are correlated with important educational outcomes, they arguably constitute a more valid conception of quality in undergraduate education than *U.S. News*’s.

Furthermore, the NSSE results point to academic and non-academic experiences that may be amenable to improvement through changes in institutional policies and practices. On the other hand, resources and academic selectivity are much harder to change and therefore may form a much more deterministic institutional identity. To the extent that an institution is actually concerned with the quality and effectiveness of the undergraduate education it provides, our findings suggest that it probably makes more sense to focus on implementing practices and experiences measured by the NSSE benchmarks than on those factors measured by *U.S. News*.

In a dynamic context grounded in an institution’s commitment to improvement, an institutional culture may arise that continuously strives to engage students in effective educational practices and experiences, thereby increasing the likelihood of improved institutional effectiveness and increased student learning and development.
A recent study by the American Association of Community Colleges shows that as many as 80% of community college leaders are expected to retire within the next five to ten years. Combined with rapid enrollment growth, that means high demand for administrative leaders. Be ready with the Doctor of Management (DM) in community college policy and administration from University of Maryland University College (UMUC). The DM program provides a better understanding of management theory, as well as higher education policy and emerging educational issues.

- The first 25 eligible students admitted to the program will receive a three-year, $27,500 scholarship (must apply by January 8)
- Courses are offered online with a two-to-three day residency each semester
- Scholarships, loans and an interest-free monthly payment plan are available

NOTE: Program is not available to Maryland residents.

**UMUC**

**DOCTOR OF MANAGEMENT**

**IN COMMUNITY COLLEGE POLICY AND ADMINISTRATION**

**SOON, THERE WILL BE PLENTY OF ROOM AT THE TOP.**

A recent study by the American Association of Community Colleges shows that as many as 80% of community college leaders are expected to retire within the next five to ten years. Combined with rapid enrollment growth, that means high demand for administrative leaders. Be ready with the Doctor of Management (DM) in community college policy and administration from University of Maryland University College (UMUC). The DM program provides a better understanding of management theory, as well as higher education policy and emerging educational issues.

- The first 25 eligible students admitted to the program will receive a three-year, $27,500 scholarship (must apply by January 8)
- Courses are offered online with a two-to-three day residency each semester
- Scholarships, loans and an interest-free monthly payment plan are available

**SCHOLARSHIPS, LOANS AND AN INTEREST-FREE MONTHLY PAYMENT PLAN ARE AVAILABLE**

**NOTE: Program is not available to Maryland residents**