



Project
MUSE[®]

Today's Research. Tomorrow's Inspiration.

<http://muse.jhu.edu>

The Review of Higher Education

Winter 2012, Volume 35, No. 2, pp. 307–333

Copyright © 2011 Association for the Study of Higher Education

All Rights Reserved (ISSN 0162-5748)

Will These Trees *Ever* Bear Fruit? A Response to the Special Issue on Student Engagement

Alexander C. McCormick and Kay McClenney

Editor's note: *The Review of Higher Education* appreciates the response of Alexander C. McCormick and Kay McClenney to the special issue (fall 2011) on student engagement. For professional fairness, the *Review* has chosen to publish the response without external review.

There is broad consensus that U.S. higher education needs to do better. Researchers, policymakers, and practitioners have called attention to a range of challenges: too many students enter college unprepared for college-level work, yet many developmental programs are little more than revolving doors; too many students who begin college never graduate, often accumulating considerable debt; the most rapid enrollment growth is among the groups that higher education has historically served least well—so institutions have

ALEXANDER C. MCCORMICK is an Associate Professor in Higher Education and Student Affairs at Indiana University Bloomington, where he also directs the National Survey of Student Engagement. KAY MCCLENNEY is Sid W. Richardson Fellow of the Community College Leadership Program at The University of Texas at Austin, where she directs the Center for Community College Student Engagement. We are grateful for the valuable suggestions and feedback offered by Victor M. H. Borden, E. Michael Bohlig, Jeffery F. Crumpley, Allison BrckaLorenz, James S. Cole, Kevin Fosnacht, Amy K. Garver, Heather A. Haeger, George D. Kuh, Amber D. Lambert, C. Nathan Marti, Thomas F. Nelson Laird, Gary R. Pike, Louis M. Rocconi, Shimon A. Sarraf, and Rick Shoup. Address queries to Alexander C. McCormick, Indiana University Center for Postsecondary Research, 1900 E. 10th Street, Suite 419, Bloomington, IN 47406; telephone : (812) 856–5824; email: amcc@indiana.edu.

to do more to ensure their students' success; students' development of generalized critical-thinking and problem-solving skills falls short of what we want and need; we are not producing enough graduates in science, technology, engineering, and math; cost escalation is unsustainable, with most of the growth occurring outside of core educational functions; and the United States is losing ground to other countries with regard to postsecondary degree attainment. And as we confront these challenges, the national understanding of college quality is dominated by beauty contests that privilege reputation and resources over teaching and learning.

The higher education research community has the capacity to contribute to our understanding of and response to these challenges. Indeed, scholars have engaged with many of them. Any could justifiably serve as the organizing theme for a special issue of one of the field's leading scholarly journals. Given the range of important topics where systematic, focused scholarly treatment could advance both research and practice, we find it curious that student engagement trumps these subjects as meriting a special issue of the Association for the Study of Higher Education's signature scholarly journal. We might be flattered that our work is seen as deserving such attention, but we are instead dismayed that the "special issue on student engagement" was in fact devoted to critiques focused exclusively on the two university-based research and service projects that we direct; that it included no contributions from scholars with a record of inquiry on student engagement; and that we had no opportunity to respond to the critique in the special issue itself so as to better advance scholarly discourse and professional practice. While our projects have always welcomed reasoned critique (continuous improvement based on feedback is a hallmark of both projects), we find these precedents worrisome. We are nevertheless grateful for the opportunity to submit this response after the fact.

In the following pages, we situate our response relative to the long-decried disconnect between higher education research and practice, a gap that our respective projects attempt to bridge. We offer brief comments about the Olivas preface, mostly to correct factual errors and omissions, and then provide more detailed responses to the substantive critiques in the articles by Porter; Dowd, Sawatzky, and Korn; Campbell and Cabrera; and Nora, Crisp, and Matthews (all 2011).

THE RESEARCH-PRACTICE DISCONNECT AND THE PURPOSES OF NSSE AND CCSSE

In a provocative 1985 *Change* article titled "Trees without Fruit: The Problem with Research about Higher Education," George Keller asserted: "If the research in medicine, agriculture, or business disappeared the consequences

would be disastrous. If the research in higher education ended, it would scarcely be missed” (p. 7). He further opined: “The research is not aimed at those who must act; it is mainly for the eyes of other researchers” (p. 8).

Five years later, Daniel Layzell asked in *The Chronicle of Higher Education*:

Why *should* policymakers pay any attention to what researchers are saying? . . . I say this regretfully, from the perspective of one who has completed graduate training in higher education and who helps formulate state policy. . . . I find myself having difficulty straddling the widening gulf between higher education research and policy research. (quoted in Terenzini, 1996, p. 6)

A decade after Keller’s article, in his ASHE presidential address, Patrick Terenzini (1996) called on higher education researchers to embrace what Ernest Boyer (1990) called the scholarship of application. In considering the development of higher education research, Terenzini argued that “it is the general tendency away from action, practice- and policy-relevant research that should concern us” and that “we must direct greater research attention to issues confronting practitioners and policymakers” (p. 8).

Two years later, in the editor’s notes to a special issue of this journal, Philip Altbach (1998) wrote, in reference to mainstream higher education research, that a “very large proportion of this research . . . does not contribute directly to the solution of problems faced by those responsible for leading academic institutions” (p. 206). He went on to attribute part of the problem to the lack of intersection between the worlds of researchers and practitioners: “The research community, including those who write for this journal and our readership, links only peripherally with practitioners” (p. 207).

And a full two decades after Keller’s article, in his own ASHE presidential address, John Braxton (2005) called for refocusing higher education research and graduate training on what he called the “scholarship of practice.” Although he framed it differently than his predecessors, his comments represent yet another call from a leader of the higher education research community for scholarship grounded in the practical problems confronting higher education leaders.

This history is vital to understanding the purposes and functions of the National Survey of Student Engagement (NSSE) and the Community College Survey of Student Engagement (CCSSE). Systematic assessments of student engagement emerged in the early years of the 21st century, led by NSSE and CCSSE. The new surveys sought to enrich the impoverished national discourse about college quality by shifting the conversation away from reputation, resources, and the preparation of entering students in favor of the student experience, especially activities and behaviors empirically linked to teaching and learning. At the same time, these surveys offered administrators and faculty members tools for examining and comparing the

prevalence of effective educational practices on their campuses and among different student populations. NSSE and CCSSE differ from other assessment tools in a number of ways:

- They are built on a foundation of research into educational practices associated with desired outcomes of higher education. In this respect, they bridge the gap between higher education research and practice.
- They are strongly focused on student and faculty *behavior*, as contrasted with satisfaction or other attitudes and beliefs.
- Because they are administered to random samples using a standard administration protocol, results are comparable across institutions. Participating institutions not only receive detailed reporting on their own students, but they can also examine their results relative to those for students at comparison group institutions.
- Much of the information provided by NSSE and CCSSE is concrete and actionable. Faculty members, department chairs, deans, provosts, and presidents can examine the results and formulate action plans to increase the prevalence of effective educational practices.

The two projects have been used at more than 2,200 different colleges and universities, and more than 90% of participating institutions administer the survey again within four years. Judged by these surveys' wide adoption and the high rate of repeat participation by institutions, U.S. colleges and universities have shown a strong appetite for the information that the surveys provide. At an early NSSE users' workshop, a dean of arts and sciences proclaimed, "Finally, a test I actually *want* to teach to!" The information provided by NSSE and CCSSE responds to a felt need, and it provides an educationally appealing alternative to "ranksteering" and "reputation management" in the pursuit of educational quality.

While NSSE and CCSSE provide detailed statistical reports and student data files to participating institutions, much of their value lies in their capacity to catalyze conversations on campus among faculty, administrators, and students. In some cases, results confirm existing perceptions of what institutions do well and where improvement is needed, and in other cases they provide an opportunity to interrogate prevailing assumptions. Both projects also devote considerable resources to helping participating institutions make use of their results, converting information to action, through a variety of print resources, workshops, webinars and other online resources, and individual consultations. Thus, both projects engage directly and constantly with "those who must act," as Keller put it. In both the "conversation starter" function and efforts to facilitate constructive use of survey results, the projects fit squarely in the action research tradition. We believe they represent exactly what past ASHE leaders and others have been calling for.

**PREFACE: AMENDMENTS, FACTUAL CORRECTIONS,
AND PUZZLEMENT**

In the preface, Olivas (2011) briefly summarizes the intellectual history behind surveys of student engagement. Conspicuous by its absence is any reference to the work of the late C. Robert Pace, from whose College Student Experiences Questionnaire (CSEQ) many NSSE and CCSSE items are drawn. Indeed, much of the research foundation on which NSSE and CCSSE were built was based on CSEQ data. This omission, as well as the mistaken reference to work by Chickering and Riesser instead of Chickering and Gamson, leads us to question Olivas's understanding of student engagement and its conceptual and empirical antecedents.

Olivas's "analysis of the SSE enterprise" lists 11 projects to indicate "how entrepreneurial and focused the Indiana group has become" (p. 2). The list purports to represent the "small empire" built by George Kuh and housed in the Indiana University Center for Postsecondary Research (CPR) (p. 1). There are three problems with this list. First, it includes the High School Survey of Student Engagement, which is based at Indiana University but has no connection to CPR. Second, it lists two independent projects that have never been based at Indiana University (Classroom Survey of Student Engagement, based at the University of Alabama; Community College Survey of Student Engagement, based at The University of Texas at Austin). Third, it lists three projects that are affiliated with CPR but that are not related to assessing student engagement (National Institute for Learning Outcomes Assessment, based at the University of Illinois; Project on Academic Success; and Strategic National Arts Alumni Project). So of the 11 projects listed by Olivas, only five qualify, including the CSEQ, which Pace created in the 1970s, two decades before the "SSE enterprise," and transferred to Indiana University in 1994. The shorter list based on readily accessible facts is less supportive of the master narrative of "empire."

Olivas lauds the appointment of a Latina scholar, Sylvia Hurtado, to succeed Alexander Astin as director of UCLA's Higher Education Research Institute. We share Olivas's enthusiasm for the appointment, but we are puzzled that he chose not to acknowledge the 2010 appointment of another Latina, Vasti Torres, to succeed Kuh as director of Indiana University's Center for Postsecondary Research. If any project housed within that center qualifies as part of the "SSE empire," then surely the ethnicity of its new director merits the same recognition as part of the salutary diversification of our field's leadership.

We conclude our discussion of the preface with a puzzle. Olivas makes much of findings in Pike, Kuh, and McCormick (2011) that "there was substantial variability across institutions in the magnitude of the relationships between learning community participation and first-year students' level of

engagement” and “a substantial amount of the variability [among institutions] in engagement-learning community relationships remained unexplained.” He bewilderingly claims that this variability represents a “major concession” (p. 11). The statements merely suggest that learning communities are not implemented in identical or equally effective ways across institutions and that some foster engagement more than others. It is unclear why Olivas interprets these statements as a concession.

RESPONSES TO THE FOUR ARTICLES

We now turn to substantive issues. A single article does not afford sufficient space to fully respond, so in the space available, we respond to three important strands of critique: the validity challenge, the asserted neglect of intercultural effort, and challenges to NSSE’s and CCSSE’s multidimensional benchmarks of effective educational practice. We place the greatest emphasis on the first of these, as we judge it to be the most serious of the challenges raised by the critics and also the one with the most far-reaching implications for the higher education research enterprise. In the following discussion, we note both the legitimate and important concerns raised by each strand of critique, and also the ways in which we believe the critique misses the mark.

Validity

As indicated by the title of his article (“Do College Student Surveys Have Any Validity?”), Porter (2011) raises a fundamental challenge to all surveys of college students and, by extension, to the considerable portion of the higher education scholarly literature that is based on survey data, as he notes (p. 70). He questions whether students can reliably respond to the questions asked by survey researchers. In pursuing this important question, he chooses a single survey—NSSE—to stand for all surveys of college students. It is curious that a scholar of Porter’s methodological sophistication would elect to generalize from a sample of one, especially when the implications of his analysis could be so far-reaching. It turns out, however, that the evidentiary base for many of his claims about NSSE is even smaller.

Points Worthy of Investigation. Porter’s critique raises some legitimate questions about how certain survey items might be reframed to facilitate recall by respondents. Some NSSE questions call upon respondents to summarize their experiences and behaviors across an extended period (typically the current academic year). We might investigate narrowing the time frame by asking about the frequency with which students have engaged in a given behavior in the most recent day of classes or over the past week. This framing is not without problems. A typical NSSE administration extends over several weeks, as multiple reminder messages are sent to nonrespondents. Thus, as the time referent for a given question narrows, students’ reported behaviors are likely

to be affected by the rhythms of the semester. For example, around midterm exams, students might report more out-of-class interaction with faculty but less writing of papers or participation in class discussions.

Porter also advocates time-use diaries to gather information about the student experience. The near-universal use of mobile devices by college students holds great potential as a platform for diary and time-sampling methods. It would be especially valuable to use such techniques in parallel with conventional survey techniques to assess differences in how students report on their learning-related behaviors and experiences. Among the possible limitations is that diaries may be more feasible for traditional-age, full-time college students than for nontraditional and commuter students who are balancing commitments to school, work, and family and who may not be willing to devote the time required for regular diary entries. (We discuss other limitations of diary methods below.)

Despite worthy considerations, we find much of Porter's analysis lacking. Given space constraints, we focus on three problematic aspects: (a) he assumes that NSSE seeks to produce precise point estimates of various quantities (number of papers written, number of books read, hours per week spent studying, college grades, etc.), and more generally he privileges criterion validity over other important validity considerations; (b) much of his argument is based on proposition and conjecture rather than evidence, sometimes overlooking contrary evidence; and (c) he offers little in the way of constructive suggestions to improve college student surveys.

1. Narrow Focus on Criterion Validity and Accuracy of Point Estimates. In a classic articulation of the complex concept of validity, the late Samuel Messick of the Educational Testing Service (1989) defined it as "an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of inferences and *actions* based on test scores and other modes of assessment" (p. 13; emphasis Messick's). He further states that "the key issues of test validity are the interpretability, relevance, and utility of scores, the import or value implications of scores as a basis for action, and the functional worth of scores in terms of social consequences of their use" (p. 13). In Messick's view, then, the validity of a measure is inextricably tied to the uses to which it is put. Kane's subsequent treatment (2006) reiterates the importance of intended use: "The evidence needed for validation necessarily depends on the claims being made. Therefore, validation requires a clear statement of the proposed interpretations and uses" (p. 23). This focus on use is central to our own view of the validity of NSSE and CCSSE, and it represents a notably different perspective than that which underlies the Porter critique.

In his treatment of validity, Porter elides the distinction between tests and surveys. The treatment of tests and surveys as comparable explains, in part, his

heavy emphasis on criterion validity and accuracy of point estimates. Let us consider more carefully the validity implications of tests versus surveys. For tests, the accuracy of scores is paramount. A test's ability to determine and discriminate mastery of the content domain, in a manner that is consistent across subgroups, is vital to its proper function and use. Test results can determine college and graduate school admission, course placement, professional licensure, and other high-stakes outcomes; and measurement error in these settings can have profound and enduring consequences. Because Porter sees a survey as analogous to a test, that perspective explains why he focuses so much attention on questions of accuracy—whether the number of papers written matches the actual number indicated on course syllabi, whether self-reported grades match institutional records, and whether a student's perceived learning gains correspond to objective pre/post test performance.

But in focusing so much attention on accuracy questions, Porter neglects the primary way that NSSE and CCSSE results are actually used: to make *relative* comparisons between groups of students. What matters is not the precise number of papers written but the fact that certain groups of students write more than others: students at a given college or university versus those at peer institutions; athletes versus non-athletes; students in some majors versus those in others; and so on. The same approach applies to the many frequency-of-behavior items on the NSSE and CCSSE surveys. What matters to institutional users is whether, for example, minority or first-generation students report lower levels of student-faculty interaction, or that they less frequently receive prompt feedback from faculty, *relative to* other students. This difference is what gets attention when institutional results are examined, and we believe that it is what *should* get attention.

Porter rightly points out that survey respondents may have difficulty recalling the frequency of a given behavior. But much of the literature on which he bases his argument about the limitations of recall involves enumerated responses: specifying a precise number or choosing from a set of numerical ranges. NSSE, by contrast, relies to a great extent on vague quantifiers (for example, "very often," "often," "sometimes"). The use of vague quantifiers does not invoke the same process of recall and tally that Porter rightly identifies as error-prone (Wanke, 2002; Wright, Gaskell, & O'Muircheartaigh, 1994). Indeed, noted survey researchers ask: "Since behavioral frequency reports are error-prone anyway, why bother asking respondents for reports that suggest more precision than they can provide?" (Sudman, Bradburn, & Schwarz, 1996, p. 226).

To be sure, these authors acknowledge that vague quantifiers have their own limitations and problems, and we concede that there is variability in how individual students interpret and apply these scales. But analyses of data from NSSE and the Beginning College Survey of Student Engagement (BCSSE) suggest that discrepancies in individuals' and groups' uses of these

response options do not meaningfully limit how the data are typically used (Cole & Korkmaz, 2011; Nelson Laird, Korkmaz, & Chen, 2009). Indeed, a cursory examination of the data from Pace and Friedlander's study presented by Porter shows that the pattern of agreement seems to approximate a normal distribution. But the important point is that research into how respondents use vague quantifiers has convincingly shown that respondents use various processes of comparison, rather than recall and tally, to situate their response (for example, Wanke, 2002; Wright, Gaskell, & O'Muircheartaigh, 1994).

2. *Conjecture, Claims without Evidence, and Overlooked Contrary Evidence.* Much of Porter's argument relies on propositions of the following form: Given what study X tells us about behavior Y, then we can't trust what college students tell us about behavior Z. We cannot disprove such claims any more than Porter has proven them.

In one version of this form of argumentation, Porter freely generalizes specific research findings to broader applications without acknowledging that such generalizations are conjectural. For example, he cites findings by Cole and Gonyea (2010) and Kuncel, Crede, and Thomas (2005) that students of higher ability report academic measures—test scores, grades, and class rank—more accurately than do lower ability students. He then generalizes this finding to the accuracy of all self-report, without any supporting evidence: "Such a finding is disconcerting, because the average level of student cognitive ability varies across colleges, implying that college comparisons using survey data may be flawed due to the correlation between cognitive ability and the accuracy of self-reports" (p. 62).

In what is surely one of the raciest passages ever to appear in the pages of this journal, Porter refers to research on college students' sexual behavior, even drawing an analogy between the salience of faculty contact and that of a particular sex act. In this analysis, he treats students' daily email diaries of sexual behavior as free of validity problems, despite the cited study's own discussion of the limitations of diary studies, including "completion bias" and "rehearsal bias" (Garry, Sharman, Feldman, Marlatt, & Loftus, 2002). In this part of his discussion, Porter seems to ignore his previous caution that "any measure used to validate another measure must itself be valid" (p. 54). Indeed, scrutiny of this analysis illustrates how very challenging criterion validation can be. For example, Porter goes so far as to assert, without evidence, that daily diaries about sexual activity should not be affected by social desirability bias, despite well-known gender differences in the reporting of sexual behavior (Brown & Sinclair, 1999) and the Garry et al. study's own acknowledgment of possible validity problems due to social desirability bias. Porter seems to cherry-pick findings that support his argument, while ignoring contrary evidence.

Speaking of contrary evidence, both NSSE and CCSSE have systematically investigated students' understanding of the survey questions and response

options. Nearly 400 students at 16 institutions have participated in two cycles of focus groups and cognitive interviews, in 2000 and 2005, to provide insight into students' responses to the NSSE survey. (The latter cycle investigated how the surveys function at both minority-serving and predominantly White institutions.) CCSSE has undertaken similar investigations with hundreds of students. In testing the NSSE questions Ouimet, Bunnage, Carini, Kuh, and Kennedy (2004) concluded: "Generally, students found the questions to be clearly worded and easy to understand. The number of items that prompted discussion [in focus groups] was relatively small, less than 10% in most focus groups" (p. 240) and the "majority of students interpreted the questions in identical or nearly identical ways" (p. 247).

Items that were revealed to be problematic in focus groups were reworded or eliminated, and the reworded items were subsequently tested through cognitive interviews. In the later study, Kuh, Kinzie, Cruce, Shoup, and Gonyea (2006) concluded from similar testing that the NSSE instrument works equally well for students of color and White students in different institutional contexts: "The cognitive interviews and focus group results suggest that the vast majority of students at different types of institutions understand what is being asked, find the directions to be clear, interpret the questions in the same way, and tend to formulate answers to questions in a similar manner" (p. 52).

Porter cites work by Pike (1995) in his arguments about the inaccuracy of self-reported data. But he neglects to mention one of Pike's important conclusions that speaks directly to our point above about relative comparisons. Pike concluded that, for the purpose of comparing groups or examining relationships to other measures, the use of self-reported data leads to nearly identical results as would be reached using more accurate institutional data. If we accept Porter's claims, students likely misunderstand NSSE's questions, cannot reliably recall the behavior asked about if they understand the question, and, even if they understand the question and correctly recall the behavior, cannot reliably interpret their recollection and convert it to the response scales. This is the logic behind his assertion that "college students only rarely report accurate information about their behaviors" (p. 49). In short, his analysis suggests that students' responses have little relationship to reality—thus, they are essentially random. This position confronts some clear evidentiary challenges. As Porter himself reports, researchers have found positive associations between NSSE engagement measures and various outcomes, including learning measures assessed in a pre/post assessment framework.¹ To this list we would add work by Pascarella, Seifert,

¹Porter dismisses these findings because the results fail to meet his subjective standard for the strength of relationship required to satisfy a claim that engagement is "highly correlated" to learning outcomes.

and Blaich (2010). Using data from 19 institutions in the Wabash National Study of Liberal Arts Education, they examined the relationship between institutional average scores on the five NSSE benchmarks with learning gains on seven measures of liberal arts outcomes. Four of five benchmarks were found to be significantly related to learning gains on at least one of the outcome measures. In addition, large-scale validation research conducted on the CCSSE instrument, encompassing three independent studies undertaken by researchers external to CCSSE, demonstrate that the broad measures of student engagement provided through the CCSSE survey are predictive of outcomes measuring academic persistence and success in community colleges. The research found positive associations between student engagement (CCSSE's five benchmarks of effective educational practice) and both the number of terms enrolled and the credit hours completed (McClenney & Marti, 2006; McClenney, Marti, & Adkins, 2007).

More to the point, though, are findings related to trends in institution-level NSSE scores. In an examination of benchmark scores at more than 200 institutions that had participated in NSSE at least four times between 2004 and 2009, 87 showed at least one positive trend for first-year students and 63 did so for seniors, while only a handful showed negative trends (NSSE, 2009). In a recent extension of this work analyzing a broader set of measures for 534 institutions that administered NSSE at least four times between 2001 and 2009, the number of institutions with detectable positive trends outnumbered negative ones by seven to one (McCormick, Kinzie, & Korkmaz, 2011). Absent a systematic (and unusually effective) campaign to influence how students fill out the survey year after year, it is difficult to square these findings with Porter's skepticism about students' ability to answer NSSE questions.²

3. Few Constructive Prescriptions to Advance the Field. Porter devotes most of his article to picking apart what he judges to be NSSE's many weaknesses. He has every right to do so, but we believe the field would be better served if he had applied more of his talents to proposing how to advance the technology of assessing the student experience. His primary recommendation appears to be time-use diaries. Although they have had limited application in the study of college students, diary- and survey-based accounts of how many hours per week students spend studying have shown a relatively high degree of consistency (McCormick, 2011). This finding could mean either that the survey data are more valid than Porter believes or that diaries are less valid than he believes. Diary-method researchers (e.g., Bolger, Davis, & Rafaeli, 2003) have noted many challenges and limitations, including memory biases,

²For the record, we are unaware of efforts by institutions to artificially influence students' survey responses. We *are* aware of widespread efforts to increase students' exposure to effective educational practices.

social desirability, personality and individual differences, reactivity effects of diary completion on behavior, and the in-depth participant training and commitment needed to fully understand diary procedures and protocols. In any case, much work needs to be done to render such approaches cost-effective and feasible for large-scale data collection. We welcome efforts by Porter and other scholars to develop new tools in the service of evidence-based improvement.

The Perfect and the Good. Implicit in much of the Porter critique is that NSSE falls short of the ideal survey: one in which all questions and response options are unambiguous, all respondents share identical understandings of all questions, and all are equally able to retrieve the information and convert it to the available response options with perfect consistency. Thus, the ideal survey requires ideal respondents as well. But student surveys operate in the real world, dependent on the good will of volunteer respondents who have other commitments. In the words of the current director of the Census Bureau, “Surveys are inherent compromises” (Groves, 1987, p. S167). In other words, the construction of surveys involves trade-offs. For example, we trade off comprehensive coverage against reasonable survey length. By design, NSSE and CCSSE ask a few questions about a wide variety of behaviors and experiences. Almost any question on either survey could be expanded into a battery of questions to deepen and fine-tune our understanding of a given phenomenon.³ But such a survey would be so long that few students would complete it without significant incentives to participate—which would merely substitute a new set of validity challenges related to data quality, as questions would be raised about students’ motivations to thoughtfully respond to our questions.

Alternatively, one could imagine a proliferation of surveys targeted to more narrowly focused topics: a national survey of active learning, a national survey of faculty expectations, a national survey of higher-order learning, and so on. Going deep rather than broad would no doubt be of great interest to researchers, but it would sharply limit the surveys’ utility for institutional users. They would have to choose which phenomena to investigate in a given year, and the splintering of topics would likely mean a much-reduced pool of available comparison institutions.

Similarly, at the item level we trade off lengthy, precise descriptions and definitions against phrasing that respondents will actually read. In its early years, NSSE experimented with more elaborated instructions (as Porter advocates on p. 59), but cognitive interviews revealed that the change had the

³Groves (1987) makes a similar point: “Any researcher who has constructed a questionnaire knows that each single question could usefully be expanded to produce an entire survey of its own” (p. S167).

opposite effect of what was intended. Many respondents skipped the longer instructions altogether (Ouimet et al., 2004). This circumstance is an apt illustration of how the real world can be an inhospitable place for idealized visions of surveys and survey research. In designing NSSE and CCSSE we chose also to focus primarily on institutional practices and student behaviors, as contrasted with attitudes, beliefs, and values, to provide both concreteness and utility of results.

We believe that imperfect information has greater decision utility than no information, and we readily acknowledge that survey research is imperfect. Paraphrasing Groves (1987), to design a survey is to make compromises. We acknowledge our compromises, but we choose them over the alternative of waiting for the development of a substantially error-free methodology. Our projects nonetheless persistently strive to make improvements, and some of Porter's critique can be helpful in this regard.

Intercultural Effort

Dowd, Sawatzky, and Korn (2011) argue that student engagement surveys, as well as other higher education assessments, fail to consider and measure what Tanaka (2002) calls intercultural effort—the additional time, energy, and psychic effort that students who experience minority status must expend in response to racism, cultural insensitivity, and other harmful practices related to their minority status. Following Tanaka, they assert that, by excluding intercultural effort, the surveys treat colleges and universities as culturally neutral spaces. They further argue that the exclusion of intercultural effort constitutes construct underrepresentation with regard to “student effort” and also that it is as important to measure bad educational practices as it is to measure good practices. The authors call for the measurement of “student experiences of racial bias on college campuses and institutional efforts in reducing institutionalized racism” (p. 18) and assessments “to assist institutions in measuring their effectiveness at being more culturally inclusive” (p. 20). NSSE and CCSSE do not consider campuses to be culturally neutral spaces. In addition, we question the construct underrepresentation argument. At the same time, we agree that the proposed expansion of assessment tools would enhance both research and practice. Whether engagement surveys are the appropriate locus for that expansion, as opposed to climate assessments, for example, is another question.

Dowd, Sawatzky, and Korn assert that student effort is “foundational” to student engagement (p. 22). While it is true that Pace's work (1980) on the CSEQ emphasized quality of student effort, and that NSSE and CCSSE trace much of their lineage to Pace's work and the many studies based on CSEQ data as well as decades of research by Astin (1984) and other scholars, student engagement involves more than student effort. As we explain above, NSSE and CCSSE were created to provide diagnostic and actionable information to

faculty and administrators who are interested in educational improvement. Thus, the surveys include questions about faculty behaviors, cognitive tasks emphasized in courses, quality of campus relationships, perceptions of institutional efforts to support students and promote their success, and so on.

It is true that one of CCSSE's five benchmarks of effective educational practice is called "student effort." Dowd, Sawatzky, and Korn imply that student effort is also embedded in NSSE's "level of academic challenge" and "active and collaborative learning" benchmarks. But just as student engagement involves more than student effort, it is broader than the benchmarks. Although these definitional issues are important, we agree to stipulate that student effort is part of student engagement so that we can focus on the central question of intercultural effort.

Dowd, Sawatzky, and Korn ground their conceptual argument in an assertion that the student effort construct is rooted in human capital theory. They apparently base this interpretation on the use of the verbs "invest" and "capitalize" in a passage that appears in the manual for the Community College Student Experiences Questionnaire (CCSEQ). Although Pace emphasized quality of student effort in his work with the CSEQ and CCSEQ, he never framed it in human capital terms, verb choice notwithstanding. Nor are we aware of any literature on student engagement that analyzes engagement behavior (asking questions in class, discussing course ideas with a classmate, collaborating with other students on group projects) in economic terms of maximizing utility and expected future value. The human capital framing leads to a somewhat tortured economic analysis of student effort and how intercultural effort can be incorporated into conventional models of the decision to pursue and continue in higher education. But for our purposes, the conceptual justification is less important than the articulation of the problem and its solution.

Useful Suggestions. We endorse the general notion that efforts can and should be made to better assess the inclusiveness of institutional environments and the experiences of students from underrepresented and disenfranchised groups. Both NSSE and CCSSE persistently exhort colleges to examine potential differences in the reported experiences of subgroups of their students, by race/ethnicity, gender, age, first-generation status, and other characteristics central to concerns of educational equity. We also support further conceptual and empirical development of the ideas advanced by Dowd, Sawatzky, and Korn and how to operationalize them in the context of a survey targeted to all students and not only to a subpopulation. That said, a distinguishing characteristic of both NSSE and CCSSE is their emphasis on activities and behaviors that prior research has shown to be related to student learning and development. Our general criteria for item development are that: (a) items need to be about student engagement; (b) they need to have an empirical base linking them to desired outcomes; (c) they need to assess

institutional practices or student behaviors, using response frames that are reliable for student self-reports; and (d) the resulting information needs to be actionable (i.e., something institutional leaders can do something about).

Before developing content to be included in student engagement surveys, a promising direction for research would be to administer one or more of the existing surveys and inventories identified by the authors at campuses that also use NSSE or CCSSE, and then jointly analyze the results. This would allow for a direct test of some of the authors' assumptions about blind spots in existing engagement constructs, while also suggesting future avenues for development. Another avenue for potential collaborative research already exists in the options that both NSSE and CCSSE provide for institutions to append additional thematic questions to the core surveys.

Challenges, Difficulties, and Disappointments. The authors' legitimate call for attention to intercultural effort is undermined in places by bold claims propped up by faulty logic. For example: "The implicit assumption, in the absence of such assessments [of racially minoritizing practices], is that institutional racism or racial bias does not exist on college campuses" (p. 20). In other words, the authors apparently take the position that survey content exhaustively defines the designers' conception of higher education. By this logic, the fact that a survey does not ask about theft or violence signifies an assumption that colleges are crime free. Another example: "By omitting measures of what campuses are doing to alienate students, existing measures have the potential to do more harm than good" (p. 20). A dramatic assertion, to be sure, but it does not follow that the failure to detect one problem renders the identification of other problems harmful.

Dowd, Sawatzsky, and Korn argue that the "theoretical construct of 'student effort' must represent 'student effort' in its entirety" (p. 23) and call for an instrument that measures "all aspects of 'student effort'" (p. 22). This appeal begs the question: Is the addition of intercultural effort sufficient? Or are there other forms of effort that would still be left out, such as efforts to connect with peers, efforts to balance multiple roles, and efforts at ethical, spiritual, and moral development, to name a few that have received the attention of scholars? Whose job will it be to articulate all aspects of student effort? The injunction to assess all aspects of student effort seems to expose us to perpetual claims of construct underrepresentation as defined by the authors. Related to this point, Dowd, Sawatzsky, and Korn imply that intercultural effort involves only racial/ethnic minorities. Should the argument apply to other dimensions of cultural difference and discrimination (social class, religion, national origin, disability, etc.)? If intercultural effort is solely about race/ethnicity, a more complete theoretical articulation specifying and justifying the boundary conditions is needed.

In their review of existing surveys, Dowd, Sawatzsky, and Korn conclude that "none specifically measures intercultural effort" (p. 32). Their review

included entire surveys focused on institutions' diversity climates, such as the Diverse Learning Environments survey of UCLA's Higher Education Research Institute. At the risk of seeming thin-skinned, we ask: Why, then, single out engagement surveys as the target of critique? In many ways, the article reads as a perfectly reasonable and appropriate articulation of "theoretical foundations and a research agenda to validate measures of intercultural effort" (the article's title), reworked somewhat awkwardly as a critique of existing surveys of student engagement.

That said, we were gratified to see the authors endorse two questions from our surveys in their discussion of existing surveys that contain promising items. We only wish they realized it. In the discussion of "content that is relevant to intercultural effort that is missing from the current constructs of student effort and engagement" (p. 32), the authors refer to research by Nelson Laird and Niskodé-Dossett (2010) that was conducted *using NSSE data* (p. 33). (The authors incorrectly identified two scales used in the study as surveys.)

The two questions quoted verbatim from that study are, in fact, questions that appear on *both* NSSE and CCSSE: "To what extent does your institution emphasize . . . helping you cope with your non-academic responsibilities (work, family, etc.)" and "Mark the box that best describes the quality of relationships with people [other students]⁴ at your institution," with responses ranging from "unfriendly, unsupportive, sense of alienation" to "friendly, supportive, sense of belonging." Because this surprising oversight leads us to question the authors' familiarity with the surveys they are critiquing, we call attention to these other relevant questions from our surveys:

- ❑ In your experience at your institution during the current school year, about how often have you . . .
 - Included diverse perspectives (different races, religions, genders, political beliefs, etc.) in class discussions or writing assignments? [NSSE]
 - Had serious conversations with students of a different race or ethnicity than your own? [NSSE and CCSSE]
- ❑ During the current school year, about how often have you done each of the following?
 - Tried to better understand someone else's views by imagining how an issue looks from his or her perspective [NSSE]
- ❑ To what extent does your institution emphasize each of the following?

⁴This is one of three questions about quality of relationships; the others focus on relationships with faculty and administrative staff, with anchors that correspond to relationships with those groups.

- Encouraging contact among students from different economic, social, and racial or ethnic backgrounds [NSSE and CCSSE]
- To what extent has your experience at this institution contributed to your knowledge, skills, and personal development in the following areas?
- Understanding people of other racial and ethnic backgrounds [NSSE and CCSSE].

These questions seem comparable to many of the others identified as promising, and we question how and why they could have been overlooked.

It is important to consider how what we have already learned about student engagement can inform this discussion. Dowd, Sawatzky, and Korn quote Tanaka's supposition that the quality of effort construct may lead researchers to overestimate the importance of effort when engagement might actually be harmful to self-worth. Tanaka's contention does not square with the evidence we have from NSSE and CCSSE. In fact, we have found the opposite. In a Lumina-sponsored project investigating student engagement and outcomes at 19 institutions, the research team found that student engagement showed stronger positive effects for underprepared students and students of color with respect to first-year GPA and retention from the first to the second year (Kuh, Cruce, Shoup, Kinzie, & Gonyea, 2008; NSSE, 2006). And a persistent finding from CCSSE has been that at-risk student populations show higher levels of student engagement. Because CCSSE is administered in the spring, these findings have been interpreted as reflecting a selection effect: the spring population represents the survivors (see, for example, Community College Survey of Student Engagement, 2005; Greene, Marti, & McClenney, 2008). This pattern is consistent with NSSE's findings of differentially positive benefits of engagement for underrepresented and underserved populations. Dowd, Sawatzky, and Korn also worry that "harmful practices can exist alongside best practices" (p. 19). While such an effect is theoretically possible, evidence from the Faculty Survey of Student Engagement (FSSE) suggests that inclusiveness is positively correlated with other effective educational practices as assessed by NSSE and FSSE (Nelson Laird & Engberg, in press).

Finally, the authors' argument raises a number of practical issues for surveys like NSSE and CCSSE. Would questions about intercultural effort be asked only of what the authors call minoritized students or of all students? If the former, how would they be identified? Would the criteria vary from campus to campus? From individual to individual? Would these questions be asked at minority-serving institutions? Of whom? These are but a few of the many practical and operational questions—some with important conceptual dimensions, as well—that would need to be answered in order to implement content related to intercultural effort in the context of a survey that seeks a representative sample of all students.

Go Forth. The authors have outlined an important program of research and development related to intercultural effort. We encourage them to develop and test survey content that taps the construct and to gather evidence of its relationship to both student engagement and educational outcomes, including academic performance, persistence, graduation, and learning. Once they have done so, scholars and practitioners can begin a serious conversation about whether and how to use the findings to incorporate these perspectives in a variety of assessment instruments, especially those that aim to assess campus climate.

NSSE and CCSSE Benchmarks of Effective Educational Practice

Next, we consider the articles examining the NSSE (Campbell & Cabrera, 2011) and CCSSE (Nora, Crisp, & Matthews, 2011) benchmarks of effective educational practice in specific institutional settings. We treat both articles together because they highlight important questions of what the benchmarks represent, what their limitations are, and the purposes they are intended to serve for our institutional users.

Both articles assess the viability of the benchmarks as latent constructs, but this assumption is a misinterpretation. The benchmarks do not represent latent constructs. They are summative indices of a range of effective educational practices. As acknowledged by Campbell and Cabrera, “The five NSSE benchmarks were created out of the NSSE survey items using a combination of theory (specific engaging practices that seem to have the most impact on student outcomes) and exploratory factor analysis” (p. 78).

A similar hybrid process produced the CCSSE benchmarks. Marti (2009) reports that a pure confirmatory factor analysis (CFA) approach yielded a nine-factor solution, which was then reduced “to a practically useful number of constructs that could be used as performance measures of institutional effectiveness” (p. 5). He distinguishes the two techniques this way:

Constructing a latent variable model with the best fit to the data and creating latent constructs useful for evaluating the engagement of a student body are clearly complementary efforts. Nevertheless, the two goals diverge, as optimal model fit requires a granular model of latent constructs whereas establishing benchmark measures is a molar endeavor that seeks to broadly classify items with less concern for the precision of model fit. (p. 5)

Marti also describes in some detail how the benchmarks represent a blend of empirical analysis and expert judgment, very similar to the process whereby NSSE’s benchmarks were created:

To establish the factor structure for the benchmarks of effective educational practice, a group of survey research experts (CCSSE’s Technical Advisory

Panel) reviewed the CFA results. The panel then assigned items to benchmarks, taking into account the results of factor analysis, reliability tests, and also applying expert judgment based on both the conceptual framework and empirical evidence related to student engagement in undergraduate learning. The objective was to create benchmarks that are reliable, useful, and intuitively compelling to community college educators. (pp. 9–10)

Taking the most extreme example of a benchmark that could not conceivably be construed to represent a latent construct, consider NSSE's "Enriching Educational Experiences" benchmark. It comprises a collection of educationally beneficial activities that do not represent a unitary underlying construct. Consider its components (wording modified slightly to convey question stems):

- Hours per week spent participating in co-curricular activities
- Participated in a learning community or some other formal program where groups of students take two or more classes together
- Participated in a practicum, internship, field experience, co-op experience, or clinical assignment
- Participated in community service or volunteer work
- Participated in foreign language coursework
- Participated in study abroad
- Completed an independent study or self-designed major
- Completed a culminating senior experience (comprehensive exam, capstone course, thesis, project, etc.)
- Frequency of serious conversations with students who are very different from you in terms of their religious beliefs, political opinions, or personal values
- Frequency of serious conversations with students of a different race or ethnicity
- Frequency of using an electronic medium (listserv, chat group, internet, instant messaging, etc.) to discuss or complete an assignment
- Degree of institutional emphasis on encouraging contact among students from different economic, social, and racial or ethnic backgrounds

Writing in reference to NSSE's online psychometric portfolio, Campbell and Cabrera note that "there is no portfolio brief on the construct validity of the five NSSE benchmarks of effective educational practices" (p. 85). The reason is that we neither assume nor claim that the benchmarks represent latent constructs. Consider the list above. If it represented a latent construct, we would expect reasonable intercorrelations among the manifest measures. But one would be hard pressed to come up with a rationale for expecting many of the items to be correlated (for example, study abroad and use of electronic media; foreign language coursework and a self-designed major; an internship and institutional encouragement of contact across difference).

A less extreme but still illustrative example is CCSSE's Student Effort benchmark. With wording modified slightly to convey question stems, its components are:

- Frequency of preparing two or more drafts of a paper or assignment before turning it in
- Frequency of working on a paper or project that required integrating ideas or information from various sources
- Frequency of coming to class without completing readings or assignments
- Number of books read on your own (not assigned) for personal enjoyment or academic enrichment
- Hours per week spent preparing for class (studying, reading, writing, rehearsing, doing homework, or other activities related to your program)
- Frequency of use: Peer or other tutoring
- Frequency of use: Skill labs (writing, math, etc.)
- Frequency of use: Computer lab

To understand why the benchmarks were constructed in this way, it is important to consider both their purpose and their intended audience. The benchmarks distill survey results—frequency distributions for a large number of survey questions that fill 15–20 pages—into a manageable, easily digested overview of results that affords a sort of dashboard display of several important facets of student engagement. They were created as a point of entry into an institution's results, one that might initiate campus conversations about the character of undergraduate education, how it compares to the educational efforts of other colleges and universities, what an institution does well, and where improvement is needed.

The audience includes presidents, deans, department chairs, faculty members, institutional researchers, student affairs administrators, and others. Benchmarks must enable institutional leaders and others to quickly assess both strengths and areas deserving of attention. For this audience, clarity, parsimony, and face validity are important considerations. Readers of this journal know that exploratory factor analysis can produce unusual combinations, including items that seem conceptually suited to one factor loading more highly on another. This may not be a problem when one's audience consists of fellow social scientists. But institutional leaders may come from a wide range of disciplinary backgrounds, and they need to grasp the conceptual underpinnings of what the benchmarks signify without the distraction of counterintuitive combinations. ("Why is 'Worked on a research project with a faculty member' under Active and Collaborative Learning and not Student-Faculty Interaction? That doesn't make any sense! I can't take this before the faculty!") Thus, for benchmarks to serve their communicative purpose, it is at least as important for them to hang together conceptually as empirically.

It is also important to note that the items that make up the benchmarks account for only about half of the engagement-related questions on the NSSE and CCSSE surveys. If we believed that the benchmarks represented all there is to know about student engagement, our surveys would be a lot shorter. Again, the benchmarks offer a way into the survey results, not a comprehensive accounting.

Given this description of what the benchmarks are and how they are constructed, it is hardly surprising that attempts to verify them as latent constructs yield unsatisfactory results. We should not expect a purely empirical procedure to match the benchmark construction process described above. But these analyses provide an important opportunity for us to provide this clarification of the purpose, construction, and use of the benchmarks of effective educational practice. We hasten to note that both NSSE and CCSSE encourage users to undertake their own local analyses to gain insight into the nature of student engagement on their campuses.

Having established our concern about treating the benchmarks as latent constructs, we now consider the methodology and findings of the two articles. The Campbell and Cabrera article is clear and generally well executed. In discussing the positive relationships between NSSE benchmarks and various liberal arts outcomes found by Pascarella, Seifert, and Blaich (2008), however, the authors incorrectly state that the upper-bound estimates in that study did not control for precollege characteristics or other NSSE benchmarks (p. 82), when in fact the other benchmarks were controlled for. Thus the upper-bound estimates should not have been so readily dismissed. With regard to the analysis sample, the decision to restrict the analysis to nontransfer seniors "because they had had four years to experience and be engaged with the institution" (p. 86) was unwarranted, because the vast majority of items in the NSSE benchmarks refer to the current academic year.⁵ Nearly half (46%) of NSSE 2009 seniors reported that they had begun college at another institution, so this exclusion substantially alters the analysis sample relative to NSSE's. A related questionable choice was to model cumulative GPA in the predictive validity analysis, given that most of the engagement experiences assessed by NSSE refer to the current academic year. Thus, the outcome variable includes at least three academic years' worth of academic performance that predates the observed engagement behavior. Cumulative GPA also artificially restricts the range of the outcome variable, because the analysis is limited only to students who made it to their senior year. A cleaner analysis, then, would have focused on senior-year GPA. But this recommendation is arguably beside the point, because there is little benefit

⁵Several items in the Enriching Educational Experiences benchmark do not make this temporal restriction, but this does not appear to be the authors' rationale.

in proceeding with a predictive validity analysis for a factor solution that is already shown to be a poor fit to the data, using explanatory variables with relatively high intercorrelations. We wish the researchers had shown more interest in understanding the relationship between student engagement and outcomes using all of the information available to them, rather than simply interrogating the integrity of the benchmarks.

The Nora, Crisp, and Matthews article is somewhat more difficult to respond to. By restricting the analysis to “college-level academic courses” (p. 114), they eliminate an appreciable slice of students who are included in the CCSSE population, namely, those enrolled in developmental/precollegiate education. Also excluded were students enrolled in pass/fail courses and students without an established GPA. Having made those exclusions, we immediately confront a comparability problem. These limitations also aptly illustrate the research-practice disconnect. Community college leaders must be apprised of the educational experiences of all of their students, and the CCSSE project defines its population to serve this aim.

Nora, Crisp, and Matthews conclude that CCSSE benchmarks have questionable validity, but they provide little detail about their factor analytic procedures. Their conclusion seems to rest on the fact that they found different results than Angell (2009) and Marti (2009). However, the three studies used different procedures, virtually guaranteeing different solutions. Nora, Crisp, and Matthews “excluded significant factors that were not representative of the five CCSSE benchmarks from further analysis” (p. 115), using an unspecified procedure, while Angell used a very low criterion for a given variable to load on a factor, which has the effect of reducing the total number of factors. Also as the authors note, Angell’s analysis included all questions from CCSSE rather than just those included in the benchmarks. Indeed, it may well be that before Nora, Crisp, and Matthews excluded some factors, their results may well have resembled Marti’s best-fitting nine-factor solution. We find that, given the researchers’ sample definition, analytic approach, and single-institution study of 393 students, the results show a higher than expected degree of correspondence to the CCSSE benchmarks.

The authors conclude by proposing (as indicated by their title) “a reconceptualization of CCSSE’s benchmarks of student engagement,” in a discussion that is oddly divorced from their empirical results. We encourage the authors to subject their proposed model to empirical examination. As with the Campbell and Cabrera article, the singular focus on testing the CCSSE benchmarks obscures potentially more interesting findings.

We remind readers that our projects’ primary purpose is to provide campus decision-makers with information that can inform educational improvement. The benchmarks simply serve as a way into the data—as a compilation of information to ignite conversations and questions about

the quality of undergraduate education. Our experience with hundreds of institutions suggests that they do just that. Regardless of whether the benchmarks hold up as latent constructs on one campus or many, we believe the more important finding is that an appreciable number of institutions show patterns of improved performance as measured by benchmarks and other indicators (McCormick, Kinzie, & Korkmaz, 2011). We believe that such improved performance is the more relevant test of our respective projects' value and impact.

CONCLUDING THOUGHTS: GROWING TREES THAT BEAR FRUIT

We conclude by reaffirming our overarching premise: Purposes matter. Responding to more than two decades' worth of calls from Keller (1985) and other scholars and policymakers, including ASHE leaders spanning decades, NSSE and CCSSE were created to help bridge the gap between research and practice in higher education and provide diagnostic, actionable data to colleges and universities. Their fundamental purpose is to promote improvement in student learning and attainment by bringing practitioners' attention to educational practices that are empirically associated with good outcomes. An important companion objective has been to help change the discourse, both in and out of the academy, about what constitutes "quality" in undergraduate education. Consistent with these purposes, both NSSE and CCSSE, as research and service projects, see our primary constituents as higher education practitioners. Our work with educators across 50 states and internationally has focused on helping them to understand data about students, including the limitations of those data, and to use survey results appropriately to inform educational improvement. These efforts have contributed to a sea change in the ways many institutions comprehend and improve their work with students.

Another fundamental value is at stake. Both NSSE and CCSSE are built on the belief that there is value in asking students about their experiences. Indeed, students are the best informants about the student experience. In a classic remark on this issue, the late Bob Pace referred to a famous Packard car commercial in the 1950s: "Ask the man who owns one." As discussed above, evidence from focus groups and cognitive interviews suggests that students understand the questions we ask and are capable of answering them.

Also as suggested above, the primary use of the NSSE and CCSSE benchmarks is heuristic, as points of entry into discussions of survey results. They point practitioners toward areas where they may choose to dig more deeply into the data, both through examination of item-level results and through explorations of the experiences of different groups of students—for example,

students of color, or part-time students, or first-generation students, or students in different academic programs. Thus, the benchmarks serve as campus conversation-starters. That quintessential process of inquiry and discussion is further guided, not by the math behind the benchmarks, but by what's important to the institution (strategic priorities? equity agenda? accreditation work? student success initiatives?); by that institution's realities (mission, student characteristics, resources); and by the ability of the results to expose questionable assumptions about students and their experiences.

CCSSE and NSSE also emphasize that the *process* of benchmarking is a process of continuous improvement—that the aim for each institution is, over time, to hone understanding of current performance, to decide what matters most as a focus for improvement, to learn where appropriate from the successful performance of other organizations, and to promote ongoing improvement in a data-informed, but not data-driven, environment.

Consistent with these clarifications of purpose, we emphasize that there are multiple definitions of “validity.” The student engagement surveys were designed for *consequential* validity—that is, to produce data that are meaningful and actionable—in other words, information that is good enough to be useful in decision-making. Most fundamentally, NSSE and CCSSE aim to transform research findings into a set of resources to help practitioners work their way through practical problems.

Toward Improvement in Undergraduate Education

Since their inception, NSSE and CCSSE have focused on facilitating the improvement of undergraduate education. A growing body of work on institutional change and improvement makes it clear that there is no silver bullet, no single lever to pull. Improving student learning and attainment requires a whole series of changes both small and large, along with passionate leadership and committed faculty and staff. Put simply, better outcomes require that we do more of what we know works.

Finally, while improvement of undergraduate education requires a long-term commitment, one of the significant problems in higher education is impatience—and the related failure to focus on and persist in implementing effective practices at scale. On the positive side, we are finally getting to the point where we can see, over time, gradual upticks in results for some key aspects of engagement. Nonetheless, as a field, we are not where students, their families and the country need us to be. Isn't changing that situation the point of this work? Isn't that where we need to be focusing scholarly effort?

REFERENCES

- Altbach, P. G. (1998). Research, policy, and administration in higher education: The Review at twenty. *Review of Higher Education*, 21(3), 205–207.

- Angell, L. (2009). Construct validity of the Community College Survey of Student Engagement (CCSSE). *Community College Journal of Research and Practice*, 33, 564–570.
- Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, 25, 297–307.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, 54, 579–616.
- Boyer, E. L. (1990). *Scholarship reconsidered: Priorities for the professoriate*. Princeton, NJ: Carnegie Foundation for the Advancement of Teaching.
- Braxton, J. M. (2005). Reflections on a scholarship of practice. *Review of Higher Education*, 28(2), 285–293.
- Brown, N. R., & Sinclair, R. C. (1999). Estimating number of lifetime sexual partners: Men and women do it differently. *Journal of Sex Research*, 36, 292–297.
- Campbell, C., & Cabrera, A. F. (2011). How sound is NSSE? Investigating the psychometric properties of NSSE at a public, research-intensive institution. *Review of Higher Education*, 35(1), 77–103.
- Cole, J. S., & Gonyea, R. M. (2010). Accuracy of self-reported SAT and ACT test scores: Implications for research. *Research in Higher Education*, 51(4), 305–319.
- Cole, J. S., & Korkmaz, A. (2011). *Estimating college student behavior frequencies: Do vague and enumerated estimation strategies yield similar results?* Paper presented at the annual meeting of the Association for Institutional Research, Toronto, Canada.
- Community College Survey of Student Engagement. (2005). *Engaging students, challenging the odds, (2005 CCSSE Findings)*. Austin: The University of Texas at Austin, Community College Leadership Program.
- Dowd, A. C., Sawatzky, M., & Korn, R. (2011). Theoretical foundations and a research agenda to validate measures of intercultural effort. *Review of Higher Education*, 35(1), 17–44.
- Garry, M., Sharman, S. J., Feldman, J., Marlatt, G. A., & Loftus, E. F. (2002). Examining memory for heterosexual college students' sexual experiences using an electronic mail diary. *Health Psychology*, 21(6), 629–634.
- Greene, T. G., Marti, N. C., & McClenney, K. (2008). The effort-outcome gap: Differences for African American and Hispanic community college students in student engagement and academic achievement. *Journal of Higher Education*, 79(5), 513–539.
- Groves, R. M. (1987). Research on survey data quality. *Public Opinion Quarterly*, 51, S156–S172.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed.) (pp. 17–64). New York: Macmillan.
- Keller, G. (1985). Trees without fruit: The problem with research about higher education. *Change*, 17(1), 7–10.
- Kuh, G. D., Cruce, T. M., Shoup, R., 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., Kinzie, J., Cruce, T., Shoup, R., & Gonyea, R. M. (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: Indiana University Center for Postsecondary Research.
- Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, 75(1), 63–82.
- Marti, C. N. (2009). Dimensions of student engagement in American community colleges: Using the Community College Student Report in research and practice. *Community College Journal of Research and Practice*, 33(1), 1–24.
- McClenney, K. M., & Marti, C. N. (2006). *Exploring relationships between student engagement and student outcomes in community colleges: Report on validation research*. Austin: Community College Leadership Program, The University of Texas at Austin.
- McClenney, K. M., Marti, C. N., & Adkins, C. (2007). *Student engagement and student outcomes: Key findings from CCSSE validation research*. Austin: Community College Leadership Program, The University of Texas at Austin.
- McCormick, A. C. (2011). It's about time: What to make of reported declines in how much college students study. *Liberal Education*, 97(1), 30–39.
- McCormick, A. C., Kinzie, J., & Korkmaz, A. (2011). *Understanding evidence-based improvement in higher education: The case of student engagement*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.) (pp. 13–103). New York: Macmillan.
- Nelson Laird, T. F., & Engberg, M. E. (in press). Establishing differences between diversity requirements and other courses with varying degrees of diversity inclusivity. *Journal of General Education*.
- Nelson Laird, T. F., Korkmaz, A., & Chen, P. D. (2009). *How often is "often" revisited: The meaning and linearity of vague quantifiers used on the National Survey of Student Engagement*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
- Nelson Laird, T. F., & Niskodé-Dossett, A. S. (2010). How gender and race moderate the effect of interactions across difference on student perceptions of the campus environment. *Review of Higher Education*, 33(3), 333–356.
- Nora, A., Crisp, G., & Matthews, C. (2011). A reconceptualization of CCSSE's benchmarks of student engagement. *Review of Higher Education*, 35(1), 105–130.
- NSSE. National Survey of Student Engagement. (2006). *Engaged learning: Fostering success for all students—Annual Report 2006*. Bloomington: Indiana University Center for Postsecondary Research.
- NSSE. National Survey of Student Engagement. (2009). *Assessment for improvement: Tracking student engagement over time—Annual results 2009*. Bloomington: Indiana University Center for Postsecondary Research.
- Olivas, M. A. (2011). If you build it, they will assess it (or, an open letter to George Kuh, with love and respect). *Review of Higher Education*, 35(1), 1–15.

- Ouimet, J. A., Bunnage, J. C., Carini, R. M., Kuh, G. D., & Kennedy, J. (2004). Using focus groups, expert advice, and cognitive interviews to establish the validity of a college student survey. *Research in Higher Education, 45*(3), 233–250.
- Pace, C. R. (1980). Measuring the quality of student effort. *Current Issues in Higher Education, 2*, 10–16.
- Pace, C. R., & Friedlander, J. (1982). The meaning of response categories: How often is “occasionally,” “often,” and “very often”? *Research in Higher Education, 17*(3), 267–281.
- Pascarella, E. T., Seifert, T. A., & Blaich, C. (2008). *Validation of the NSSE benchmarks and deep approaches to learning against liberal arts outcomes*. Paper presented at the Annual Meeting of the Association for the Study of Higher Education, Jacksonville, FL.
- Pascarella, E. T., Seifert, T., & Blaich, C. (2010). How effective are the NSSE benchmarks in predicting important educational outcomes? *Change, 42*(1), 16–22.
- Pike, G. R. (1995). The relationships between self reports of college experiences and achievement test scores. *Research in Higher Education, 36*, 1–22.
- Pike, G. R., Kuh, G. D., & McCormick, A. C. (2011). An investigation of the contingent relationships between learning community participation and student engagement. *Research in Higher Education, 52*(3), 300–322.
- Porter, S. R. (2011). Do college student surveys have any validity? *Review of Higher Education, 35*(1), 45–76.
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking about answers*. San Francisco: Jossey-Bass.
- Tanaka, G. (2002). Higher education’s self-reflexive turn: Toward an intercultural theory of student development. *Journal of Higher Education, 73*(2), 263–296.
- Terenzini, P. T. (1996). Rediscovering roots: Public policy and higher education research. *Review of Higher Education, 20*(1), 5–13.
- Wanke, M. (2002). Conversational norms and the interpretation of vague quantifiers. *Applied Cognitive Psychology, 16*, 301–307.
- Wright, D. B., Gaskell, G. D., & O’Muircheartaigh, C. A. (1994). How much is “quite a bit”? Mapping between numerical values and vague quantifiers. *Applied Cognitive Psychology, 8*, 479–496.