

RUNNING HEAD: The Validity of NSSE

Investigating the Validity of the National Survey of Student Engagement

Peter Swerdzewski

James Madison University

B.J. Miller

Eastern Mennonite University

Randy Mitchell

James Madison University

## The Context for the Assessment of Student Learning, Skills, and Personal and Social Development

There has been a growing demand over the past two decades by the public, policy makers, parents, and students for greater fiscal and educational accountability among the nation's colleges and universities (American College Personnel Association, 2007). Stakeholders report concerns about rising tuition costs (Dickeson, 2004; National Center for Public Policy and Higher Education, 2004), the quality of the education provided by the nation's postsecondary institutions (U.S. Department of Education, 2006), and the ability of the nation's higher education system to meet the demands of society (Kirsch, Braun, Yamamoto, & Sum, 2007). Implicit to the conversation about the quality of the nation's higher education system is the need for strong programs of assessment. Assessment is generally defined as the systematic process by which inferences are made about student learning and development (Erwin, 1991), and institutions that employ comprehensive assessment strategies equip themselves with the data necessary to make informed programmatic decisions for the betterment of the institution.

Tantamount to a comprehensive campus assessment strategy is the reliance on one or more instruments that address the various programs and avenues through which students learn and develop. This includes skills gained through general classroom experiences (Association of American Colleges and Universities, 2002), the knowledge learned through structured programs such as a university's general education program (Reynolds et al., 1998), and the development that takes place through the co-curricular non-classroom-based programming that is often crafted by college student affairs divisions (R. P. Keeling, Dungy, American College Personnel Association, & National Association of Student Personnel Administrators, 2004; R. P. Keeling, American College Personnel Association, & National Association of Student Personnel

Administrators, 2006). An assessment program that recognizes the multidimensional nature of student learning and development can meet not only external demands for accountability, but can also harness the power of assessment data for the purpose of program improvement.

#### The NSSE: One Approaching to Collecting Assessment Data

One approach to collecting data for accountability and program improvement purposes is the National Survey of Student Engagement (NSSE; for detailed information on the NSSE, see <http://nsse.iub.edu/index.cfm>). The NSSE began in 2000 with a grant from the Pew Charitable Trusts and quickly grew; it was estimated that one quarter of U.S. four-year colleges and universities administered NSSE in 2005 (Dwyer, Millett, & Payne, 2006). NSSE is currently operated as a commercially-available instrument in conjunction with the Center for Postsecondary Research at Indiana University Bloomington. NSSE has recently gained prominence in the national conversation about postsecondary accountability due to its inclusion as an instrument of choice for the measurement and reporting of student experiences and perceptions in the Voluntary System of Accountability system being crafted by the American Association of State Colleges and Universities and the National Association of State Universities and Land-Grant Colleges (AASCU & NASULGC, 2007).

Although NSSE is reportedly a powerful tool for understanding the behaviors of students on college campuses (National Survey of Student Engagement, 2004), it is important to understand what NSSE purports *to measure* and *not to measure*. The NSSE “is designed to obtain, on an annual basis, information from scores of colleges and universities nationwide about student participation in programs and activities that institutions provide for their learning and personal development. The results will provide an estimate of how undergraduates spend their time and what they gain from attending college” (National Survey of Student Engagement,

2007b). Data are obtained using a paper- or Web-based survey in which students are asked to select various types and degrees of involvement that span a wide breadth of university activities and programs. The NSSE is *not* a direct measure of either student learning or development (Dwyer et al., 2006). Instead, NSSE is predicated upon the empirically-tested assumption that if students report being more engaged at a given university, that university is of a higher quality than other universities (National Survey of Student Engagement, 2004). Although NSSE in no way directly tests students' cognitive abilities or assesses the developmental changes that students experience as a result of attending college, it is claimed that "student engagement results from NSSE are a direct indicator of what students put into their education and an indirect indicator of what they get out of it. That is, NSSE data show how engaged various types of students are in effective educational practices during the first and last years of college" (National Survey of Student Engagement, 2007a).

### *Three Advantages to NSSE*

Despite the sometimes-conflicting claims as to the precise constructs measured by the NSSE, the instrument is widely recognized as having various strengths that are of great value to institutions of higher education. These include (1) the ability to benchmark results against other institutions, (2) an understanding of the degree to which students are engaged in a university's general education program, and (3) the provision of information about students' personal and social development outside the classroom.

*Benchmarks.* One strength of the NSSE is the ability of a participating institution to make cross-institutional comparisons on five "Benchmarks of Effective Educational Practice." These benchmarks, derived from 42 items on the survey, are entitled (1) Level of Academic Challenge, (2) Active and Collaborative Learning, (3) Student-Faculty Interaction, (4)

Supportive Campus Environment, and (5) Enriching Educational Experiences (National Survey of Student Engagement, 2007b). A participating institution has the option to choose, from among the pool of other participating institutions, a small number of schools to which the institution will be compared (or “benchmarked against”) on the five Benchmarks of Effective Educational Practice. If, for example, administrators at an institution find that their students report lower student-faculty interaction than the institution’s chosen comparison institutions, the administrators should implement programming and services to augment the degree to which students and faculty interact on that campus. The ability for administrators to benchmark their institution against other institutions is an important advantage of NSSE because few institutions could alone accumulate this type of data without the aid of a cross-institutional assessment program such as NSSE.

The five benchmarks were constructed with “a blend of theory and empirical analysis” (National Survey of Student Engagement, 2007b). In 2005, NSSE researchers initially employed principal components analysis (PCA) with oblique rotation on a national sample of student respondents to understand which items’ responses differentially correlate or “hang together.” They then “examined factors produced using an oblique rotation,” which suggests exploratory factor analysis. Theory was then used in conjunction with the factor analysis results to formally assign each of 42 items to one of the five factors. Institution-specific scores for each of the five benchmarks are calculated using the items identified from this 2005 PCA. Although the benchmarks were at least partially constructed using PCA, the NSSE literature (National Survey of Student Engagement, 2007b) consistently refers to these benchmarks as “factors” and implies that these benchmarks represent latent traits rather than simple distillations of data as would be accomplished by a PCA approach to data reduction.

*Engagement in General Education.* The NSSE literature (National Survey of Student Engagement, 2007b) indicates that the educational and personal growth items that comprise question 11 of the 2005 form of NSSE yielded three components: personal and social development, practical competence, and general education. A second strength of the NSSE, then, is the ability for administrators at an institution to relatively easily collect a reflection of the degree to which students are engaged in their university's general education program. Inferences about student engagement in general education are made from items comprising the general education component extracted from question 11 data. These items inquire about the degree to which a student's experience at the school has contributed to his or her broad general knowledge and ability to write and speak clearly (National Survey of Student Engagement, 2007b). Students are asked to respond on a four-point Likert-type scale ranging from "Very Much" to "Very Little." It is thought that if students report their experiences at the institution "very much" contributed to their general knowledge, writing, and speaking abilities, the institution's general education program is effective in addressing these three outcomes. Although the administrators at a given institution could similarly survey their own students about their institution's general education program, NSSE provides the added benefits of conducting all the data collection and scoring on behalf of the institution and providing the institution with comparative data.

*Personal and Social Development.* NSSE is currently often thought of as a comprehensive campus assessment instrument; however, the instrument was originally commonly used exclusively by student affairs professionals. Many of the NSSE items address co-curricular or extracurricular programs and services that are normally designed to foster learning and development apart from that which is usually nurtured *within* college classrooms. Several of these items are included in question 11 and comprise the personal and social

development component (National Survey of Student Engagement, 2007b). Specifically, student affairs professionals are often tasked with creating programs and services that augment non-cognitive “soft skills” among student participants. For example, residence hall programs may create an environment conducive to increasing students’ appreciation for community. Similarly, holding a leadership role in a student organization could increase a student’s ability to navigate complex, real-world problems. The development of personal and social skills and abilities is an important aspect of college that is often not assessed through other means, thus NSSE’s claim that it assesses these non-cognitive skills is advantageous to college administrators.

#### Purpose of the Current Study

As noted above, there are numerous benefits (three of which are explicitly discussed above) for the administrators at an institution to employ NSSE as a tool to garner evidence that students are engaged at the institution and thus that students’ lives are positively changed as a result of their experiences at the college or university. NSSE addresses key questions that are frequently posited by postsecondary stakeholders such as “how does this campus compare to another?,” “to what degree are students’ skills improving as a result of the institution’s general education program?,” and “are students’ non-cognitive personal and social skills becoming greater due to their experiences at the institution?” Furthermore, NSSE is a cost-effective alternative to hiring a campus-based assessment practitioner to collect data for accountability and program improvement purposes. Yet despite the advantages of NSSE, one must be cautious in blindly accepting the institution-specific results reported for NSSE or any other similar survey without first collecting validity evidence for the inferences to be made from the survey results (cf. American Educational Research Association, American Psychological Association, National

Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.), 1999; Messick, 1989; Shadish, Cook, & Campbell, 2002).

The purpose of the current study was to collect evidence for or against using the NSSE to make programmatic and policy changes at James Madison University (JMU). Data collected at JMU during the 2005 administration of NSSE was subjected to a three-stage strong program of construct validation (Benson, 1998). Benson's first stage of a strong program of construct validation specifies the exploration of the substantive (i.e. theoretical) underpinnings of an instrument. It is assumed that the extensive research conducted on student engagement theory (cf. American College Personnel Association, 1996; Astin, 1968; Astin, 1973; Kuh, 1991; National Survey of Student Engagement, 2007b; Pascarella & Terenzini, 1991; Pascarella & Terenzini, 2005) satisfies the need for evidence at this first stage.

The second stage of Benson's (1998) strong program of construct validation, known as the "internal phase," requires an exploration of an instrument's psychometric properties. These properties include a study of an instrument's reliability, factor structure, and item bivariate correlations given data from a specific population of interest. At this stage, one is generally interested in whether or not items relate to one another in hypothesized ways.

The third stage of Benson's (1998) strong program of construct validity is the "external stage," in which items or factors from the instrument are related in hypothesized ways to other constructs that are external to the central construct(s) measured by the instrument of interest. If scores on the instrument do not predictably relate with similar constructs, there is evidence that the instrument of interest is *not* measuring what it is purported to measure. On the other hand, if the instrument of interest *does* strongly correlate (either positively or negatively) with a construct to which it should not be related, there is likewise evidence that the instrument of interest is *not*

measuring what it is purported to measure. By relating the instrument of interest—in this case the NSSE—to other constructs we can build evidence in the form of a “nomological net” that the instrument is measuring what we think it is measuring.

### *Research Questions*

The current study follows a practical inquiry approach to research (Yousey, 2006). Specifically, the research questions identified in the current study were pursued because they are of practical utility to the assessment practices at the university. Rather than conducting a full validation of the NSSE using data from the campus addressed in this study, only three specific questions are addressed due to their value to the campus.

*Research Question 1.* Will an internal study per Benson’s (1998) second phase of a strong program of construct validity yield acceptable evidence for the factor structure of the five cross-institutional benchmarks specified for the NSSE?

*Research Question 2.* Will an internal study per Benson’s (1998) second phase of a strong program of construct validity yield acceptable evidence for the factor structure of the general education, personal and social development, and practical competence factors specified for the NSSE?

*Research Question 3.* Will an external study per Benson’s (1998) third phase of a strong program of construct validity yield relationships between NSSE factors and external data relationships that are in hypothesized directions and of expected magnitudes? Specifically, students’ scores from the NSSE general education component should positively and strongly relate with students’ cognitive test scores from a quantitative and scientific reasoning general education assessment as well as test scores from a psychology / sociology general education assessment. Furthermore, students’ scores from the NSSE general education component should

positively and moderately relate with students' cumulative grade point average (GPA). Additionally, students' NSSE scores from the personal and social development factor should positively and moderately relate with scores on a diversity understanding instrument and a metacognition instrument, and should positively and strongly relate with scores on a sense of belonging instrument and a feelings of morale [for being part of a group] instrument.

## Method

### *Participants*

The NSSE data used in the current study were collected during 2005. A total of 1309 students at James Madison University (JMU) responded to the Web-based version of the survey, yielding a response rate of 49%. Analyses in the current study were based only on the responses from 495 full-time first-year students, of whom 72% were female and 83% were Caucasian. This sample represents approximately 15% of the entire first-year cohort. All data were originally collected by NSSE staff members and a report and raw data file (used in this study) were later transmitted back to the university.

### *Measures*

Each spring semester, JMU conducts campus-wide assessment days during which sophomores are measured on constructs related to the general education experience as well as various student development constructs (used for the assessment of student affairs programs). The test battery includes both cognitive and developmental measures. Many of the first-year students who participated in the NSSE in 2005 also participated in the spring 2006 assessment day. Because student identification numbers were collected on both occasions, this presented an opportunity to examine correlations between NSSE responses and scores on the various general

education instruments. For the current study, student scores on three developmental and two cognitive tests were considered.

*MGUDS-S.* The short form of the Miville-Guzman Universality-Diversity Scale (MGUDS-S; Fuertes, Miville, Mohr, Sedlacek, & Gretchen, 2000) is a measure of personal and social development with respect to diversity. The MGUDS-S is composed of three 5-item subscales (Diversity of Contact, Relativistic Appreciation, and Comfort with Differences). A total score is also obtained by summing the three subscale scores. Students respond using a 6-point scale where 1 indicates strong disagreement and 6 indicates strong agreement, thus higher scores are associated with higher levels of development with respect to diversity.

*MAI.* The Metacognitive Awareness Inventory (MAI; Schraw & Dennison, 1994) is a 52-item instrument assessing two components of metacognition – knowledge of cognition and regulation of cognition. At JMU, only the 35 items representing regulation of cognition are administered. Students use a 5-point response scale in which 1 indicates that the behavior described in the statement is never true of the student and 5 indicates that the behavior is always true of the student. In this way, higher MAI scores are associated with higher levels of metacognition.

*PCS.* The Perceived Cohesion Scale (PCS; Bollen & Hoyle, 1990) is a 6-item instrument measuring group cohesion according to a two-dimensional conceptualization of the construct. That is, people perceive their cohesion to a group in terms of both their sense of belonging (SOB) to the group and their feelings of morale (FOM) as a member of the group. Students respond using a 9-point scale where 1 indicates strong disagreement and 9 indicates strong agreement, so higher PCS scores are associated with higher levels of cohesion on each of the two scales.

*Cognitive Tests.* The Socio-Cultural Domain Assessment (SDA) is a 40-item multiple-choice test that assesses the five learning goals in the “Social and Cultural Processes” cluster of JMU’s general education program. The Natural World (NW) is a 65-item multiple-choice test that assesses the eight general education learning goals in scientific and quantitative reasoning. Both tests were developed by JMU faculty who teach courses in the respective general education areas.

*GPA.* In addition to the tests administered on assessment day, cumulative grade point average was used as a measure of student achievement. Although there is debate as to whether or not GPA actually measures performance (Lei, Bassiri, & Schultz, 2001), GPA is nonetheless a commonly-accepted index of student performance.

#### *Data Analyses*

*Factor Analyses.* A confirmatory factor analysis (CFA) and a principal components analysis (PCA) were conducted to verify the reported data structure of the NSSE. Specifically, LISREL 8.72 (Jöreskog & Sörbom, 1996) was used to analyze responses to the 42 items specified as comprising the NSSE benchmarks. A five-factor model corresponding to the five benchmark scores was tested. In addition, a PCA with oblique rotation was conducted to extract three components from responses to the 16 parts of NSSE Question 11.

*Correlations.* Criterion measures were correlated with survey responses. NSSE component scores, as opposed to item scores, were used when psychometrically supported.

### Results

*NSSE Benchmarks.* The five-factor model corresponding to NSSE’s benchmark scores did not fit the data according to commonly accepted criteria. Specifically, a multiple-index approach was employed to evaluate the fit of the five-factor model to the data (Brown, 2006).

The chi square goodness of fit test is reported; however, the chi square is sensitive to sample size, so caution must be taken in interpreting the value of the chi square (Kline, 2005). Three additional indices are used to assess model-data fit (for a discussion of these and other fit indices, see Hu & Bentler, 1995; Hu & Bentler, 1998): the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Standardized Root Mean Squared Residual (SRMR). Hu and Bentler (1999) suggest that RMSEA values less than 0.06 indicate good fit, CFI values greater than 0.95 indicate good fit, and SRMR values less than 0.08 indicate good fit. Table 1 shows the fit statistics obtained for the CFA model. All values are outside the recommended boundaries for good fit, and the CFI's substantially-low value indicates the possible presence of complex model misspecification (Hu & Bentler, 1998), or the lack of relationships among items across various factors.

Table 1

*Fit Statistics for the Five-Factor CFA Model Corresponding to the NSSE Benchmarks*

Fit Statistic	Value
$\chi^2(809)$	4295.31, $p < .000$
RMSEA	.09
CFI	.53
SRMR	.09

*NSSE Question 11: General Education and Personal / Social Development.* For the 16 parts of NSSE Question 11, a three-component solution explained 65% of the variance in item responses. However, only two eigenvalues were greater than 1, and the scree plot indicated that a two-component solution may have been more appropriate. Table 2 presents standardized pattern coefficients that met the .30 cut-off value. Despite several cross-loadings, the three-component solution was, for the most part, interpretable within the NSSE framework. Specifically, Component 1 (items 11.a – 11.h) seemed to correspond to NSSE's "general education,"

Component 2 (items 11.i – 11.k) seemed to correspond to NSSE’s “personal and social development,” and Component 3 (items 11.l – 11.p) seemed to correspond to NSSE’s “practical competence.” Although these items reasonably mapped to the three components also named by NSSE researchers, it is critical to note that many of the items did *not* map back to the same components specified in the NSSE documentation (National Survey of Student Engagement, 2007b). Again, given that assessment scores (and thus construct validity) are inherently sample-specific, a three-factor solution with the item mappings specified in Table 2 may be most appropriate for the population under study.

Table 2

*Standardized Pattern Coefficients for Items Comprising NSSE Question 11*

Item #	<i>Institutional contribution to the following areas:</i>	Per NSSE*	Component		
			1	2	3
11.a	Acquiring a broad general education	GenEd	0.62	0.31	
11.b	Acquiring job or work-related knowledge and skills	PC	0.51		0.31
11.c	Writing clearly and effectively	GenEd	0.91		
11.d	Speaking clearly and effectively	GenEd	0.86		
11.e	Thinking critically and analytically	PC	0.86		
11.f	Analyzing quantitative problems	PC	0.69		
11.g	Using computing and information technology	PC	0.66		
11.h	Working effectively with others	PC	0.62	0.30	
11.i	Voting in local, state (provincial), or national (federal) elections	PSD		0.72	
11.j	Learning effectively on your own	PSD		0.67	
11.k	Understanding yourself	PSD		0.54	0.33
11.l	Understanding people of other racial and ethnic backgrounds	PSD			0.71
11.m	Solving complex real-world problems	PSD			0.64
11.n	Developing a personal code of values and ethics	PSD			0.65
11.o	Contributing to the welfare of your community	PSD			0.69
11.p	Developing a deepened sense of spirituality	PSD			0.86

\* NSSE literature specifies the item maps to the following factor: GenEd = General Education,

PSD = Personal and Social Development, PC = Practical Competence

*Correlations.* Because the five-factor NSSE benchmark model did not fit the data, benchmark scores were not calculated and so could not be correlated with any of the criterion

measures. Component scores were calculated by summing the responses to Question 11 according to the obtained three-component solution. The “general education” component scores were correlated with the two cognitive tests (SDA and NW) and with GPA. The “personal and social development” component scores were correlated with the developmental measures (MGUDS-S, MAI, SOB, and FOM). Tables 3 and 4 present means, standard deviations, reliabilities and correlations for the two component scores.

Table 3

*N, Mean, SD, Reliability, and Correlations for the General Education Component*

	<i>N</i>	<i>Mean</i>	<i>SD</i>	$\alpha$	<i>r</i>
General Education	352	12.8	3.65	.84	
SDA	110	28.6	5.08	.79	-.23
NW	81	49.0	8.17	.86	-.32
GPA	352	3.1	0.39	--	-.09

Table 4

*N, Mean, SD, Reliability, and Correlations for the Personal and Social Development Component*

	<i>N</i>	<i>Mean</i>	<i>SD</i>	$\alpha$	<i>r</i>
Personal & Social	203	24.6	4.77	.88	
MGUDS-S	101	66.5	9.71	.88	.14
MAI	203	127.3	15.91	.88	.17
SOB	199	21.1	5.38	.95	.32
FOM	199	21.2	5.53	.89	.31

*Note.* *N*, *Mean*, and *SD* in Tables 3 and 4 are based on students with matching records for NSSE and listed measure; however, coefficient alpha for criterion measures was calculated using all available assessment day test scores.

## Discussion

In an exploration of the second and third phases of Benson’s (1998) strong program of construct validity, little evidence was amassed for the use of the NSSE for the sample of interest

(students at JMU) on the items and factors of interest (those items/factors relating to the benchmarks, general education, and personal/social development).

Specifically, a CFA of the benchmarks as specified by NSSE literature produced poor model fit. For the sample of students examined for this study, the five-factor benchmark model supported by NSSE was not upheld, thus a five-factor solution is uninterpretable for the sample. A comparison of benchmark scores from this sample to scores from a sample at another university should not be made, as the benchmark scores from this sample were not empirically supported. Until further studies can provide evidence for an interpretable set of benchmarks for this and similar samples, one should not make policy or programmatic decisions based on the benchmark scores. In short, the internal phase of Benson's (1998) strong program of construct validity did not yield evidence of an interpretable factor structure for the benchmarks.

Additionally, the factor structure of the general education and personal/social development indicators was not supported for this sample using a PCA factor analysis. Again adhering to the internal phase of Benson's (1998) strong program of construct validity, an interpretable solution was found, yet this solution did not parallel that espoused in NSSE literature. Given that the components found for the sample examined in this study did yield an interpretable solution, these components may provide some utility for programmatic and policy decisions, given that in the future the components are replicated using NSSE data from additional samples from the same population. Although it is hoped that the same interpretable solution is again found, we are apprehensive that such a solution will be recovered due to various measurement issues inherent to the NSSE, such as the restricted four-point response scale used for these items.

Finally, using the interpretable three-component solution found for research question two, the general education and personal/social development components were correlated with various criteria in an attempt to produce external validity evidence for this section of the NSSE (again per Benson's strong program of construct validity). Surprisingly, the general education component correlated *negatively* with the general education test scores for students at the university. Said another way, students who reported on the NSSE that the institution contributed greatly to their development of general education-related competencies actually scored *lower* on the institution's measures of performance in two of the five general education divisions (known as "clusters") and these students also had *lower* GPAs. Evidence was thus amassed from this study that *refutes* the external validity of the general education component for this sample of students. The relationship between the personal and social development component and the non-cognitive measures collected at the university were in the hypothesized direction, but were not of the hypothesized magnitude. The correlation between the personal/social development NSSE component and the MGUDS-S (diversity scale) and the MAI (metacognition scale) were extremely weak, thereby providing little evidence for the external phase of this validity study. However, moderate correlations were found between the perceived cohesion scales (the SOB and the FOM) and the personal/social development NSSE component, yet these correlations are not nearly as strong as initially hypothesized.

#### *Implications for Practice*

These findings provide evidence that assessment practitioners should be cautious in interpreting NSSE data, and even further cautious in basing policy decisions upon the data until sufficient psychometric work yields acceptable results that address Benson's (1998) internal and external phases of a strong program of construct validity. Recognizing that the sample for this

study is likely very similar to the vast majority of the students at the nation's traditional four-year institutions at which NSSE is commonly administered, it is possible that the results found in this study will similarly replicate at other traditional four-year institutions. NSSE is widely-used throughout the United States and its validity is seemingly blindly accepted. College and university administrators must remain diligent in addressing the psychometric properties of NSSE just as they would any other instrument. Given the sample-specific nature of construct validity, studies similar to this must be periodically conducted to ensure (1) the items are congruent with the theory of interest, (2) the items related to one another in predicted ways, and (3) the items / components relate to external criteria in predicted ways. Without adherence to a strong program of construct validity, inferences made from instruments that are even as well-known as NSSE must be made with extreme vigilance.

## References

- AASCU & NASULGC. (2007). *The voluntary system of accountability (VSA): Summary and background materials*. Unpublished manuscript. Retrieved September 9, from <http://www.nasulgc.org/NetCommunity/Document.Doc?id=170>
- American College Personnel Association. (1996). The student learning imperative: Implications for student affairs. *Journal of College Student Development*, 37(2), 118-122.
- American College Personnel Association. (2007). *ASK standards: Assessment skills and knowledge content standards for student affairs practitioners and scholars*. Washington, D.C.: American College Personnel Association.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (1999). *Standards for educational and psychological testing*. Washington: DC American Educational Research Association.
- Association of American Colleges and Universities. (2002). *Greater expectations: A new vision for learning as a nation goes to college. National panel report*. Washington, D.C.: Association of American Colleges and Universities.
- Astin, A. W. (1968). Undergraduate achievement and institutional "excellence." Traditional indices of institutional quality do not appear to contribute to student achievement. *Science*, 161(842), 661-668.

- Astin, A. W. (1973). Measurement and determinants of the outputs of higher education. In L. C. Solmon, & P. J. Taubman (Eds.), *Does college matter? Some evidence on the impacts of higher education*. New York: Academic Press.
- Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice*, 17(1), 10.
- Bollen, K. A., & Hoyle, R. H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social Forces*, 69, 479-504.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Dickeson, R. C. (2004). *Collision course: Rising college costs threaten America's future and require shared solutions* (Policy Brief). Indianapolis, IN: Lumina Foundation.
- Dwyer, C. A., Millett, C., M., & Payne, D. G. (2006). *A culture of evidence: Postsecondary assessment and learning outcomes. Recommendations to policymakers and the higher education community*. Princeton, NJ: Educational Testing Service.
- Erwin, T. D. (1991). *Assessing student learning and development: A guide to the principles, goals, and methods of determining college outcomes*. San Francisco: Jossey-Bass.
- Fuertes, J. N., Miville, M. L., Mohr, J. J., Sedlacek, W. E., & Gretchen, D. (2000). Factor structure and short form of the Miville-Guzman Universality-Diversity Scale. *Measurement and Evaluation in Counseling and Development*, 33, 157-169.
- Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications*. (pp. 76-99) Sage Publications, Inc.

- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological methods*, 3(4), 424-453.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1.
- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Chicago: Scientific Software International.
- Keeling, R. P., Dungy, G. J., American College Personnel Association, & National Association Of Student Personnel Administrators. (2004). *Learning reconsidered : A campus-wide focus on the student experience*. Washington, D.C: ACPA & NASPA.
- Keeling, R. P., American College Personnel Association, & National Association Of Student Personnel Administrators. (2006). *Learning reconsidered 2: Implementing a campus-wide focus on the student experience*. Washington, D.C.: ACPA & NASPA.
- Kirsch, I., Braun, H., Yamamoto, K., & Sum, A. (2007). *America's perfect storm: Three forces changing our nation's future* (Policy Information Report). Princeton, NJ: Educational Testing Service.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
- Kuh, G. D. (1991). *Involving colleges : Successful approaches to fostering student learning and development outside the classroom* (1st ed.). San Francisco: Jossey-Bass Publishers.

- Lei, P., Bassiri, D., & Schultz, E. M. (2001). *Alternatives to the grade point average as a measure of academic achievement in college* (ACTRR No. 2001-4). Iowa City, IA: ACT, Inc.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement*. (3rd ed., pp. 13-103). New York, NY: Macmillan Publishing Co, Inc.
- National Center for Public Policy and Higher Education. (2004). *Measuring up 2004 : The national report card on higher education*. San Jose, CA: National Center for Public Policy and Higher Education.
- National Survey of Student Engagement. (2004). *Student engagement pathways to collegiate success : The NSSE annual report 2004*. Bloomington: Indiana University Center for Postsecondary Research and Planning.
- National Survey of Student Engagement. (2007a). *Accreditation toolkit: Mapped to 2007 NSSE survey questions. southern association of colleges and schools (SACS)*. Unpublished manuscript. Retrieved October 1, from <http://nsse.iub.edu/institute/documents/accred/NSSE%20Toolkit%20SACS.pdf>
- National Survey of Student Engagement. (2007b). *NSSE: National survey of student engagement*. Retrieved August 28, 2007, from <http://nsse.iub.edu/index.cfm>
- Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students : Findings and insights from twenty years of research* (1st ed.). San Francisco: Jossey-Bass Publishers.

- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students : A third decade of research* (1st ed.). San Francisco: Jossey-Bass.
- Reynolds, C. W., Allain, V. A., Erwin, T. D., Halpern, L. C., McNallie, R., & Ross, M. K. (1998). Looking backward: James Madison University's general education reform. *Journal of General Education*, 47(2), 149.
- Schraw, G. and Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology*, 19, 460-475.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. higher education. A report of the commission appointed by secretary of education Margaret Spellings*. Washington, D.C.: U.S. Department of Education.
- Yousey, K. M. (2006). Assessment matters—The best of both worlds: Coupling research and assessment. *About Campus*, 11(4), 23-25.