

Measuring Quality:

A Comparison of U. S. News Rankings and NSSE Benchmarks

Gary R. Pike
Director of Institutional Research
Mississippi State University
P. O. Box EY
Mississippi State, MS 39762-5708
gpike@ir.msstate.edu

Paper presented at the annual meeting of the Association for Institutional Research, Tampa, FL (May, 2003).

Abstract

College rankings and guidebooks have become big business. The prominent role played by rankings and guidebooks is problematic because the criteria used to evaluate institutions have little to do with the quality of education students receive. Designed as an alternative to college rankings, NSSE assesses student engagement in activities that contribute to learning and success during college. This study compared the NSSE scores for 14 AAU public research universities with their rankings by U. S. News and World Report.

Measuring Quality:

A Comparison of U. S. News Rankings and NSSE Benchmarks

According to Hossler (2000), attempts to rank colleges and universities on the basis of educational quality are a relatively new phenomenon, growing out of public interest in accountability and assessment. Despite the fact that rankings and guidebooks are a relatively recent development, they have become big business. More than 100 different guidebooks and rankings are available (Hunter, 1995), and more than 6.7 million copies are sold each year (McDonough, Antonio, Walpole, & Perez, 1998). U. S. News and World Report, for example, sells approximately 2.2 million copies of its rankings each year, reaching nearly 11 million people (Dichev, 2001).

The importance of college guidebooks and rankings extends far beyond numbers sold. They are among the most frequently used print and media sources of information used in selecting a college (Hossler & Foley, 1995). McDonough, Antonio, Walpole, and Perez (1998) estimated that each year 400,000 prospective students and their parents make use of guidebooks and rankings in selecting a college. Prospective students, and their parents, tend to accept the information provided by guidebooks and rankings at face value (Hunter, 1995) and use that information to confirm their choices (McDonough, Antonio, Walpole, & Perez, 1998). In addition, some middle- and upper-class students use guidebooks and rankings to eliminate or include colleges in the choice process (Hossler & Foley, 1995). Guidebooks and rankings also are perceived to be important in selecting a college by admission officers (Hossler & Foley, 1995). Many college and universities regularly include information from guidebooks and rankings in their promotional materials (Hossler, 2000; Hunter, 1995).

The prominent role in the college-choice process played by guidebooks and rankings is problematic for several reasons. Many of these publications provide little, if any, information about how data are obtained and ratings calculated (Hossler & Litten, 1993), and there is even some evidence to suggest that college and universities have provided inaccurate data to improve their rankings (Hossler, 2000; Pollock, 1992; Stecklow, 1995). Even when the data are accurate, the results may be suspect. Dichev (2001) found that as little as 10% of the variation in an institution's U. S. News scores over time was due to changes in the quality of the institution. The remainder of the variation was primarily due to changes in how the ratings were calculated.

Other scholars have questioned the validity of college ratings, noting that the criteria used to evaluate institutions have little or nothing to do with the quality of the education a student receives (Hossler, 2000; Pascarella, 2001). As Pascarella (2001) concluded:

The national magazine rankings should stop pretending that they are actually identifying the “best colleges” with respect to undergraduate education. Since their measures of what constitutes “the best” in undergraduate education are based primarily on resources and reputation, and not on the within-college experiences that we know really make a difference, a more accurate, if less marketable, title for their enterprise might be “America’s Most Advantaged Colleges” (p. 21).

Developed, in part, as an alternative to reputation- and resource-based rankings, the National Survey of Student Engagement (NSSE) was designed to assess the extent to which students are engaged in educationally purposeful activities that contribute to their learning and success during college (Kuh et al., 2001). Half a century of research has shown that student engagement is an important factor in student learning and success (Astin, 1977, 1993; Feldman & Newcomb, 1969; Pascarella & Terenzini, 1991). Most recently, several studies found that measures of

student engagement, similar to those used by NSSE, were significantly related to students' reported gains in learning (Astin, 1993; Kuh, Pace, & Vesper, 1997; Pike, Kuh, & Gonyea, in press) and achievement test scores (Pike, 1995).

The conceptual framework underpinning NSSE is drawn from Chickering and Gamson's (1987) "Seven Principles of Good Practice in Undergraduate Education" (Kuh et al., 2001). NSSE's survey, The College Student Report, asks students about their engagement in activities that reflect good practice in undergraduate education. Five institutional benchmarks have been developed using items from the survey (Kuh, 2001b). These benchmarks are (1) Level of Academic Challenge, (2) Active and Collaborative Learning, (3) Student Interaction with Faculty Members, (4), Enriching Educational Experiences, and (5) Supportive Campus Environment. These benchmarks serve as proxy measures to identify opportunities for improving undergraduate education (Kuh, 2001a).

Although the NSSE benchmarks are touted as an alternative to the ratings and rankings of guidebooks and news magazines, there is not published research showing that the two approaches for evaluating college and universities yield different results. The present research compares the NSSE benchmark scores for 14 AAU public research universities with those same institutions' rankings by U. S. News and World Report. The focus of this study was on the relationships, if any, between the five NSSE benchmarks and the six categories of academic excellence used by U.S. News.

Research Methods

Although it is tempting to assess the relationships between U. S. News rankings and NSSE benchmarks by simply correlating rankings of academic excellence with institutional

benchmarks, this approach ignores the fact that the data used to derive the U. S. News rankings of academic excellence are collected at the institutional level, whereas NSSE benchmark scores are based on the responses of students nested within institutions. Moreover, the characteristics of student populations differ substantially across institutions and can confound the interpretation of benchmark scores (Kuh et al., 2001). As a result, hierarchical linear modeling was used to accommodate the multi-level data in this study.

Data Sources

The data for this study were the 2000 U. S. News rankings for 14 AAU public research universities and the results from the 2000 NSSE survey for those same institutions. All of the data were obtained from the AAU data exchange. During the Spring 2000 semester, 3587 seniors completed The College Student Report at the 14 participating institutions.

Approximately 58% of the seniors were female and 21% were members of a minority group (1% Native American, 8% Asian American, 4% African American, 4% Hispanic, and 4% other or multiracial). Approximately 16% of the participants were 24 years of age or older, 73% began college at their institutions, and 88% attended college full time. Slightly more than 17% of the seniors were members of a fraternity or sorority, and 16% lived on campus. Only 6% of the students majored in the humanities; 26% majored in mathematics or the natural sciences; 28% majored in pre-professional programs, 12% majored in the social sciences, and 28% had multiple majors. Respondents were generally typical of seniors at their institutions, except that Asian Americans tended to be overrepresented and African Americans tended to be underrepresented in the sample of NSSE respondents.

Measures

Three sets of measures were used in the present research. The five engagement benchmarks from The College Student Report were the first set of measures used in this study. Self-report data, such as the NSSE benchmarks have been widely used in research on college effects, and the validity and credibility of these data have been extensively studied (Baird, 1976; Berdie, 1971; Pace, 1985; Pike, 1995; Pohlmann & Beggs, 1974). Research has shown that self reports are likely to be valid under five conditions:

1. the information requested is known to the respondents;
2. the questions are phrased clearly and unambiguously;
3. the questions refer to recent activities;
4. the respondents think the questions merit a serious and thoughtful response; and
5. answering the questions does not threaten, embarrass, or violate the privacy of the respondent or encourage the respondent to respond in socially desirable ways (Kuh et al., 2001, p. 9).

Empirical evidence indicates that The College Student Report satisfies these conditions (Kuh, 2001b; Kuh et al., 2001). In addition, research has found that test-retest reliability for students is quite high (0.83), and that mode of administration effects for paper and web administration are quite small (Kuh, 2001b; Kuh et al., 2001). Moreover, follow-up studies found few meaningful differences between respondents and non-respondents in terms of their levels of engagement (Kuh, 2001b).

Following procedures similar to those outlined by Kuh et al. (2001), students' raw scores for the individual items comprising a benchmark were all put on the same scale. Raw benchmark scores for each student were then calculated by summing scores for the items comprising a

benchmark. Finally, the raw benchmark scores were put on a 100-point scale. Level of Academic Challenge was calculated from 10 survey items focusing on time spent preparing for class, amount of reading and writing, and expectations for academic performance ($\alpha = 0.68$). Active and Collaborative Learning included seven questions dealing with class participation, working with other students, and discussions of ideas with others ($\alpha = 0.64$). Student Interactions with Faculty Members was based on six questions concerning students' discussions with faculty, feedback from faculty, and opportunities to participate in research ($\alpha = 0.72$). Enriching Educational Experiences included 11 questions dealing with interacting with diverse groups of students, using electronic technology in classes, and participating in a variety of curricular and co-curricular activities ($\alpha = 0.63$). The final benchmark, Supportive Campus Environment, included six items about the extent to which students perceived the institution helped them succeed and promoted positive relationships with faculty, peers, and administration ($\alpha = 0.74$).

The second set of measures was the background characteristics of the NSSE respondents. Gender was a dichotomous variable indicating whether students were female (1) or male (0). Age also was a dichotomous variable indicating whether respondents were non-traditional (i.e., 24 year of age or older) or traditional students. Ethnicity was a dichotomous item indicating whether students were members of a minority group (1) or white (0). Although some research has shown that college experiences can differ substantially across minority groups, preliminary analyses revealed that four of the five benchmark scores did not differ significantly across the minority groups. Differences on the Supportive Campus Environment benchmark accounted for less than 1% of the variance in scale scores.

Where students had attended college also was a dichotomous item indicating whether respondents were native (1) or transfer (0) students at their present institution. Enrollment status was represented by a dichotomous item indicating whether respondents were enrolled full (1) or part (0) time. A dichotomous item indicating whether students were 91) or were not (0) members of a Greek-letter organization was used to represent fraternity/sorority membership, and campus living arrangements were represented by a dichotomous item indicating whether students lived on (1) or off (0) campus. Academic discipline was represented by four items indicating whether students were majoring in science, engineering, and mathematics, pre-professional programs (business, health professions, etc.), social sciences, or multiple disciplines. Students coded as not majoring in any of these areas were in the humanities and fine arts.

The third set of measures consisted of six rankings of academic excellence from U. S. News in 2000. The first of these measures, the rankings for academic reputation, was based on presidents', provosts', and chief admission officers' ratings of academic programs at national universities. The second measure was the ranking of national universities based on their six-year graduation and freshman-to-sophomore retention rates. The third measure, faculty resources, was based on a variety of factors including class size, faculty pay, proportion of faculty with terminal degrees, and proportion of full-time faculty. Student selectivity was the fourth measure of academic excellence and included factors such as average ACT and SAT scores, acceptance rates, and yield rates. The fifth measure, financial resources, was derived from per student expenditures for instruction, research, student services, and related educational expenses. The final measure, alumni giving, was based on the percentage of alumni giving to their school.

Data Analysis

The data for this study were analyzed using the HLM 5.05 computer program (Raudenbush, Bryk, Cheong, & Congdon, 2001), and the steps used in the analysis followed the procedures outlined by Bryk and Raudenbush (1992) and Ethnington (1997). The first step in the hierarchical linear modeling process involved determining whether there was sufficient variance in NSSE institutional benchmark scores to warrant the use of HLM procedures. To answer this question, baseline models for each NSSE benchmark were specified and tested. The models included an intercept for the student level model and no other parameters. The models were equivalent to a oneway analysis of variance in which institution was the independent variable and students' benchmark scores were the dependent variables. Dividing the variance in the intercepts (i.e., institutional means) by itself plus the pooled variance within institutions (i.e., the total variance in benchmark scores) provided estimates of the proportions of variance in the five benchmarks that were attributable to institutions. These estimates of explained variance were equivalent to traditional eta-squared coefficients produced by ANOVA.

The second step in the HLM process involved the within-institution regression of students' benchmark scores on student background characteristics (gender, minority status, major, etc.). For these analyses, students' background characteristics were centered about their grand means. That is, the grand mean (across institutions) for an independent variable was subtracted from each student's observed score for that variable. As a result of centering, intercepts represented institution benchmark means that were net the effects of student differences across institutions (see Bryk & Raudenbush, 1992).

As with traditional OLS regression, tests of the effect parameters provided an indication of whether the student-background variables were significantly related to benchmark scores. In

addition, variances in the level-1 intercepts provided an indication of whether there was sufficient variability in institutional benchmark means, after controlling for student characteristics, to warrant developing an institution-level model. Two tests of the variances were utilized. First, chi-square significance tests were calculated to determine if group differences existed. Second, reliability coefficients were examined to determine if the observed differences among institutions were meaningful. Reliability coefficients of 0.60 or greater were considered to be indicators of meaningful differences. Examining changes in the pooled within-school variances (i.e., residuals) from the first to the second step in the analysis provided an indication of the explanatory power of the student characteristics. Specifically, dividing the decrease in the pooled within-institution variance components by the within-institution variance components from the first step in the analysis identified the proportion of the student-level variance that was explained by student characteristics.

The final step in the HLM analysis involved specifying and testing two-level models that included the models used in the previous step, plus U. S. News rankings as variables explaining the variances in level-1 intercepts (i.e., institutional benchmark means that were adjusted for differences in students' characteristics). Only benchmarks for which there was statistically significant variance in the level-1 intercepts were included in this step. Significance tests for the effect parameters identified the U. S. News rankings that were associated with differences in institutional benchmark scores.

Results

Analysis of the baseline models revealed that there were statistically significant differences among the institutional means for all five NSSE benchmarks. The results, which are presented in Table 1, also indicated that the explanatory power of the institutional differences was extremely

modest. Estimates of explained variance ranged from less than 1% for Student Interaction with Faculty Members to 6% for Enriching Educational Experiences.

Insert Table 1 about here

The second set of analyses revealed that students' background characteristics were significantly related to their NSSE scores. The results of these analyses are presented in Table 2. An examination of the effect parameters revealed that perceptions of the Level of academic challenge at an institution were positively related to being female (1.98) and being enrolled full time (5.73). Overall, student characteristics accounted for 8% of the variance in level of Academic Challenge scores. Student characteristics accounted for 6% of the variance in Active and Collaborative Learning scores. These scores were positively related to being a full-time student (3.98), majoring in science or mathematics (3.58), majoring in a pre-professional program (7.59), and having multiple majors (5.25). Only 3% of the variance in Student Interaction with Faculty Members was explained by student characteristics. Being enrolled full time was positively related to faculty-student interaction (2.78), whereas majoring in a pre-professional program was negatively related to interaction with faculty (-4.55).

Approximately 13% of the variance in Enriching Educational Experiences was related to student characteristics. Being female (2.10) and a member of a minority group (3.55) were positively related to engagement in enriching experiences, as were being a native student (2.26), being a full-time student (4.11), being a member of a fraternity or sorority (2.71), and living on campus (1.90). Being 24 years of age or older was negatively related to students' Enriching Educational Experiences scores (-2.95). Student characteristics explained 5% of the variance in

the Supportive Campus Environment scores. Majoring in a pre-professional program was positively related to perceptions of a supportive environment (4.49).

Insert Table 2 about here

Table 3 presents the results of tests of the variances in intercepts (i.e., adjusted benchmark means) across institutions. An examination of the coefficients presented in the table revealed that there was significant, and meaningful, variation in the intercepts for four of the five NSSE benchmarks. Only Student Interaction with Faculty Members did not vary across institutions after controlling for differences in students' backgrounds.

Insert Table 3 about here

Table 4 presents the results of the final set of analyses which examined the relationships between U. S. News rankings of academic excellence and NSSE benchmark scores, after controlling for differences in student characteristics. An examination of these results revealed that U. S. News rankings and NSSE benchmarks were not related to one another, with one exception. The U. S. News rankings for student selectivity were related to Enriching Educational Experiences means (-0.08). Given that a higher ranking had a smaller numerical value than a lower ranking (i.e., 1 was higher than 99), the direction of this relationships was such that students at more selective institutions reported higher levels of engagement in activities that could be classified as educationally enriching.

Insert Table 4 about here

Discussion

This study found modest, but statistically significant, differences in mean NSSE benchmark scores for 14 AAU public research universities. For the most part, institutional differences persisted after controlling for differences in the characteristics of students attending those institutions. Most important, institutional mean benchmark scores generally were not related to U. S. News and World Report rankings of academic excellence. The only statistically significant relationship identified in this study was between the U. S. News selectivity ranking and the NSSE Enriching Educational Experiences benchmark. Specifically, students attending more selective research universities reported higher levels of engagement in enriching activities than did their counterparts at less selective research universities.

Limitations

Care should be taken not to over generalize these results. The findings of this study were based on a subset of research universities ranked by U. S. News who also participated in the National Survey of Student Engagement. At most, these results can be generalized to large public research universities. The relatively small number of institutions in the study also limited the power of the statistical tests and raises questions about the stability of the parameter estimates in the level-2 analyses. Tests based on a larger sample of institutions might identify other important relationships between college rankings and NSSE scores.

The homogeneity of the institutions included in the present research represents a second limitation on the generalizability of the results. Limiting this study to AAU public research universities may explain why the analyses found only modest differences among universities on the NSSE benchmarks. Future research should focus on a more diverse group of institutions. However, the diversity of institutions is limited by the fact that U. S. News provides separate rankings for different types of institutions (e.g., national universities, national liberal arts colleges, and regional universities).

This research also was limited by its reliance on only one set of rankings—those published by U. S. News and World Report. Although it may be appropriate to say that there are few relationships between NSSE scores and U. S. News rankings, it cannot be said with confidence that the NSSE benchmarks are not related to other reputation- and resource-based rankings. Future research should focus on a more diverse set of college guides, as well as a larger and more diverse set of institutions. Finally, the present study is a snapshot in time. The data are limited to NSSE scores and U. S. News rankings from a single year. As Dichev (2001) observed, U. S. News rankings can and do vary substantially from year to year. Averaging scores over several years would likely produce more stable results.

Implications

Despite these limitations, the current research has important implications for theory and practice. First and foremost, this study indicates that the quality of a student's education is not synonymous with the results and reputation of an institution. In fact, educational quality seems to have little to do with resources and reputation. For educational researchers, this study raises important questions about the validity of resource- and reputation-based rankings as indicators of

academic quality. For academic leaders, this study's findings suggest that efforts to garner additional resources and enhance institutional reputation have little effect on the quality of their students' educations. Parents of prospective students should also be aware that institutional resources and reputations may be objective and easily quantified, but they do not automatically translate into good educational experiences for their children.

The relationship between institutional selectivity and the NSSE Enriching Educational Experiences benchmark is intriguing. It is certainly possible that more selective institutions provide a stronger focus on student engagement in the types of activities that are the focus of the Enriching Educational Experiences benchmark (e.g., internships, community service, study abroad, culminating senior experiences, etc.). It is also possible that high-ability students, who are more likely to be admitted to and attend selective institutions, are more likely to have the time and the inclination to participate in a broad range of educationally enriching activities. This finding may also be the result of a spurious correlation. More selective institutions may also be relatively small institutions where students have more opportunities to become involved in educationally purposeful activities. To the extent that the relationship between selectivity and engagement is real, and to the extent that engagement leads to learning and development, the challenge for less selective institutions is to identify strategies that allow and encourage students to become more involved in activities that lead to learning. Given current demands on institutional resources and student time, embedding practical experiences and service components in existing courses may prove to be a cost-effective strategy for improving student engagement and learning.

The relationships between student characteristics and engagement scores that were identified in the second set of analyses also have important implications for theory and practice. For

example, conventional wisdom suggests that minority students attending predominantly white institutions are less likely to be involved than their white counterparts. Consistent with recent findings by Pike, Kuh, and Gonyea (in press), this study found that white males tend to be less involved than other groups. This finding is ironic given that colleges were designed with white males in mind and that the male perspective has tended to dominate the college student development literature (see Heath, 1968; Katz & Associates, 1968; Perry, 1970; Sanford, 1962). Institutions interested in improving engagement on their campuses may want to refocus their attention on the experiences of white males.

The results of this study also underscore the importance of attending school full time. Full-time enrollment has a consistent positive relationship with NSSE benchmarks, probably because full-time students have more opportunities to become engaged in educationally purposeful activities. Institutions with large part-time, nontraditional populations should consider innovative ways to encourage part-time students to remain on campus. This finding also underscores the importance of taking into consideration the characteristics of student populations when comparing institutions. Institutions with large numbers of full-time students are more likely to have high engagement scores, not because of actions by the institution, but because of the characteristics of the students. However, not all student characteristics are controlling. Contrary to conventional wisdom and previous research (see Chickering, 1974; Pike, Kuh, & Gonyea, in press), living on campus was not associated with higher levels of engagement. At least at the institutions included in this study, commuter students were able to overcome the disadvantages of less access to activities that are frequently attributed to living off campus. As a result, commuter students were just as likely as students living in residence halls to engage in collaborative learning and hold positive perceptions of the campus environment.

The findings of the present research also indicate that how students become engaged and how they learn differs by discipline. Similar findings have emerged from other studies (Biglan, 1973; Pike & Killian, 2001; Smart & Hagedorn, 1994), but these findings are too frequently not accounted for in research on college students. This author's research is certainly not immune to this criticism. Here again, institutions interested in improving student engagement on their campuses should carefully consider the academic interests of their students and develop programs that are consistent with the academic demands of the disciplines in which students are enrolled. Likewise, parents and prospective students should be aware that differences in institutional missions can have significant effects on levels and types of engagement. Simple comparisons of which institution is best may beg the question: Best for what?

Conclusion

Given the popularity of college guidebooks and rankings, as well as their apparent importance in the college-choice process, it is very appropriate to ask questions about the validity of the recommendations being made by the popular press. The results of the present research raise serious questions about whether the criteria used by guidebooks and ratings provide appropriate information to parents and prospective students about the academic quality of colleges and universities. Likewise, college and university leaders who strive for high reputation- and resource-based rankings may be short changing their students by focusing their efforts on institutional characteristics that are irrelevant to a high-quality education. In short, the rankings in U. S. News and World Report may truly reflect "America's Best Colleges." The question is, what do we mean by best?

References

- Astin, A. (1977). Four critical years. San Francisco: Jossey-Bass.
- Astin, A. W. (1993). What matters in college? Four critical years revisited. San Francisco: Jossey-Bass.
- Baird, L. (1976). Using self-reports to predict student performance. New York: The College Board.
- Berdie, R. (1971). Self-claimed and tested knowledge. Educational and Psychological Measurement, 31, 629-636.
- Biglan, A. (1973). The characteristics of subject matter in different academic areas. Journal of Applied Psychology, 57, 195-203.
- Bryk, A. S., & Raudenbush, S. W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.
- Chickering, A. (1974). Commuting versus residential students: Overcoming educational inequities of living off campus. San Francisco: Jossey-Bass.
- Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. AAHE Bulletin, 7(3): 3-7.
- Dichev, I. (2001). News or noise? Estimating the noise in the U. S. News university rankings. Research in Higher Education, 42, 237-266.
- Ethington, C. A. (1997). A hierarchical linear modeling approach to studying college effects. In J. Smart (Ed.), Higher education: Handbook of theory and research (vol. 12, pp. 165-194). New York: Agathon.
- Feldman, K. A., & Newcomb, T. (1969). The impact of college on students. San Francisco: Jossey-Bass.

Heath, D. (1968). Growing up in college. San Francisco; Jossey-Bass.

Hossler, D. (2000). The problem with college rankings. About Campus, 5(1): 20-24.

Hossler, D., & Foley, E. M. (1995). Reducing the noise in the college choice process: The use of college guidebooks and ratings. In R. D. Walleri & M. K. Moss (Eds.), Evaluating and responding to college guidebooks and rankings (New Directions for Institutional Research Series No. 88, pp. 21-30). San Francisco: Jossey-Bass.

Hossler, D., & Litten, L. H. (1993). Mapping the higher education landscape. New York: College Entrance Examination Board.

Hunter, B. (1995). College guidebooks: Background and development. In R. D. Walleri & M. K. Moss (Eds.), Evaluating and responding to college guidebooks and rankings (New Directions for Institutional Research Series No. 88, pp. 5-12). San Francisco: Jossey-Bass.

Katz, J. & Associates. (1968). No time for youth: Growth and constraint in college students. San Francisco: Jossey-Bass.

Kuh, G. D. (2001a). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. Change, 33(3): 10-17.

Kuh, G. D. (2001b). The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties. Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Kuh, G. D., Hayek, J. C., Carini, R. M., Ouimet, J. A., Gonyea, R. M., & Kennedy, J. (2001). NSSE technical and norms report. Bloomington, IN: Indiana University Center for Postsecondary Research and Planning.

Kuh, G. D., Pace, C. R., & Vesper, N. (1997). The development of process indicators to estimate student gains associated with good practices in undergraduate education. Research in Higher Education, 38, 435-454.

McDonough, P. M., Antonio, A. L., Walpole, M. B., & Perez, L. X. (1998). College rankings: Democratized knowledge for whom? Research in Higher Education, 39, 513-538.

Pace, C. R. (1985). The credibility of student self-reports. Los Angeles: Center for the Study of Evaluation, University of California Los Angeles.

Pascarella, E. T. (2001). Identifying excellence in undergraduate education: Are we even close? Change, 33(3): 18-23.

Pascarella, E., & Terenzini, P. (1991). How college affects students: Findings and insights from twenty years of research. San Francisco: Jossey-Bass.

Pike, G. R. (1995). The relationship between self reports of college experiences and achievement test scores. Research in Higher Education, 36, 1-21.

Pike, G., & Killian, T. (2001). Reported gains in student learning: Do academic disciplines make a difference? Research in Higher Education, 42, 429-454.

Pike, G. R., Kuh, G. D., & Gonyea, R. M. (in press). The relationship between institutional mission and students' involvement and educational outcomes. Research in Higher Education.

Pohlmann, J., & Beggs, D. (1974). A study of the validity of self-reported measures of academic growth. Journal of Educational Measurement, 11: 115-119.

Pollock, C. R. (1992) College guidebooks—users beware. Journal of College Admissions, 135, 21-28.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. (2001). HLM5: Hierarchical linear and nonlinear modeling (2nd ed.). Chicago: Scientific Software International.

Sanford, N. (Ed.) (1962). The American College. A psychological and social interpretation of the higher learning. New York: Wiley.

Smart, J. C., & Hagedorn, L. S. (1994). Enhancing professional competencies in graduate education. Review of Higher Education, 17, 241-258.

Stecklow, S. (1995, April 5). Colleges inflate SATs and graduation rates in popular guidebooks. Wall Street Journal, pp. A1, A4, A8.

Table 1:

Variance Components for the Baseline Models

NSSE Benchmark	Variance Component	Total Variance	R ²
Level of Academic Challenge	6.46***	215.05	0.03
Active and Collaborative Learning	3.29***	242.61	0.01
Student Interaction with Faculty Members	1.49*	321.60	0.00
Enriching Educational Experiences	16.80***	270.99	0.06
Supportive Campus Environment	4.78***	316.84	0.02

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 2:

Effects of Student Background Characteristics on NSSE Benchmark Scores

Background Characteristic	LAC	ACL	SIF	EEE	SCE
Intercept	54.39***	46.13***	35.54***	47.93***	49.95***
Gender (Female)	1.98**	0.66	0.14	2.10**	-1.61
Ethnic (Minority)	1.68	0.49	0.88	3.55***	0.10
Age (24 and over)	1.09	0.82	-0.48	-2.95**	0.97
Native Student	-0.57	1.57	0.27	2.26**	1.29
Full-Time Student	5.73***	3.98**	2.78*	4.11***	2.02
Greek Student	0.65	1.63	2.26	2.71**	1.55
Live On Campus	-1.36	-1.42	-0.87	1.90*	0.66
Math and Science	0.87	3.58*	-3.13	0.43	0.93
Pre-Professional	-1.88	7.59***	-4.55*	-0.08	4.49*
Social Science	-1.73	0.29	-3.43	0.22	0.94
Multiple Majors	0.32	5.25***	-1.40	3.26	1.05
R ²	0.08	0.06	0.03	0.13	0.05

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

LAC = Level of Academic Challenge; ACL = Active and Collaborative Learning; SIF = Student Interaction with Faculty Members; EEE = Enriching Educational Experiences; SCE = Supportive Campus Environment

Table 3:

Significance Tests for the Level-1 Intercepts

NSSE Benchmark	Variance	
	Component	Reliability
Level of Academic Challenge	6.25***	0.79
Active and Collaborative Learning	3.66***	0.67
Student Interaction with Faculty Members	1.02	0.31
Enriching Educational Experiences	3.34***	0.85
Supportive Campus Environment	4.70***	0.66

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 4:

Relationships Between U. S. News Rankings and Adjusted NSSE Benchmark Scores (Level-1 Intercepts)

U. S. News Ranking	LAC	ACL	SIF	EEE	SCE
Academic Reputation	-0.05	0.00	---	-0.04	-0.02
Graduation/Retention	-0.03	-0.03	---	-0.01	-0.02
Faculty Resources	-0.01	0.00	---	0.00	0.03
Student Selectivity	-0.01	0.00	---	-0.08*	0.00
Financial Resources	-0.02	0.02	---	0.03	-0.02
Alumni Giving	0.01	0.01	---	0.02	-0.01

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

LAC = Level of Academic Challenge; ACL = Active and Collaborative Learning; SIF = Student Interaction with Faculty Members; EEE = Enriching Educational Experiences; SCE = Supportive Campus Environment