

STUDENT ENGAGEMENT AND STUDENT LEARNING: Testing the Linkages*

Robert M. Carini,**,‡ George D. Kuh,*** and Stephen P. Kleint†

.....

This study examines (1) the extent to which student engagement is associated with experimental and traditional measures of academic performance, (2) whether the relationships between engagement and academic performance are conditional, and (3) whether institutions differ in terms of their ability to convert student engagement into academic performance. The sample consisted of 1058 students at 14 four-year colleges and universities that completed several instruments during 2002. Many measures of student engagement were linked positively with such desirable learning outcomes as critical thinking and grades, although most of the relationships were weak in strength. The results suggest that the lowest-ability students benefit more from engagement than classmates, first-year students and seniors convert different forms of engagement into academic achievement, and certain institutions more effectively convert student engagement into higher performance on critical thinking tests.

.....

KEY WORDS: student engagement; critical thinking; value added; NSSE; student learning.

BACKGROUND

Various approaches have been used to estimate collegiate quality. For instance, some analyses focus on quantitative characteristics of institutions and programs, such as graduation rates, minority access, and faculty characteristics. Other approaches emphasize outcomes such as performance on general education and major field tests, expert judgments by accreditors (which are usually based usually on inputs and resources), and reputational ratings such as those published by *US News*

*A version of this paper was presented at the annual meeting of the American Educational Research Association, San Diego, April 2004.

**Department of Sociology, University of Louisville, Louisville, Kentucky, USA.

***Center for Postsecondary Research, Indiana University, Bloomington, Indiana, USA.

†Senior Research Scientist, The RAND Corporation, Santa Monica, California, USA.

‡Address correspondence to: Robert Carini, Department of Sociology, University of Louisville, 103 Lutz Hall, Louisville, KY 40292, USA. E-mail: rmcari01@louisville.edu

and World Report. Still others employ surveys of student experiences and self-reports of their learning and personal development. Each method has advantages and limitations as well as champions and critics (Chun, 2002; Klein, 2001, 2002; Kuh and Pascarella, 2004; Pascarella, 2001).

Comprehensive assessment systems typically use multiple indicators to tap quality (Astin, 1991; Ewell, 1984, 1988; Gentemann, Fletcher, and Potter, 1994; Halpern, 1987; Jacobi, Astin, and Ayala, 1987; Terenzini, 1989; Vandament, 1987). Student surveys are particularly popular because they are generally less expensive and easier to develop and administer than tests of student knowledge and proficiencies. When well-crafted, student surveys can provide insights into the student experience that other sources of information cannot, such as estimates of one's ability to interact effectively with others on an individual basis or in small groups, and the degree to which one's values and ethics have developed since starting college (Kuh, 2001, 2003; Pace, 1984).

A fair amount of evidence indicates that student self-reports are valid and reliable under certain conditions (Baird, 1976; Pace, 1984; Pohlmann, 1974): (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 thoughtful response, (5) the information requested is potentially verifiable, and (6) the question asks for information that is known to those answering the questions and does not threaten, embarrass, or violate their privacy or encourage the respondent to respond in socially desirable ways. However, even when surveys are designed to meet these conditions, as with the National Survey of Student Engagement (NSSE) (Kuh et al., 2001), it does not necessarily mean that the surveys ask about behaviors that are linked to desired outcomes, such as whether institutions that engage students at higher levels are adding value to the student experience.

Student engagement is generally considered to be among the better predictors of learning and personal development. The premise is deceptively simple, perhaps self-evident: The more students study or practice a subject, the more they tend to learn about it. Likewise, the more students practice and get feedback on their writing, analyzing, or problem solving, the more adept they should become (Kuh, 2003). The very act of being engaged also adds to the foundation of skills and dispositions that is essential to live a productive and satisfying life after college. That is, students who are involved in educationally productive activities in college are developing habits of the mind and heart that enlarge their capacity for continuous learning and personal development

(Shulman, 2002). The widespread use of the NSSE survey—more than 850 different four-year colleges have used it at least once since 2000—underscores the need to understand the degree to which—and the processes whereby student engagement contributes to more favorable outcomes for college students.

In one recent study, using data from six South Dakota public colleges and universities, the National Center for Higher Education Management Systems (NCHEMS) found some positive links between student engagement and ACT CAAP scores, though the strength of the statistically significant associations were modest at best (Ewell, 2002). However, the NCHEMS researchers emphasized that the measures of student engagement and the CAAP test scores were obtained at different points in time, in many instances a year or more separated the two measurement points. Other studies are more promising, such as those by Hughes and Pace (2003) and by institutional researchers at individual colleges and universities (e.g., Indiana University Bloomington, St. Xavier University, University of Montana, and the University of Wisconsin-Green Bay) who report positive relationships between NSSE results and persistence, as well as academic performance represented by college grade point average (GPA). Even so, questions remain about whether collegiate grades and tests such as the ACT CAAP are adequate measures of how much students learn, and to what degree the institution or general intellectual ability shapes GPA.

To address these questions, RAND and the Council for the Aid to Education are developing, testing and refining new measures of college-level learning. The goal is to provide richer indicators of important dimensions of student learning than are tapped by existing instruments (Klein, 2002; Klein et al., 2005). By adjusting for student ability prior to matriculation (such as through SAT scores and background characteristics), it is anticipated that the new RAND measures will be more sensitive to program and institutional effects than traditional tests of general educational abilities (Bohr et al., 1994; Pascarella and Terenzini, 1991; Pascarella, Bohr, Nohra, and Terenzini, 1994; Winter, McClelland, and Stewart, 1981), thereby promising value-added estimates of institutional performance (Benjamin and Hersh, 2002).

PURPOSES

This study examines the forms of student engagement associated with learning as measured by the RAND tests, the new essay prompts on the Graduate Record Examination (GRE), and college GPA. More specifically, we sought to determine:

- (1) the extent to which student engagement is related to experimental and traditional measures of academic performance;
- (2) the forms of student engagement that are most closely related to student performance;
- (3) whether the relationships between student engagement and academic performance are “conditional,” or vary depending on student characteristics; and
- (4) whether certain four-year colleges and universities are more effective than others in converting student engagement into stronger academic performance.

ANALYTIC APPROACH

Data Sources

The NSSE survey instrument, *The College Student Report* (hereafter, *The Report*), is composed of about 70 items that assess the extent to which students devote time and energy to educationally purposeful activities. A large majority of items from *The Report* deal with behaviors that have been linked empirically to favorable outcomes of college in prior studies (Kuh, 2001, 2003). The psychometric properties of the NSSE survey are reported in detail by Kuh et al. (2001) and at http://www.iub.edu/~nsse/html/psychometric_framework_2002.htm. NSSE results have been treated by some postsecondary institutions as a proxy for student learning, although the survey was designed primarily to provide process indicators, or measures that could help colleges and universities identify areas of student performance and institutional practices that could be improved to enhance the overall quality of the student experience (Ewell and Jones, 1996).

The measures developed by RAND consist of a series of newly developed cognitive and performance tests (Klein et al., 2005). Specifically, the RAND tests consist of various combinations of six different 90-minute critical thinking and performance problems. The two performance tests were modeled after the performance section of the bar exam (Klein, 1996). The four critical thinking tests were derived from “Tasks in Critical Thinking” designed by the New Jersey Department of Education (Ewell, 1994). Subject areas for the critical thinking tasks included science, social science, and arts and humanities. In addition, RAND researchers administered two essay prompts from the GRE (Klein et al., 2005). The 45-minute make-an-argument prompt asked students to take and argue a position on a topic. The 30-minute break-an-argument prompt asked

students to critically analyze a given position on an issue. The RAND and GRE tasks were presented to students in open-ended formats. In addition, RAND researchers were able to obtain cumulative GPA and total SAT scores for most students in the 2002 field test.

Sample and Administration

RAND researchers administered *The Report* and their own cognitive tests to 1352 students at 14 four-year colleges and universities during the fall and spring of 2002. Students from all four undergraduate class levels were recruited and paid \$20–25 per hour of testing time as an incentive. SAT scores were not available for 149 students, paring our analytic sample to 1203 cases. Missing student responses on other control variables reduced the analytic sample further to 1058. Of the 1058 students with complete data, 940 completed the RAND measures and 506 completed the GRE tests. The NSSE instrument was administered in a non-standard method. Instead of exclusively designating a portion of time to administer *The Report*, students were directed to begin the survey if they completed their open-ended written tasks before time was called. They were then given additional time to complete the survey. It is not known what effect, if any, this administration procedure may have had on student responses.

Respondent Characteristics

The colleges that were invited to participate represented an intentional mix of institutional characteristics—public/private, urban/nonurban location, admissions selectivity, region of the country, large/small enrollments, and varied student characteristics such as race/ethnicity. Students at participating institutions self-selected themselves by responding to their university's invitations, such as announcements in student or local newspapers. The four class levels were almost evenly represented (i.e., the sample consisted of 28, 25, 23, and 23% first-year students through seniors, respectively). Approximately 98% of participants were enrolled full-time, and about 57% were women. About 73% classified themselves as exclusively White, while 9% were Black, 7% were Asian, 3% were Latina/o, and another 7% were multiracial. The percentage of Black respondents was bolstered by the participation of an HBCU institution. Nearly 5% of respondents considered themselves international students, and 8% had begun college at an institution other than their current one. Full-time students and international students were overrepresented compared with the universe of college students (NSSE, 2003),

while transfer students were underrepresented. Ten of 14 colleges (71%) were private as compared with 63% of U.S. colleges (NSSE, 2004). To the extent that it was possible to do so with 14 colleges, the sample roughly approximated U.S. colleges with respect to region of the country, Carnegie Classification, enrollment, and urbanicity (NSSE, 2004). Klein et al. (2005) provide further details on the characteristics of participating colleges.

Measures

Student learning was assessed by academic performance as measured by the RAND and GRE test scores, and college-reported GPA. A set of engagement and self-reported outcome measures was selected from the NSSE survey for the analysis. Five of these measures are what NSSE publications have termed “clusters of effective educational practice”—each made up of 6–11 related items. The five clusters are: level of academic challenge, active and collaborative learning, student–faculty interaction, enriching educational experiences, and supportive campus environment (Kuh, 2001). A number of other scales more focused on particular engagement areas were also employed: reading and writing, quality of relationships, institutional emphases on good practices, higher-order thinking, student–faculty interaction concerning coursework, and integration of diversity into coursework. Appendix A.1 lists specific survey items that contributed to each scale and provides reliability coefficients. In some instances, there is overlap between items in these scales and the five clusters of effective educational practice (Appendix A.2). We also explored relationships between academic performance measures and self-reported gains associated with attending college in three areas: (1) general education, (2) personal and social development, and (3) practical competence. Finally, we probed the possible link between student engagement and satisfaction with their collegiate experience at their present institution. Descriptions and summary statistics for measures of academic performance, student engagement, and self-reported outcomes are presented in Table 1.

Analytic Strategy

RAND and GRE test scores were standardized by converting each to an SAT metric. The standardization was performed by converting the RAND and GRE scores to distributions that had the same mean and standard deviation as the students’ SAT scores. To facilitate direct comparisons of GPA across students from different institutions (and to

TABLE 1. Means, Standard Deviations and Descriptions for Academic Performance, Student Engagement Scales, and Self-reported Outcomes

Measure	Description	Metric	Mean	SD ^a	N
<i>Academic Performance</i>					
RAND Test Score	Composite performance and critical thinking score	Standardized to SAT distribution	1186.13	177.55	940
GRE Test Score	Composite GRE prompt score	Standardized to SAT distribution	1120.93	157.77	506
GPA	Cumulative college grade point average	Z-scored to within school distribution	.02	.98	1004
<i>Student Engagement</i>					
Level of academic challenge	Nature and amount of academic work performed	Sum of 11 items	18.98	4.33	1034
Active and collaborative learning	Frequency of class participation and collaborative learning	Sum of 7 items	10.51	3.35	1045
Student–faculty interaction	Frequency of student interactions with faculty	Sum of 6 items	8.43	3.85	1043
Enriching educational experiences	Degree if participation in educationally fruitful activities	Sum of 11 items	6.55	1.75	1023
Supportive campus climate	Degree to which the institution is perceived to be supportive	Sum of 6 items	11.12	3.08	1049
Reading and writing	Amount of assigned reading and writing	Sum of 3 z-scored items	-.001	2.25	1052
Quality of relationships	Quality of your relationships with people at your institution	Sum of 3 items	16.21	2.84	1054
Institutional emphases on good practices	Degree institution perceived to uphold good practices	Sum of 5 items	13.15	3.32	1047

TABLE 1. (Continued)

Measure	Description	Metric	Mean	SD ^a	N
Higher-order thinking	Frequency higher-order skills included in coursework	Sum of 4 items	12.43	2.51	1052
Student-faculty interaction concerning coursework	Degree of interaction involving coursework	Sum of 3 items	7.94	1.97	1051
Integration of diversity into coursework	Degree coursework involved diverse perspectives and ideas	Sum of 5 z-scored items	-0.01	3.39	1046
<i>Self-reported Outcomes</i>					
Gains: general education	Self-reported gains in writing, speaking, and thinking critically	Sum of 4 items	8.42	2.50	1053
Gains: personal and social	Self-reported gains related to personal and community issues	Sum of 7 items	12.02	4.38	1048
Gains: practical competence	Self-reported gains related to job-related skills	Sum of 4 items	7.53	2.53	1051
Satisfaction	Degree satisfied with institution	Sum of 2 items	6.50	1.35	1053

^aSD = Standard deviation.

adjust for possible differences in grading standards across institutions), the GPAs within a school were converted to a score distribution that had the same mean and standard deviation as the SAT scores at that school. Once RAND, GRE, and GPA measures were standardized to the SAT scale, we tested the relationships between various student engagement measures and academic performance. We also juxtaposed self-reported gains and satisfaction from the NSSE—what some consider to be meaningful outcomes in their own right (Shulman, 2002)—with measures of student engagement.

We attempted to tease out conditional effects of student engagement on performance by examining these effects for students of different classes, ability levels (as measured by the SAT), sexes, races/ethnicities, fraternity or sorority affiliation, and transfer status, any of which might arguably affect student learning. Searching for conditional effects is important, because it is likely that the impact of college on student learning and personal development will be underestimated if conditional effects are neglected (Pascarella and Terenzini, 1991, 2005).

In an attempt to parse out precollege academic level, we statistically controlled for students' total SAT score in most analyses. We also controlled for a host of student demographic variables and collegiate experiences that could reasonably influence learning as well as student engagement: class, gender, residence, enrollment status, race/ethnicity, number of parents with a bachelor's degree or higher, major field, amount of unassigned reading, commute time, time spent caring for dependents, and hours worked off campus. We did this to more accurately isolate possible effects of student engagement on "learning" as represented by the RAND, GRE, and GPA measures.

The literature on student learning is replete with concerns about regression-based analyses to assess change in an outcome over time (Baird, 1988; Cronbach and Furby, 1970; Pascarella and Wolniak, 2004; Pike 1992, 2004), especially attempts to attribute changes in a given outcome to a particular experience such as participation in a program or involvement with an organization. In particular, standard approaches to measure change such as gain scores, residual scores, and repeated measures are notoriously unreliable (Baird, 1988; Pike 1992). Many of these so-called "value added" studies rely on the measurement of the outcome at more than one point in time, a minimum of a "pretest" and a "posttest." With the exception of total SAT score, our data are cross-sectional in nature, in that initial "pretest" measures of RAND, GRE, or GPA at college entrance were not available. In that sense, ours is not a study of changes in student performance over time. However, because we can use total SAT as a control variable, and because SAT is

correlated at .48 and .55 with RAND and GRE scores, respectively, we are able to explain substantial portions of the variance on RAND and GRE scores with the SAT score, rendering the unexplained portion to potentially be attributed to the net effect of student engagement. Following the lead of Pike (2004), the central aim of our study is whether students with identical SAT scores (and other student characteristics) but different engagement levels exhibit different learning outcomes. Finally, despite statistically controlling for many student characteristics, this effort—as with all non-experimental research—cannot eliminate the possibility that an unmeasured variable may confound the patterns reported here.

An earlier study using these data reported that students at certain institutions had higher RAND and GRE scores than would be expected given the typical relationship of these scores with their SAT scores (Klein et al., 2005). This finding suggests that, relative to other participating institutions, certain colleges added more value than others in our sample. Within this study, we consider the possibility that student engagement is a key process by which institutions add value for their students. Indeed, institutions may differ in their learning productivity, or how effectively they convert student engagement into learning outcomes (Kuh and Hu, 2001). We examined each institution's partial correlations between student engagement measures and academic performance to determine whether students at particular institutions realized more or less academic payoff from their engagement in effective educational practices. We ranked institutions according to the strength of the associations between engagement and learning, categorizing them as high, mid-range, or low in terms of learning productivity. We then regressed each performance measure on each engagement scale, the institution's learning productivity score (converting engagement into learning outcomes), and an interaction term between engagement and learning productivity.

One drawback to using the student as the unit of analysis when estimating institutional differences is that standard errors will be biased too small, increasing the chance that coefficients will be statistically significant (Raudenbush and Bryk, 2002). Thus, we relied on effect sizes to gauge whether institutional differences on fostering favorable engagement–performance linkages were substantively important. Should comparable data become available involving a greater number of institutions, future researchers should investigate institutional effects with multilevel modeling techniques.

RESULTS

Relationships among RAND, GRE, GPA, and SAT

RAND and GRE scores correlated with total SAT at .48 and .55, respectively, indicating a healthy relationship between where a student falls in the SAT distribution and where they fall in the RAND and GRE distributions. Among first-year students, RAND and GRE scores correlated with SAT scores at .52 and .54, respectively. Among seniors, RAND and GRE scores correlated with SAT scores at .38 and .46, respectively. The drops in both RAND-SAT (.52 to .38) and GRE-SAT (.54 to .46) correlations between the first and senior years are consistent with the hypothesis that institutions may influence student learning in the areas covered by these tests. A curvilinear specification between RAND and SAT (but not GRE and SAT or GPA and SAT) yielded a better-fitting model than a linear one. Specifically, higher SAT scores were related to higher RAND scores only up to an SAT score of 1434—at which point higher SAT scores yielded slightly declining RAND scores. However, this inflection point is very high—only about 7% of the students tested had scores above 1434. Finally, RAND and GRE scores correlated at .28 and .27, respectively with GPA. Controlling for SAT resulted in partial correlations of .18 and .15 for RAND-GPA and GRE-GPA, respectively.

Student Engagement Scales and Academic Performance

Table 2 presents bivariate and partial correlations between student engagement scales, RAND, GRE, and GPA measures, and self-reported outcomes. As shown in Table 2, we found a number of small but statistically significant positive correlations between student engagement and scores on the RAND and GRE tests, both before and after controls were added for a host of student characteristics. Level of academic challenge, supportive campus climate, reading and writing, quality of relationships, institutional emphases on good practices, and general education gains yielded positive partial correlations with RAND of .10 or greater (Column 2). None of the 15 partial correlations with RAND scores was negative. RAND scores had somewhat higher partial correlations with engagement and self-reported outcomes (Column 4) than did GRE scores, but there were no statistically significant negative partial correlations with GRE scores. Reading and writing and gains in practical competence were correlated with GRE scores at .16 and .13, respectively.

TABLE 2. Correlations Between Student Engagement Scales, RAND, GRE, and GPA Measures, and Self-reported Outcomes

Scale	RAND		GRE		GPA	
	bivariate correlation	partial correlation	bivariate correlation	partial correlation	bivariate correlation	partial correlation
<i>Student Engagement</i>						
Level of academic challenge	.11***	.10**	.10*	.09*	.09**	.07*
Active and collaborative learning	.02	.02	.06	.08*	.13***	.13***
Student-faculty interaction	.03	.01	.06	.04	.13***	.13***
Enriching educational experiences	.09**	.02	.13**	.07	.10**	.04
Supportive campus climate	.08**	.13***	-.03	-.02	.06*	.08**
Reading and writing	.12***	.10***	.16***	.16***	.08**	.06*
Quality of relationships	.14***	.14***	.01	-.02	.08**	.08**
Institutional emphases on good practices	.03	.10**	-.06	-.02	.04	.07*
Higher-order thinking	.08*	.06*	.08*	.07	.06*	.03
Student-faculty interaction concerning coursework	.03	.02	.02	.01	.11***	.10***
Integration of diversity into coursework	.10***	.06*	.12**	.09*	.11***	.07*
<i>Self-reported Outcomes</i>						
Gains: general education	.09**	.10***	.01	.01	.11***	.12***
Gains: personal and social	-.02	.04	-.09*	-.04	.10***	.11***
Gains: practical competence	-.02	.03	-.15***	.13**	-.01	.04
Satisfaction	.14***	.08*	.07	-.04	.09**	.05

* $p < .05$, ** $p < .01$, *** $p < .001$ 1-tailed test.

Note: Controls for partial correlations included total SAT score, squared SAT (RAND only), class, gender, residence, enrollment status, race/ethnicity, number of parents with a bachelor's or higher, major field, amount of unassigned reading, commute time, time spent caring for dependents, and hours worked off campus.

Student Engagement and GPA

Columns 5 and 6 of Table 2 show that levels of student engagement were often positively related to GPA. Very modest but statistically significant positive partial correlations (Column 6) were found for 9 of the 11 engagement scales: level of academic challenge (.07), active and collaborative learning (.13), student–faculty interaction (.13), supportive campus climate (.08), reading and writing (.06), quality of relationships (.08), institutional emphases on good practices (.07), student–faculty interaction concerning coursework (.10), and integration of diversity into coursework (.07). In addition, significant positive correlations were noted between self-reported outcomes and GPA: general education gains (.12) and personal-social gains (.11). *None* of the 15 scales examined were negatively correlated with GPA. These relationships are almost certainly understated due to the lagged nature of cumulative rather than current semester GPA. Overall, the relationships between the NSSE scales, GPA, RAND, and GRE scores were comparable in strength.

In a separate analysis not shown here in tabular format, models that regressed RAND, GRE, and GPA on student characteristics, including total SAT, explained about 32% of the variance in RAND, 37% of the variance in GRE, and 13% of the variance in GPA. Using the results from these three student-level OLS regression models, we computed residuals for each student. These residual scores represented the amount of student over- or underperformance relative to predicted scores, with negative residuals denoting that a student underperformed relative to similar counterparts in the sample. We then entered the residuals for RAND, GRE, and GPA as dependent variables, and inserted a block of 11 student engagement measures (see Table 1) as predictors into each of the three equations. These 11 engagement measures explained 2.9, 1.3, and 3.1% of the variance in the residuals for RAND, GRE, and GPA, respectively. When entered into the model separately, the most powerful predictors of RAND residuals were supportive campus environment (1.8%), quality of relationships (1.8%), institutional emphases on good practices (1.0%), and reading and writing (.8%).

Student Engagement Items and RAND

To probe even further into the specific forms of engagement associated with higher academic performance, we computed partial correlations between RAND and individual items on the NSSE survey for first-year students and seniors. Many of these items were included in the 15 scales presented in Table 1, but others were not. Table 3 reports only those

TABLE 3. Selected Partial Correlations between Student Engagement Items, RAND Scores, and Self-reported Outcomes, First-year Students and Seniors

Item	Partial Correlation (First-year Students)	Partial Correlation (Seniors)
<i>Student Engagement</i>		
Prepared two or more drafts of a paper or assignment before turning it in	.11*	-.01
Came to class having completed readings and assignments	.16**	.01
Worked with other students on projects during class	-.01	.18*
Put together ideas or concepts from different courses when completing assignments or during class discussions	.09	.20*
Received prompt feedback from faculty	.11*	.19***
Worked harder than you thought you could to meet instructors' expectations	.14*	-.08
Discussed coursework outside of class	.14*	.12
Had serious conversations with students who were very different from you (religious beliefs, political opinions, or values)	-.14*	.07
Number of papers of fewer than 5 pages	.17**	-.04
Quality of academic advising received at your institution	.04	.18*
Quality of relationships with faculty	.16**	.14
Quality of relationships with administrative personnel and offices	.16**	.03
Extent to which your institution emphasized providing support that you need to succeed academically	.12*	.13
Extent to which your institution emphasized encouraging contact among students of different backgrounds	-.07	.16*
Extent to which your institution emphasized attendance of campus events and activities	-.01	.16*
<i>Self-reported Outcomes</i>		
Gains at this institution with respect to a general education	.15*	.03
Gains at this institution with respect to writing clearly and effectively	.12**	.07

TABLE 3. (Continued)

Item	Partial Correlation (First-year Students)	Partial Correlation (Seniors)
Gains at this institution with respect to thinking critically and analytically	.14*	.11
Gains at this institution with respect to understanding people of other racial and ethnic backgrounds	-.15*	-.04

* $p < .05$, ** $p < .01$, *** $p < .001$ 1-tailed test.

Note: Controls included total SAT score, squared SAT, gender, residence, enrollment status, race/ethnicity, number of parents with a bachelor's degree or higher, major field, amount of unassigned reading, commute time, time spent caring for dependents, and hours worked off campus.

partial correlations that achieved statistical significance for one or both classes.¹ The largest positive correlations for first-year students (Column 1) were for the number of papers of fewer than 5 pages (.17), came to class having completed readings and assignments (.16), quality of relationships with faculty (.16) and administrative personnel and offices (.16), and worked harder than you thought you could to meet instructors' expectations (.14). The only significant and negative correlation between engagement and RAND for first-year students involved having serious conversations with students who were very different from you (-.14). While correlations between self-reported outcomes and RAND were generally positive, gains with respect to understanding people of other racial and ethnic backgrounds were negatively related (-.15). Correlations for seniors (Column 2) yielded several differences. In particular, seniors appeared to benefit less than first-year students from working harder, coming to class more prepared, preparing drafts of papers, writing more small papers, and having high quality relationships with collegiate employees. In contrast, seniors benefited more from working with other students on projects during class, integrating ideas from different courses, receiving high quality academic advising, and being at institutions that emphasize contact among students of different backgrounds, as well as attendance of campus events and activities.

Conditional Effects of Pre-college Ability

We examined the linkages between student engagement, RAND, GRE, and GPA measures, and self-reported outcomes for those who scored in the lowest and highest quintiles of the SAT distribution. The

lowest quintile included those who scored below 1030 on the SAT, while the highest quintile consisted of those who scored greater than 1330. Table 4 presents partial correlations between student engagement, academic performance, and self-reported outcomes for both the low ability and high ability groups—those who scored in the lowest quintile and highest quintile on the SAT, respectively. The results suggest that low ability students benefited more from engagement than high ability counterparts, particularly in terms of their RAND and GRE scores, and to a lesser extent, their GPA.

Column 1 displays correlations between student engagement and RAND scores for low ability students. Low ability students appeared to receive the greatest payoffs from quality of relationships (.26), a supportive campus climate (.23), integration of diversity into coursework (.20), student–faculty interaction concerning coursework (.18), and reading and writing (.18). In addition, gains in general education (.19) were significantly correlated with RAND scores for this group. In contrast, high ability students benefited considerably less than their low ability counterparts in that none of the 15 correlations presented in Column 2 were significant and positive, and the majority were negative. This pattern generally persists for GRE scores in Columns 3 and 4, yet the differences between low and high ability students diminishes somewhat in Columns 5 and 6 for GPA. Taken as a whole, these findings suggest that institutional interventions to boost student engagement may have the greatest payoff for those most at-risk for leaving college prematurely.

One interpretation of the different payoffs on RAND tests for students of different abilities might be that the RAND tests tap general intellectual ability to a large degree. Highly intelligent students may simply require less effort to achieve similar academic results. This might explain the large number of modestly negative correlations for high ability students. We find some evidence to support the notion that high ability students need to expend less effort in college to do well on the RAND tests. Although the correlation between scores on the RAND tasks and the amount of reading and writing students do in college was the strongest of the 15 correlations reported for high ability students (.09), it was still less than that for low ability students (.18). In supplementary analyses not shown in Table 4, we found that low ability students had a .17 correlation between total time spent preparing/studying for class and RAND score, while high ability students sported a correlation of only .01. Similarly, working harder than you thought you could to meet instructors' expectations was correlated at .13 ($p = .09$) with RAND score for low ability students, and only at .01 for high ability students.

TABLE 4. Partial Correlations between Student Engagement Scales, RAND, GRE, and GPA Measures, and Self-reported Outcomes for the Lowest and Highest Ability Students

Scale	Partial Correlation for Lowest Ability RAND	Partial Correlation for Highest Ability RAND	Partial Correlation for Lowest Ability GRE	Partial Correlation for Highest Ability GRE	Partial Correlation for Lowest Ability GPA	Partial Correlation for Highest Ability GPA
<i>Student Engagement</i>						
Level of academic challenge	.15	.07	.24**	-.04	.18*	.13
Active and collaborative learning	.17*	-.10	.10	.05	.17*	.12
Student-faculty interaction	.14	-.08	.04	-.02	.20**	.13
Enriching educational experiences	.12	-.04	.02	.17	.03	.11
Supportive campus climate	.23**	-.01	.09	-.02	.09	.04
Reading and writing	.18*	.09	.25**	-.06	.13	-.001
Quality of relationships	.26***	.01	.065	-.09	.12	.08
Institutional emphases on good practices	.14	-.002	.05	.02	.03	.09
Higher-order thinking	.05	.03	.22*	.002	.16*	.002
Student-faculty interaction concerning coursework	.18*	-.03	.06	-.18	.21**	.08
Integration of diversity into coursework	.20*	-.09	.05	.12	.24**	.13
<i>Self-reported Outcomes</i>						
Gains: general education	.19*	.01	.11	-.06	.14	.09
Gains: personal and social	.13	-.03	.07	-.08	.13	.12
Gains: practical competence	.11	-.11	-.002	-.19	.19**	-.02
Satisfaction	.12	.002	.07	.05	.20**	.09

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed).

Note: Controls included class, gender, residence, enrollment status, race/ethnicity, number of parents with a bachelor's degree or higher, major field, amount of unassigned reading, commute time, time spent caring for dependents, and hours worked off campus.

Of particular interest is that the low ability group appears to benefit disproportionately from perceptions of a nurturing environment, such as a supportive campus climate and high quality relationships. Because this finding is consistent with the research on student retention (Pascarella and Terenzini, 1991, 2005; Tinto, 1993), perhaps the apparent conditional effect of student ability is simply an artifact of selection processes, in that many alienated low ability students had already left institutions by this time. However, our data do not support the differential selection hypothesis. Restricting the analysis to first-year students of low ability, before the greatest risk of drop-out has occurred, continued to yield positive correlations with scores on the RAND tasks: supportive campus environment (.23) and quality of relationships (.10).

Other Conditional Effects

Earlier we reported some differential effects for first-year students and seniors (Table 3). We also tested whether engagement led to differential effects on RAND tests for women and men, whites and nonwhites, native and transfer students, and students with Greek organizational affiliations. A series of multivariate regression analyses did not reveal conditional effects for any of these measures except for level of academic challenge, which favored women. Further, we found little evidence of conditional effects by gender *within* particular institutions.

Institutional Differences

Klein et al. (2005) reported that students at some institutions had higher RAND and GRE scores than would be expected given the typical relationship of these scores with their SAT scores. This finding suggests that, relative to other participating institutions, certain colleges added more value than others. A final aim of our study was to determine if different institutions more effectively converted student engagement into academic performance, as measured by RAND tests. To this end, we examined the partial correlations within each institution, controlling for a curvilinear specification of SAT. Three institutions (a liberal arts college, a general liberal arts college, and a HBCU) had a large number of substantial positive associations between engagement and RAND performance, while the majority had few or none. Three institutions exhibited many negative relationships (two master's level universities and one doctoral extensive university). To determine whether these observed within-institution differences were statistically significant, we conducted multivariate regressions that included an

interaction term between each engagement scale and a dummy variable that was coded as “1” if the student was from one of the three institutions identified as the most effective at promoting favorable engagement–performance linkages, and “0” if the student was at an institution that promoted the least favorable or even detrimental engagement–performance linkages. Table 5 displays interaction terms if the interaction coefficient achieved statistical significance at the .05 level.

Table 5 shows the statistically significant differences between students at the highest and lowest schools in the learning productivity distribution. An effect size (ES) was calculated for each scale if the interaction coefficient was significant. Initially, fully-standardized coefficients were computed for the most and least effective institutions, then the absolute difference calculated to yield the boost in RAND scores (in standard deviation units) accruing to a standard deviation increase in student engagement within most versus least effective institutions. As compared to institutions in the lowest learning productivity groups, institutions in the highest learning productivity group were more able to convert 8 of the 11 engagement scales into higher scores on the RAND tests:

- level of academic challenge (ES = .22)
- student–faculty interaction (ES = .25)
- supportive campus environment (ES = .29)
- reading and writing (ES = .12)
- quality of relationships (ES = .24)
- institutional emphases on good practices (ES = .20)
- higher-order thinking (ES = .24), and
- student–faculty interaction concerning coursework (ES = .27)

Another way to interpret these findings – given that the overall standard deviation for the RAND test was 177.55 (Table 1) – is that each standard deviation increase in level of academic challenge is linked to an additional 39 point boost ($.22 \times 177.55$). Moreover, the highest learning productivity institutions had significantly stronger relationships between RAND and student-reported outcomes (general education gains [ES = .29], personal–social gains [ES = .27], practical competence gains [ES = .20], and satisfaction [ES = .15]).

We also contrasted the top three institutions in terms of learning productivity with the eight institutions that performed near the middle of the sample; this latter set of schools included several selective liberal arts colleges, two master’s level universities, and one moderately selective doctoral degree-granting university (Table 6). As expected, we found fewer statistically significant interaction coefficients than in

TABLE 5. OLS Regressions of RAND Scores on Interactions Involving Student Engagement or Self-reported Outcomes and Institutional Learning Productivity (Highest versus Lowest)

Scale	Interaction Terms ^a
<i>Student Engagement</i>	
Level of academic challenge	-3.79
	-160.85*
	12.04***
Active and collaborative learning Student-faculty interaction	ES = .22
	NS ^b
	-5.43
Enriching educational experiences Supportive campus environment	-52.36
	14.17***
	ES = .25
Reading and writing	NS
	-2.88
	-12059*
Quality of relationships	17.07**
	ES = .29
	.53
Institutional emphases on good practices	64.51
	14.17*
	ES = .12
Higher-order thinking	-.66
	-225.66**
	18.35***
Student-faculty interaction concerning coursework	ES = .24
	-.58
	-53.78
Integration of diversity into coursework	9.57*
	ES = .20
	-8.15*
	-172.68*
	19.46**
	ES = .24
	-12.415*
	-155.03*
	28.09***
	ES = .27
	NS

TABLE 5. (Continued)

Scale	Interaction Terms ^a
<i>Self-reported Outcomes</i>	
General education gains	-7.60
	-93.65
	19.83**
Personal-social gains	ES = .29
	-6.10*
	-43.36
Practical competence gains	9.80**
	ES = .27
	-8.74
Satisfaction	-36.87
	13.27*
	ES = .20
	-10.60
	-105.40
	27.50*
	ES = .15

* $p < .05$, ** $p < .010$, *** $p < .001$ (two-tailed) N s ranged from 361 to 378.

Note: Controls included total SAT score and squared SAT.

^aFirst, second, and third coefficients correspond to student engagement or self-reported outcome, institutional learning productivity dummy, and multiplicative interaction, respectively.

^bInteraction coefficient not statistically significant at the .05 level.

Table 5, even though the larger N s at the student level (738 to 760 rather than 361 to 378) afforded additional statistical power. Even so, the institutions that appeared to perform the best were somewhat more effective at converting student-faculty interaction (ES = .12), enriching educational experiences (ES = .13), supportive campus environment (.09), reading and writing (ES = .05), quality of relationships (ES = .08), and student-faculty interaction concerning coursework (ES = .10) into greater RAND scores.

In supplementary analyses not shown in tabular format, we found little evidence of varied institutional effectiveness for the GRE. However, this finding is based only on the six institutions that had sufficient numbers of students who completed the GRE tests. With respect to GPA, four institutions (HBCU, one selective liberal arts college, two master's level institutions (one public, one private) appeared to more effectively convert student engagement into higher grades.

TABLE 6. OLS Regressions of RAND Scores on Interactions Involving Student Engagement or Self-reported Outcomes and Institutional Learning Productivity (Highest versus Mid-range)

Scale	Interaction Terms ^a
<i>Student Engagement</i>	
Level of academic challenge	NS ^b
Active and collaborative learning	NS
Student–faculty interaction	.55 –25.26 7.85*
Enriching educational experiences	ES = .12 –4.87 –92.89 20.10**
Supportive campus environment	ES = .13 5.65** –50.02 8.07*
Reading and writing	ES = .09 4.02 35.34** 12.09*
Quality of relationships	ES = .05 6.79** –129.49 10.70*
Institutional emphases on good practices	ES = .08 NS
Higher-order thinking	NS
Student–faculty interaction concerning coursework	2.52 –62.34 12.68*
Integration of diversity into coursework	ES = .10 NS
<i>Self-reported Outcomes</i>	
General education gains	NS
Personal–social gains	NS
Practical competence gains	NS
Satisfaction	NS

* $p < .05$, ** $p < .010$, *** $p < .001$ (two-tailed) *Ns* ranged from 738 to 760.

Note: Controls included total SAT score and squared SAT.

^aFirst, second, and third coefficients correspond to student engagement or self-reported outcome, institutional learning productivity dummy, and multiplicative interaction, respectively.

^bInteraction coefficient not statistically significant at the .05 level.

CONCLUSIONS

On balance, the results of this study corroborate what many other researchers have found: that student engagement is linked positively to desirable learning outcomes such as critical thinking and grades. Although the relationships between engagement and academic performance were not as robust as we might hope, they were more conclusive than those reported by Ewell (2002). To the degree that student experiences, engagement, and academic performance change over the course of a collegiate career (and there is considerable evidence suggesting that this is so), the asynchrony in data collection points likely accounts for the inconclusive patterns found in the South Dakota data.

Our findings along with others (Ewell, 2002; Klein et al., 2005; Pascarella and Terenzini, 2005) underscore the fact that learning outcomes stem from a variety of sources, of which student engagement is only one. Indeed, the positive relationships between engagement and outcomes described in this paper are relatively small in magnitude. A large portion—and in some cases a majority—of the variance in key outcomes remains to be explained by yet undiscovered factors. We hope to learn more about the relationships between student engagement and learning from a Pew-sponsored National Forum on College-Level Learning project that co-administered NSSE, RAND, and the National Assessment of Adult Literacy at about 50 colleges and universities in 2003–2004.

In addition to the generally positive but relatively weak effects of engagement on critical thinking skills, the NSSE–RAND analyses point to two evocative findings. First, college students with the lowest SAT scores appeared to benefit more from student engagement than those with the highest SATs. While this pattern dovetails with some recent research, we cannot rule out that these findings may be in part attributable to ceiling effects or the phenomenon of regression to the mean. A future study might use multiple measures to control for student ability prior to matriculation to increase their stability. Such a strategy would decrease the probability that the very lowest or highest measures of pre-college ability were due to chance or unusual circumstances unlikely to repeat at a later measurement point.

Second, using these same data, Klein et al. 2005 suggested that certain institutions appear to add more value than others, as measured by the RAND critical thinking tests. Our analysis indicates that certain institutions more effectively convert engagement into stronger student performance as represented by RAND scores. Thus, the findings suggest that student engagement constitutes a constellation of institutional

processes that may “add value” to student learning. Although this study is based on only 14 institutions, institutional differences were discernable, particularly between institutions most and least effective in converting student engagement into performance. It would be instructive to conduct intensive case studies of high performing institutions to learn what they do or have that seem to contribute to greater student learning productivity, i.e., to extend the works of Kuh et al. (1991, 2005).

In closing, it is important to note that we were able to examine the complex relationships between engagement and learning outcomes only because we were able to analyze student-level data across multiple institutions. The integration of multiple measures (SAT scores, RAND/CAE scores, grades, student engagement results from NSSE) was made possible by the cooperation of the participating colleges and different research teams. Such efforts are few and far between. The results of this study illustrate the insights that can accrue from such collaborative research endeavors.

ENDNOTES

1. Three institutions were dropped from this subanalysis by class, because their students' classes were measured differently, i.e., a semester prior to completion of *The Report* rather than during the same term. However, the inclusion of these three schools did not appreciably alter the patterns reported here by class.

APPENDICES

APPENDIX A.1 Survey Items Contributing to Student Engagement and Self-reported Outcome Scales

I. Level of Academic Challenge (Cronbach's $\alpha = .70$)

- Item 1. Number of hours per week spent on preparing for class (studying, reading, writing, rehearsing, and other activities related to your academic program)
2. Frequency worked harder than you thought you could to meet instructors' standards or expectations during the current school year
 3. Number of assigned textbooks, books, or book-length packs of course readings during the current school year
 4. Number of written papers or reports of 20 pages or more during the current school year
 5. Number of written papers or reports between 5 and 19 pages during the current school year
 6. Number of written papers or reports of fewer than 5 pages during the current school year
-

APPENDIX A.1 (*Continued*)

7. During the current school year, the extent coursework emphasized analyzing the basic elements of an idea, experience, or theory
8. During the current school year, the extent coursework emphasized synthesizing and organizing ideas, information, or experiences into new, more complex interpretations and relationships
9. During the current school year, the extent coursework emphasized making judgments about the value of information, arguments, or methods
10. During the current school year, the extent coursework emphasized applying theories or concepts to practical problems or in new situations
11. The extent the institution emphasized spending significant amounts of time studying and on academic work

II. Active and Collaborative Learning (Cronbach's alpha = .62)

- Item 12. Frequency asked questions in class or contributed to class discussions during the current school year
13. Frequency made class presentations during the current school year
 14. Frequency worked with other students on projects during class during the current school year
 15. Frequency worked with classmates outside of class to prepare class assignments during the current school year
 16. Frequency tutored or taught other students (paid or voluntary) during the current school year
 17. Frequency participated in a community-based project as part of a regular course
 18. Frequency discussed ideas from your readings or classes with others outside of class (students, family members, coworkers, etc.) during the current school year

III. Enriching Educational Experiences (Cronbach's alpha = .56)

- Item 19. Frequency used an electronic medium (list-serv, chat group, Internet, etc.) to discuss or complete an assignment
20. Frequency had serious conversations with students of a different race or ethnicity than your own
 21. Frequency had serious conversations with students who differed from you in terms of their religious beliefs, political opinions, or personal values
 22. Have done or plan to complete a practicum, internship, field experience, co-op experience, or clinical assignments
 23. Have done or plan to do community service or volunteer work
 24. Have done or plan to take foreign language coursework
 25. Have done or plan to study abroad
 26. Have done or plan to complete an independent study or self-designed major
 27. Have done or plan to complete a culmination senior experience (comprehensive exam, capstone course, thesis project, etc.)
-

APPENDIX A.1 (*Continued*)

28. Number of hours per week participated in co-curricular activities (organizations, campus publications, student government, social fraternity or sorority, intercollegiate or intramural sports, etc.)
29. The extent the institution emphasized contact among students from different economic, social, and racial or ethnic backgrounds

IV. Student-faculty Interaction (Cronbach's alpha = .71)

- Item 30. Frequency discussed grades or assignments with an instructor during the current school year
31. Frequency talked about career plans with a faculty member or advisor during the current school year
 32. Frequency discussed ideas from your readings or classes with faculty members outside of class during the current school year
 33. Frequency worked with faculty members on activities other than coursework (committees, orientation, student life activities, etc.) during the current school year
 34. Frequency received prompt feedback from faculty on your academic performance (written or oral) during the current school year
 35. Have done or plan to work on a research project with a faculty member outside of course or program requirements

V. Supportive Campus Environment (Cronbach's alpha = .75)

- Item 36. The extent the institution emphasized providing the support you needed to succeed academically
37. The extent the institution emphasized helping you cope with non-academic responsibilities (work, family, etc.)
 38. The extent the institution emphasized providing the support you needed to thrive socially
 39. Quality of Relationships with other students at your institution
 40. Quality of Relationships with faculty members at your institution
 41. Quality of Relationships with administrative personnel and offices at your institution

VI. Reading and Writing (Cronbach's alpha = .61)

- Item 3. Number of assigned textbooks, books, or book-length packs of course readings
5. Number of written papers or reports between 5 and 19 pages
 6. Number of written papers or reports of fewer than 5 pages

VII. Quality of Relationships (Cronbach's alpha = .60)

- Item 39. Quality of Relationships with other students at your institution
40. Quality of Relationships with faculty members at your institution
 41. Quality of Relationships with administrative personnel and offices at your institution
-

APPENDIX A.1 (*Continued*)

VIII. Institutional Emphases on Good Practices (Cronbach's alpha = .78)

- Item 29. The extent the institution emphasized contact among students from different economic, social, and racial or ethnic backgrounds
36. The extent the institution emphasized providing the support you needed to succeed academically
 37. The extent the institution emphasized helping you cope with non-academic responsibilities (work, family, etc.)
 38. The extent the institution emphasized providing the support you needed to thrive socially
 42. The extent the institution emphasized the attendance of campus events and activities

IX. Higher-Order Thinking (Cronbach's alpha = .74)

- Item 7. During the current school year, the extent coursework emphasized analyzing the basic elements of an idea, experience, or theory
8. During the current school year, the extent coursework emphasized synthesizing and organizing ideas, information, or experiences into new, more complex interpretations and relationships
 9. During the current school year, the extent coursework emphasized making judgments about the value of information, arguments, or methods
 10. During the current school year, the extent coursework emphasized applying theories or concepts to practical problems or in new situations

X. Student-faculty Interaction Concerning Coursework (Cronbach's alpha = .68)

- Item 30. Frequency discussed grades or assignments with an instructor during the current school year
32. Frequency discussed ideas from your readings or classes with faculty members outside of class during the current school year
 34. Frequency received prompt feedback from faculty on your academic performance (written or oral) during the current school year

XI. Integration of Diversity into Coursework (Cronbach's alpha = .71)

- Item 8. During the current school year, the extent coursework emphasized synthesizing and organizing ideas, information, or experiences into new, more complex interpretations and relationships
32. Frequency discussed ideas from your readings or classes with faculty members outside of class
 43. Frequency worked on a paper or project that required integrating ideas or information from various sources
 44. Frequency included diverse perspectives (different races, religions, genders, political beliefs, etc.) in class discussions
 45. Frequency put together ideas or concepts from different courses when completing assignments or during class discussions
-

APPENDIX A.1 (Continued)

XII. Gains in General Education (Cronbach's alpha = .77)

- Item 46. The extent your college experience contributed to writing clearly and effectively
47. The extent your college experience contributed to speaking clearly and effectively
48. The extent your college experience contributed to acquiring broad general education
49. The extent your college experience contributed to thinking critically and analytically

XIII. Gains in Personal and Social Development (Cronbach's alpha = .80)

- Item 50. The extent your college experience contributed to developing a personal code of values and ethics
51. The extent your college experience contributed to understanding people of other racial and ethnic backgrounds
52. The extent your college experience contributed to understanding yourself
53. The extent your college experience contributed to learning effectively on your own
54. The extent your college experience contributed to working effectively with others
55. The extent your college experience contributed to acquiring broad general education
56. The extent your college experience contributed to thinking critically and analytically

XIV. Gains in Practical Competence (Cronbach's alpha = .66)

- Item 57. The extent your college experience contributed to acquiring job or work-related knowledge and Skills
58. The extent your college experience contributed to analyzing quantitative problems
59. The extent your college experience contributed to using computing and information technology
60. The extent your college experience contributed to (your) working effectively with others

XV. Satisfaction (Cronbach's alpha = .75)

- Item 61. How would you evaluate your entire educational experience at this institution?
62. If you could start over again, would you go to the same institution you are now attending?
-

APPENDIX A.2: Degree of Overlap Between Survey Items Contributing to Student Engagement and Self-reported Outcome Scales^a

ITEM	SCALE														
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
1	X														
2	X														
3	X					X									
4	X														
5	X					X									
6	X					X									
7	X								X						
8	X								X		X				
9	X								X						
10	X								X						
11	X														
12		X													
13		X													
14		X													
15		X													
16		X													
17		X													
18		X													
19			X												
20			X												
21			X												
22			X												
23			X												
24			X												
25			X												
26			X												
27			X												
28			X												
29			X					X							
30				X						X					
31				X											
32				X						X	X				
33				X											
34				X						X					
35				X											
36					X			X							
37					X			X							
38					X			X							

APPENDIX A.2 (Continued)

ITEM	SCALE														
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
39					x		x								
40					x		x								
41					x		x								
42								x							
43										x					
44										x					
45										x					
46											x				
47											x				
48											x				
49											x				
50												x			
51												x			
52												x			
53												x			
54												x			
55												x			
56												x			
57														x	
58														x	
59														x	
60														x	
61															x
62															x

^aSee Appendix A.1 for the denotation of item and scale numbers.

REFERENCES

- Astin, A. W. (1991). *Assessment for Excellence: The Philosophy and Practice of Assessment and Evaluation in Higher Education*, American Council on Education/Macmillan, New York.
- Baird, L. L. (1976). Biographical and educational correlates of graduate and professional school admissions test scores. *Educational and Psychological Measurement* 36(2): 415-420.
- Baird, L. L. (1988). Value Added: Using student gains as yardsticks of learning. In: Adelman, C. (ed): *Performance and Judgement: Essays on Principles and Practice in the Assessment of College Student Learning*, U.S. Government Printing Office, Washington, DC, pp. 205-216.
- Benjamin, R., and Hersh, R. H. (2002). Measuring the difference that college makes. *Peer Review* 4(2/3): 7-10.

- Bohr, L., Pascarella, E. T., Nora, A., Zusman, B., Jacobs, M., Desler, M., and Bulakowski, C. (1994). Cognitive effects of two-year and four-year colleges: A preliminary study. *Community College Review* 22(1): 4–11.
- Chun, M. (2002). Looking whether the light is better: A review of the literature on assessing higher education quality. *Peer Review* 4(2/3): 16–25.
- Cronbach, L. J., and Furby, L. (1970). How we should measure “change” – or should we? *Psychological Bulletin* 74(1): 68–80.
- Ewell, P. T. (1984). *The Self-regarding Institution: Information for Excellence*, National Center for Higher Education Management Systems, Boulder, CO.
- Ewell, P. T. (1988). Outcomes, assessment, and academic improvement: In search of usable knowledge. In: Smart, J. C. (eds.), *Higher Education: Handbook of Theory and Research*, 4, Agathon Press, New York, pp. 53–108.
- Ewell, P. T. (1994). *A Policy Guide for Assessment: Making Good Use of the Tasks in Critical Thinking*, Educational Testing Service, Princeton, NJ.
- Ewell, P. T. (2002). *An Analysis of Relationships between NSSE and Selected Student Learning Outcomes Measures for Seniors Attending Public institutions in South Dakota*, National Center for Higher Education Management Systems, Boulder, CO.
- Ewell, P. T., and Jones, D. P. (1996). *Indicators of “Good Practice” in Undergraduate Education: A Handbook for Development and Implementation*, National Center for Higher Education Management Systems, Boulder, CO.
- Gentemann, K. M., Fletcher, J. J., and Potter, D. L. (1994). Refocusing the academic program review on student learning. In: Kinnick, M. K. (eds.), *Providing Useful Information for Deans and Department Chairs. New Directions for Institutional Research*, Jossey-Bass, San Francisco, Vol. 84, pp. 31–46.
- Halpern, D. F. (1987). Recommendations and caveats, In Halpern, D. F. (ed.), *Student Outcomes Assessment: What Institutions Stand to Gain. New Directions for Higher Education*, Vol. 59, pp. 109–111.
- Hughes, R., and Pace, C. R. (2003). Using NSSE to study student retention and withdrawal. *Assessment Update* 15(4): 1–2.
- Jacobi, M., Astin, A., and Ayala, F. (1987). *College Student Outcomes Assessment: A Talent Development Perspective. ASHE-ERIC Higher Education Report No. 7*, Association for the Study of Higher Education, Washington, DC.
- Klein, S. P. (1996). The costs and benefits of performance testing on the bar examination. *The Bar Examiner* 65(3): 13–20.
- Klein, S. P. (2001). *Rationale and Plan for Assessing Higher Education Outcomes with Direct Constructed Response Measures of Student Skills*, Council for Aid to Education, Higher Education Policy Series, Number 3, New York, NY.
- Klein, S. P. (2002). Direct assessment of cumulative student learning. *Peer Review* 4(2/3): 26–28.
- Klein, S. P., Kuh, G. D., Chun, M., Hamilton, L., and Shavelson, R. (2005). An approach to measuring cognitive outcomes across higher education institutions. *Research in Higher Education* 46(3): 251–276.
- Kuh, G. D. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Change* 33(3): 10–17, 66.
- Kuh, G. D. (2003). What we’re learning about student engagement from NSSE. *Change* 35(2): 24–32.
- Kuh, G. D., Hayek, J. C., Carini, R. M., Ouimet, J. A., Gonyea, R. M., and Kennedy, J. (2001). *NSSE Technical and Norms Report*, Indiana University Center for Postsecondary Research and Planning, Bloomington, IN.
- Kuh, G. D., and Hu, S. (2001). Learning productivity at research universities. *Journal of Higher Education* 72: 1–28.

- Kuh, G. D., Kinzie, J., Schuh, J. H., and Whitt, E. J., et al. (2005). *Student Success in College: Creating Conditions that Matter*, Jossey-Bass and American Association for Higher Education, San Francisco.
- Kuh, G. D., and Pascarella, E. T. (2004). What does institutional selectivity tell us about educational quality? *Change* 36(5): 52–58.
- Kuh, G. D., Schuh, J. H., and Whitt, E. J., et al. (1991). *Involving Colleges: Successful Approaches to Fostering Student Learning and Personal Development Outside the Classroom*, Jossey-Bass, San Francisco.
- National Survey of Student Engagement (2003). *Converting Data into Action: Expanding the Boundaries of Institutional Improvement*, Indiana University Center for Postsecondary Research and Planning, Bloomington, IN.
- National Survey of Student Engagement (2004). *NSSE 2004 Overview*, Indiana University Center for Postsecondary Research and Planning, Bloomington, IN.
- Pace, C. R. (1984). *Measuring the Quality of College Student Experiences*, University of California, Higher Education Research Institute, Los Angeles.
- Pascarella, E. T. (2001). Identifying excellence in undergraduate education: Are we even close? *Change* 33(3): 19–23.
- Pascarella, E. T., and Terenzini, P. T. (1991). *How College Affects Students*, Jossey-Bass, San Francisco.
- Pascarella, E. T., and Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research*, 2, Jossey-Bass, San Francisco.
- Pascarella, E. T., and Wolniak, G. C. (2004). Change or not to change: Is there a question? A response to Pike. *Journal of College Student Development* 45(3): 353–355.
- Pascarella, E. T., Bohr, L., Nora, A., and Terenzini, P. T. (1994). *Is Differential Exposure to College linked to the Development of Critical Thinking?* National Center on Postsecondary Teaching, Learning, and Assessment, Illinois Univ., Chicago.
- Pike, G. R. (1992). Lies, damn lies, and statistics revisited: A comparison of three methods of representing change. *Research in Higher Education* 33(1): 71–84.
- Pike, G. R. (2004). Lord's paradox and the assessment of change during college. *Journal of College Student Development* 45(3): 348–352.
- Pohlmann, J. T. (1974). A description of effective college teaching in five disciplines as measured by student ratings. *Research in Higher Education* 4(4): 335–346.
- Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd Ed.), Sage Publications, Thousand Oaks, CA.
- Shulman, L. S. (2002). Making differences: A table of learning. *Change* 34(6): 36–45.
- Terenzini, P. T. (1989). Assessment with open eyes: Pitfalls in studying student outcomes. *Journal of Higher Education* 60: 644–664.
- Tinto, V. (1993). *Leaving College: Rethinking the Causes and Cures of Student Attrition*, (2nd Ed.) University of Chicago Press, Chicago.
- Vandament, W.E. (1987). A state university perspective on student outcomes assessment. In Halpern, D. F. (ed.). *Student outcomes assessment: What institutions stand to gain. New Directions for Higher Education*, Vol. 59, pp. 25–28.
- Winter, D. G., McClelland, D. C., and Stewart, A. J. (1981). *A New Case for the Liberal Arts*, Jossey-Bass, San Francisco.

Received August 30, 2004.