*Research in Higher Education, Vol. 47, No. 5, August 2006 (© 2006)*
DOI: 10.1007/s11162-005-9006-z

# INSTITUTIONAL STRUCTURES AND STUDENT ENGAGEMENT

Stephen R. Porter*,**

A common finding in the literature is that institutional structures have little to no impact on student engagement and development. I argue that theory suggests peer ability (as measured by selectivity), institutional density, the differentiation of the curriculum, and the research orientation of the institution should all affect student engagement. Using the nationally representative Beginning Post-secondary Student survey, a non-linear selection on observables correction for selection bias, and a multilevel modeling approach, I find that institutional structures do affect student engagement in predictable and substantively significant ways.

**KEY WORDS:** student engagement; institutional effects; multilevel models.

Student engagement has become a much-studied topic in higher education because engagement is highly correlated with learning and personal development (Astin, 1993; Pascarella and Terenzini, 1991). The level of educationally purposeful activities such as student–faculty interaction and active and collaborative learning has also been emphasized as an alternative measure of collegiate quality (Kuh, 2003). Little research has been conducted, however, on how institutional structures affect student engagement; in addition, the literature suggests a minimal role for institutional structures in student development (Pascarella and Terenzini, 1991, 2005, see also the review in Toutkoushian and Smart, 2001).

Given that competition in higher education, emphasis on student outcomes by accreditors, and scrutiny by legislators and the public

*Department of Educational Leadership and Policy Studies, Iowa State University, Ames, IA, USA.

**Address correspondence to: Stephen R. Porter, Department of Educational Leadership and Policy Studies, Iowa State University, N243 Lagomarcino Hall, Ames, IA 5001, USA. E-mail: srporter@iastate.edu

regarding college costs and outcomes have been increasing in recent years, it is more important than ever to understand what structural aspects of college enhance student outcomes. Does the small size and increased contact with faculty at liberal arts colleges justify their high cost? Does the emphasis on faculty research and graduate teaching at research and doctoral universities come at the expense of undergraduate development? Is the frenzy around admissions to selective colleges and universities simply the result of a misguided emphasis on collegiate rankings by students and their families? Only by understanding how institutional structures affect students can we begin to answer these questions, as well as guide future changes in how we organize our institutions.

Although there is a substantial literature on student engagement and development, research on the effect of institutions is limited for several different reasons. First, research in this area uses surveys based on convenience samples such as the National Survey of Student Engagement (NSSE), the Cooperative Institutional Research Program (CIRP) survey and the College Student Experiences Questionnaire (CSEQ) rather than samples designed to be nationally representative (Astin and Lee, 2003; Hu and Kuh, 2002, 2003a; Kim, 2002b; Kuh and Hu, 2001; Toutkoushian and Smart, 2001; Zhao and Kuh, 2004). Schools self-select to participate in these studies, and the effect of this self-selection is unknown. This decision is clearly not random; for example, public institutions are overrepresented in the NSSE. Only a representative sample of institutions and students can yield results that are both generalizable to all schools and that accurately estimate the relationship between institutional attributes and student outcomes.

Second, as Astin and Lee (2003) suggest, the cross-sectional approach used by much of the engagement literature does not take into account the effect of pre-college characteristics on institutional outcomes. Because institutions differ in their student inputs, they will also differ in their outputs; without appropriate measures of student pre-college characteristics, we cannot tell how much of the variation in engagement outcomes across institutions is due to differences in student bodies, and how much is due to the institutions themselves.

While their theoretical argument is sound, their empirical analysis is weakened by the use of student data averaged to the institutional level. As Robinson (1950) has demonstrated, correlations of aggregated individual-level data can be highly misleading; in some cases, the signs of correlations estimated on aggregated data can be the reverse of the correlations estimated on individual-level data (in his terminology, such correlations may lead to an "ecological fallacy"). Thus, it is still unclear as to what extent pre-college characteristics affect engagement in general, and more importantly, how they affect the estimates of the impact of other variables in the typical student engagement model.

Third, research in this area does not account for selection bias. Students are not randomly assigned to colleges and universities as in an experiment; as Astin and Lee (2003) point out, schools differ in their inputs, and it should be emphasized that these differences are due to the selection process. Selection bias is not an obscure technical issue to be debated among statisticians; it is instead one of the biggest econometric problems faced by higher education researchers today (Ehrenberg, 2004).

Fourth, with the exception of Toutkoushian and Smart (2001), most studies use only a few school-level variables in their models, such as public/private status, selectivity, or Carnegie classification (Hu and Kuh, 2002, 2003a, b; Kuh and Hu, 2001; Lundberg and Schreiner, 2004; Zhao, Kuh, and Carini, 2005), while others also include enrollment size (Kim, 2002a, b; Zhao and Kuh, 2004). Expenditures are rarely examined. Only three studies include measures of expenditures (Kim, 2002a; Smart, Ethington, Riggs, and Thompson, 2002; Toutkoushian and Smart, 2001). Many of these variables are intended only as controls, but a more complex set of institutional variables is necessary if we are to understand how institutions affect students.

Fifth, many studies use the Carnegie classification to explain variation in engagement between institutions. The Carnegie classification has several weaknesses, including an oversimplification of institutional differences, a poor measure of mission, and a reliance on arbitrary cutoffs for categories (McCormick, 2000). Clearly, a broader range of continuous variables will do a better job of measuring institutional differences; hence the new Carnegie classification system is adopting a multidimensional approach (McCormick, 2004).

Another drawback to using the Carnegie classification is that the classification conflates student body size with research orientation (McCormick, 2000). The 1994 definitions, for example, classify institutions not only by the number of degrees offered but also by the amount of federal grant support (Carnegie Foundation for the Advancement of Teaching, 1994). If we find that that student outcomes tend to be higher at liberal arts colleges than research universities (Pascarella, Wolniak, Cruce, and Blaich, 2004), it is difficult to understand why this is the case, because these two Carnegie groups differ not only by size, but also by how much research their faculty conduct. Thus, the Carnegie classification does not explain why student outcomes vary; it merely raises additional questions.

Sixth, many scholars use multiple regression (or ordinary least squares, OLS) rather than hierarchical linear modeling to study engagement (Hu and Kuh, 2003a; Kuh and Hu, 2001; Pascarella et al., in press, 2004; Toutkoushian and Smart, 2001; Zhao and Kuh, 2004). Use of OLS on grouped data is known to result in biased coefficients and standard errors.

This paper uses the 1996 Beginning Post-secondary Student survey (BPS) and a multilevel modeling approach to understand the impact of institutions on student engagement. It differs from previous research in four ways. First, the BPS is a nationally representative survey. Second, the analysis includes precollege characteristics such as engagement in high school. Third, the analysis controls for selection bias through a non-linear selection on observables approach. Fourth, the paper uses a different and more theoretically relevant set of institutional variables to understand why student engagement varies between students and colleges.

## CONCEPTUAL FRAMEWORK AND LITERATURE REVIEW

Several theoretical models for student outcomes have been developed in the literature (e.g., Astin, 1993; Pascarella, 1985; Perna, 2004); in essence, these models posit that student outcomes such as engagement are affected by the human, social and cultural capital students bring to college, as well as their experiences on campus and aspects of the institution such as size and selectivity. This paper adopts a similar approach of using student- and institutional-level variables to understand engagement. Because the focus of this paper is on the effect of institutional structures, the theoretical discussion focuses on institutional structures rather than student attributes.

While a significant amount of discussion has focused on the theoretical underpinnings of student characteristics and their effect on student behavior, less attention has been paid as to how and why institutional structures should affect student behavior. Astin's (1993) well-known input-environment-outcome model cannot in many respects be considered a theory of student behavior (Pascarella and Terenzini, 2005), while Pascarella's (1985) General Model shows that structural characteristics of an institution affect student development, but does not address how specific aspects of these structures affect student development. This paper focuses on why the three aspects of institutions used most often in quantitative studies of student development and engagement should affect student outcomes: selectivity, size, and research orientation. Figure 1 provides an overview of how these institutional structures should affect student engagement, along with measures of these structures.

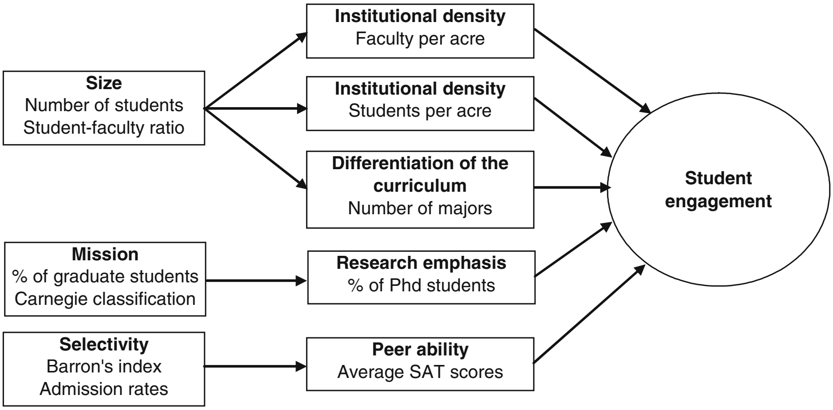


**FIG. 1.** How institutional structures affect student engagement.

## Selectivity and the Theory of Peer Effects

The role of peer groups in student development has been the subject of a large body of research (Pascarella and Terenzini, 1991, 2005; Renn and Arnold, 2003). College peers have been shown, for example, to affect intellectual self-confidence and degree aspirations (Antonio, 2004), interracial interaction and diversity related activities (Antonio, 2001; Milem, Umbach, and Liang, 2004), and self-reported learning (Lundberg, 2003). While there is some debate as to the exact process by which peers influence student outcomes (Renn and Arnold, 2003), it is clear that peer effects result from interactions with fellow students (Winston and Zimmerman, 2004).

The role of peer effects in higher education is important, because it explains why institutional selectivity affects student outcomes in general and engagement in particular. With the exception of Pascarella et al. (in press), there has been little theoretical discussion in the engagement literature as to why selectivity should affect engagement. At first glance, we would conclude that selectivity should not affect engagement once we have controlled for students' academic background and institutional resources, two variables highly correlated with selectivity. The theory of peer effects argues the opposite; the theory asserts that by attending college with high quality students, a student's behavior and academic performance will be higher than if they attended college with lower quality students (Winston and Zimmerman, 2004). A student may spend more time studying, for example, because she sees that her fellow students spend much of their time studying. But if this same student had attended school with students who spend little time studying, she would, in turn, study less and have poorer academic outcomes, such as a lower GPA or salary after graduation. Because selectivity is viewed as a measure of average student quality, we can conclude that *ceteris parabis*, student behavior such as engagement and outcomes such as graduate school attendance will be higher at more selective institutions, due to the presence of peer effects.

Research on secondary schools confirms the presence of peer effects at both school (Dills, 2004) and classroom levels (Hoxby, 2001). Post-secondary research is somewhat mixed, in part due to the difficulty in accurately estimating such effects (Manski, 1993). Analyses relying on the natural experiment of random roommate assignment have consistently found peer effects. They show that students rooming with someone with a higher academic ability have better grades than students living with someone of similar academic ability (Sacerdote, 2001; Winston and Zimmerman, 2004; Zimmerman, 2003). Studies using a variety of quasi-experimental approaches have also found that selectivity has a positive impact on retention (Titus, 2004), salaries after college (Black and Smith, 2003; Brewer Eide, and Ehrenberg, 1999; James, Alsalam, Conaty, and To, 1989; Loury and Garman, 1995; Monks, 2000; Rumberger and Thomas, 1993; Thomas, 2000, 2003) and graduate school attendance (Eide, Brewer, and Ehrenberg, 1998).

In contrast, as noted by Pascarella and Terenzini (2005), the literature on college student development and engagement has generally found little to no effect for institutional selectivity. For example, recent studies have found no effect for selectivity on student engagement (Hu and Kuh, 2002, 2003a; Kuh and Hu, 2001; Pascarella et al., in press) or self-reported abilities, gains and intellectual development (Hu and Kuh, 2002; Kim, 2002a, b; Kuh and Hu, 2001; Strauss and Volkwein, 2002).

These null findings are in strong contrast to the predictions of theory and other research on the presence of peer effects at the secondary and post-secondary levels. If students' grade-point average and post-graduation outcomes are affected by their peers, other behaviors such as student–faculty interaction should also be affected. The null findings are most likely due to two issues: poor measures of peer ability and unrepresentative data sources.

First, many scholars (Hu and Kuh, 2002, 2003a,b; Kuh and Hu, 2001; Lundberg and Schreiner, 2004) have used Barron's selectivity index, a six-point scale created by Barron's based on SAT/ACT scores, high school rank and admission rates, or admission rates by themselves (Strauss and Volkwein, 2002), as measures of an institution's selectivity. While both items may measure an institution's selectivity, the theory of peer effects asserts that average student *quality* will affect student outcomes, not selectivity *per se*. While selectivity and student quality are certainly highly correlated, SAT scores are a better and more accurate measure of average student ability than any construct based on admission rates, which are more a measure of institutional admission processes than student background.

In addition, when we refer to the effect of peers we are, in essence, referring to a contextual effect. As described by Raudenbush and Bryk (2002), "such effects are said to occur when the aggregate of a person level-characteristic $\bar{X}_{.j}$, is related to the outcome $Y_{ij}$, even after controlling for the effect of the individual characteristic $X_{ij}$" (p. 139). In other words, to correctly estimate a contextual effect we need to have similar measures at both the individual level and the college level. So for both theoretical and statistical reasons, average SAT scores are a better measure to use to understand the impact of selectivity on engagement.

Second, as stated previously, the student development and engagement literature is based on convenience samples rather than nationally representative samples. It is unclear if only certain types of schools choose to participate in these studies; in addition, substantial truncation of variance in the dependent and independent variables (e.g., engagement and institutional selectivity) may occur. In addition to the selection process, the institutional and student survey responses tend to be lower than nationally representative studies. The 2003 administration of the NSSE, for example, had an institutional participation rate of only 29% and a student survey response rate of 43% (National Survey of Student Engagement, 2003), while the national surveys administered by the federal government generally have institutional and student response rates in the 75%–90% range.[1] The use of convenience samples does not necessarily preclude the correct estimation of the impact of selectivity on student outcomes, but the nature of these data raises this question. It is interesting that positive findings for selectivity are all found within analyses that are based on nationally representative studies such as Baccalaureate and Beyond, High School and Beyond, the National Longitudinal Study of the High School Class of 1972, the National Longitudinal Survey of Youth, and the Survey of Recent College Graduates (e.g., Brewer et al., 1999; Eide et al., 1998; James et al., 1989; Loury and Garman, 1995; Monks, 2000; Rumberger and Thomas, 1993; Thomas, 2003; Titus, 2004), while the null findings for the effect of selectivity are concentrated within studies based on convenience samples (e.g., Hu and Kuh, 2002; Kim, 2002a, b; Kuh and Hu, 2001; Pascarella et al., in press).

In sum, there are strong theoretical reasons for believing that peer effects exist in higher education and that they might have an impact on student engagement. Natural experiments using roommate assignments and econometric analyses of nationally representative surveys have found the existence of peer effects across several different student outcomes. A review of the literature on student engagement and development suggests that peer effects on engagement may be found if direct measures of ability are used with a nationally representative dataset.

## Size, Redundancy and Differentiation of the Curriculum

Besides selectivity, one of the most common independent variables used in quantitative studies of student outcomes is student body size, either directly measured or indirectly proxied through the use of Carnegie classification. In terms of student engagement and development, institutional size does not appear to have much of an effect (Pascarella and Terenzini, 2005). Some studies have found that students are less engaged at research universities and more engaged at liberal arts colleges (Hu and Kuh, 2002; Kuh and Hu, 2001; Pascarella et al., 2004), but this could be due to research orientation rather than size.

Despite its prevalence in the literature, there has been surprisingly little discussion as to why size should matter. Instead, most discussions of size refer to its deleterious effects in a variety of areas; in other words, large institutions are associated with negative outcomes (e.g., Astin, 1993). Chickering and Reisser (1993) offer one explanation as to why large institutional size may have a negative effect on student outcomes. They make a distinction between physical settings, such as classrooms and dormitories, and people. They show that as institutions increase in size, the number of people increase faster than the number of settings. This results in "redundancy," in which the number of people begin to outnumber the possibilities for interaction and participation. As they describe it,

> When the number of people is small, each person has more opportunities to participate and derives more satisfaction from the experience. In task-oriented settings, some functions impose obligations on the participants. When few people are available, each participant has to assume more responsibilities and each becomes the focus for more obligations. ... If the setting is important as part of a larger context, external pressures will increase as the number of participants diminishes. There will be more invitations or demands, and the social rewards for contributions will increase. At the same time, requirements for admission or for certain kinds of positions will become more liberal (p. 304).

Thus, the causal mechanism is not institutional size, but size combined with geography; in other words, institutional *density*. They implicitly recognize this when they later refer to "the ratio of persons to settings" (p. 305), as have other scholars when they discuss, for example, "opportunities for students to become involved" (Pascarella and Terenzini, 1991, p. 654), because the number of opportunities for involvement is dependent on the ratio of people to settings. So we can conclude that as institutions become more dense, student outcomes such as engagement and development suffer.

Yet we must be careful when we discuss the ratio of people to settings, because the natural question arises, which people? Clearly Chickering and Reisser have students in mind, such that small numbers of students per setting are to be recommended. Yet faculty are overlooked in this formulation, and indeed the literature on student–faculty interaction would argue the opposite: that more faculty per setting rather than less would be beneficial to students, because more faculty per setting increases the probability of students interacting with faculty, especially outside the classroom. Understanding that student interaction with other students and student interaction with faculty are affected by the density of the institution rather than size sheds light on the common finding that size and student–faculty ratio appear to have little impact on student outcomes. It could be that they are both poor measures of institutional density and how it affects interactions within an institution.

Size may also affect student outcomes in ways other than the density of students and faculty. There is a substantial literature at the secondary level that documents the negative effect of high schools with large numbers of students on student learning and engagement (Johnson, Crosnoe, and Elder, 2001; Lee and Smith, 1997; Lee, Smith, and Croninger, 1997). Some researchers speculate that high school size is a proxy for other factors such as a differentiated curriculum (Lee and Smith, 1997). Larger high schools offer a wider variety of courses outside the core curriculum; students can fill their coursework with courses outside the core and thus learn less.

A similar process may be working at the post-secondary level, as one of the distinguishing differences between large and small schools is the variety of courses available. Small colleges tend to focus on the liberal arts, while large universities offer many courses outside the liberal arts, in a variety of vocational and professional areas. How does a wider variety of courses affect engagement? Hu and Kuh (2002) have found that engaged students are less likely to state that their institution emphasizes vocational and practical matters, while the NSSE has consistently found that students majoring in areas such as business, education, engineering are less engaged than students in the humanities, social sciences and natural sciences (National Survey of Student Engagement, 2003). It may be that faculty in professional and vocational areas emphasize active and collaborative learning and interact with students less than faculty in the traditional liberal arts disciplines. If so, then one reason size may negatively affect engagement is due to the differentiated curriculum, in addition to the impact size has on student relationships with other students and faculty.

## Research Orientation and Faculty Time Allocation

Why should the research orientation of an institution affect student engagement? Most discussions about the relationship between teaching and research debate whether the two are mutually reinforcing or in competition; research indicates that faculty who focus on research tend to do so at the expense of teaching (Fairweather, 1996, 2002; Fox, 1992). In part, the issue boils down to time: there is only so much time during the day, and time spent on research is time taken from other activities. Fairweather (1996), for example, has shown that the number of refereed publications by a faculty member is inversely related to the reported amount of time spent on teaching.

The issue of time is important, as Chickering and Reisser (1993) note that accessibility is one of the four components of positive student–faculty relationships. Other components, such as knowledge of students and the ability to communicate with students, undoubtedly increase as the amount of time spent with students increase. Thus, faculty at institutions that emphasize research will devote large portions of their time to research, due to institutional reward structures. This in turn means less time spent with students, affecting student–faculty interaction in a variety of ways, from advising to socializing with students outside of the classroom.

Surprisingly, research on student outcomes indicates little effect by Carnegie type (Pascarella and Terenzini, 2005), although some recent research indicates that students at liberal arts colleges are more engaged than students at research universities (Pascarella et al., 2004). These mixed results may be in part due to how research orientation is measured. As stated previously, the Carnegie classification is a poor measure of mission, and may be a crude indicator of faculty accessibility and knowledge about their undergraduate students. Other scholars have used the percentage of graduate students in the student body (Toutkoushian and Smart, 2001), which more closely proxies the demands of research on faculty time. Faculty at an institution with few graduate students have less to distract them from undergraduates, while faculty at institutions with a large proportion of graduate students will face more demands on their time.

Yet this measure may also be too crude, as faculty in a department with substantial doctoral students face huge demands on their time in terms of advising dissertations, while faculty whose graduate students are working on a professional degree may face fewer demands for their time. If true, this suggests that the type of graduate student is as important as their proportion, and may be a better measure of how research affects student–faculty relationships.

### Summary

While reviews of the literature indicate that institutional structures have little impact on student outcomes (Pascarella and Terenzini, 1991, 2005; Toutkoushian and Smart, 2001), I argue that an institution's selectivity, size and emphasis on research should all affect student outcomes, particularly engagement. The theory of peer effects suggests that institutional selectivity should positively affect engagement, while Chickering and Reisser's theory of persons and settings implies that institutional density should also affect engagement. Research at the high school level suggests that a differentiated curriculum should negatively affect engagement, while the faculty productivity literature indicates that engagement should be lower at institutions that emphasize faculty research. The remainder of the paper is devoted to testing these propositions.

## METHOD

### Data

This paper uses the BPS 96:01, a panel study of college students conducted by the National Center for Education Statistics (NCES) beginning in 1996, combined with data from the 1995–1996 IPEDS surveys, and Barron's and Peterson's college guidebook data for 1995. The BPS survey is designed to be nationally representative of full-time beginning students (FTB), and the base year survey has an institutional participation rate of 91% and a FTB student survey response rate of 78% (National Center for Education Statistics, 1997).[2] Another strength of the survey is that the BPS dataset includes data about high school experiences derived from the questionnaires that accompany the SAT and ACT standardized tests. Thus while the base survey year is 1996, it also includes survey items from the SAT and ACT that were collected during high school when the student took one of these exams. The analysis is limited to students who were first-time beginning students at a 4-year not-for-profit school classified as a research, doctoral, comprehensive or baccalaureate institution who participated in the 1996 base year computer assisted telephone interview (CATI).

### Statistical Approach

The models are estimated using a random-intercept hierarchical linear model (HLM), in which a separate intercept is estimated for each school. All independent variables are grand-mean centered. The BPS weight for the base year survey is also used (Thomas and Heck, 2001). In general, HLM is the appropriate statistical technique for analyzing nested data, such as students nested within schools (Heck and Thomas, 2000; Raudenbush and Bryk, 2002). While use of OLS on grouped data is known to result in biased coefficients and standard errors, it is worth describing in detail two particular issues that arise in this situation that make OLS problematic.

Use of OLS with grouped data increases the probability of Type I error (Raudenbush and Bryk, 2002) for the institution-level variables, as the hypothesis tests for the effects of institution-level variables use the student N for degrees of freedom, instead of the much smaller institution N. Thus, many school variables may appear statistically significant when they are not, leading to erroneous conclusions about the effect of institutional structures.

In addition, multiple regression fails to take into account the grouped nature of the data. These samples are not simple random samples; they are instead cluster samples, where schools are first selected and then students within those schools are administered the survey instrument. In essence, the convenience samples used by most of the engagement literature have sample designs that are more similar to the surveys conducted by the National Center for Education Statistics (NCES) than a simple random sample. Clustered samples cannot be treated as a simple random sample, because standard errors derived from these samples tend to be larger than those derived from simple random samples (Groves et al., 2004; Thomas and Heck, 2001). In general, this clustered design must be taken into account when estimating standard errors, otherwise the estimated standard errors will be too small, affecting hypothesis tests at the student level.

Besides dealing with the clustered nature of the data, another statistical issue is how to correct for selection bias. Following Dale and Krueger (2002), suppose that we have $i$ students enrolled in $j$ schools. Suppose further that admission committees use two variables to admit applicants, $X_1$ and $X_2$. $X_1$, say SAT score, is termed an "observable" characteristic because it is observable to the researcher in their dataset. $X_2$, say the applicant's essay, is termed "unobservable" to the researcher because it not in the researcher's dataset.

A common goal in the student outcome literature is to understand if the student outcome $Y_i$ is affected by institutional selectivity, notated as $SAT_{.j}$. If the equation below could be estimated,

$$Y_{ij} = \mathrm{B}_0 + \mathrm{B}_1 SAT_{.j} + \mathrm{B}_2 X_{1ij} + \mathrm{B}_3 X_{2ij} + \varepsilon_{ij} \quad (1)$$

then the coefficient for institutional selectivity ($\mathrm{B}_1$) would be an unbiased estimate of the effect of selectivity on the student outcome. Unfortunately with many datasets in higher education it is not possible to estimate Equation (1), because of unobservable data. Instead, researchers estimate

$$Y_{ij} = \mathrm{B}_0 + \mathrm{B}_1 SAT_{.j} + \mathrm{B}_2 X_{1ij} + \varepsilon_{ij} \quad (2)$$

which results in a biased estimate of $\mathrm{B}_1$, because $\mathrm{B}_1$ is picking up the effect of the omitted variable(s). Because students with high values on the omitted variable(s) are more likely to be admitted to schools with high average SAT scores, the general result will be an overestimate of the impact of institutional selectivity. Thus, any estimated effect we find for selectivity may simply be due to high ability students self-selecting into these schools, rather than the impact of peer effects.

A wide variety of econometric techniques have been proposed to deal with the problem of selection on unobservables (Black and Smith, 2003; Brewer et al., 1999; Dale and Krueger, 2002; Eide et al., 1998). The logic underlying these approaches is simple: the researcher does not have enough data about the student to correctly estimate Equation (1), therefore an econometric technique such as instrumental variables or propensity scores must be used. Even if the researcher did have enough data, a second problem arises: linearity in the model. Simply including these variables in a standard regression model assumes their functional form is linear, while there is evidence that the relationships for these variables may be non-linear; thus Black and Smith (2003, p. 100) conclude that "reliance on the linear functional form seems heroic."

I address the issue of selection bias in two ways. First, unlike previous researchers, I have in the BPS an almost complete set of data on student characteristics used in the college admission process. Besides common student variables such as gender, race, and SAT score, I include other variables used by institutions to select students, such as first-generation college student, whether on financial aid (as many schools are not need blind in admissions), whether the student was a legacy at the BPS institution, high school varsity athlete (athletic ability has been shown to have a major impact on college admissions (Shulman and Bowen, 2001)), high school grade point average, rigor of the coursework taken in high school, and the extent of extracurricular activities in high school.

What is missing from this list are some of the unobservables mentioned by Dale and Krueger, such as the student's essay and teacher recommendations. While these may be unobserved, they play a small role in the college admission process. A national survey of admission practices in 1992 asked admission offices how important various factors were in the admission decision (with the BPS, students would have been admitted in spring of 1994 or before). Given a five-point response scale (not considered, of minor importance, moderately important, very important, single most important), the percentage of institutions choosing either "very important" or "single most important" was 85% for high school grade point average, 62% for test scores, 51% for pattern of high school course work, 24% for letters of recommendation, and 19% for essays (Breland, Maxey, Gernand, Cumming, and Trapani, 2002). And while letters and essays may play a minor role, their quality is likely to be highly correlated with standardized test scores, high school grade point average and coursework, and extent of extracurricular activities. This is an important point, because as Dale and Kreuger (2002) note, the greater the correlation between $X_1$ and $X_2$, the lesser the bias in our estimate of the impact of selectivity on student outcomes.

Second, unlike previous research relying on a selection on observables approach (e.g., James et al., 1989; Monks, 2000), I allow for non-linear relationships in the models. SAT score is included, as well as squared and cubed terms. High school GPA and rigor of the student's coursework are interval scales in BPS; these scales are entered as dummy variables in the models to allow their impact to vary as the scale increases. Extent of extracurricular activities is a simple count variable which is also entered as a series of dummy variables. Together, the expanded set of student characteristics commonly used in the admission process along with a non-linear specification should largely correct for selection bias in the estimated equations.

## Dependent Variable

The dependent variable is a factor scale based on seven academic engagement items from the BPS. Because the base year of the BPS contains only FTB students, the analysis looks at student engagement in the first year of college. Students were given a three-point response scale (never, sometimes, or often) and asked how often during the 1995–1996 academic year they:

- Attended academic or career-related lectures, conventions, or field trips.
- Attended study groups outside of the classroom.
- Had informal or social contacts with an advisor or other faculty members outside of classrooms and offices.
- Met with an advisor concerning academic plans.
- Participated in school clubs (e.g. student government, religious clubs, service activities).
- Attended music, choir, drama, or other fine arts activities.
- Talked with faculty about academic matters outside of class time.

These items are similar to several items on two well-known engagement surveys, the CSEQ and NSSE. The alpha for this scale is .70. The intraclass correlation is .18, indicating that 18% of the variation in the engagement scale lies between schools. This is fairly substantial, given that many intraclass correlations in higher education research are .10 or less (Porter, 2005).

## Independent Variables

The independent variables in the analysis can be divided into four groups: student background variables, high school experiences, college experiences, and college institutional characteristics. Summary data and the source for these variables are presented in Table 1.

### *Student Variables*

Variables measuring student background include age in years, a dummy variable for females, and dummy variables for Blacks, Hispanics, Asians, other/unknown race/ethnicity, and non-resident aliens (reference category is Whites). Parental influences include a first-generation college student dummy variable indicating that neither parent graduated from college, and a legacy dummy variable indicating that one of the student's

**TABLE 1. Summary Statistics and Sources for Independent Variables**

| Variable | Mean | SD | Min | Max | Source |
|---|---|---|---|---|---|
| *Student level* | | | | | |
| Age | 18.666 | 1.933 | 16 | 76 | BPS (AGE) |
| Female | 0.561 | 0.496 | 0 | 1 | BPS (GENDER) |
| Black | 0.091 | 0.288 | 0 | 1 | BPS (RACE2) |
| Hispanic | 0.064 | 0.244 | 0 | 1 | BPS (RACE2) |
| Asian | 0.052 | 0.222 | 0 | 1 | BPS (RACE2) |
| Other | 0.010 | 0.099 | 0 | 1 | BPS (RACE2) |
| Non-resident alien | 0.005 | 0.073 | 0 | 1 | BPS (RACE2) |
| First generation | 0.452 | 0.498 | 0 | 1 | BPS (PAREDUC) |
| Legacy | 0.025 | 0.156 | 0 | 1 | BPS (PARNATT) |
| SAT score (100's) | 9.520 | 2.057 | 4.0 | 15.5 | BPS (TESATDER) |
| HS GPA: A to A– | 0.365 | 0.481 | 0 | 1 | BPS (HCGPAREP) |
| HS GPA: A– to B | 0.425 | 0.494 | 0 | 1 | BPS (HCGPAREP) |
| HS GPA: B to B– | 0.119 | 0.323 | 0 | 1 | BPS (HCGPAREP) |
| HS GPA missing | 0.116 | 0.321 | 0 | 1 | BPS (HCGPAREP) |
| HS coursework: highly rigorous | 0.165 | 0.371 | 0 | 1 | BPS (CTAKING) |
| HS coursework: moderately rigorous | 0.185 | 0.389 | 0 | 1 | BPS (CTAKING) |
| HS coursework: slightly rigorous | 0.400 | 0.490 | 0 | 1 | BPS (CTAKING) |
| HS coursework: only met New Basics | 0.043 | 0.204 | 0 | 1 | BPS (CTAKING) |
| HS coursework missing | 0.159 | 0.366 | 0 | 1 | BPS (CTAKING) |
| HS engagement: five areas | 0.022 | 0.147 | 0 | 1 | BPS (EXCURSUM) |
| HS engagement: four areas | 0.114 | 0.318 | 0 | 1 | BPS (EXCURSUM) |
| HS engagement: three areas | 0.168 | 0.374 | 0 | 1 | BPS (EXCURSUM) |

| | | | | | |
|---|---|---|---|---|---|
| HS engagement: two areas | 0.126 | 0.331 | 0 | 1 | BPS (EXCURSUM) |
| HS engagement: one area | 0.099 | 0.298 | 0 | 1 | BPS (EXCURSUM) |
| HS athlete | 0.333 | 0.471 | 0 | 1 | BPS (EXCSPORT) |
| On financial aid | 0.746 | 0.435 | 0 | 1 | BPS (AIDTYPE) |
| Full-time attendance | 0.862 | 0.345 | 0 | 1 | BPS (ATTNPTRN) |
| Hours worked per week | 12.587 | 13.155 | 0 | 60 | BPS (J1HOURY1) |
| Reside on campus | 0.688 | 0.463 | 0 | 1 | BPS (HTENRLY1) |
| Reside off campus | 0.117 | 0.322 | 0 | 1 | BPS (HTENRLY1) |
| Major: none/unknown | 0.314 | 0.464 | 0 | 1 | BPS (SEMAJ2Y1) |
| Major: humanities | 0.087 | 0.282 | 0 | 1 | BPS (SEMAJ2Y1) |
| Major: social sciences | 0.075 | 0.263 | 0 | 1 | BPS (SEMAJ2Y1) |
| Major: natural science or mathematics | 0.123 | 0.328 | 0 | 1 | BPS (SEMAJ2Y1) |
| *School level* | | | | | |
| Expenditures per student ($1,000's) | 20.111 | 18.676 | 3.180 | 212.370 | IPEDS Finance 1996 and Fall Enrollment 1995 |
| Location: urban | 0.556 | 0.497 | 0 | 1 | IPEDS Institutional Characteristics 1995 |
| Location: urban fringe | 0.168 | 0.374 | 0 | 1 | IPEDS Institutional Characteristics 1995 |
| Location: missing | 0.030 | 0.171 | 0 | 1 | IPEDS Institutional Characteristics 1995 |
| Barrons selectivity index | 2.231 | 1.139 | 0 | 5 | Barron's (1997) |
| % Graduate students | 0.183 | 0.129 | 0.000 | 0.707 | IPEDS Fall Enrollment 1995 |
| Student body size (1000's) | 11.101 | 8.947 | 0.478 | 37.711 | IPEDS Fall Enrollment 1995 |

**TABLE 1.** (*Continued*)

| Variable | Mean | SD | Min | Max | Source |
|---|---|---|---|---|---|
| Student–faculty ratio | 15.384 | 4.596 | 3 | 34 | Barron's (1997) |
| Carnegie: research | 0.312 | 0.463 | 0 | 1 | IPEDS Institutional Characteristics 1995 |
| Carnegie: doctoral | 0.138 | 0.344 | 0 | 1 | IPEDS Institutional Characteristics 1995 |
| Carnegie: comprehensive | 0.348 | 0.476 | 0 | 1 | IPEDS Institutional Characteristics 1995 |
| Average SAT score (100's) | 9.864 | 1.312 | 5.6 | 13.7 | Barron's (1997) and Peterson's (1996) |
| Faculty per acre (logged) | 0.487 | 1.084 | −3.071 | 5.098 | IPEDS Fall Staff 1995 and Peterson's (1996) |
| Students per acre (logged) | 3.175 | 1.033 | −0.350 | 7.271 | IPEDS Fall Enrollment 1995 and Peterson's (1996) |
| Number of majors | 56.460 | 27.837 | 9 | 145 | IPEDS Completions 1996 |
| % PhD degrees | 0.021 | 0.028 | 0 | 0.176 | IPEDS Completions 1996 |
| % MA degrees | 0.189 | 0.117 | 0 | 0.631 | IPEDS Completions 1996 |
| % 1st prof. degrees | 0.027 | 0.051 | 0 | 0.439 | IPEDS Completions 1996 |

Note: BPS 96:01 variable name in parentheses. Average class size data are calculated on reduced sample.

parents is a graduate of the BPS institution. This variable, along with the high school experiences variables, helps to control for selection bias.

The second set of student-level variables measure various aspects of high school experiences. Three variables measure the student's academic ability. The first is combined SAT score, which is based on either the SAT or a converted ACT score. Most of these scores were obtained from either ETS, ACT or the BPS institution. Squared and cubed terms are included to allow for a non-linear relationship. The second variable is self-reported high school grade-point average. In the BPS it is a seven-point scale that ranges from D/D− to A/A−. Few students had a grade point average below a C, so three dummy variables are used in the models: high school grade-point average of A to A−, A− to B, and B to B−, with the reference category a grade-point average below B−. The third variable is the rigor of the student's high school coursework, a five-point scale in the BPS based in part on the New Basics curriculum standards (National Commission on Excellence in Education, 1983). Following the NCES definition for this scale, four dummy variables were created (see Table 2). The reference category is that the student's coursework did not meet the New Basics standards. Both sets of dummy variables are derived from the questionnaire that accompanies the SAT and ACT.

**TABLE 2. High School Course Rigor Dummy Variable Definitions**

| Variable | Definition |
|---|---|
| Highly rigorous | Minimum curriculum of: 4 years each English and math; 3 years each foreign language, science, social science; one AP or honors class or AP test score in any subject; and student had taken all of the following: pre-calculus, biology, chemistry, and physics. |
| Moderately rigorous | Minimum curriculum of: 4 years English; 2 years foreign language; 3 years each math and science; and student had taken all of the following: Algebra 2, biology, chemistry, and physics. |
| Slightly rigorous | Minimum curriculum of: 4 years English; 1 year foreign language; 3 years each of math and science; and student had taken two of the following: biology, chemistry, and physics. |
| Only met New Basics | Minimum curriculum of New Basic standards: 4 years English, 3 years each social science, math, science. |

*Source*: BPS 96:01 electronic codebook.

A dummy variable indicating self-reported high school varsity athlete status is included, as well as dummy variables for the BPS summated scale measuring engagement in high school. This scale indicates the number of different high school extracurricular areas in which a student reported participating, and ranges from zero (no areas) to five (all five areas). Dummy variables indicating the number of areas in which the student participated are included, with no extracurricular activities as the reference category. The five areas used in the scale are academic (e.g., honor societies, foreign exchanges, foreign languages, or math and science), art (e.g., art, dance, instrumental music, or theatre), community service, civic expression (e.g., debate, government and politics, or journalism and literature), and ethnic organizations. Both of these are derived from the questionnaire that accompanies the SAT and ACT.

The third set of variables measure a student's college experiences and include a dummy variable indicating that the student was on financial aid, a dummy variable indicating that a student attended school full-time during the 1995–1996 academic year, the average number of hours spent working per week, and two dummy variables indicating that a student resided on campus or resided off campus without family (with residing off campus with family or relatives as the reference category). Academic major is measured with five dummy variables for no/unknown major, humanities major, social sciences major, and natural sciences or mathematics major (with professional major as the reference category).

### *Institutional Variables*

The final set of variables measures aspects of the BPS institution attended by the student. Four variables are included in all models. Financial resources are measured by expenditures per student, defined as the total educational and general expenditures and transfers from the IPEDS Finance survey divided by the number of students. School location is measured by three dummy variables indicating that the school is in an urban area (defined by IPEDS as large or mid-size city) an urban fringe area (defined by IPEDS as the urban fringe of a large or mid-size city), or area type missing/unknown. The reference category for these variables is large towns, small towns and rural areas combined. Expenditures per student control for differences in resources between schools, while the location dummy variables control for urban status, which is important because many dense schools (as defined below) are located in urban settings.

The remaining institutional variables consist of two groups, the first mirroring the standard set of variables used in the student outcomes literature, and the second reflecting the theoretical discussion about how institutional structures should affect student engagement. The standard set of variables include the number of students in the 1995–1996 academic year, Barron's (1997) selectivity index, the percentage of the student body composed of graduate students, student–faculty ratio, and three dummy variables for Carnegie research, doctoral and comprehensive institutions. These are some of the most commonly used variables in the student engagement and development literature (e.g., Hu and Kuh, 2002, 2003a; e.g., Kuh and Hu, 2001; Toutkoushian and Smart, 2001).

Seven variables are included to test the propositions in the theoretical discussion. Selectivity is measured by the average SAT score of the incoming class at the BPS institution in 1995. For those institutions in the college guidebooks that report only ACT scores for their class, a concordance table was used to convert these scores to SAT scores (Marco, Abdel-fattah, and Baron, 1992).

Institutional density is measured by two variables, faculty per acre and students per acre. Student and faculty counts are taken from IPEDS, and the number of acres comprised by the campus from the Peterson's (1996) electronic database for 1995. Both of these variables are logged because of their distributions.

Faculty per acre is a proxy for the probability that a student will meet a faculty member she knows as she walks across campus; at dense schools the value for this variable will be high, as there are many faculty in a relatively small space. Thus we would expect to see a positive relationship between faculty density and student engagement, as the opportunities for interacting with faculty increase in a dense faculty environment. Students per acre, on the other hand, is a proxy for what the faculty member sees as they walk across campus: a few students or a crowd of students. It is also a measure of opportunities to participate in campus life, because as Chickering and Reisser note, as the number of students per setting (or acre) increases, the number of opportunities for participation decreases. This variable should then have a negative relationship with engagement, as students become disengaged in dense student environments.

National data on college curricula are not available. I measure the differentiation in the college curriculum by the number of different undergraduate majors graduating in the 1995–1996 academic year, using the six-digit Classification of Instructional Program codes (National Center for Education Statistics, 1990) in the IPEDS Completions survey to count the number of different majors at an institution. This variable should have a negative impact on student engagement.

Rather than Carnegie classification or percentage of graduate students, research emphasis is measured by the size of the doctoral, master's and first professional programs at each BPS school. Unfortunately national data on the size of these programs are not available. Instead, I use the percentage of degrees at an institution awarded as PhDs, MAs and first professional degrees to measure differences in the size of graduate programs between institutions, taken from Completions data for the 1995–1996 academic year. While these are not precise measures of these types of programs because time to degree and dropout rates vary by degree type, they should be highly correlated with the proportion of doctoral, master's and first professional students at these schools.

## Missing Data

While missing data is not often discussed in quantitative higher education research, recently scholars have provided greater detail as to the extent and source of missing data in their analyses (e.g., Perna, 2004; Titus, 2004). Such detail is necessary for replication of findings, and is especially important in an analysis such as this one, as missing data for both students and schools in a national study is unavoidable. While some scholars have advocated mean substitution for missing data when possible (Perna, 2004), others have argued that listwise deletion, where a case with missing data on any variable is deleted, can often be the preferable approach (Allison, 2002). Both approaches are used in the paper.

The base sample of CATI respondents in the BPS 96:01 who attended a research, doctoral, comprehensive or baccalaureate institution comprises 6870 students in 412 schools. Some respondents did not answer all engagement items for the dependent variable. If one item was missing, mean substitution was used for that item, but if the respondent was missing data for more than one item out of the seven engagement items, they were deleted from the analysis. This resulted in a loss of 2% of the sample. There is missing data for several of the independent dummy variables such as college residence and legacy data; listwise deletion of these cases resulted in a further reduction of 6%. Two ordinal independent variables, high school GPA and rigor of high school coursework, each have substantial missing data (15% and 20%, respectively). The median value of these scales were used for those students with missing data; in addition, dummy variables indicating missing data for these students are included in the models estimated. Finally, 5% of the sample are missing either SAT or ACT test scores. Because NCES went to substantial effort to obtain these scores (they first received data on standardized scores from Educational Testing Service and ACT, then they queried the student's institution, and finally they queried the students themselves), it is clear that missing data for this variable are truly missing and these cases were deleted. This process left 6011 students in 403 schools in the sample.

Unfortunately there is substantial missing data at the school level, due either to IPEDS non-response or incomplete data in college guidebooks. Recent simulations using the High School and Beyond survey indicate that listwise deletion at the group level is a better strategy than imputation for multilevel models (Gibson and Olejnik, 2003). Missing data for the school variables results in a further reduction of cases to 5,114 students in 329 schools; the average number of students per school is 16.

## RESULTS

Table 3 presents the multilevel model results. Model 1 is similar to many models in the student engagement and development literature, in that there are no precollege characteristics other than student demographics, and the institutional-level covariates include expenditures per student, location, Barron's selectivity index, size of the student body, student–faculty ratio, percentage of graduate students in the student body, and Carnegie dummy variables.

The student-level results indicate that *ceteris paribus*, females, Blacks, Hispanics, students on financial aid, full-time students, on-campus residents, and humanities and science majors are more engaged, while first generation, working students, and students who do not have a major are less engaged. The school-level results are interesting, in that they indicate that several variables such as selectivity, student body size, and student–faculty ratio have a statistically significant effect on engagement, suggesting that some of the null findings in the engagement literature may be due to differences in samples. More selective, smaller schools with low student–faculty ratios have higher levels of engagement, as well as schools classified as baccalaureate institutions.

Model 2 answers the question raised by Astin and Lee (2003), do precollege characteristics matter? Starting with the variance explained statistics, the answer appears to be yes, as the variance explained by Model 2 increases slightly at both the student and school levels. A comparison of the student-level coefficients with Model 1, however, reveals a remarkable stability: our substantive conclusions about the effects of student demographics and college experiences remain unchanged, although the

**TABLE 3. Random Intercept Multilevel Regression Results**

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| *Student level* | | | | |
| Intercept | 0.0265 | −1.7084* | −1.4594$^{+}$ | −1.7317* |
| Age | −0.0119$^{+}$ | −0.0070 | −0.0065 | −0.0005 |
| Female | 0.1615** | 0.1121** | 0.1154** | 0.1242** |
| Black | 0.2956** | 0.3044** | 0.3112** | 0.2667** |
| Hispanic | 0.1913** | 0.1838** | 0.1793** | 0.1798** |
| Asian | −0.0205 | −0.0649 | −0.0619 | −0.0677 |
| Other | 0.1827 | 0.1455 | 0.1624 | 0.0862 |
| Non-resident alien | 0.0265 | 0.1705 | 0.1629 | 0.1154 |
| First generation | −0.1511** | −0.1281** | −0.1260** | −0.1198** |
| Legacy | | 0.0005 | −0.0117 | 0.0290 |
| SAT score (100's) | | 0.5260* | 0.5318* | 0.6201* |
| SAT score (100's): squared | | −0.0546* | −0.0561* | −0.0653* |
| SAT score (100's): cubed | | 0.0018* | 0.0019* | 0.0022* |
| HS GPA: A to A− | | 0.1210* | 0.1181* | 0.1420* |
| HS GPA: A− to B | | 0.1094* | 0.1093* | 0.1439** |
| HS GPA: B to B− | | 0.0736 | 0.0729 | 0.1102$^{+}$ |
| HS GPA missing | | −0.0385 | −0.0490 | −0.0337 |
| HS courses: highly rigorous | | 0.0020 | 0.0054 | −0.0129 |
| HS courses: moderately rigorous | | 0.0576 | 0.0588 | 0.0501 |
| HS courses: slightly rigorous | | 0.0235 | 0.0319 | 0.0216 |
| HS courses: only met New Basics | | −0.0056 | 0.0126 | 0.0308 |
| HS courses missing | | −0.0470 | −0.0504 | −0.0712 |
| HS engagement: five areas | | 0.3893** | 0.3864** | 0.4097** |
| HS engagement: four areas | | 0.3192** | 0.3079** | 0.2985** |
| HS engagement: three areas | | 0.1977** | 0.1814** | 0.1925** |
| HS engagement: two areas | | −0.0270 | −0.0445 | −0.0446 |
| HS engagement: one area | | −0.1560** | −0.1722** | −0.1846** |
| HS athlete | | 0.0466 | 0.0452 | 0.0295 |
| On financial aid | 0.0984** | 0.0799** | 0.0863** | 0.0836** |
| Full-time attendance | 0.1612** | 0.1378** | 0.1260** | 0.1220** |
| Hours worked per week | −0.0061** | −0.0058** | −0.0055** | −0.0055** |
| Reside on campus | 0.3591** | 0.3394** | 0.3204** | 0.3374** |
| Reside off campus | 0.2067** | 0.1792** | 0.1614** | 0.1729** |
| Major: none/unknown | −0.1099** | −0.0937** | −0.0906** | −0.0919** |
| Major: humanities | 0.1073* | 0.1037* | 0.1080* | 0.1089* |
| Major: social sciences | −0.0027 | −0.0080 | −0.0061 | 0.0187 |
| Major: natural science or math | 0.0905* | 0.0825* | 0.0791$^{+}$ | 0.0483 |
| *School level* | | | | |
| Expenditures per student ($1,000's) | −0.0021 | −0.0026$^{+}$ | −0.0033* | −0.0037$^{+}$ |
| Location: urban | 0.0097 | −0.0076 | 0.0458 | 0.0497 |

**TABLE 3.** (*Continued*)

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| Location: urban fringe | 0.1729** | 0.1558* | 0.1436* | 0.1346* |
| Location: missing | 0.0398 | 0.0232 | 0.0277 | 0.0700 |
| Barrons selectivity index | 0.0928** | 0.0720** | | |
| Student body size (1000's) | −0.0103** | −0.0106** | | |
| Student–faculty ratio | −0.0141** | −0.0132* | | |
| % Graduate students | 0.3328 | 0.3055 | | |
| Carnegie: research | −0.2067$^{+}$ | −0.1874$^{+}$ | | |
| Carnegie: doctoral | −0.2366* | −0.2227* | | |
| Carnegie: comprehensive | −0.2424** | −0.2423** | | |
| Average SAT score (100's) | | | 0.0707** | 0.0649** |
| Faculty per acre (logged) | | | 0.2622** | 0.3082** |
| Students per acre (logged) | | | −0.3541** | −0.3854** |
| Number of majors | | | −0.0026** | −0.0024* |
| % PhD degrees | | | −2.6920* | −1.9914 |
| % MA degrees | | | −0.2178 | −0.3787$^{+}$ |
| % 1st professional degrees | | | 0.7375$^{+}$ | 0.6750 |
| Variance explained: student level | 5.6% | 7.9% | 8.0% | 8.0% |
| Variance explained: school level | 65.1% | 67.7% | 67.9% | 66.5% |

Note: $^{+}p < .10$. $*p < .05$. $**p < .01$.
N for Models 1–3 = 5114; N for Model 4 = 4451.

size of the effects drops somewhat. The main difference between the two models lies at the school level. Expenditures per student is now statistically significant, while the selectivity coefficient drops in value by 22%. With the exception of selectivity, the results of the two models are quite similar.

All three SAT terms are statistically significant, and indicate that as SAT score moves from low to high, engagement increases, levels off, and increases again. Students with a high school grade point average of A and A− are more engaged in college, as are students who were engaged in extracurricular activities in high school.

Model 3 includes all the student-level, expenditure per students, and location variables, and the group of institutional variables based on the conceptual framework discussion. The coefficient for expenditures indicates that a $10,000 increase in spending per student would actually result in an overall decrease in student engagement of .03 SD, a modest but negative effect. While counterintuitive, it is line with other research showing negative relationships between expenditures and student outcomes (Smart et al., 2002; Toutkoushian and Smart, 2001). The location variables indicate that students at schools in suburban areas are slightly more engaged than students at schools in urban or rural areas.

Average SAT score is positive and significant, indicating that selectivity does have an impact on engagement. The density variables affect engagement as predicted, with faculty density increasing engagement, while student density decreases it. Density, however, is not the only impact of size; the number of majors variable is statistically significant and negative. The percentage of degrees awarded as PhDs is negative as predicted, while percentage of degrees awarded as MAs has no effect and percentage of degrees awarded as first professional has a positive effect ($p < .06$). Model 3 explains approximately the same proportion of variance as Model 2 at both levels.

Because there are some extreme values for the two density measures due to very small or very large reported campus sizes, Model 4 was estimated on a dataset that excluded schools and their students if the schools had a value on one of the two density measures above the 95th percentiles or below the 5th percentiles for the two measures. This approach tests how robust the density results are to outliers in the dataset. As can be seen, the substantive results for the two density measures remain unchanged; indeed, the effect of these variables is now slightly stronger.

Table 4 illustrates the substantive impact of the school-level variables in Model 3. It shows the actual change in student engagement, given a one standard deviation change in the independent variable, as well as the change over the entire range of the independent variable. At first glance, the effect of these variables seems small, especially given Cohen's well-known rule of thumb that an effect size of .10 is small, .30 is moderate, and .50 large. But as McCartney and Rosenthal (2000) note, researchers rarely achieve effect sizes of .50, in part because of issues of measurement error and quasi-experimental design, and more

**TABLE 4. Effects of Institutional Structures on Student Engagement**

| | Change in Variable (1 SD) | Change in Engagement | Change in Engagement, $\Delta$min to max |
|---|---|---|---|
| Average SAT score | 130 pts. | 0.09 | 0.57 |
| Faculty per acre | – | – | 2.14 |
| Students per acre | – | – | −2.70 |
| Number of majors | 28 majors | −0.07 | −0.35 |
| % PhD degrees | 2.8% | −0.07 | −0.47 |
| % 1st prof. degrees | 5.1% | 0.04 | 0.32 |

importantly, because of the context of the analysis, which must be taken into account before effects of any size can be interpreted. The context of higher education is that college student behavior and outcomes are quite difficult for institutions to change. Students enter our institutions with eighteen or more years of personal and educational experiences, and research has demonstrated that the human, social and cultural capital that students bring to college have a large impact (Perna, 2004). In this context, even small effects can be important, given the difficulty in affecting the dependent variable.

From Table 4 we can see that a student moving from a school with an average SAT score of 1000 to one with an average SAT score of 1130 would see their engagement increase by about 1/10 of a standard deviation, about the same as attending full versus part-time or reducing the number of hours worked per week by 16, while moving from the least to most selective institution would increase engagement by over one half a standard deviation.

Because logging the two density variables results in a non-linear relationship between density and engagement, the effects of these two variables are shown in Fig. 2. As faculty density increases, student engagement sharply increases and then begins to level off; as student density increases, student engagement drops sharply and then begins to level off. The impact of these variables is quite large; moving from the most to the least dense institution can change engagement by 2.1–2.7 SD. The third size-related measure, number of available majors, reduces an individual student's engagement by .07 SD for each additional 28 majors added to the curriculum.

The impact of percentage of degrees awarded as PhDs has a similar effect, −.07 SD. The range of change statistic is interesting, because schools with no PhDs are largely liberal arts and masters institutions, while those with the maximum percentage of PhDs are research universities. Mean student engagement between these institutions differs by almost 1/2 SD due to the presence of doctoral programs, while first professional programs have a smaller, positive effect (.04 SD).

## LIMITATIONS

There are several limitations to this analysis that should be kept in mind. First, unlike the very comprehensive NSSE or other student engagement surveys, I cannot examine the impact of student or institutional characteristics on specific types of engagement. The NSSE, for example, contains many items that can be used to create several more focused scales of engagement such as academic challenge (contains

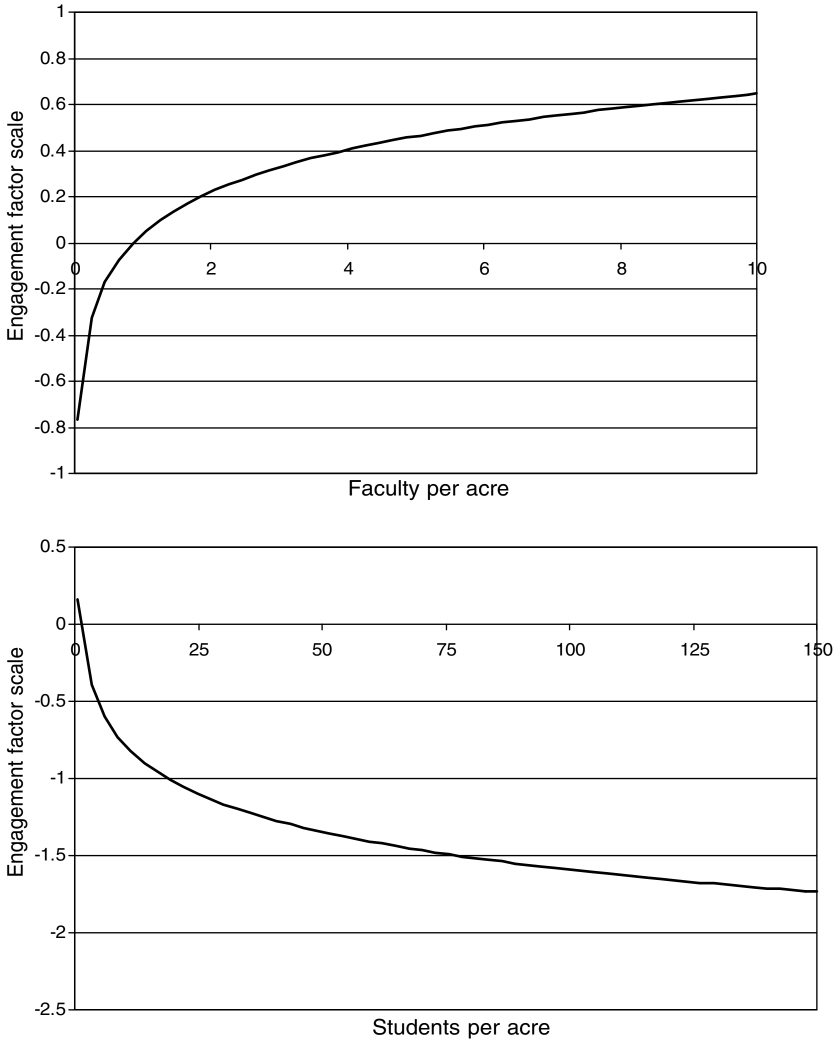


**FIG. 2.** Impact of institutional density on engagement.

eleven items), active and collaborative learning (seven items), student–faculty interaction (six items), and enriching educational experiences (12 items). In contrast, the BPS contains only seven items related to student engagement, taken from three different areas: active and collaborative learning, student–faculty interaction, and enriching educational experiences. Thus because of the design of the BPS questionnaire, I can only analyze overall engagement, as opposed to more specific areas of engagement. This is a major limitation, because the effect of institutional characteristics may vary with the type of engagement.

Second, as with almost any analysis of institutional-level effects, multicollinearity is an issue. Although the variance inflation factor is often used to assess multicollinearity, a better approach is the condition index (Belsley, 1991). For Model 3 in Table 4, the largest condition index is 1217, far higher than the recommended value of 30, and is due to the multiplicative SAT terms at the student level. Rerunning the model without the cubed and squared terms reduces the highest condition index to 109, while the model results, especially at the institutional level, remain relatively unchanged.

With this exclusion, the next two highest values for the condition index are 53 and 109, and are due to the two density variables, which is not surprising given that they share a common denominator. As Kennedy (2003) notes, there simply are not very many solutions to multicollinearity. One option is to drop the correlated variables from the model, but as student and faculty density are two of the main variables of interest, they cannot be removed from the model. In addition, dropping relevant variables can lead to specification errors. The second option is to do nothing. Generally researchers adopt this approach when the correlated variables are statistically significant; recall that with multicollinearity, fixed coefficient estimates are unbiased but variances are inflated (Kennedy, 2003; Shieh and Fouladi, 2003). Thus the major effect of multicollinearity is less powerful hypothesis testing, which is not an issue for these two variables.

Third, there is substantial missing data in the analysis. An anonymous referee suggested that with so much missing data, influential outliers might pose a problem for the analysis. Rerunning Model 3 in Table 4 after excluding the 5% of the observations with the highest values of Cook's *D* statistic resulted in very similar coefficient estimates, indicating that outliers may not be an issue for this analysis. Much of the school-level missing data is due to IPEDS non-response. Because the Department of Education has begun to sanction schools that do not submit IPEDS data, analyses using more recent data should not have this problem.

Fourth, the data are relatively old and the results of the analyses can only be generalized to FTB students in 1995. Student behavior and the interaction between students and their institutions may have changed since then. Fifth, the analysis analyzes the impact of the overall campus environment and does not examine how this affects subenvironments on campus, and how subenvironments may affect student engagement. More refined measures of institutional density are needed.

## DISCUSSION

Using the nationally representative Beginning Post-secondary Student survey, a non-linear selection on observables correction for selection bias, and a multilevel modeling approach, contrary to previous research I find that institutional structures do affect student engagement in predictable and substantively significant ways. Perhaps the most surprising finding is the limited impact that the inclusion of precollege characteristics has on model results, compared with the findings of Astin and Lee (2003). Substantive conclusions about most of the other student and school-level variables did not change with the inclusion of precollege characteristics.

The explanation for these results lies in the models estimated. While the base model estimated in this analysis did not include explicit measures of precollege characteristics, it did include student-level measures that in essence control for differences in the makeup of the student body between institutions. By including measures of student demographics and college experiences, Model 1 has in large part controlled for differences in pre-college characteristics, because these variables are correlated with pre-college characteristics.

Leaving aside the issue of ecological correlations, the primary focus of the Astin and Lee analysis is on variance explained. That is, they estimated how much additional variance was explained by controlling for a precollege characteristic, and concluded that because the variance explained increased, then much of the variation in outcomes at the institutional-level were due to student characteristics rather than institutional characteristics. Yet their regressions did not include any student-level controls such as gender, race, SES, or on-campus residence, variables that we know are strongly correlated with student outcomes (Walpole, 2003; Whitt, Pascarella, Nesheim, Marth, and Pierson, 2003). Inclusion of these variables would have undoubtedly reduced their variance explained statistics.

More importantly, Astin and Lee focus on the wrong yardstick. There are two problems with using the *R*-square to judge the impact of an independent variable. First, the *R*-square is, in general, a misleading and uninformative statistic: it cannot be compared across different dependent variables, and it cannot be compared across different datasets (King, 1986; Lewis-Beck and Skalaban, 1990). Second, and more importantly, the focus of our models is not prediction, which is what the

R-square measures. It is instead hypothesis testing. That is, we ask ourselves, can we conclude that this variable actually affects students in general? And if so (and more importantly), what is the extent of that impact? Hypothesis tests and regression coefficients can answer these questions; measures of variance explained cannot. The analyses presented here indicate that precollege characteristics can indeed increase variance explained, but their effects on our substantive conclusions may be minimal if we include other student-level measures such as student background and college experiences that are correlated with precollege characteristics.

The second finding at odds with previous research is the positive and substantive impact of selectivity on student engagement. Given the positive results for the Barron's selectivity index in Model 1, the most likely explanation is differences in samples. This result could also be the result of selection bias that was not controlled for through the non-linear selection on observables correction, but it is interesting that selection bias should also be affecting other research in the engagement literature. Future research should investigate how the samples used in higher education research may be affecting our results.

The selectivity result shows that the increased competition to gain entry into selective institutions is not simply a matter of status seeking. Student outcomes do differ if a student attends Harvard rather than a school with open admissions, and the difference is due to factors other than differences in resources. Peers exert an affect on college students, and we can see that attending school with high ability students will affect how engaged a student is. When we consider that attending a selective school means that all else being equal, a student will be more engaged, have a higher likelihood of retention (Titus, 2004) and graduate school attendance (Eide et al., 1998), and also higher salaries after graduation (Black and Smith, 2003; Brewer et al., 1999; James et al., 1989; Loury and Garman, 1995; Monks, 2000; Rumberger and Thomas, 1993; Thomas, 2000, 2003), it makes sense for students and their families to focus on admittance to the most selective school possible. The results here also fit with findings in the college choice literature, which show that students prefer to attend a school with a higher average SAT score than their own (Fuller, Manski, and Wise, 1982; Toutkoushian, 2001).

The presence of peer effects in the form of selectivity can also explain certain aspects of institutional behavior. Peer effects explain why schools care so much about the quality of the students they admit, and the curious fact that elite schools create long lines of applicants by setting their tuition far below the full cost of a year of education. They do this because student quality is a key input in their production of educational services; that is, high quality students improve an institution's education through peer effects, and institutions attract and select these students by setting tuition below costs and by failing to expand in size to meet demand (Winston, 1999; Winston and Zimmerman, 2004).

The graduate student result is interesting, because it is another piece of evidence that implies faculty research does indeed come at the expense of the undergraduate experience. While the post-WWII development of the American research university has made America preeminent in the sciences (Graham and Diamond, 1997), the emphasis on research and the economy of scales necessary to maintain research universities may come at the expense of the undergraduate experience. While this is not a new criticism of research universities (e.g., Brooks, 1994), the evidence presented in this paper is new, and raises fresh questions as to the best way to design our institutions. Most importantly, the finding that doctoral programs have a negative effect on student engagement, rather than master's or first-professional programs, indicates that it is institutional emphasis on research rather than the presence of graduate students that leads to decreased engagement at the undergraduate level.

While abolishing doctoral programs is not a feasible strategy for institutions, changing the faculty reward structure is. The percentage of time faculty spend on research and research productivity have been increasing over the past several decades (Dey, Milem, and Berger, 1997; Milem, Berger, and Dey, 2000), and there appears to be little evidence that this is changing. Even faculty at traditional liberal arts institutions now face expectations of publishing. While faculty research productivity may lead to increased institutional prestige (Porter and Toutkoushian, in press), it may also negatively impact student outcomes. When institutions explicitly reward faculty publishing activity but not activities such as student–faculty interaction, it should come as no surprise that student engagement suffers.

In terms of future research, it is clear that more detailed information about faculty behavior is needed to understand the relationship between the research emphasis of an institution and student engagement. Degrees awarded by degree type can only crudely proxy differences in faculty behavior between institutions. More ambitious research programs, such as the combination of the Faculty Survey of Student Engagement with the NSSE, will undoubtedly yield greater insights into the relationship between faculty behavior and the level of student engagement at an institution.

The two measures of density proposed here operationalize the concepts of size and student–faculty interaction in a more theoretically meaningful way than traditional measures such as number of students and student–faculty ratio. Both faculty per acre and students per acre showed strong, opposite effects on engagement as predicted. In terms of future research, these findings indicate that our focus should not only be on the effects of size, but also of density. How exactly does density affect student and faculty interactions on a campus? How do internal aspects of an institution, such as residence halls, affect engagement? Future research should consider more refined measures of density using different settings within an institution rather than the macro measures used in this analysis.

The strong effect of size also shows the need for honor colleges and learning communities within large universities. If engagement is promoted by smaller, more intimate surroundings such as learning communities (Zhao and Kuh, 2004), creating smaller structures within larger institutions is one way to reap the benefits of smallness. Yet is should be emphasized that the positive effect of learning communities may be due not to their small size, but to their providing a less dense student environment and more dense faculty environment within an larger institutional environment of greater student density and lesser faculty density.

The negative impact of curriculum differentiation (proxied by number of majors) raises serious questions as to how our institutions should be structured. Given that all institutions offer courses in the liberal arts, one interpretation of this result is that as institutions move away from a liberal arts curriculum, student engagement suffers. Compared with selectivity, density, and the presence of doctoral programs, the curriculum is probably the easiest characteristic for an institution to change. Yet institutions face pressures to move in the opposite direction, as increasing the number of professional majors may yield more tuition revenue than increasing liberal arts majors. It is no surprise that the for-profit University of Phoenix has focused on professional baccalaureate degree programs rather than the liberal arts, while one analysis found that over two-thirds of institutions claiming to focus primarily on the liberal arts (through college catalog announcements) awarded a majority of their degrees in professional fields (Delucchi, 1997). While the curriculum may appear easy to change, financial pressures may limit institutional abilities to do so.

Finally, the findings here shed light on recent research that student outcomes due differ by institution type, and that liberal arts colleges tend to have better student outcomes than larger institutions (Pascarella et al., 2004; Siegfried and Getz, 2003). The standard argument about the differences between these institutions has focused on faculty research and the presence of graduate students (Cech, 1999), but the results presented here indicate that density and the differentiation of the curriculum may also explain why student outcomes differ between liberal arts colleges and research universities.

## ENDNOTES

1. The institutional participation rate was calculated by dividing the 476 four-year institutions reported in the NSSE 2003 Overview by the 1533 doctoral, master's and baccalaureate institutions reported in the 2003 Digest of Education Statistics (National Center for Education Statistics, 2004).
2. Because the BPS uses the National Post-secondary Student Aid Study as its base sample, the response rates cited are from that study.

## REFERENCES

Allison, P. D. (2002). *Missing Data*, Sage Publications, Thousand Oaks, CA.

Antonio, A. L. (2001). Diversity and the influence of friendship groups in college. *Review of Higher Education* 25(1): 63–89.

Antonio, A. L. (2004). The influence of friendship groups on intellectual self-confidence and educational aspirations in college. *Journal of Higher Education* 75(4): 446–471.

Astin, A. W. (1993). *What Matters in College? Four Critical Years Revisited*, Jossey-Bass, San Francisco, CA.

Astin, A. W., and Lee, J. J. (2003). How risky are one-shot cross-sectional assessments of undergraduate students? *Research in Higher Education* 44(6): 657–672.

Barron's., (1997). *Barron's Profiles of American colleges*, (21st Ed.), Barron's Educational Series, Inc, Hauppauge, NY.

Belsley, D. (1991). *Conditioning Diagnostics: Collinearity and Weak Data in Regression*, John Wiley and Sons, New York.

Black, D. A., and Smith, J. A. (2003). How robust is the evidence on the effects of college quality? *Evidence from matching*. *Journal of Econometrics* 121: 99–124.

Breland, H., Maxey, J., Gernand, R., Cumming, T., and Trapani, C. (2002). *Trends in College Admission 2000: A Report of a National Survey of Undergraduate Admission Policies, Practices, and Procedures*: ACT, Inc., Association for Institutional Research, the College Board, Educational Testing Service, and the National Association for College Admission Counseling.

Brewer, D. J., Eide, E., and Ehrenberg, R. G. (1999). Does it pay to attend an elite private college? Cross-cohort evidence on the effects of college type on earnings. *Journal of Human Resources* 43(1): 104–123.

Brooks, H. (1994). Current criticisms of research universities. In: Cole, J. R., Barber, E. G., and Graubard, S. R. (eds.), *The Research University in a Time of Discontent*, Johns Hopkins University Press, Baltimore, MD.

Carnegie Foundation for the Advancement of Teaching. (1994). *A Classification of Institutions of Higher Education*, Princeton, NJ.

Cech, T. R. (1999). Science at liberal arts colleges: A better education? *Daedelus* 128(1): 195–216.

Chickering, A. W., and Reisser, L. (1993). *Education and Identity* (2nd Ed.), Jossey-Bass, San Francisco.

Dale, S. B., and Krueger, A. B. (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *Quarterly Journal of Economics* 117(4): 1491–1527.

Delucchi, M. (1997). "Liberal arts" colleges and the myth of uniqueness. *Journal of Higher Education* 68(4): 414–426.

Dey, E. L., Milem, J. F., and Berger, J. B. (1997). Changing patterns of publication productivity: Accumulative advantage or institutional isomorphism? *Sociology of Education* 70(4): 308–323.

Dills, A. K. (2004). Does cream-skimming curdle the milk? A study of peer effects. *Economics of Education Review* 24(1): 19–28.

Ehrenberg, R. G. (2004). Econometric studies of higher education. *Journal of Econometrics* 121: 19–37.

Eide, E. R., Brewer, D. J., and Ehrenberg, R. G. (1998). Does it pay to attend an elite private college? Evidence on the effects of undergraduate college quality on graduate school attendance. *Economics of Education Review* 17(4): 371–376.

Fairweather, J. S. (1996). *Faculty Work and Public Trust: Restoring the Value of Teaching and Public Service in American Academic life*, Allyn and Bacon, Boston.

Fairweather, J. S. (2002). The mythologies of faculty productivity: Implications for institutional policy and decison making. *Journal of Higher Education* 73(1): 26–48.

Fox, M. F. (1992). Research, teaching and publication productivity: Mutuality versus competition in academia. *Sociology of Education* 65(4): 293–305.

Fuller, W. C., Manski, C. F., and Wise, D. A. (1982). New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources* 17(4): 477–498.

Gibson, N. M., and Olejnik, S. (2003). Treatment of missing data at the second level of hierarchical linear models. *Educational and Psychological Measurement* 63(2): 204–238.

Graham, H. D., and Diamond, N. (1997). *The Rise of American Research Universities: Elites and Challengers in the Postwar Era*, Johns Hopkins University Press, Baltimore, MD.

Groves, R. M., Fowler, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., and Tourangeau, R. (2004). *Survey Methodology*, Wiley, Hoboken, NJ.

Heck, R. H., and Thomas, S. L. (2000). *An Introduction to Multilevel Modeling Techniques*, Lawrence Erlbaum Associates, Mahwah, NJ.

Hoxby, C. M. (2001). *Peer Effects in the Classroom: Learning from Gender and Race Variation*. National Bureau of Economic Research.

Hu, S., and Kuh, G. D. (2002). Being (dis)engaged in educationally purposeful activities: The influences of student and institutional characteristics. *Research in Higher Education* 43(5): 555–574.

Hu, S., and Kuh, G. D. (2003a). Diversity experiences and college student learning and personal development. *Journal of College Student Development* 44(3): 320–334.

Hu, S., and Kuh, G. D. (2003b). Maximizing what students get out of college: Testing a learning productivity model. *Journal of College Student Development* 44(2): 185–203.

James, E., Alsalam, N., Conaty, J. C., and To, D.-L. (1989). College quality and future earnings: Where should you send your child to college? *American Economic Review* 79(2): 247–252.

Johnson, M. K., Crosnoe, R., and Elder, G. H. (2001). Students' attachment and academic engagement: The role of race and ethnicity. *Sociology of Education* 74: 318–340.

Kennedy, P. (2003). *A Guide to Econometrics*, The MIT Press, Cambridge, MA.

Kim, M. M. (2002a). Cultivating intellectual development: Comparing women-only colleges and coeducational colleges for educational effectiveness. *Research in Higher Education* 43(4): 447–481.

Kim, M. M. (2002b). Historically Black vs. White institutions: Academic development among Black students. *Review of Higher Education* 25(4): 385–407.

King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science* 30(3): 666–687.

Kuh, G. D. (2003). *The National Survey of Student Engagement: Conceptual framework and overview of psychometric properties*, Indiana University Center for Postsecondary Research, Bloomington, IN.

Kuh, G. D., and Hu, S. (2001). The effects of student–faculty interaction in the 1990s. *Review of Higher Education* 24(3): 309–332.

Lee, V. E., and Smith, J. B. (1997). High school size: Which works best and for whom? *Educational Evaluation and Policy Analysis* 19(3): 205–227.

Lee, V. E., Smith, J. B., and Croninger, R. G. (1997). How high school organization influences the equitable distribution of learning in mathematics and science. *Sociology of Education* 70(2): 128–150.

Lewis-Beck, M. S., and Skalaban, A. (1990). The R-squared: Some straight talk. *Political Analysis* 2: 153–170.

Loury, L. D., and Garman, D. (1995). College selectivity and earnings. *Journal of Labor Economics* 13(2): 289–308.

Lundberg, C. A. (2003). The influence of time-limitations, faculty, and peer relationships on adult student learning: a causal model. *Journal of Higher Education* 74(6): 665–688.

Lundberg, C. A., and Schreiner, L. A. (2004). Quality and frequency of faculty-student interaction as predictors of learning: An analysis by student race/ethnicity. *Journal of College Student Development* 45(5): 549–565.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60: 531–542.

Marco, G. L., Abdel-fattah, A. A., and Baron, P. A. (1992). *Methods Used to Establish Score Comparability on the Enhanced ACT Assessment and the SAT*, College Board, New York, NY.

McCartney, K., and Rosenthal, R. (2000). Effect size, practical importance, and social policy for children. *Child Development* 71(1): 173–180.

McCormick A. C. (2000). What's ahead for the Carnegie classifications? *AAHE Bulletin*.

McCormick A. C. (2004). The 2005 revision of the Carnegie classification system. *Paper presented at the Washington Higher Education Secretariat*, Washington, DC.

Milem, J. F., Berger, J. B., and Dey, E. L. (2000). Faculty time allocation: A study of change over twenty years. *Journal of Higher Education* 71(4): 454–475.

Milem, J. F., Umbach, P. D., and Liang, C. T. H. (2004). Exploring the perpetuation hypothesis: The role of colleges and universities in desegregating society. *Journal of College Student Development* 45(6): 688–700.

Monks, J. (2000). The returns to individual and college characteristics: Evidence from the National Longitudinal Survey of Youth. *Economics of Education Review* 19: 279–289.

National Center for Education Statistics. (1990). *Classification of Instructional Programs*, Washington, DC.

National Center for Education Statistics. (1997). *National Postsecondary Student Aid Study, 1995–96 (NPSAS:96) Methodology Report (No. 98-073)*, Washington, DC.

National Center for Education Statistics. (2004). *Digest of Education Statistics 2003* (No. 2005-025). Washington, DC.

National Commission on Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform*, Washington, DC.

National Survey of Student Engagement (2003). *2003 Overview*, Indiana University Center for Postsecondary Research, Bloomington, IN.

Pascarella, E. T. (1985). College environmental influences on learning and cognitive development: A critical review and synthesis. In: Smart, J. (ed.), *Higher Education Handbook of Theory and Research* (Vol. 1), Agathon, Edison, NJ.

Pascarella, E. T., Cruce, T. M., Umbach, P. D., Wolniak, G. C., Kuh, G. D., Carini, R. M., et al. (in press). Institutional selectivity and good practices in undergraduate education: How strong is the link? *Journal of Higher Education*.

Pascarella, E. T., and Terenzini, P. T. (1991). *How College Affects Students: Findings and Insights from Twenty Years of Research*, Jossey-Bass, San Francisco.

Pascarella, E. T., and Terenzini, P. T. (2005). *How College Affects Students: A Third Decade of Research* (Vol. 2), Jossey-Bass, San Francisco.

Pascarella, E. T., Wolniak, G. C., Cruce, T. M., and Blaich, C. F. (2004). Do liberal arts colleges really foster good practices in undergraduate education? *Journal of College Student Development* 45(1): 57–74.

Perna, L. W. (2004). Understanding the decision to enroll in graduate school: Sex and racial/ethnic group differences. *Journal of Higher Education* 75(5): 487–527.

Peterson's., (1996). *Peterson's Undergraduate Database*, The Thomson Corporation, Lawrenceville, NJ.

Porter, S. R. (2005). What can multilevel models add to institutional research? In: Coughlin, M. A. (ed.), *Applications of Advanced Statistics in Institutional Research*, Association of Institutional Research, Tallahassee, FL.

Porter, S. R. and Toutkoushian, R. K. (in press). Institutional research productivity and the connection to average student quality and overall reputation. *Economics of Education Review*.

Raudenbush, S. W., and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2nd Ed.), Sage Publications, Thousand Oaks, CA.

Renn, K. A., and Arnold, K. D. (2003). Reconceptualizing research on college student peer culture. *Journal of Higher Education* 74(3): 261–291.

Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *Sociological Review* 15: 351–357.

Rumberger, R. W., and Thomas, S. L. (1993). The economic returns to college major, quality and performance: A multilevel analysis of recent graduates. *Economics of Education Review* 12(1): 1–19.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *Quarterly Journal of Economics* 116(2): 681–704.

Shieh, Y.-Y., and Fouladi, R. T. (2003). The effect of multicollinearity on multilevel modeling parameter estimates and standard errors. *Educational and Psychological Measurement* 63(6): 951–985.

Shulman, J. L., and Bowen, W. G. (2001). *The Game of Life: College Sports and Educational Values*, Princeton University Press, Princeton, NJ.

Siegfried, J. J., Getz, M. (2003). *Where Do the Children of Professors Attend College?*, Unpublished manuscript, Department of Economics, Vanderbilt University.

Smart, J. C., Ethington, C. A., Riggs, R. O., and Thompson, M. D. (2002). Influences of institutional expenditure patterns on the development of students' leadership competencies. *Research in Higher Education* 43(1): 115–132.

Strauss, L. C., and Volkwein, J. F. (2002). Comparing student performance and growth in 2- and 4-year institutions. *Research in Higher Education* 43(2): 133–161.

Thomas, S. L. (2000). Deferred costs and economic returns to college major, quality, and performance. *Research in Higher Education* 41(3): 281–313.

Thomas, S. L. (2003). Longer-term economic effects of college selectivity and control. *Research in Higher Education*. 44(3): 263–299.

Thomas, S. L., and Heck, R. H. (2001). Analysis of large-scale secondary data in higher education research: Potential perils associated with complex sampling designs. *Research in Higher Education* 42(5): 517–540.

Titus, M. (2004). An examination of the influence of institutional context on student persistence at 4-year colleges and universities: A multilevel approach. *Research in Higher Education* 45(7): 673–699.

Toutkoushian, R. K. (2001). Do parental income and educational attainment affect the initial choices of New Hampshire's college-bound students. *Economics of Education Review* 20: 245–262.

Toutkoushian, R. K., and Smart, J. C. (2001). Do institutional characteristics affect student gains from college? *Review of Higher Education* 25(1): 39–61.

Walpole, M. (2003). Socioeconomic status and college: How SES affects college experiences and outcomes. *Review of Higher Education* 27(1): 45–73.

Whitt, E. J., Pascarella, E. T., Nesheim, B. S. E., Marth, B. P., and Pierson, C. T. (2003). Differences between wome and men in objectively measured outcomes, and the factors that influence those outcomes, in the first three years of college. *Journal of College Student Development* 44(5): 587–610.

Winston, G. C. (1999). Subsidies, hierarchy and peers: The awkward economics of higher education. *Journal of Economic Perspectives* 13(1): 13–36.

Winston, G. C., and Zimmerman, D. J. (2004). Peer effects in higher education. In: Hoxby, C.M. (ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, University of Chicago Press, Chicago.

Zhao, C.-M., and Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education* 45(2): 115–138.

Zhao, C.-M., Kuh, G. D., and Carini, R. M. (2005). A comparison of international student and American student engagement in effective educational practices. *Journal of Higher Education* 76(2): 209–231.

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics* 85(1): 9–23.

Received May 3, 2005.