REVIEW OF NCES RESEARCH ON FINANCIAL AID AND COLLEGE PARTICIPATION

AND

OMITTED VARIABLES AND SAMPLE SELECTION ISSUES IN THE NCES RESEARCH ON FINANCIAL AID AND COLLEGE PARTICIPATION

REPORTS PREPARED FOR THE ADVISORY COMMITTEE ON STUDENT FINANCIAL ASSISTANCE BY

DONALD E. HELLER
ASSOCIATE PROFESSOR AND SENIOR RESEARCH ASSOCIATE
CENTER FOR THE STUDY OF HIGHER EDUCATION
THE PENNSYLVANIA STATE UNIVERSITY

AND

WILLIAM E. BECKER
PROFESSOR OF ECONOMICS
INDIANA UNIVERSITY – BLOOMINGTON

SEPTEMBER 2003

INTRODUCTION

The Advisory Committee was created by the Higher Education Amendments of 1986 to serve as an independent source of advice and counsel to Congress and the Secretary of Education on student financial aid policy. The most important statutory charge of the Advisory Committee is to make recommendations that will lead to the maintenance and enhancement of access to postsecondary education for low- and middle-income students. This congressional mandate specifically requires the Committee not only to review and assess legislation, policy proposals, and regulations that impact access to the federal student assistance programs, but also to "recommend to Congress and the Secretary . . . studies, surveys, and analyses of student financial assistance programs, policies, and practices," particularly those that affect the needs of low-income students.

Since 1999 the Advisory Committee has been engaged in an effort to assess the condition of access to postsecondary education for low- and middle-income students in preparation for the reauthorization of the Higher Education Act of 1965. In pursuit of this goal, the Committee has held public meetings on access at the University of Mississippi, Boston University, the University of Vermont, and the University of Texas at Brownsville, as well as participated in a symposium on access sponsored by the Harvard Graduate School of Education. These forums enabled the Committee to hear testimony from students, administrators, researchers, and others, as well as to gather research and information critical to fulfilling the Committee's congressional mandate.

As Congress prepares to reauthorize the Higher Education Act, changes to which could affect access to postsecondary education for tens of thousands of young Americans, the Advisory Committee finds that the gap in college participation rates among students from different socioeconomic backgrounds remains one of the most pressing issues in education and social policy. The Advisory Committee's most recent reports on access, *Access Denied: Restoring the Nation's Commitment to Equal Educational Opportunity* and *Empty Promises: The Myth of College Access in America*, demonstrate that excessive unmet need is a critical factor in the decision-making process of even college-qualified low-income students. Furthermore, the Committee's analysis finds that failing to offer access to college-qualified students represents a significant loss to the national economy over the next several decades.

The United States invests in higher education—in human capital—because the potential economic benefits, such as increased productivity, a flexible workforce able to respond to a changing economy, and an increased standard of living for workers, are important both for the nation and its citizens. This investment also produces an educated electorate and a more informed democracy. These returns have motivated a very large federal investment in student aid since 1965, designed to ensure that students who otherwise could not afford to attend college have the financial resources to enroll and persist through degree completion.

This investment, while essential to all students, is most critical for low- and middle-income high school graduates who are academically prepared to meet the admissions criteria of a four-year college based on completed curriculum, grades, class rank, and test scores. For these students, a

shortage of family financial resources constitutes the most important barrier to college. This is the rationale for the creation of the federal student aid programs. This is the reason federal and state policies that promote access to college for low- and middle-income students are essential in an era of educational reform.

What can be done to ensure that reasonably qualified young Americans will have the opportunity to attain a bachelor's degree now and in the future is one of the most pressing questions facing Congress in the next reauthorization. Good policymaking requires good research, and the Committee has undertaken its review and assessment of the condition of college access in order to provide Congress with detailed and well-researched analyses and recommendations.

After reviewing recent federally funded access research for its two most recent reports, the Committee identified inconsistencies and contradictions in this research, and felt that a systematic review of the most important federal access studies was warranted. Consequently, the Committee commissioned the two papers that follow:

- First, Dr. Donald Heller, Associate Professor and Senior Research Associate with the Center for the Study of Higher Education at the Pennsylvania University, reviews recent key access studies by NCES, and finds that four serious statistical errors led to faulty inferences that could mislead policymakers.

- Second, Dr. William Becker, Professor of Economics, Indiana University – Bloomington, provides an econometric assessment of the errors identified by Heller, and shows how the data must be reanalyzed in order to yield valid conclusions.

Drs. David Breneman, Michael McPherson, Morton Owen Schapiro, and Sandy Baum served as reviewers of the two papers. The papers were presented at a colloquium on access research at Macalester College on June 12, 2003. (See Appendix A for the colloquium agenda, participants and additional reviewers.)

Taken together, these two papers reflect a careful methodological assessment of recent NCES access research and the degree to which the statistical errors uncovered affect the validity of the studies' findings. Several major implications can be drawn for future research and policy:

- Financial aid (or some measure of net price) must be included as an independent variable or covariate in multivariate analyses purporting to estimate the likelihood of college enrollment and persistence. Otherwise, valid inferences about the effects of family income, and other variables related to family income such as parents' education and academic preparation, are impossible.

- Equal access cannot be defined as equality in the rates of enrollment of college-qualified low- and middle-income high school graduates who tested and applied to a four-year college. This definition screens out over three-quarters of low-income students and, accordingly, most of the effects of family income and financial aid. Such a definition constitutes selection bias and leads inevitably to faulty inferences and policy conclusions.

- The effects of family income and financial aid on student academic qualifications and educational expectations and plans must be accounted for explicitly in analyses of access. Failure to account for these effects, readily observable in national data, results in a significant underestimation of the effects of finances on access and persistence.

- The effects of family income and financial aid on taking the steps of testing for and applying to a four-year college must also be accounted for explicitly. Large, income-related differences in four-year college enrollment cannot be explained away by simply observing that many college-qualified low-income high school graduates do not test or apply without ascertaining why. Rather, the decision not to test and apply must be interpreted as a rational economic response of low-income high school graduates confronted by a severe shortage of financial aid and record-high high net prices.

- Lastly, in addition to the effects of family income and financial aid, the effects of differences between low- and middle-income students' expected rate of return to college must be accounted for explicitly. Such differences can exacerbate the effects of financial aid shortages and record-high net prices.

It is imperative that in future research on these issues the rich data maintained by NCES be reanalyzed using more sophisticated analytical models and appropriate statistical techniques capable of estimating the effects of student aid on enrollment and persistence.

REVIEW OF NCES RESEARCH ON FINANCIAL AID AND COLLEGE PARTICIPATION

ADVISORY COMMITTEE ON STUDENT FINANCIAL ASSISTANCE

MARCH 2003

REVIEW OF NCES RESEARCH ON FINANCIAL AID AND COLLEGE PARTICIPATION

Report Prepared for the Advisory Committee on Student Financial Assistance

by Donald E. Heller
Associate Professor and Senior Research Associate
Center for the Study of Higher Education
The Pennsylvania State University

March 2003

EXECUTIVE SUMMARY

In recent years the National Center for Education Statistics (NCES), the statistical data collection and analysis section of the U.S. Department of Education, has issued a series of reports focusing on access to postsecondary education. Using a number of analytical tools and textual descriptions, these reports describe the relationship between data on student and institutional characteristics and postsecondary outcome data, such as student enrollment in college, type of college attended, and persistence to degree attainment. At the request of the Advisory Committee on Student Financial Assistance, this study examines in detail four primary NCES reports that focus on college participation, and analyzes the nature of the conflict between the results of the tabular and multivariate analyses in those reports. It also compares those results with the findings of other research on college access.

This study found much agreement between the findings of the NCES tabular analyses, which are primarily descriptive, and previous research on college access. As most research concludes, there is a strong relationship between family income and college participation; lower-income students are less likely to attend college than their peers from wealthier families and when they do, they are less likely to be enrolled in a four-year institution. Similarly, students whose parents had higher levels of educational attainment were more likely to enroll in college and persist to degree once there.

However, the findings from the NCES multivariate analyses, which are causal in nature and from which policy inferences are drawn, provide contradictory evidence of the relationship between family income, parental education, and college participation. While the tabular analyses show great differences in college participation among students from different income groups, as well as among students whose parents have differing levels of educational attainment, the multivariate analyses purport to show that these outcomes are greatly minimized or even eliminated when controls for other factors are included. This conclusion of the multivariate analyses is in conflict with a large body of economic research on college access.

This study's key conclusion is that the discrepancies between NCES tabular and multivariate analyses, and between NCES multivariate analyses and previous research are due to four methodological errors made in the studies, described as follows and displayed in Figure 6.

- **Omitted Variable Bias**: In none of the studies was total financial aid or any measure of net price used as an explanatory variable in the multivariate analyses. This led to a systematic underestimate of the effect of family income on enrollment and persistence.

- **Selection Bias**: In the first and most important study, a college-qualification index was created that, together with the additional condition that students must have tested for and applied to a four-year college, screened out the majority of low-income students in the sample and, accordingly, eliminated the effects of family income and financial aid.

- **Endogeneity Bias**: In three of the studies, the effects of family income and financial aid on college qualification and certain academic behaviors such as testing and applying were

not considered, although they were readily apparent in the study data, leading once again to a statistical underestimate of their effects on enrollment and persistence.

- **Multicollinearity:** In none of the studies were the strong relationships between family income and most other explanatory variables, including parents' education, adequately explored or taken into consideration before drawing conclusions about the relative importance of family income. This methodological error likely contributed directly to a finding that family income was statistically insignificant in two of the studies.

While a reanalysis of the data is required to fully assess the consequences of the above errors, faulty methodology almost certainly led NCES to systematically underestimate the effect of family income and financial aid on the enrollment and persistence of low-income students. Consequently, these studies may mislead policymakers toward erroneous conclusions regarding the role of family income and parental education in determining college participation. In turn, these faulty conclusions could inadvertently undermine the demand for policies designed to reduce the gap in college participation among students from different socioeconomic groups.

This study concludes with some suggestions for improving NCES research on college access in order to address the methodological issues identified. In addition, recommendations for further research on the topic are included.

TABLE OF CONTENTS

INTRODUCTION

In recent years the National Center for Education Statistics (NCES) has issued a series of reports focusing on access to postsecondary education. Using a number of analytical tools and textual descriptions, these reports describe the relationship between a number of factors, such as student and institutional characteristics, and other factors, such as whether students enroll in college, what type of college they attend, and whether they persist to degree attainment.

Some of the NCES findings provide contradictory evidence of the relationship between financial characteristics, including family income, tuition prices, and the availability of financial aid, and the college participation rate of low-income students. This is described in the request from the Advisory Committee on Student Financial Assistance (Advisory Committee) that forms the basis of this report:

> The Advisory Committee is interested in analyzing a significant conflict between the descriptive and causal analyses of access and persistence by the National Center for Education Statistics. Of particular interest to the Committee is why NCES's *tabular* analyses, which are descriptive, are consistent with the widespread and well-documented conclusion that financial aid matters greatly to the enrollment and persistence of low-income high school graduates, but NCES *multivariate* analyses, which are causal in nature, contradict that conclusion.

With Congress and the Administration due to take up the reauthorization of the Higher Education Act of 1965 (HEA) in the next year, it is important to understand this conflict. Title IV of the HEA authorizes the federal student aid programs, including the Pell and Supplementary Educational Opportunity Grants, the Perkins, Ford Direct, and Family Education Loan programs, and the College Work Study program. These programs together made available over $54 billion in aid to students in the 2001-2002 academic year, or 61 percent of all financial aid (College Board, 2002).[1] Thus, the federal government makes a substantial investment and has a vested interest in ensuring that these funds are used to promote the goals articulated in HEA over 35 years ago:

> It is the purpose of this part to provide, through institutions of higher education, educational opportunity grants to assist in making available the benefits of higher education to qualified high school graduates of exceptional financial need, who for lack of financial means of their own or of their families would be unable to obtain such benefits without such aid ("Higher Education Act of 1965," 1965).

The goal of this study is to examine in detail four primary NCES reports that focus on college participation, and analyze the nature of the conflict, if any, between the results of the tabular and multivariate analyses in these reports, as well as between those results and the findings of other researchers on college access.[2] The specific questions addressed in this report include:

- whether the NCES tabular analyses in these studies are consistent with prevailing theoretical and empirical views of the importance of family income, net price,

unmet need, and financial aid on college-going behavior of low-income high school graduates;

- whether the NCES multivariate analyses and findings are consistent with those views and the NCES tabular analyses; and

- whether the NCES multivariate analyses and findings are methodologically sound.

The results of this study will help explain this conflict, as well as provide recommendations for further research that can help explain the effectiveness of financial aid in promoting college participation for students with different socioeconomic characteristics.

Following this introduction, the second section provides a brief overview of the existing empirical research, other than the four NCES reports, on the relationship between a number of variables and college participation. The third section summarizes the findings of the NCES tabular analyses, and compares those findings to other research on college participation. The fourth section summarizes the findings of the NCES multivariate analyses, and in similar fashion, compares those findings to other research on college participation.

The fifth section of the report provides a more detailed critique of the methodological approach used by NCES in its multivariate analyses, with a focus on understanding why the results conflict with the findings in the tabular analyses. The sixth and final section makes some recommendations for future research that could provide more evidence of the role of financial aid in promoting college participation.

PREVIOUS RESEARCH ON COLLEGE PARTICIPATION

A broad body of research on college participation in the United States exists, much of which focuses on the relationship between student and family characteristics and the decision to enroll in and persist through college. Many of these studies also examine the role of financial aid in helping students to overcome the cost barriers that inhibit them from participating in postsecondary education. This section briefly summarizes this research and its key findings.

Tabular Analyses

Concerns over the relationship between family financial resources and college participation long preceded the HEA. In a study in the early part of the 20th century, Morey (1928) described the potential discouragement effect on college enrollment of fees charged at public institutions, and the need for financial aid to overcome that effect. After World War II, President Harry Truman's Commission on Higher Education expressed concerns similar to those addressed nearly twenty years later in the HEA: "For the great majority of our boys and girls, the kind and amount of education they may hope to attain depends, not on their own abilities, but on the family or community into which they happened to be born or, worse still, on the color of their skin or religion of their parents" (President's Commission on Higher Education, 1947).

Analyses of the college participation rates of students from different socioeconomic groups has documented the long-standing gaps that exist among these groups. These gaps have persisted at all measurement points of the postsecondary education attainment process: in high school graduation rates,[3] in college entry rates, in college persistence rates, and in degree attainment rates. For example, while students from all income groups have seen gains in these rates in the three decades since passage of the HEA, the gaps between high-income and low-income students have stubbornly persisted.

Thomas G. Mortenson, in his newsletter *Postsecondary Education OPPORTUNITY*, has long tracked these gaps using Census Bureau data. His most recent analysis (Figure 1), describes the gaps at some of these points (Mortenson, 2001b). Low-income students are less likely to continue through each point in the educational pipeline, and the gap between low- and high-income students increases through later stages of the pipeline. The Census Bureau's own reports confirm these gaps; for example, the high school dropout rate of students from families making less than $20,000 in 1999 was more than three times the rate of students from families making over $40,000 (United States Bureau of the Census, 2001).

Figure 1: Educational attainment rates of highest and lowest income quartiles groups, 2000.

Measure	Highest Income Quartile	Lowest Income Quartile	Difference
High school graduation	92%	65%	27 points
College entry from high school	82%	54%	28 points
Bachelor's degree attainment	52%	7%	45 points

Dependent students age 18 to 24
Source: Mortenson (2001b)

The gaps in postsecondary participation and attainment found among individuals from different income groups are found also when students with other socioeconomic characteristics are compared. Race and ethnicity is another important correlate of educational attainment. Mortenson (2001a) also examines this relationship and reports large differences in college participation among the groups. Among dependent 18 to 24 year olds, the college participation rates are: whites, 64 percent; Asian/Pacific Islanders, 78 percent; blacks, 46 percent; and Hispanics, 40 percent.[4] As with the gaps in participation among different income groups, these racial/ethnic gaps have similarly persisted for decades (Clotfelter, 1991; Heller, 1999; Koretz, 1990). That Asian American and white families have much higher incomes, on average, than black and Hispanic families is, no doubt, an important factor in explaining the racial gap in college participation; thus, the strong correlation between race and income in the United States.

Mortenson (1999b) also documented the relationship between parental education levels and college participation. As the education level of a student's mother, father, or guardian increases, the probability that the student would enroll in college also increases.

Multivariate Analyses of the Relationship Between Socioeconomic Status and College Participation

There are strong relationships among the various measures that are often included under the label "socioeconomic status." As described above, race and income are strongly correlated in the United States. Similarly related are the relationship between educational attainment and income. As has been documented in numerous reports, people with higher levels of education earn more money.[5] And as has been described above, students from families with more money are more likely to participate in college.

In order to separate out the effects of these collinear relationships, numerous researchers have conducted multivariate analyses of the relationship between socioeconomic status and college participation. Using a number of factors, including such measures as race, family income, pre-college academic achievement, parental education, tuition prices, and financial aid offers,

researchers attempt to gauge the effect each has on the probability that a given individual will enroll in or persist through college.

The results of these studies are fairly consistent; controlling for other factors, researchers generally have found the following factors related to college participation:

- higher levels of family income are related to a higher probability of college participation (Jackson, 1989; Manski & Wise, 1983; St. John, 1991);

- higher levels of parental education are related to a higher probability of college participation (Ellwood & Kane, 2000; Jackson, 1989); and

- higher levels of academic achievement are related to a higher probability of college participation (Behrman, Kletzer, McPherson, & Schapiro, 1992; Rouse, 1994; St. John, 1990).

It is important to note that the findings regarding the relationship between socioeconomic status and college access hold even when the student's academic preparation is taken into account. In an analysis of data from the National Education Longitudinal Study (NELS) of 1988, Kane (1999) divided students into quartiles based on their score on math tests administered as part of that study. Even for students in the top test score quartile, i.e., those who were the most academically qualified, he found a large gap in the probability that the student would enroll in college when comparing those from the lowest family income group and those in the highest family income group. This gap between rich and poor still existed when he used class rank as the indicator of academic achievement, rather than test scores.[6]

In addition to these findings, researchers have examined the relationship between tuition prices, financial aid, and postsecondary participation. Three reviews of this literature in the last three decades (Heller, 1997; Jackson & Weathersby, 1975; Leslie & Brinkman, 1988), which cumulatively examined over 150 studies, have reached the following conclusions:

- the college participation rate of low-income students is most responsive to increases in tuition prices; high-income students show little responsiveness to higher tuition prices in their college *entry* decisions, though their college *choice* decisions can be influenced by changing prices; and

- the awarding of financial aid, and, in particular, grants, is related to higher probability of college participation for low-income students (i.e., grant aid can offset at least some of the impact of rising tuition prices); as income increases, the enrollment responsiveness of students to financial aid offers decreases. Financial aid can affect the college *choice* decisions of higher income students, however.

REVIEW OF NCES TABULAR ANALYSES

This review examines the following NCES reports on college access:

1. Berkner, L., & Chavez, L. (1997). *Access to postsecondary education for the 1992 high school graduates* (NCES 98-105).

2. Choy, S. P. (2001). *Students whose parents did not go to college: Postsecondary access, persistence, and attainment* (NCES 2001-126).

3. Horn, L., & Nuñez, A.-M. (2000). *Mapping the road to college: First-generation students' math track, planning strategies, and context of support* (NCES 2000-153).

4. Wei, C. C., & Horn, L. (2002). *Persistence and attainment of beginning students with Pell Grants* (NCES 2002-169).

For the sake of brevity, each report in the remainder of this review will be referred to as Report 1 through 4.[7]

The sources of the data for the analyses in these NCES reports are the following longitudinal surveys conducted for NCES:

- NELS of 1988, which included students in the eighth grade in 1988, with follow-up surveys in 1990, 1992, 1994, and 2000 (Reports 1, 2, and 3);

- Beginning Postsecondary Students (BPS) Longitudinal Study, which included students beginning postsecondary education in either the 1989-90 or 1995-96 academic years. The first cohort was surveyed again in 1992 and 1994, and the second cohort was resurveyed in 1998 (Reports 1 and 4); and

- Baccalaureate and Beyond Longitudinal Study, which included students completing a bachelor's degree in the 1992-93 academic year, with follow-up surveys in 1994 and 1997 (Report 2).

In addition, supplementary information from the National Postsecondary Student Aid Study (NPSAS) was used to provide comparison data in some of the reports.

The Relationship Between Family Income and College Participation

The tabular analyses of the relationship between family income and college participation conducted by NCES confirm the findings of the research reported in section two of this report: students from high-income families are more likely to enter college than are students from low-income families.[8] In Report 1, students graduating from high school in 1992 are divided into three income groups: those from families making less than $25,000 (28 percent of all students); from families making $25,000 to $74,999 (57 percent); and from families making $75,000 or more (15 percent). While 37 percent of low-income students had not enrolled in any form of postsecondary education within two years of high school graduation, only 21 percent of middle-income students and 7 percent of high-income students had not entered higher education within

that same time period. These differences are also evident in the type of institution attended for those students who did enroll in postsecondary education. Eighty-two percent of high-income students who enrolled in college attended a four-year institution, while only 51 percent of low-income students were enrolled in this sector.

Detailed comparisons of these figures with research conducted by others on this relationship is difficult, because of differences in time periods studied, as well as differences in defining family income groups. However, the pattern reported by NCES in its tabular analyses is consistent with that of other researchers: low-income students are less likely to attend college than their peers from wealthier families, and when they do, they are less likely to be enrolled in a four-year institution.

The Relationship Between Parental Education and College Participation

As with income, parental education is a strong correlate of college participation. NCES Reports 1, 2, and 3 show the following percentages of students graduating high school in 1992 who did *not* enroll in college within two years: students whose parents were high school graduates or less, 41 percent; some college, 25 percent; college graduates, 8 percent.[9] Again, these results are consistent with the bivariate analyses conducted by others and described in section two of this report.

The Relationship Between Financial Aid and College Participation

The four NCES reports reviewed in this study provide limited tabular data regarding financial aid and college participation. For example, Report 4 focuses exclusively on Pell Grant recipients, so by nature of its sample it excludes many middle- and most all high-income students. In addition, it only analyzes those students who enrolled in college. Report 1 includes financial aid data only for low-income students (below $25,000), but it does not distinguish between students who did or did not receive a financial aid offer from a university and whether the students enrolled in college or not. In the many reports issued based on data from the NPSAS surveys (Berkner, 1998; Berkner, Berker, Rooney, & Peter, 2002; Tuma & Geis, 1995), NCES does provide much detailed data on the distribution of financial aid to students with different socioeconomic characteristics. But because these studies only examine students already enrolled in college, they are unable to provide information on the relationship between financial aid and college participation.

REVIEW OF NCES MULTIVARIATE ANALYSES

As described in section two of this report, multivariate analysis allows the researcher to examine the simultaneous effects of a number of characteristics on a chosen outcome, or in the vernacular of research, examine the effect of one factor while controlling for others. Multivariate analysis is particularly powerful when these factors, or predictors of the outcome, are interrelated, a condition quite common among student background characteristics such as race, family income, and levels of parental education.

Reports 1, 3, and 4 all provide multivariate analyses of an outcome related to college participation.[10] Reports 1 and 3 focus on enrollment in college within two years of high school graduation as the outcome, while Report 4 focuses on continuous enrollment in college through 1998 for those students who began in the 1995-96 academic year. Two predictors, family income and parental education level, appear in the analyses in all three reports.[11] Other predictors appear in only one or two of the three reports, and include such variables as race, gender, an index of "college qualification," educational expectations, and whether the student took a college entrance examination.

It is clear from the organization of the NCES reports that the multivariate analysis is not the centerpiece of the analysis; in each of the three reports, the multivariate analysis is relatively brief and is the final section before a concluding chapter. Nevertheless, the findings of the multivariate analyses are often discussed in the textual description of the reports, including the executive summary or highlights sections.

Multivariate Findings Regarding Income, Parental Education, and College Participation

Since income and parental education level are factors included in all three reports, these findings will be discussed first. As reported in section three of this report, the NCES tabular analyses report large differences in college entry (Reports 1 and 3) and college persistence (Report 4) among students from different income groups as well as among students whose parents had differing levels of educational attainment themselves.

The approach of the NCES multivariate analyses is to measure the predicted outcome (college enrollment or persistence) as one factor among a number of predictors, and then report "adjusted percentages" of the outcome for each characteristic after controlling for these other factors. The adjusted percentages can then be compared to the raw, or unadjusted outcome percentages, for each factor. Figure 2 summarizes the adjusted percentage for income and parental education in these three reports.

Figure 2: Adjusted college participation percentage by income and parental education level

	Report 1		Report 3		Report 4
	Attended community college	Attended four-year institution	Attended other than four-year institution†	Attended four-year institution	Persisted through 1998
Income					
Low	21.9*	42.2*	49.8	44.3*	54.8*
Middle	**28.6**	**44.7**	58.3	44.8*	**59.3**
High	22.9*	52.3*	**57.8**	**56.9**	‡
Parental education					
HS graduate or less	23.0	41.0*	**49.3**	**42.3**	56.6
Some college	28.5*	42.8*	56.9*	43.6	54.7*
College graduate	**24.9**	**50.6**	61.9*	51.1*	**59.4**
Advanced degree					65.2*

Notes

Referent groups are shown in bold; * p≤ .05 (compared to referent group)

Other variables included in the multivariate models varied in each report, but included measure such as race, gender, age, college-qualification index, taking steps toward four-year college (entrance examinations and applying), type of high school (public or private), and parents' educational aspirations for their children

The income groups are as follows:

Reports 1 and 3, low: <$25,000; middle: $25,000 - $74,999; high: >$75,000

Report 4, dependent students: low, <$25,000; middle: $25,000 - $69,999

Report 4, independent students: low, <$6,000; middle: $6,000 - $24,999

† For those who did not enroll in a four-year institution

‡ Because this report focused on Pell Grant recipients, it included only low- and middle-income students.

These results indicate that, controlling for other factors, family income and parental education are still predictors of college participation. For example, Report 1 indicates that high-income students were more likely to attend a four-year institution within two years of high school graduation, and low-income students less likely than their middle-income peers. Report 3 indicates that both low- and middle-income students were less likely to attend a four-year institution than their high-income peers. Report 4 shows that low-income students were less likely to persist for three years continuously than their middle-income counterparts.[12]

The effects of parental education are similar. Reports 1 and 3 show that students whose parents were college graduates were more likely to enroll in a four-year institution than were those whose parents had less education. The results regarding the effect of parental education on persistence were more mixed. Students whose parents had an advanced degree were more likely

to persist than those whose parents had only a bachelor's degree, and those whose parents had attended college without attaining a bachelor's degree were less likely to persist. Interestingly, those students whose parents had never attended college had persistence rates that were not statistically different from those whose parents held a bachelor's degree. A possible explanation for this result (though evidently not tested in Report 4) could be that students with parents who had no college experience had overcome such high barriers just to get to college that the drive and motivation to be successful once there was as great as that of students whose parents had higher levels of educational attainment.

Report 2 focuses exclusively on the role of parental education in explaining postsecondary participation as well as post-baccalaureate outcomes. While the report draws largely on analyses conducted in the other three reports, the way that the analyses are recounted may be misunderstood by some readers. For example, on page seven a highlighted statement indicates that, "the likelihood of enrolling in postsecondary education is strongly related to parents' education *even when other factors are taken into account*" (emphasis added). Yet the data provided to support this claim are tabular analyses of the relationship between parental education and college entry from Reports 1 and 3 that *do not* control for other factors. In fact, the multivariate analysis in Report 3 indicates that when the outcome is entry into a four-year college or university, there is no statistical difference between students whose parents had never attended college and students whose parents had some college experience, but had not attained a bachelor's degree, a finding that is acknowledged later in Report 2.

In another section of Report 2, the author summarized the findings regarding parental education and college persistence, stating that, "students whose parents did not attend college remain at a disadvantage with respect to staying enrolled and attaining a degree…again controlling for other related factors" (p. 4). Yet Report 4 demonstrates that this is not true, as Table 17 of that report shows that there is no statistical difference in three-year persistence rates between students whose parents had no college experience and those whose parents had a bachelor's degree after controlling for other factors, including family income.[13] This also may lead readers to draw an incorrect conclusion regarding the relationship between parental education and college participation when other factors are taken into account.

Multivariate Findings Regarding College Costs, Financial Aid, and College Participation

The NCES reports have very little to say regarding the role of financial aid and the cost of college in encouraging or discouraging college participation.[14] In some cases, this appears to be because of a limitation of the surveys used and resulting data elements available to the report authors. For example, Report 1, which uses data from NELS, notes in the multivariate analysis chapter that, "financial aid was not included as a variable because the amounts are known only for those who enrolled [in college]" (p. 67). In other words, data about financial aid awards was only collected for those NELS students who enrolled in college, and not for those who may have been offered financial aid but chose not to enroll in college. Similarly, information about the cost of college was available only for students who enrolled in college, but not for those students who may have been accepted at one or more colleges, but chose not to enroll.

In the one report where detailed financial aid and college cost information was available, the authors chose not to include the breadth of data available to them when conducting the

multivariate analyses. In the multivariate analysis of persistence in Report 4, which included low- and middle-income students, the only measure of college costs or financial aid that was included was the receipt of a Pell Grant during the first year of college. No other financial aid variables, such as loans, state grants, institutional grants, or other aid, were included as a control variable, nor was any measure of the cost of college, such as tuition, cost of attendance, or net price, included. The authors do not state why these variables were omitted from the analysis.

The omission of financial aid and college cost variables from the multivariate analyses is troubling given other information in these reports. Report 1 has a chapter on differences regarding concerns about paying for college among students from different socioeconomic groups. For example, Table 27 in this report indicates that while only 20 percent of high-income students and 16 percent of their parents were "very concerned about college costs and availability of financial aid," 69 percent of low-income students and 79 percent of their parents were similarly concerned. So while the impact of finances on students' college enrollment decisions is not clear because the authors did not report on this linkage, it is clear that there are differences in the expressed concern about finances among these different groups of students. While the NELS study has very little information about financial aid, the BPS study does provide detailed financial aid information for each student. This information could have been used to expand the analyses in Report 4 to better account for the role of different forms of financial aid on persistence.

ANALYSIS OF NCES METHODOLOGICAL APPROACH

The NCES College Qualification Index

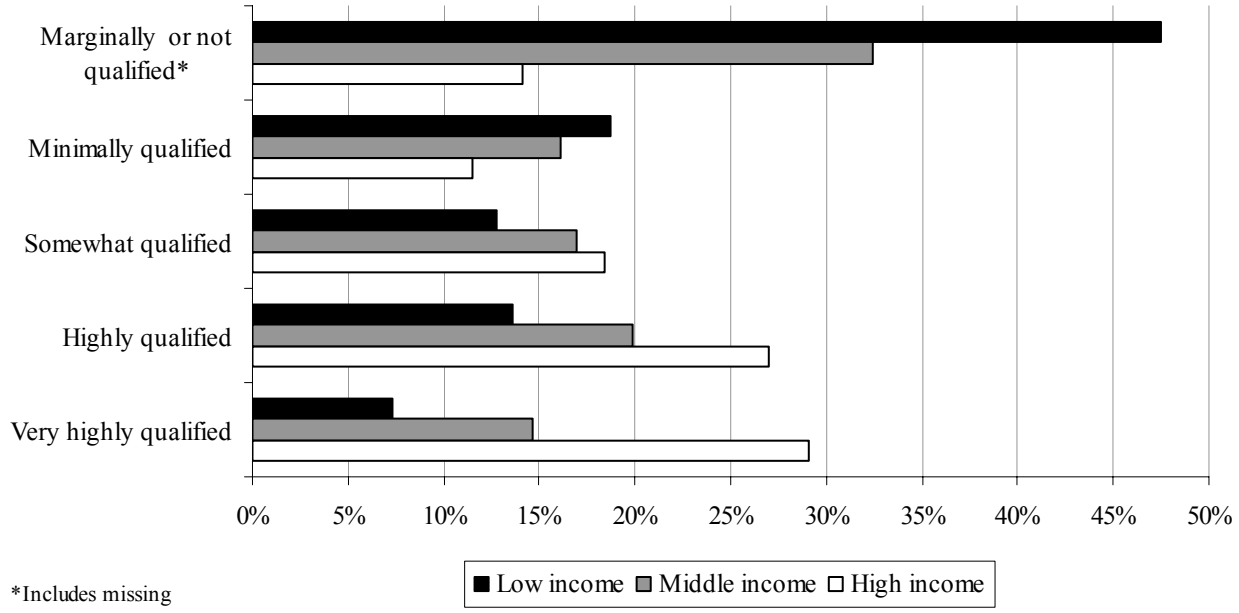
In analyzing the effect of socioeconomic status on college participation, the NCES and the authors of its reports have attempted to address an important consideration: all students are not equally qualified, nor necessarily equally motivated, to attend postsecondary education. In order to understand the impact of a number of these factors on college participation it is important to try to separate those students who could not or would not attend college because of other reasons, and focus on the remaining students.

The approach that NCES has taken to perform this separation is to label students as "college qualified" based on high school grades, class rank, courses taken, and aptitude test scores, including tests administered for the NELS survey as well as SAT and ACT tests. In addition, Report 1 also includes two additional steps by examining "those students who had the initiative to take a college entrance exam and submit an application for admission to a four-year institution" (p. 1).

The relationship between family income and each of the steps toward becoming college-qualified can be gleaned from some of the NCES reports. Report 1 creates a five-level scale based on a composite of the academic measures described earlier. The distribution of all 1992 high school graduates in the three income groups is shown in Figure 3.[15] Close to half of all low-income students (income below $25,000) were considered only marginally or not qualified for a four-year institution and thus excluded from being labeled "college-qualified," in contrast to fewer than 15 percent of the high-income students ($75,000 or more).

Similar differences were reported by NCES when Report 1 examines the steps taken toward attending college, taking a college entrance examination, and applying to a four-year institution by those students who were college-qualified. For example, while 62 percent of low-income students accomplished both of these steps, 91 percent of high-income students did both. Nineteen percent of low-income students took neither step, while only three percent of high-income students failed to do either (Report 1, Table 22).

Figure 3: Proportion of 1992 high school graduates by income and college qualification index.



*Includes missing

Source: Report 1, Table 14

Even among those students who were in the marginally or not college-qualified category (all of whom the authors label "not college-qualified"), the high-income students were more successful at finding their way into some form of postsecondary education. Seventy percent of high-income students who were *not* college-qualified still attended some form of postsecondary education within two years of high school graduation, and almost half of those attended a four-year institution. Less than half of low-income students who were not college-qualified attended college, and less than a quarter of those who did enrolled in a four-year institution (Report 1, Table 33).

The Problem With Focusing on "College-Qualified" Students

Controlling for this separation of college-qualified students from those who were not college-qualified, a key conclusion reached by the authors of the NCES reports can be found in the highlights section of Report 1:

> High school graduates whose parents have low levels of income and education are able to attend four-year colleges at the same rates as students from middle-income families, if they do what four-year colleges expect them to do. That is, if low-income students have an academic record and aptitude test scores which

> demonstrate even the minimal qualifications for admission to a four-year institution, if they take a college entrance examination, and if they submit an application for admission, the majority of low-income students enroll in postsecondary education, and over 83 percent attend a four-year college or university (p. iii).

Report 3 echoes this finding:

> After adjustment, however, low- and middle-income students enrolled [in four-year institutions] at similar rates. This last finding may reflect the leveraging effect of financial aid in providing access to college for low-income students (p. 54).

These passages say that those low-income students who managed to get themselves "college-qualified" by taking the steps and achieving the academic standards outlined in the reports had college participation rates similar to those of their middle-income peers.[16] This implies that college finances, i.e., the cost of college and availability of financial aid, are not a barrier to the participation of low-income students, at least as compared to middle-income students.

This is an important consideration given the concern about college finances among students from different income groups that was reported earlier.[17] However, the analysis in the NCES reports does not attempt to incorporate this concern and measure its impact on college participation. The multivariate analyses of the impact of financial aid on college participation that have been conducted by other researchers and summarized in section two have generally found that financial aid *does* influence the enrollment and persistence of low-income students, even controlling for academic skills and abilities.

The first problem posed by the analytic approach chosen by NCES is that its methodology ignores the role that college finances may have on the decisions made or efforts performed by students to make themselves "college-qualified." In other words, if students and their parents are discouraged early in their high school careers from attending college because they believe it is financially out of reach, then they may not take the steps necessary to put themselves into this college-qualified pool. This series of sequential steps, many dependent upon successful completion of the earlier ones, sets up a screening mechanism that may exclude all but the most determined students who somehow are able to overcome the price signals they receive from the higher education market.

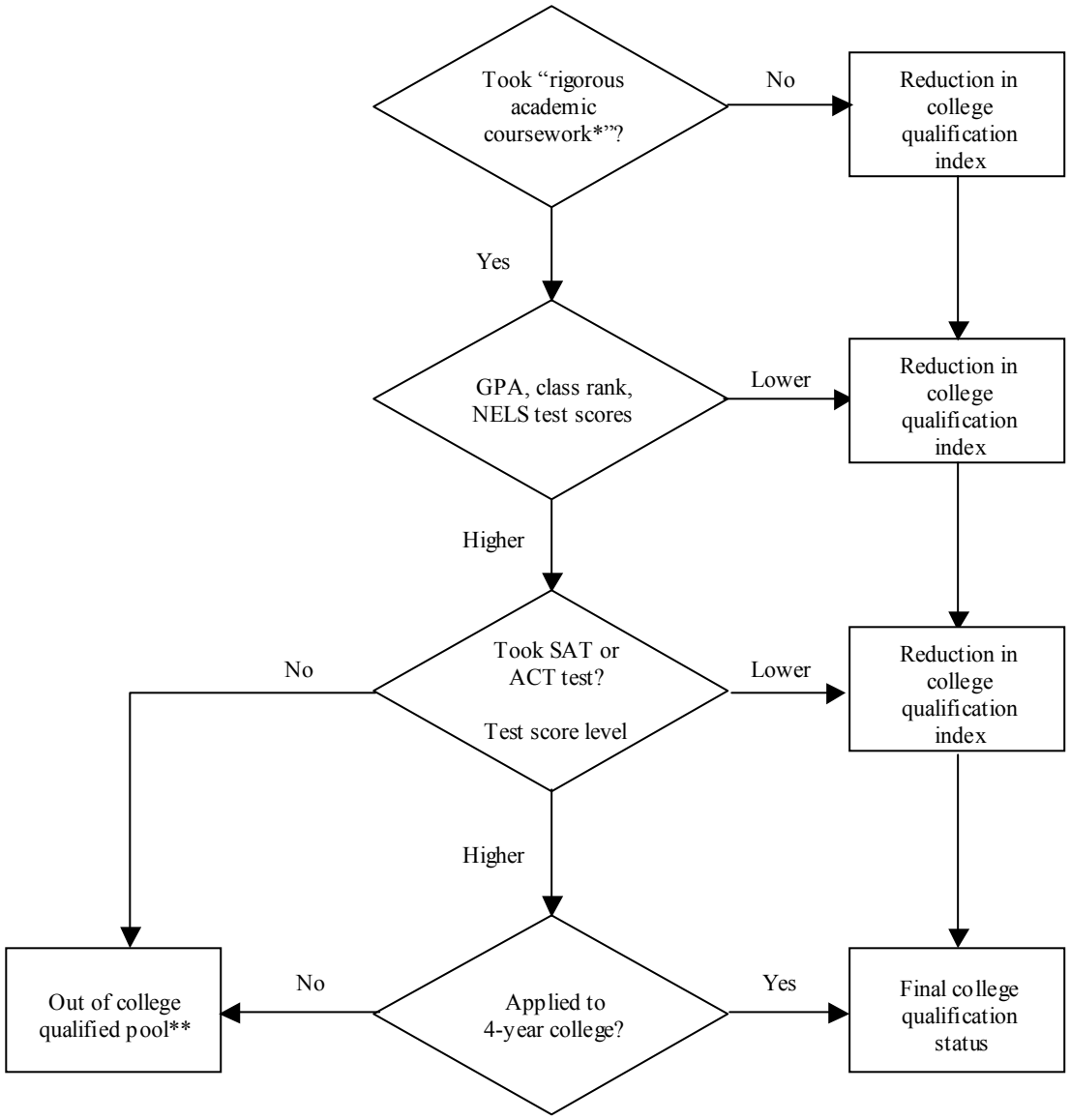
For example, if high tuition prices and lack of information about financial aid as early as the middle school years discourage a low-income student from considering college as an option, then she is not likely to take the college preparatory course sequence defined by NCES. If the student does not take this course sequence, then it is unlikely she will, one, score at the level necessary on the aptitude tests administered by NCES to satisfy the test score criteria for college qualification, and, two, be encouraged to take a college entrance examination. And if she does

not take a college entrance examination, or score at a sufficient level on one of these tests, then it is unlikely she will be encouraged to apply to a four-year institution.

The implications of the NCES approach can be seen in Figure 4. At any of these points, students who did not believe they could ever go to college because of financial barriers, despite their aptitude to be successful in college, could make decisions or perform academically in ways that would serve to lower their final college qualification ranking. These decisions or performance levels could have cumulative effects toward lowering the index value. The decision not to apply to a four-year college or to take a college entrance examination tags a student as not qualified for college in some of the NCES analyses. It is important to remember again that there were large differences among income groups in these college qualification indices. While 86 percent of high-income students were at least minimally "college qualified," only 53 percent of low-income students achieved this standard (Report 1, Table 15).

A second problem with the approach taken by NCES is that it ignores the unmeasured differences between those students who managed to get themselves "college-qualified" (including those who also took the SAT or ACT test and applied to a four-year institution) and students who could not achieve this standard. These unmeasured variables may include factors such as internal or external motivation or "push" to attend college[18], assistance on college planning from peers, teachers, counselors, or others[19], and competitive spirit. It is reasonable to expect that those students who were able to make themselves college-qualified by the NCES standard, and, in particular, low-income students, differed not just in their academic talents, but also in possessing a higher level of something a generation or two ago called "gumption." The unmeasured differences between these students and the resulting impact on the multivariate analyses are described by researchers as "selectivity" or "self-selection" bias.

Figure 4: Sequential steps in NCES definition of being "college qualified"



* At least four years of English, three years each of science, math, and social studies, years of foreign language (Report 1, p. 24)

** Some of the NCES analyses exclude students who did not complete these two steps.

By including the successful completion of these hurdles as a criterion for later measurement as part of the college participation pool, the NCES methodology introduces a serious limitation in its analysis. The decisions of students to overcome these hurdles, or their ability to do so, cannot simply be included as a minimal threshold before their postsecondary experiences can be related to their background characteristics, such as family income and parental education. This methodology treats the steps toward college-qualification (defined by academic achievement, taking college tests, and applying to a four-year college) as exogenous variables that are independent of the other factors that help determine whether one goes to college or not. But as described earlier, they are not exogenous, but, rather, endogenous to the decision to enroll in college.

By not attempting to measure the impact that concern over college finances may have on low-income students, or by not mentioning more prominently the potential of this impact, the authors of these reports may inadvertently be misleading readers about the role of family income on college participation. From the quote excerpted from Report 1 above, it would be reasonable for a reader to conclude that all the efforts of federal, state, or private programs should focus on getting these low-income students college qualified; i.e., if we could just solve that problem, then the gaps in college participation outlined in section two of this report could be eliminated.[20] Yet it is not known from these reports whether lessening the concerns regarding college finances of low-income families – by lowering college costs, by increasing the availability of financial aid, by providing better information about aid, and the like – would have a similar or perhaps even greater impact on eliminating the college participation gap.

The third problem with the NCES college-qualification index is its assumption that the index represents the steps necessary for a student to be prepared for enrollment in a four-year college or university. While this may be in part true, it represents a very traditional path toward college entry, a path that has undergone great change in recent years and is likely to change even more in the future.

A report by the National Center for Fair & Open Testing (Rooney & Schaeffer, 1998) listed almost three hundred schools that have eliminated the requirement of students submitting SAT or ACT scores, made them optional, or deemphasized their use in the admissions process for at least some entering students. Included are major university systems such as the public university system in Texas and the California State University system. In addition, many colleges have alternative admissions programs for some students who may score well below the institutional norms on the standard criteria of high school grades, class rank, and college entrance examination tests. While many of these changes have been made subsequent to the cohort of students analyzed in the NCES reports (the high school graduating class of 1992), it is still important to note that the ways in which many students become "college-qualified" today are quite different from those assumed by the NCES methodology.

The Problem With Focusing on Four-Year College Entry

While the NCES reports have some information about enrollment in less than four-year institutions, the focus is primarily on entry into four-year institutions. This approach tends to deemphasize the fact that over 40 percent of all first-time freshmen in degree-granting institutions enroll in a community college (National Center for Education Statistics, 2002, Table

182). Some of these students go on to enroll in a four-year institution and attain a bachelor's degree. Understanding the role of family income, academic preparation, and the other factors that influence college participation is equally important for these students and for these institutions throughout the nation.

This issue is particularly critical since community colleges are an important entry point into postsecondary education for low-income youth. McPherson and Schapiro (1998) analyzed data from the 1994 American Freshman Survey of the UCLA Higher Education Research Institute to determine the enrollment of students from different income groups across higher education sectors. While 31 percent of all full-time freshmen were enrolled in community colleges[21], 47 percent of freshmen from the lowest income group (family income of less than $20,000 in 1994) were enrolled in this sector (Table 5.1). In contrast, fewer than 14 percent of students from families making over $100,000 were in community colleges. Thus, by focusing on four-year college entry as an outcome, the NCES reports pay little attention to the experiences of many low-income students who see community college as the only postsecondary option available to them.

The Statistical Approach of the NCES Multivariate Analyses

As described earlier, three of the Reports (1, 3, and 4) present multivariate analyses of the impact of several factors on college entry or persistence. Each report includes different sets of independent variables as predictors of the outcome of entry or persistence. Report 1 has the most parsimonious model; it includes only race, family income, parental education, college qualification index, and the two steps toward four-year college as predictors. The analyses in the other two reports include a broader set of variables as predictors.

The NCES reports provide little information about some characteristics of these statistical models. For example, no measures of model fit, or the explanatory value of the models, are provided. Without these measures, it is impossible to tell how much of the variation in the outcome (college entry or persistence) is predicted by the independent variables, and thus, it is difficult to tell if the independent variables taken together are important predictors or have very little impact on the outcome.

The reports also provide only final versions of the regression models and do not provide intermediate models that show the joint effects of conceptually grouped sets of predictors. In reporting multivariate results, it is a common convention to provide individual models that show the results of these groups of predictors on the outcome, building from a model with only one set of predictors up to a fully-specified model. In the models in these reports, such an approach would entail showing first, for example, the effect of student background characteristics, such as race, gender, family income, and parental education, on college participation. The next model would then add to these background traits the students' academic characteristics, such as the measures that make up the college qualification index, and show how the statistical fit or predictive value of this second model was improved over the first. This process would continue until all the variables were included in a fully-specified model. A process like this would provide the reader a sense of the importance of each group of variables in predicting the outcome.

Related to this point is the minimal information provided in the reports explaining why certain variables were included or excluded from the multivariate models. It is impossible to tell why the model in Report 1 included only the five variables described above, while excluding any of the other variables available in the NELS survey shown by other research to be related to college entry. These were, in fact, reported on in the tabular analyses of that report. As noted earlier, the report did explain that financial aid variables were not included in the analysis because financial aid data were not available for students who did not enroll in college. However, inclusion of information about financial aid and the cost of college could provide valuable information about the choice of institutions for students who did enroll. For example, it might show whether students who received financial aid were more likely to attend a four-year institution than a community college.

Report 1 has a section of tabular analyses, described earlier, on the relationship between a number of background characteristics and the concern over college finances. That section of the report also explores similar relationships between students' background characteristics and steps taken to obtain information about financing a college education. It would be logical to ask whether these financial concerns, or the steps taken to obtain information about college financing, were predictors of college entry after controlling for other factors or vice-versa. Yet these variables were excluded from the multivariate analysis, and the authors provide no rationale for this decision.

As described earlier, the multivariate analysis of three-year persistence rates in Report 4 included no information about college costs and financial aid, other than the receipt of a Pell Grant, even though detailed information is available in the BPS Longitudinal Study. Given the existing research that has documented the effects of college prices and financial aid on the persistence of low-income students, it is unclear why the authors chose to exclude these variables from their analysis.

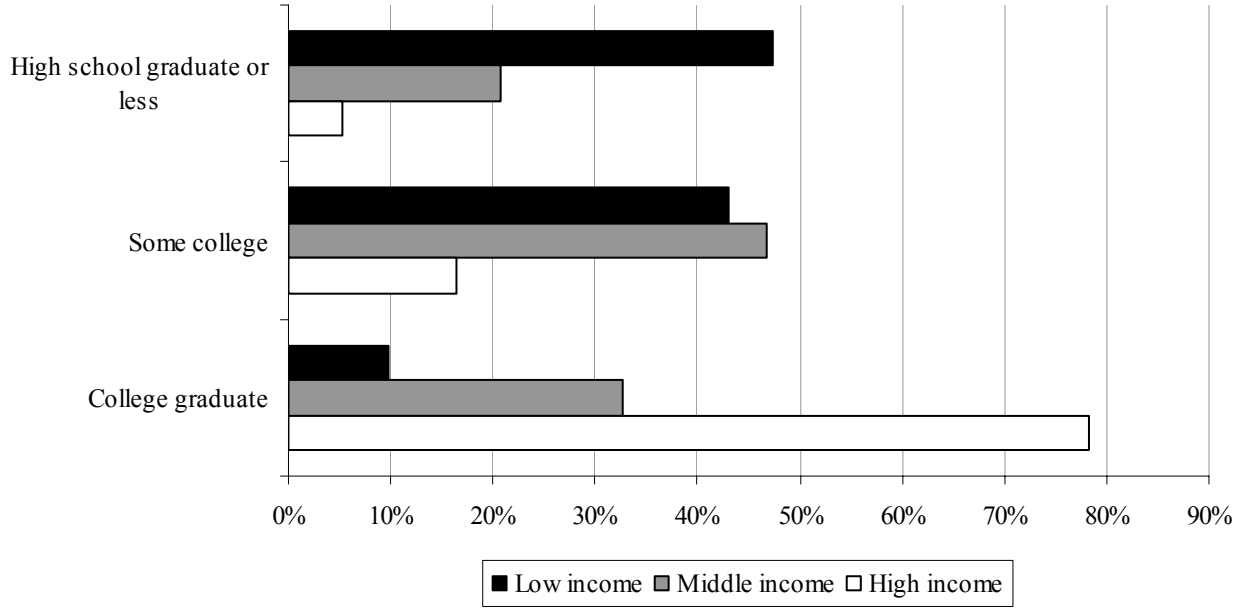
The final issue related to the statistical approach used in the NCES multivariate analyses is the problem of collinearity, also known as multicollinearity, or the correlation between the independent variables in the models. When independent variables are highly correlated, or statistically related to one another, the coefficients of the resulting model may be misestimated.[22] Researchers normally take steps to minimize the impact of collinearity on models by first measuring the correlation between independent variables. When two or more of these variables are highly correlated, standard procedure calls for one or more of them to be excluded from the model.[23] Failure to correct for collinearity may lead to a bias in the parameter estimates for the collinear variables. A common result of collinearity is to bias downward the estimates of the coefficients of correlated variables, thus leading one to conclude that a variable is not as important in predicting the outcome as it truly is.[24]

The NCES reports do not provide a correlation matrix of the variables used in the multivariate models, so it is impossible to gauge the exact impact that collinearity may have on the multivariate models. However, information provided in the NCES tabular analyses indicates that there may be a high degree of correlation between at least some of the predictors used in the models. For example, Table 1 of Report 1 shows that there appears to be a strong correlation between the key background characteristics used in the multivariate models in Reports 1, 3, and 4. While 52 percent and 54 percent of the Hispanic and black 1992 high school graduates,

respectively, were from low-income families, only 34 percent of Asian/Pacific Islanders and 21 percent of whites were from this same group. Similarly, the nature of the relationship between family income and parental education can be seen in Figure 5. While almost half of low-income students had parents who had never attended college, only five percent of high-income students were in this category. At the other end of the parental education scale, only ten percent of low-income students had at least one parent who graduated from college, while over three-quarters of high-income students were the children of college graduates.

Without analyzing the NELS data to measure the extent of the statistical correlation between these variables, it is impossible to determine the impact of collinearity on the multivariate analyses conducted by NCES. However, the relationships between some of the predictors that are demonstrated in the tabular analyses appear strong enough to question whether the results of the multivariate models may be biased by the effects of collinearity. One questionable result is that the parameter estimates of one or more of the correlated variables (for example race, family income, and parental education) may be biased downward, i.e., the statistical relationship between the outcome (college entry or persistence) and the predictor may actually be stronger than reported by NCES.

Figure 5: Relationship between family income and parental education, 1992 high school graduates



Source: Report 1, Table 1

The Impact of the Conflict Between the NCES Tabular and Multivariate Analyses

A key difference between the findings in the NCES tabular and multivariate analyses is in the relationship between family income and college participation. As noted in section three of this report, the NCES tabular analyses found that while 37 percent of low-income students had not enrolled in any form of postsecondary education within two years of high school graduation, only 21 percent of middle-income students and 7 percent of high-income students had not entered higher education within that same time period. In the NCES multivariate analyses, however, these differences were greatly reduced or even eliminated (see Figure 2).

The impact of this conflict between the tabular and multivariate results, as well as others that have been described in this report, is summarized in Figure 6. Shown are the conclusions from the tabular analyses, how each issue is dealt with in the multivariate analyses, the statistical problem with this treatment, and the resulting contradictory conclusion.

Figure 6: Contradictions between NCES tabular and multivariate analyses

Conclusion From Tabular Analyses	Treatment in Multivariate Analyses	Related Statistical Issue	Resulting Contradictory Conclusion
Financial barriers are important determinants of college enrollment for low-income students	Exclude college costs, financial aid, and unmet need as predictors and/or covariates	Omitted variable bias	Financial barriers are not important determinants of enrollment for low-income students
Access to college is unequal among low-, middle-, and high-income students	Focus exclusively on use of college-qualification index, as well as requirement of taking SAT/ACT tests and applying to four-year institution	Selection bias	There are small, if any, differences in college access for students who are college-qualified and take necessary steps toward enrollment
Finances discourage low-income, college-qualified high school graduates from taking the necessary steps toward enrollment	Ignore the effects of finances on steps toward enrollment, and include the steps as independent variables	Endogeneity bias	Taking steps toward enrollment enables low-income students to attend college at the same rates or close to those of their high-income peers
Family income is a major barrier to access and persistence for high school graduates	Attribute the joint effects of family income and parental education to parental education alone	Collinearity	Parental education is the major barrier to access and persistence for high school graduates

It cannot be known from the information presented in the NCES reports whether correcting the methodological problems outlined in Figure 6 would resolve the contradictory results between the tabular and multivariate analyses. However, until the methodological problems are addressed, it is misleading to accept the conclusions of the reports that develop from the multivariate analyses.

CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

The four NCES reports reviewed in this study provide valuable information regarding the relationships between a number of variables and college participation in this country. However, because of the limitations of some of the data sets used, along with some decisions by NCES and its contractors regarding the focus of the reports, the four provide very limited information about the impact of college costs and financial aid on college participation.

The four reports require very careful readings by experienced researchers to understand the complex nature of these relationships. In an attempt to simplify the presentation of the results and keep the details of the analyses to a minimum, the reports may lead to a misinterpretation of some key findings. Given the importance and visibility of the work of NCES in informing postsecondary education policy throughout the country and, particularly, in light of the reauthorization of the HEA that will be taken up by Congress and the Administration this year, it is critical that the work of NCES achieve the highest standards of research.

The policy implications of the methodological problems outlined in this study should not be overlooked. For example, the conclusion in the NCES reports that differences in college-going rates are largely attributable to differences in parental education levels, rather than income, could lead to the conclusion that there is little that federal or state governments, or institutions can do to help close the gap in college participation between rich and poor. Levels of parental educational attainment are largely immutable, at least in the short run. However, if the differences in college entry rates are at least in part a factor of differences in resources among these groups – a conclusion that is not just plausible, but likely given the findings of other researchers – then there *is* a role for government and higher education institutions in closing the gap. The policy levers of financial aid and tuition levels can be utilized to help overcome these differences in resources.

There are a number of actions that NCES or other researchers, provided they are given access to the data used by NCES, can take to add to our understanding of the dynamics of college entry and persistence presented in these four reports. One step is to provide more details of the process followed in conducting the existing multivariate analyses, as described in the previous section of this study. This additional information would allow researchers to gauge the statistical validity of the analyses conducted by NCES, and thus provide more evidence of the value these models may have for informing higher education policy and practice. These should include:

- more information about the statistical fit of the models;
- presentation of the intermediate models leading up to the fully-specified models shown in the existing reports; and
- more information about the correlation of predictor variables and the tests for and potential effects of collinearity in the multivariate models.

A second action that could be undertaken by NCES would be to perform a reanalysis of the existing data to include variables in the multivariate models that were excluded in the initial work. These variables should be included based on the conceptual and empirical work

conducted by other researchers that has explored the relationship between different factors and the outcomes of college entry and persistence. The NCES researchers should provide a thorough explanation of the rationale for including or excluding each variable. In particular, the researchers should consider the inclusion of more variables related to financial aid and college costs.

A third area of study is for NCES to examine its use of the college qualification index and the additional two steps toward entry described as taking the SAT or ACT and applying to college. As described earlier, such an approach introduces the issue of selectivity bias into the analyses and does not account for the impact of college prices and financial aid on the decisions of students to make themselves "college-qualified." At a minimum, NCES should present the same type of tabular analyses found in Report 1, which focus almost exclusively on college-qualified students, for those students who were not college-qualified. Since low-income students are disproportionately found in this latter group, more information should be provided about their pre-collegiate experiences.[25]

A fourth area for reexamination is the focus on the experiences of students entering four-year institutions, and the steps taken by students to qualify themselves for entry into these institutions. As noted earlier, over 40 percent of all first-time freshmen are enrolled in community colleges. Understanding more about the predictors of entry into these institutions could help inform policy.

Finally, NCES should analyze the need for a new nationally representative longitudinal study of high school graduates. Such a study should combine the detailed data about the high school and in some cases, middle school experiences found in NELS, with the postsecondary information found in the NPSAS. The strength of NELS is that it includes data from surveys of students, as well as their teachers, parents, and school administrators. It also includes high school transcript data and the administration of aptitude tests to the respondents. The NPSAS surveys contain detailed information from student interviews, as well as student-record data from the students' postsecondary institutions.

The NELS cohort of high school graduates is already over ten years old. Much has changed in both secondary and postsecondary education in the last decade, and in both public and institutional policy. A new longitudinal survey that combines the level of detail found in NELS and NPSAS would provide valuable data for researchers to answer many of the questions regarding the impact of financial aid and college costs on college participation, questions that NCES has had difficulty answering given the limitations of the existing data sets and the focus chosen by NCES in these reports.

END NOTES

[1] Some of this loan and work-study aid goes to graduate students; the focus of this report is solely on college participation by undergraduate students.

[2] The four reports are described in section three of this report.

[3] A high school diploma or GED is the minimal credential required for entry into most postsecondary education institutions.

[4] These figures represent the proportion of each group enrolled in college, using the average for the years from 1997 to 2000.

[5] See for example Levy and Murnane (1992) and Mortenson (1995, 1999a).

[6] It also should be noted here that Kane conducted these analyses with and without controlling for the educational level of the student's parents. In both methods family income was still a large indicator of whether the student would enroll in college, even among these most academically talented students.

[7] These four reports were selected by the Advisory Committee staff for review. While each report was written by contractors, rather than NCES staff, because of the oversight role of NCES they are being referred to here as "NCES reports."

[8] Because Reports 1 and 3 analyze data from students who graduated high school in 1992, the analyses there are restricted to dependent students, often labeled "traditional college students." Report 4 includes all beginning students, dependent and independent alike.

[9] The percentages across the three reports varied slightly (less than one percentage point in each category), due most likely to slight differences in the samples of students included in the analysis.

[10] Report 2 is primarily a summary of other NCES analyses, and thus, does not present original analyses of its own.

[11] These reports measure parental education as the highest level achieved by either parent or guardian of the student.

[12] Because Reports 1 and 3 did not include multivariate analysis of entry into *any* form of postsecondary education, we cannot tell the impact of family income on this outcome.

[13] This same table presents the counter-intuitive finding described earlier that students whose parents had some college experience, but no bachelor's degree, had persistence rates that were *below* those of students whose parents had no college, thus further calling into question the value of parental education in predicting college persistence.

[14] The phrases "financial aid" and "college costs" are used here broadly to include such measures as tuition prices, cost of attendance, net prices, effective family contribution, and unmet need.

[15] This analysis does not include those students who had not graduated from high school by 1992. Because low-income students are more likely to drop out of high school, they would be disproportionately excluded before the point at which this analysis was conducted.

[16] Report 1 does not mention in the highlights section that the college entry rates of low-income students still lagged behind those of their high-income peers by ten percentage points, controlling for other factors (table 34).

[17] This is reinforced by the fact that students and parents tend to *overestimate* the cost of college, and the overestimation tends to be greatest among low-income families (Ikenberry & Hartle, 1998; National Center for Education Statistics, 2001).

[18] See for example Hossler, Schmit, and Vesper (1999) and McDonough (1997).

[19] While measures of some of these assistance factors are available in the NELS survey, they evidently were not included in the multivariate analyses in Report.

[20] It should be noted here that the contractors hired to write the NCES reports are generally prohibited from discussing the policy implications of their findings in those reports.

[21] This proportion is less than that reported by NCES because the American Freshman Survey includes only full-time students, while the NCES figures include all students in community colleges.

[22] See Kennedy (1992) and Kleinbaum, Kupper, and Muller (1988) for more on the problems of collinearity in multivariate models.

[23] Another alternative is to combine the correlated variables into a "composite" variable, similar to what NCES has done in creating the college qualification index.

[24] A related problem is that of endogeneity, where predictor variables are related to one another though not linearly correlated. Endogeneity can result in a similar downward bias in the coefficient estimates. The NCES reports do not provide the information necessary to judge the degree that endogeneity may be affecting the results of the multivariate analyses.

[25] Report 3 does provide more information about all high school graduates, not just those who were college qualified, but its scope is more limited than Report 1.

REFERENCES

Behrman, J. R., Kletzer, L. G., McPherson, M. S., & Schapiro, M. O. (1992). *The college investment decision: Direct and indirect effects of family background on choice of postsecondary enrollment and quality* (DP-18). Williamstown, MA: Williams Project on the Economics of Higher Education.

Berkner, L. (1998). *Student financing of undergraduate education: 1995-96* (NCES 98-076). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Berkner, L., Berker, A., Rooney, K., & Peter, K. (02). *Student financing of undergraduate education: 1999-00* (NCES 02-167). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Clotfelter, C. T. (1991). Financial Aid and Public Policy. In C. T. Clotfelter & R. G. Ehrenberg & M. Getz & J. J. Siegfried (Eds.), *Economic Challenges in Higher Education*. Chicago: University of Chicago Press.

College Board. (2002). *Trends in student aid, 02*. Washington, DC: Author.

Ellwood, D. T., & Kane, T. J. (2000). Who is getting a college education? Family background and the growing gaps in enrollment. In S. Danziger & J. Waldfogel (Eds.), *Securing the future: Investing in children from birth to college*. New York: Russell Sage Foundation.

Heller, D. E. (1997). Student price response in higher education: An update to Leslie and Brinkman. *Journal of Higher Education, 68*(6), 624-659.

Heller, D. E. (1999). Racial equity in college participation: African American students in the United States. *Review of African American Education, 1*(1), 5-29.

Higher Education Act of 1965, Pub. L. No. 89-329 (1965).

Hossler, D., Schmit, J., & Vesper, N. (1999). *Going to college: How social, economic, and educational factors influence the decisions students make*. Baltimore, MD: The Johns Hopkins University Press.

Ikenberry, S. O., & Hartle, T. W. (1998). *Too little knowledge is a dangerous thing: What the public thinks about paying for college*. Washington, DC: American Council on Education.

Jackson, G. A. (1989). *Responses of black, Hispanic, and white students to financial aid: College entry among recent high school graduates*. College Park, MD: National Center for Postsecondary Governance & Finance, U. of Maryland.

Jackson, G. A., & Weathersby, G. B. (1975). Individual demand for higher education. *Journal of*

Higher Education, 46(6), 623-652.
Kane, T. (1999). *The price of admission: Rethinking how Americans pay for college*. Washington, DC: Brookings Institution Press.

Kennedy, P. (1992). *A guide to econometrics*. Cambridge, MA: The MIT Press.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariate methods* (2nd ed.). Boston: PWS-KENT Publishing Company.

Koretz, D. (1990). *Trends in the postsecondary enrollment of minorities*. Santa Monica, CA: The RAND Corporation.

Leslie, L. L., & Brinkman, P. T. (1988). *The economic value of higher education*. Washington, DC: American Council on Education.

Levy, F., & Murnane, R. J. (1992). U.S. earnings levels and earnings inequality: A review of recent trends and proposed explanations. *Journal of Economic Literature, 30*, 1333-1381.

Manski, C., & Wise, D. (1983). *College choice in America*. Cambridge, MA: Harvard University Press.

McDonough, P. M. (1997). *Choosing colleges: How social class and schools structure opportunity*. Albany: State University of New York Press.

McPherson, M. S., & Schapiro, M. O. (1998). *The student aid game: Meeting need and rewarding talent in American higher education*. Princeton, NJ: Princeton University Press.

Morey, L. (1928). Student fees in state universities and colleges. *School and Society, 28*(712), 185-192.

Mortenson, T. (1995). Educational attainment by family income 1970 to 1994. *Postsecondary Education OPPORTUNITY, 41*, 1-8.

Mortenson, T. G. (1999a). Family income by educational attainment 1956-1997. *Postsecondary Education OPPORTUNITY, 82*, 11-16.

Mortenson, T. G. (1999b). Parental educational attainment and higher educational opportunity. *Postsecondary Education OPPORTUNITY, 79*, 1-14.

Mortenson, T. G. (2001a). College participation by family income, gender and race-ethnicity for dependent 18 to 24 year olds 1996 to 00. *Postsecondary Education OPPORTUNITY, 114*, 1-8.

Mortenson, T. G. (2001b). Family income and higher education opportunity 1970 to 00. *Postsecondary Education OPPORTUNITY, 112*, 1-9.

National Center for Education Statistics. (2001). *The condition of education 01* (NCES 01-072). Washington, DC: U.S. Department of Education.

National Center for Education Statistics. (2002). *Digest of education statistics, 01*. Washington, DC: U.S. Department of Education.

President's Commission on Higher Education. (1947). *Higher education for American democracy*. New York: Harper & Brothers.

Rooney, C., & Schaeffer, B. (1998). *Test scores do not equal merit: Enhancing equity & excellence in college admissions by deemphasizing SAT and ACT results*. Cambridge, MA: National Center for Fair & Open Testing.

Rouse, C. E. (1994). What to do after high school: The two-year versus four-year college enrollment decision. In R. G. Ehrenberg (Ed.), *Contemporary Policy Issues in Education*. Ithaca, NY: ILR Press.

St. John, E. P. (1990). Price response in persistence decisions: An analysis of the high school and beyond sophomore cohort. *Research in Higher Education, 31*(4), 387-403.

St. John, E. P. (1991). What really influences minority attendance? Sequential analyses of the High School and Beyond sophomore cohort. *Research in Higher Education, 32*(2), 141-158.

Tuma, J., & Geis, S. (1995). Student financing of undergraduate education, 1992-93 (NCES 95-2). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

United States Bureau of the Census. (2001). School enrollment in the United States - Social and economic characteristics of students (P-533). Washington, DC: Author.

OMITTED VARIABLES AND SAMPLE SELECTION ISSUES IN THE NCES RESEARCH ON FINANCIAL AID AND COLLEGE PARTICIPATION

ADVISORY COMMITTEE ON STUDENT FINANCIAL ASSISTANCE

JUNE 2003

Omitted Variables and Sample Selection Issues in the NCES Research on Financial Aid and College Participation

Report prepared for the
Advisory Committee on Student Financial Assistance
Submitted for Presentation on June 12, 2003
Macalester College
St Paul, Minnesota

William E. Becker*
Professor of Economics
Indiana University – Bloomington

June 2003

*** In the preparation of this report extensive use was made of Chapters 17, 21 and 22 of William H. Greene, *Econometric Analysis* (Prentice Hall Fifth Edition, 2003). Personal communication with Greene on technical issues related to the nature of the omitted-variable problem in nonlinear probability models is gratefully acknowledged. Likewise, communication with Wilbert van der Klaauw on the use of nonparametric and semiparametric estimation of financial aid effects is gratefully acknowledged. Suzanne Becker, William Goggin, Donald Heller, Edward St. John, John Siegfried, and Michael Watts are thanked for their critical reading of earlier versions.**

EXECUTIVE SUMMARY

The National Center for Education Statistics (NCES) recently commissioned four major reports on the determinants of initial college enrollment and persistence toward a baccalaureate degree. In essence, the authors of these multivariate analyses give the impression that parental education levels and college preparatory work, and not family income, college costs or the availability of financial aid, are paramount in the college-going decisions of students. A recent paper by Donald Heller at Pennsylvania State University called attention to the shortcomings of the analyses supporting such conclusions. In particular, the authors of these NCES studies actually omitted the relevant financial variables from their analyses and/or restricted their analyses to only those students who were already college-qualified (completed relatively rigorous academic high school courses, achieved sufficient grades/class rank, took the SAT/ACT, and applied to college.)

In this paper I provide an econometric assessment of the consequences of omitting relevant financial variables from a multivariate analysis of college-going decisions, ignoring the sample selection issues and related endogeneity issues associated with focusing only on those who are college-qualified, and not adequately considering the implications of highly related variables that are believed to influence college enrollment and persistence decisions. I also provide suggestions and examples of how the data should be re-analyzed to provide consistent estimators of the relevant parameters in student-choice models of the college-going decision.

The four NCES reports considered here are limited by some of the data sets employed and by some decisions that NCES and its contractors made about the importance of including financial aid variables that other researchers have found to influence college-going decisions. NCES should commission a complete reanalysis of the existing data to include financial variables in the multivariate models that were excluded in the initial work but which others have found to influence college-going decisions. A thorough explanation of the rationale for including or excluding each variable must be provided based on the existing literature. In addition to addressing the omitted-variable problems, NCES must reexamine its use of the college qualification index (and the additional two steps toward entry into four-year institutions of taking the SAT or ACT and applying to college). Use of this index introduces a sample selection problem that cannot be overcome by simply adding more variables or more observations.

It is absolutely essential that the studies and findings advanced by the NCES not be based on methodologies that do not adequately control for the high school student's expectations of the net cost of college and future earnings. Without including relevant financial aid measures in studies of college access and without adequately controlling for sample selection in the college-going decision, conclusions based on regression analyses about the importance of other explanatory variables cannot be taken seriously.

TABLE OF CONTENTS

INTRODUCTION

This report is an extension of Donald Heller's "Review of NCES Research on Financial Aid and College Participation," which was prepared for the US Department of Education Advisory Committee on Student Financial Assistance (Draft: March 2003). Heller reviewed four NCES studies that deal with college access:

1. Berkner, L., and Chavez, L. (1997). *Access to postsecondary education for the 1992 high school graduates* (NCES 98-105).
2. Choy, S. P. (2001). *Students whose parents did not go to college: Postsecondary access, persistence, and attainment* (NCES 2001-126).
3. Horn, L., and Nuñez, A.-M. (2000). *Mapping the road to college: First-generation students' math track, planning strategies, and context of support* (NCES 2000-153).
4. Wei, C. C., and Horn, L. (2002). *Persistence and attainment of beginning students with Pell Grants* (NCES 2002-169).

In accordance with the task specified by the Advisory Committee on Student Financial Assistance, I have read Heller's report on these four studies, and I take as given his interpretation of the statistical analyses employed, his conclusions reached based on that analysis, and the errors contained in these studies. My report extends the Heller report by providing an econometric assessment of the consequences of omitting relevant financial variables from the multivariate analysis, ignoring the sample selection problems and related endogeneity issues, and not adequately considering the implications of highly related variables that are believed to influence college enrollment and persistence decisions. I also provide suggestions on how the data should be re-analyzed to provide consistent estimators of the relevant parameters in student-choice models of the college-going decision.

Before addressing the technical issues associated with omitting relevant financial variables from an assessment of the college-going decision and focusing only on those who are college-qualified and associated endogeneity problems, it may be helpful to consider an analogy involving a contest of skill between two types of contestants: Type A and Type B. There are 8 of each type who compete against each other in the first round of matches. The 8 winners of the

first set of matches compete against each other in a second round, and the 4 winners of that round compete in a third. Type A and Type B may compete against their own type in any match after the first round, but one Type A and one Type B manage to make it to the final round. In the final match they tie. Should we conclude, on probabilistic grounds, that Type A and Type B contestants are equally skilled? How is your answer affected if I tell you that on the first round 5 Type As and only 3 Types Bs won their matches and only the one Type B was successful in the second and third round? This additional information should make clear that we have to consider how the individual matches are connected and not just look at the last match. But before you conclude that Type As had a superior attribute only in the early contests and not in the finals, consider another analogy provided by Thomas Kane.[1]

Kane has a hypothetical series of races between 8 greyhounds and 8 dachshunds. In the first race, the greyhounds enjoy a clear advantage with 5 greyhounds and only 3 dachshunds finishing among the front-runners. These 8 dogs then move to the second race, when only one dachshund wins. This dachshund survives to the final race when it ties with a greyhound. Kane asks: "Should I conclude that leg length was a disadvantage in the first two races but not in the third?" And answers: "That would be absurd. The little dachshund who made it into the third race and eventually tied for the win most probably had an advantage on other traits – such as a strong heart, or an extraordinary competitive spirit – which were sufficient to overcome the disadvantage created by its short stature."

These analogies demonstrate all three sources of bias found in the National Center of Education Statistics recent studies of college-going and persistence decisions: sample selection bias, endogeneity, and omitted variables. For example, the length of the dogs' legs not appearing to be a problem in the final race (financial aid not appearing important among those who jump the hurdles to become college-qualified) reflects the sample selection issues resulting if the researcher only looked at that last race. Looking only at the last race (corresponding to those who apply to college) would be legitimate if the races were independent (high school and college educational decisions were independent), but they are sequentially dependent; thus, the endogeneity problem. As Kane points out, concluding that leg length (income/expense variables) was important in the first two races (high school) and not in the third (going to

college) reveals the omitted-variable problem: a trait such as heart strength or competitive motivation (known availability of financial aid) might be overriding short legs and thus should be included as a relevant explanatory variable in the analyses. The mathematics of selection, endogeneity and relevant omitted variables are well known, and they are the focus of my report.

A PRIMER ON OMITTED RELEVANT VARIABLES AND RELATED PROBLEMS

Reports 1 (Berkner and Chavez) and 3 (Horn and Nuñez) focus on enrollment in college within two years of high school graduation as the outcome, whereas report 4 (Wei and Horn) focuses on continuous enrollment in college through 1998 for those students who began in the 1995-1996 academic year. (Report 2 by Choy is a summary of other studies.) Heller notes that family income and parental education appear as explanatory variables in all of the studies. Either because of the lack of data (in the National Education Longitudinal Study) or explicit omission, however, only the multivariate analysis of persistence in report 4 (which included low- and middle-income students) made any attempt to include financial data in the explanation of the college decision.[2] But even in report 4, whether the student received a Pell Grant or not in the first year of college was the only financial variable used as a regressor (other than family income). There were no other financial variables (for tuition, cost of attendance, net price, loans, state grants, institutional grants, or the like) included.

The omission of financial data other than family income renders these NCES studies suspect in ways even more severe than those recognized by Heller. From the early work of Griliches (1957) and Theil (1957), the consequence of omitting relevant explanatory variables has been well known. The bias that results from excluding an explanatory variable with available or unavailable data can be seen in the bivariate choice to enroll in college. The i^{th} potential student's decision to enroll in college ($Y_i = 1$) or not enroll ($Y_i = 0$) can be related to sets of variables represented in two matrices: $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$, where the subscript i indicates the i^{th} student's record in the i^{th} row of the two matrices. The first matrix, $\mathbf{X}_{1i}$, contains a column of ones and sets of explanatory variables related to the student's characteristics (SAT/ACT score, grade point average/class rank, etc.), family characteristics (parent income, education, etc.), environmental factors (peers, social category, etc.). The second matrix, $\mathbf{X}_{2i}$, contains the financial variables related to college cost (tuition, cost of attendance, loans, state grants, institutional grants, etc.). The linear probability model is then written

$$Y_i = \mathbf{X}_{1i}\boldsymbol{\beta}_1 + \mathbf{X}_{2i}\boldsymbol{\beta}_2 + \varepsilon_i \qquad (1)$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of parameters to be estimated that correspond to the variables in the

$\mathbf{X}_1$ and $\mathbf{X}_2$ matrices Each of the epsilon error terms ε_i in the vector of error terms $\boldsymbol{\varepsilon}$ is assumed to have an expected value (mean) of zero and be unrelated to the variables in $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2$; i.e., $E(\boldsymbol{\varepsilon} \mid \mathbf{X}) = 0$ and $E(\mathbf{X}'\boldsymbol{\varepsilon}) = 0$. Thus,

$$E(Y_i \mid \mathbf{X}_i) = \mathbf{X}_{1i}\boldsymbol{\beta}_1 + \mathbf{X}_{2i}\boldsymbol{\beta}_2 = \text{Prob}(Y_i = 1) \qquad (2)$$

because $E(Y_i) = (1)[\text{Prob}(Y_i = 1)] + (0)[1 - \text{Prob}(Y_i = 1)] = \text{Prob}(Y_i = 1)$.

Although the error terms in the linear probability model are distributed as binomial random variables and do not have constant variance, as required for hypothesis testing with ordinary least squares estimators of the β coefficients, the coefficients in the $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ vectors can be estimated without bias if both $\mathbf{X}_1$ and $\mathbf{X}_2$ are included.[3] But if the college financial variables in $\mathbf{X}_2$ are omitted ($Y_i = \mathbf{X}_{1i}\boldsymbol{\beta}_1 + \varepsilon_i^r$), then the expected value of the ordinary least squares estimator $\mathbf{b}_1^r$ of the $\boldsymbol{\beta}_1$ vector is

$$E(\mathbf{b}_1^r) = \boldsymbol{\beta}_1 + [\mathbf{X}_1'\mathbf{X}_1]^{-1}\mathbf{X}_1'\mathbf{X}_2\boldsymbol{\beta}_2 \qquad (3)$$

The second term in equation (3) shows that unless all of the parameters in $\boldsymbol{\beta}_2$ are zero, or $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal (unrelated regressors), the parameters in the $\boldsymbol{\beta}_1$ vector are estimated with bias by $\mathbf{b}_1^r$. That is, $E(\mathbf{b}_1^r) = \boldsymbol{\beta}_1$ only if $\boldsymbol{\beta}_2 = 0$ or $\mathbf{X}_1'\mathbf{X}_2 = 0$; or in English, the bias depends on the values of the omitted variables, given the included variables, and the parameters of the omitted variables.

Because this point will be critical when we consider the maximum likelihood estimators of probit and logit index models of college enrollment, the bias or lack of bias in the ordinary least squares estimation $\mathbf{b}_1^r$ does not depend on the distribution of epsilon. It only depends on $\boldsymbol{\beta}_2$ and $\mathbf{X}_1'\mathbf{X}_2$. The may be easier to appreciate by considering a simple case of two explanatory variables: say parental income (which we will label x_1) from the larger data set in matrix $\mathbf{X}_1$ and financial aid (labeled x_2) from the larger omitted data set in matrix $\mathbf{X}_2$.[4] That is, as in equation (1), the true linear probability model is now

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \qquad (4)$$

with critical error term assumptions $E(\varepsilon_i \mid x_{1i}, x_{2i}) = 0$ and $E(\varepsilon_i x_{ji}) = 0$. Financial aid is

dependent on family income. Let this relationship be given by

$$x_{2i} = \delta_0 + \delta_1 x_{1i} + \eta_i \qquad (5)$$

where the δ's are parameters and η_i is the well-behaved error term associated with the i^{th} student's financial aid – that is, its mean is zero, $E(\eta_i \mid x_{1i}) = 0$; it has constant variance, $E(\eta_i^2 \mid x_{1i}) = \sigma_\eta^2$; and it is not related to x_1, $E(x_{1i}\eta_i) = 0$. But if the college financial variable x_2 is omitted ($Y_i = \beta_0 + \beta_1 x_{1i} + \varepsilon_i^r$), then the expected value of the income coefficient estimator b_1^r is

$$E(b_1^r) = \beta_1 + \delta_1\beta_2 \qquad (6)$$

If family income and financial aid are not related ($\delta_1 = 0$), then b_1^r is an unbiased estimator of the family income effect on college enrollment, even though the financial aid variable was omitted. But if family income and financial aid are negatively related ($\delta_1 < 0$), then b_1^r is a biased estimator of the family income effect on college enrollment when the financial aid variable is omitted. In particular, if family income and financial aid are negatively related, excluding the financial aid variable from the college-going decision implies that the effect of family income is underestimated.

In the case of enrollment decisions, all the financial aid variables that enter into the "net price" for a college education (which are excluded from the four NCES studies reviewed by Heller) are clearly related to parental income and other explanatory variables included in the explanation of the college-going decisions in matrix $\mathbf{X}_1$. Thus, the parameters estimated in these studies are biased. In particular, to the extent that the excluded financial variables are negatively (positively) correlated with the included variable, the estimated coefficients can be expected to under-(over-)estimate the parameters of the included variables.

Worth noting is that the inclusion of parental income, parental education, and college financial variables (e.g., net price) in an equation aimed at explaining the college-going decision may make it difficult to estimate the individual effect of these variables, as Heller mentions in his report, because these variables can be highly related. This problem of multicollinearity,

however, does not justify excluding some of these variables from the regression. Omitting them implies that correlation has just been built into the error term and the included regressors, because the effect of the excluded variable(s) is relegated to the error term.[5] When included regressors are highly correlated little can be done to untangle the detrimental effects on estimated coefficient standard errors without new sample data or outside information that can be used to drive the determinant of the $\mathbf{X'X}$ away from zero or otherwise affect the variance covariance matrix. As suggested by Heller, to assess the influence of multicollinearity reporting pair-wise correlations or other measures of regresser dependence might be helpful. One might also consider conducting an *F* test for sets of coefficients of potential explanatory variables that are suspected of being highly collinear and ignoring individual *t* statistics.[6]

Omitted variables are not the only source of regressor and error term correlation. Including regressors that are jointly dependent with the variable to be explained (endogeneity) is another source. For example, in the NCES study by Berkner and Chavez (report 1), attending a college is made a function of being "college-qualified," with this designation used as a zero-one covariate in their two- and four-year college enrollment regressions. But the factors that go into planning to attend college are the same factors that enter the decision to become college-qualified; thus, becoming college-qualified is said to be endogenous in an explanation of the decision to attend college – it is not an independent explanatory variable – it is at least in part determined along with the amount of education to be pursued.

There are several ways in which the designate "college-qualified" can be shown to be endogenous. For example, as recognized by the NCES researchers, becoming college-qualified is itself a function of many factors, which the NCES researchers arbitrarily restrict to completing relatively rigorous academic high school courses, achieving sufficient grades/class rank, taking the SAT/ACT, and applying to college. Instead of this subjective definition of being college-qualified, the full set of factors that determine whether the i^{th} student is truly college-qualified can be written as

$$(collegequalified)_i = f_i(\text{many factors}) + \varphi_i \qquad (7)$$

where φ_i is the error term reflecting the uncertainty in the ith student's qualification, and the

enrollment (Y_i), again written for simplicity as a linear probability model, is

$$Y_i = \beta f_i(many\ \ factors) + ... + \varepsilon_i \qquad (8)$$

But, as in the Berkner and Chavez regression, if enrollment is specified as

$$Y_i = \beta(collegequalified)_i + ... + \varepsilon_i * \qquad (9)$$

then $\varepsilon_i * = \varepsilon_i - \beta\varphi_i$, by substituting equation (7) into equation (8). A positive shock to the error term φ_i in the college-qualified equation (7), produces a like positive move in being college-qualified in both equations (7) and (9) and a negative move in $\varepsilon_i *$. Thus, "$college qualified$" and $\varepsilon *$ are related in equation (9); college-qualified is endogenous, and β would be estimated with bias.

Finally, although of no consequence in our discussion of multicollinearity and endogeneity, implicit in the linear probability model is a heterogeneity problem that is caused by the variance of ε_i depending on the values in the data matrix $\mathbf{X}_i$. This heterogeneity can be removed with a generalized least squares routine.[7] More critically, even though $E(Y_i \mid \mathbf{X}_i) = \text{Prob}(Y_i = 1)$ is between 0 and 1, there is no assurance that the predicted probability of college enrollment in a linear probability model will fall between 0 and 1. For this reason, the linear probability model is best viewed as a starting and comparison point for estimating the probability of enrolling in college. Because of its simplicity, it is ideal for showing the bias introduced to parameter estimation when omitted relevant variables are related to the included explanatory variables or problems of endogeneity are suspected.

LATENT REGRESSION, LOGITS AND PROBITS, AND OMITTED VARIABLES

Consider the student's decision to enroll in college. Classical microeconomic theory states that the student will enroll if the net utility or net benefit of enrolling is positive. It is intuitively appealing, although not necessary, to interpret this net utility as the unobservable latent variable y*. For the i^{th} student,

$$y_i^* = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i \qquad (10)$$

where **X** is again the data matrix of explanatory variables of all students; $\boldsymbol{\beta}$ is the vector of related parameters; the error term vector is $\boldsymbol{\varepsilon}$; and the subscript *i* denotes the appropriate row for the i^{th} student. The error term ε_i is again assumed to have a mean of zero. If a logit model is specified, then y_i^* = Ln[Prob(Y_i = 1)/Prob(Y_i = 0)]. If a standardized probit model is specified, then $y_i^* = -z_i$ (where *z* is a standard normal score from which probability is calculated) and ε_i has a variance of $\sigma_\varepsilon^2 = 1$. (With no loss in generality, the unit variance for ε_i is achieved by interpreting the beta coefficients as divided by the standard deviation of epsilon for scaling. This scaling issue becomes critical when the omitted-variable problem is considered in what follows.)

Both the logit and probit models ensure that the predicted probabilities of college enrollment lie between 0 and 1, but these models greatly increase the mathematical complexity of parameter estimation via maximum-likelihood, nonlinear-iterative routines that require a properly specified population model from which the data are believed to be generated. Other than for reasons of computation, which current computer programs handle with equal ease, there is typically little reason to prefer a probit or logit model. The main difference between probit and logit models is that the conditional probability of enrolling in college approaches the extreme values of zero or one at a slightly slower rate in a logit than in a probit because the logistic distribution has slightly fatter tails. The implications of omitted relevant explanatory variables on the consistent estimators of parameters (estimates that collapse on their true expected values of the betas as the sample size goes to infinity) are similar in the logit and probit models, but as we will see, different than in the linear probability model.

If the student enrolls in college, then we observe $Y_i = 1$ and infer that $y_i^* > 0$. If there is no college enrollment, then $Y_i = 0$ is observed and $y_i^* \leq 0$ is inferred. Making the distinction between the college financial variables $\mathbf{X}_2$ and the other columns in the data matrix $\mathbf{X}$ gives the basic college enrollment model as:

$$y_i^* = \mathbf{X}_{1i}\boldsymbol{\beta}_1 + \mathbf{X}_{2i}\boldsymbol{\beta}_2 + \varepsilon_i \tag{11}$$

$Y_i = 1$ (observed college enrollment), if $y_i^* > 0$ (unobserved)

$Y_i = 0$, if $y_i^* \leq 0$

For simplicity in algebra, and to make explicit the nature of the omitted variables problem in a latent regression model of binary choice, consider only a two explanatory variable model for the propensity to enroll in college:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i \tag{12}$$

$Y_i = 1$, if $y_i^* > 0$ and $Y_i = 0$, if $y_i^* \leq 0$

If x_2 and x_1 are related linearly as before in equation (5), then equation (12) can be rewritten as

$$y_i^* = \beta_0 + \beta_1\delta_0 + (\beta_1 + \beta_2\delta_1)x_{1i} + \varepsilon_i + \beta_2\eta_i \tag{13}$$

If ε and η are independently distributed as normal random variables, conditioning on x_1, then equation (13) is a bivariate probit model, where again $y_i^* > 0$ for $Y_i = 1$ and $y_i^* \leq 0$ for $Y_i = 0$.

In the linear probability model, any bias in the estimation of the coefficients of the included variables depends on the values of the omitted variables, given the included variables, and the parameters of the omitted variables. Unlike this linear probability model, estimation of β_1 in a latent regression model depends on the assumed distribution of the error term $\varepsilon_i + \beta_2\eta_i$. There can be bias in the estimation of β_1 even if the excluded and included explanatory variables are unrelated because the iterative estimation process depends on the error term distribution (intuitively, think of it as draws from the error term distribution to generate the unobserved y_i^* values). For the i^{th} student in equation (13), the error term $\varepsilon_i + \beta_2\eta_i$ has a mean of zero and a

variance of $\beta_2^2\sigma_\eta^2 + \sigma_\varepsilon^2$. Assuming the error term is a standard normal random variable for maximum-likelihood estimation of β_2 in the probit specification implies that β_2 is involved in the scaling of β_1. From Yatchew and Griliches (1985), we know that maximum likelihood estimators of β_0 and β_1 converge to

$$\frac{\beta_0 + \beta_2\delta_0}{\sqrt{\beta_2^2\sigma_\eta^2 + \sigma_\varepsilon^2}} \text{ and } \frac{\beta_1 + \beta_2\delta_1}{\sqrt{\beta_2^2\sigma_\eta^2 + \sigma_\varepsilon^2}} \qquad (14)$$

From the second ratio in (14), omitting the financial aid variable x_2 from the estimation of the propensity to enroll in college, equation (12), has two effects on the estimation of family income coefficient β_1. First, as in the linear probability model, there is the bias in the family income coefficient that is equal to the coefficient of the omitted financial aid variable ($\beta_2 > 0$), times the coefficient of the income variable in the regression of the omitted financial variable on the included parent income variable ($\delta_1 < 0$). Second, and unlike the linear probability model, there is a rescaling effect determined by the standard deviation $\sqrt{\beta_2\sigma_\eta^2 + \sigma_\varepsilon^2}$ in the denominator of equation (14), which does not vanish even if there is no relationship between family income and financial aid ($\delta_1 = 0$). That is, omitted relevant variables in a probit model result in biased estimation of the coefficient of the included explanatory variables regardless of the relationship between the included and excluded variables.

Although individual coefficients cannot be estimated without bias when relevant variables are omitted from a probit model, relative effects can be if the omitted variables are not correlated with the included variables.[8] For example, the estimate of the slope relative to the intercept is the ratio of the two ratios in (14), which is (β_1 / β_0), if β_2 is zero. Of course, if the omitted and included variables are related ($\beta_2 \neq 0$), then as in a nonlinear probability model individual and relative effects cannot be calculated without bias.

The omitted-variable problems found in the estimation of probit models also exist in other nonlinear probability models. In particular, excluding the financial variables from a logit model

produces a bias in the maximum-likelihood parameter estimates for the included explanatory variables, regardless of the relationship among the included and excluded variables. From the early work of Lee (1982), the existence of this bias in the logit model has been known: "In the standard linear (probability) model, if the omitted variable and the included variable are independent, the coefficient of the included variable will not be affected. But this is not so for the logit model." (p. 208) Biased estimation of the coefficient of included explanatory variables in either a probit or logit model of enrollment will occur even if the included and excluded variables are independent.

SAMPLE SELECTION

Reports 1 and 3 provide multivariate analyses only for college-qualified students (i.e., those who have completed relatively rigorous academic high school courses, achieved sufficient grades/class rank, took the SAT/ACT, and applied to college) who desire to enroll and/or persist in college.

Family income and parental education level are explanatory variables considered in all four reports. Other explanatory variables (such as race, gender, an index of "college qualification," educational expectations, and whether the student took a college entrance examination) appear in only one or two of the reports. As already addressed, one of the two college participation outcomes considered by the NCES and addressed by Donald Heller is students' initial enrollment in college. The second outcome is whether the students persist to degree attainment. For pedagogical ease, in this section I now treat these as one observable outcome (y_i^p) generated by a continuous random variable that measures the amount of time the i^{th} student invests or persists in college. For example, y_i^p could be measured by the number of terms completed, with 0 reflecting a young person who never attended college and an undefined upper limit. The problems associated with sample selection and endogeneity can be demonstrated with the model

$$y_i^p = \mathbf{X}_i\boldsymbol{\beta} + \varepsilon_i = \beta_1 + \sum_{j=2}^{k} \beta_j x_{ji} + \varepsilon_i \qquad (15)$$

where again $\mathbf{X}$ is the data set of explanatory variables, $\mathbf{X}_i$ is the row of x_{ji} values for the relevant variables believed to explain the i^{th} student's decision to enroll in and continue in college, the β_j's are the associated slope coefficients in the vector $\boldsymbol{\beta}$, and ε_i is the individual random shock (caused by unknown events, for example) that affect the i^{th} student's persistence. In empirical work, the exact nature of y_i^p is critical. For example, to model the truncation issues in the distribution of epsilon a Tobit model can be specified for y_i^p. As we have already seen, to explicitly model only the college going decision as a "yes" or "no" choice a logit or probit model can be specified. These refinements do not alter the basic issues regarding sample selection and endogeneity that I address in this section.

In accordance with the National Center for Education Statistics, let $T_i = 1$, if the i^{th} student is "college-qualified" (i.e., completed relatively rigorous academic high school courses, achieved sufficient grades/class rank, took the SAT/ACT, applied to college), and let $T_i = 0$, if not. Assume that there is an unobservable continuous dependent variable T_i^* underlying the i^{th} student's decision to jump the NCES hurdles to be labeled college-qualified. Call this latent variable T_i^* the student's propensity to be college-qualified.

For the population of N students, let $\mathbf{T}^*$ be the vector of all students' propensities to be college-qualified. Let $\mathbf{H}$ be the matrix of explanatory variables that are believed to drive these propensities, which includes directly observable things (e.g., parental income and parental education), expected values (expected college net cost and expected future earnings), attitude variables (personal motivation and risk aversion) and environmental items (peer pressure, counseling, family support). Let $\boldsymbol{\alpha}$ be the vector of corresponding slope coefficients. The individual random shocks that affect each student's propensity to become qualified for college are contained in the error term vector $\boldsymbol{\omega}$. The i^{th} student's propensity to become college-qualified can now be written

$$T_i^* = \mathbf{H}_i\boldsymbol{\alpha} + \omega_i \qquad (16)$$

where

$T_i = 1$, if $T_i^* > 0$, and student i is college-qualified, and

$T_i = 0$, if $T_i^* \leq 0$, and student i is not qualified.

For estimation purposes, the error term ω_i is assumed to be a standard normal random variable that is independently and identically distributed with the other students' error terms in the $\boldsymbol{\omega}$ vector. As already discussed, this probit model for the propensity to be college-qualified can be estimated using the maximum-likelihood routine in programs such as LIMDEP or STATA.

Studies supported by the NCES and others are aimed at establishing the effect of variables believed to influence decisions related to attending and succeeding in college. The effect of not

including students who have not completed academic high school courses, do not have sufficient grades or class rank, do not take the SAT/ACT, and do not apply to college $(T_i = 0)$ and an adjustment for the resulting bias caused by excluding these students from the NCES studies of these college-related decisions can be illustrated with a two-equation model formed by the selection equation (16) and the i^{th} student's persistence into and through college, persistence equation (15).[9] Each of the disturbances in vector $\boldsymbol{\varepsilon}$, equation (15), are assumed to be distributed bivariate normal with the corresponding disturbance term in the $\boldsymbol{\omega}$ vector of the selection equation (16). Thus, for the i^{th} student we have

$$(\varepsilon_i, \omega_i) \sim \text{bivariate normal}(0,0,\sigma_\varepsilon,1,\rho) \tag{17}$$

and for all perturbations in the two-equation system we have

$$E(\boldsymbol{\varepsilon}) = E(\boldsymbol{\omega}) = 0, \;\; E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma_\varepsilon^2 \mathbf{I}, \;\; E(\boldsymbol{\omega}\boldsymbol{\omega}') = \mathbf{I}, \text{ and } \mathrm{E}(\varepsilon\omega') = \rho\sigma_\varepsilon \mathbf{I} \tag{18}$$

That is, the disturbances have zero means, unit variance, and no covariance among students, but there is covariance between selection in getting the college-qualified status and persistence into and through college for each student.

The difference in the functional forms of the selection equation (16) and the persistence equation (15) ensures the identification of equation (15) but ideally other restrictions would lend support to identification. Estimates of the parameters in equation (15) are desired, but the i^{th} student's college persistence y_i is observed in the NCES studies for only the subset of students for whom $T_i = 1$. The regression for this censored sample of $n_{T=1}$ students is

$$E(y_i^p \mid \mathbf{X}_i, T_i = 1) = \mathbf{X}_i\boldsymbol{\beta} + E(\varepsilon_i \mid T_i^* > 0); \; i = 1,2,\ldots n_{T=1} \text{ (for } n_{T=1} < N) \tag{19}$$

Similar to omitting a relevant variable from a regression, selection bias is a problem because the magnitude of $E(\varepsilon_i \mid T_i^* > 0)$ varies across individuals and yet is not included in the estimation of equation (15). To the extent that ε_i and ω_i (and thus T_i^*) are related, the estimators are biased.

The persistence regression can be adjusted for those who elected never to become college-qualified in several ways. An early Heckman-type solution to the sample selection problem is to rewrite the omitted variable component of the regression so that the equation to be estimated is

$$E(y_i^p \mid \mathbf{X}_i, T_i = 1) = \mathbf{X}_i\boldsymbol{\beta} + (\rho\sigma_\varepsilon)\lambda_i;\ i = 1, 2, \ldots n_{T=1} \qquad (20)$$

where $\lambda_i = f(-T_i^*)/[1 - F(-T_i^*)]$, and $f(.)$ and $F(.)$ are the normal density and distribution functions. The inverse Mill's ratio (or hazard) λ_i is the standardized mean of the disturbance term ω_i, for the i^{th} student who was college-qualified; it is close to zero only for those well above the $T = 1$ threshold. The values of λ are generated from the estimated probit selection equation (16) for all students. Each student in the persistence regression gets a calculated value λ_i, with the vector of these values serving as a shift variable in the persistence regression. The estimates of both ρ and σ_ε and all the other coefficients in equations (15) and (16) can be obtained simultaneously using the maximum-likelihood routine in LIMDEP.

The Heckman type selection model represented by equations (15) and (16) makes clear the nature of the sample selection problem inherent in establishing determinants of the college-going decision. Estimation of the parameters in this model, however, requires cross-equation exclusion restrictions (variables that affect selection but not enrollment and persistence), differences in functional forms, and/or distributional assumptions for the error terms. Parameter estimates are typically sensitive to these model specifications.

Alternative nonparametric and semiparametric methods are being explored for assessing treatment effects in nonrandomized experiments (Heckman, 1990; Manski, 1990; and Newey, et al., 1990) but these methods have been slow to catch on in education research. Exceptions, in the case of financial aid and the enrollment decision, are the works of Wilbert van der Klaauw and Thomas Kane. Van der Klaauw (2002) estimates the effect of financial aid on the enrollment decision of students admitted to a specific east coast college, recognizing that this college's financial aid is endogenous because competing offers are unknown and thus by definition are omitted relevant explanatory variables in the enrollment decision of students considering this college.

The college investigated by van der Klaauw created a single continuous index of each student's initial financial aid potential (based on a SAT score and high school GPA) and then classified

students into one of four aid level categories based on discrete cut points. The aid assignment rule depends at least in part on the value of a continuous variable relative to a given threshold in such a way that the corresponding probability of receiving aid (and the mean amount offered) is a discontinuous function of this continuous variable at the threshold cut point. A sample of individual students close to a cut point on either side can be treated as a random sample at the cut point because on average there really should be little difference between them (in term of financial aid offers received from other colleges and other unknown variables). In the absence of the financial aid level under consideration, we should expect little difference in the college-going decision of those just above and just below the cut point. Similarly, if they were all given the financial aid, we should see little difference in outcomes, on average. To the extent that some actually get it and others do not, we have an interpretable treatment effect. (Intuitively, this can be thought of as running a regression of enrollment on financial aid for those close to the cut point, with an adjustment for being in that position.) In his empirical work, van der Klaauw obtained credible estimates of the importance of the financial aid effect without having to rely on arbitrary cross-equation exclusion restrictions and functional form assumptions. His estimates suggest that an additional $1,000 in financial aid results in a 4 to 5 percentage point increase in the probability of the mean student attending this university.

Kane (2003) uses an identification strategy similar to van der Klaauw but does so for all those who applied for the Cal Grant Program to attend any college in California. Eligibility for the Cal Grant program is subject to a minimum GPA and maximum family income and asset level. Like van der Klaauw, Kane exploits discontinuities on one dimension of eligibility for those who satisfy the other dimensions of eligibility. His results suggest that additional financial aid dollars have a large impact on students' decisions even when provided late in the schooling process: "Financial aid applicants were 4 to 6 percentage points more likely to enroll in college as a result of the receipt of a Cal Grant A award, even after they have already made the investment of filing a federal financial aid form and applied to college." (p. 26) For comparison purposes if nothing else, there is value in pursuing these alternative approaches to the sample selection and endogeneity issues that van der Klaauw and Kane recognize as intrinsic to the analyses of the college-going decision.

CONCLUSION

The four NCES reports reviewed by Heller, as he states, are limited by some of the data sets employed and by some decisions that NCES and its contractors made about the importance of including financial aid variables that other researchers have found to influence college-going decisions. NCES should commission a complete reanalysis of the existing data to include financial variables in the multivariate models that were excluded in the initial work but which others have found to influence college-going decisions. A thorough explanation of the rationale for including or excluding each variable must be provided based on the existing literature.

In addition to addressing the omitted-variable problems, NCES must reexamine its use of the college qualification index (and the additional two steps toward entry into four-year institutions of taking the SAT or ACT and applying to college). Use of this index introduces a sample selection problem that cannot be overcome by simply adding more variables or more observations. The problem of sample selection in the college-going decision arises because youths who elect to pursue a certain high school experience are those who expect schooling to have a favorable outcome for them. If expected outcomes are related to observed ones, then the outcomes experienced by youth who choose to become college-qualified would differ from those that non-college-qualified youth would have experienced if they had become college-qualified. From the early work of Nobel laureate James Heckman (1979) this sample selection problem and methods of adjusting for it are well known and should not have been overlooked by the NCES. Furthermore, Dominitz and Manski (1996) and Betts (1993) document the fact that youth from low income families greatly underestimate the return to a college education and thus can be expected not to pursue the steps required to become college-qualified:

> "One of the most interesting patterns is that students whose parents' income was less than $50,000 tended to make significantly lower estimates of earnings of college graduates . . .young people form beliefs about the returns to education by observing workers in their neighborhoods. To the extent that families segregate themselves by income, students in low-income neighborhoods should systematically underestimate the return to education." (Betts, 1993, p. 37)

We can only assume that low-income students likewise would underestimate the financial aid

available to them if they were to start down the more rigorous college prep path and succeeded in becoming college-qualified.[10]

It is absolutely essential that the studies and findings advanced by the NCES not be based on methodologies that do not adequately control for the high school student's expectations of the net cost of college and future earnings. Without including relevant financial aid measures in studies of college access and without adequately controlling for sample selection in the college-going decision, conclusions based on regression analyses about the importance of other explanatory variables cannot be taken seriously.

END NOTES

[1] Kane's example is from a letter sent to Jacqueline King (June 13, 2002) at the American Council on Education and reproduced here with his permission (April 14, 2003, email).

[2] Edward St. John (2002) may have been the first education researcher to call attention to the NCES studies ignoring the effect of financial-aid variables when analyzing the cause of disparity in college access.

[3] Because the error terms in a linear probability model are heteroskedastic, are not distributed normally, and predicted probabilities can lie outside the 0-1 interval, use of this specification is criticized other than for preliminary investigation and comparison purposes. Caudill (1988), however, argues that the linear probability model has an advantage over the probit and logit models when all members of a subgroup have the same outcome. For example, we may be interested in the effect of having completed high school calculus on the college-going decision. If every student who took high school calculus went to college, the coefficient on the high school calculus dummy is not estimable in either a logit or probit model, but it can be estimated in a linear probability model. Heckman and Snyder (1997) provide a general derivation for the linear probability model as a representation of a random utility model.

[4] The financial-aid variables are typically treated as exogenous in single equation enrollment models. Without adequate control variables included in the regression this is a dubious assumption.

[5] The building of models based on data-mining routines such as stepwise regression are doomed by the omitted variable problem. If a relevant variable is omitted from a regression in an

early step, and if it is related to the included variables then the contribution of the included variables is estimated with bias. It does not matter with which of the related explanatory variables you start; the contribution of the included variables will always be biased by the excluded.

[6] The null hypothesis that a financial aid variable (or sets of variables) has no effect can never be accepted for there is always another hypothesized value, in the direction of the alternative hypothesis, that cannot be rejected with the same sample data and level of significance. The Type II error inherent in accepting the null hypothesis is well known but often ignored by researchers.

[7] In the linear probability model heteroskedasticity of the error term does not lead to inconsistent estimation of the regression parameters, but the standard errors will be wrong (inconsistently estimated). That is why heteroskedasticity-corrected standard errors must be reported.

[8] In semiparametric estimation of the discrete choice models, as seen in Klein and Spady (1993), the inability to identify the intercept is common. In those papers the focus is only on the relative parameter values.

[9] Although y_i^p is treated as a continuous variable this is not essential. For example, a bivariate choice (probit or logit) model can be specified to explicitly model only the college-going decision as a "yes" or "no" for students who enrolled within two years of high school graduation as in L. Berkner and L. Chavez, "Access to Postsecondary Education for the 1992 High School Graduates" (NCES 1997, 98-105) and L. Horn and A. –M. Nuñez , "Mapping the Road to College: First-Generation Students' Math Track, Planning Strategies and Context

Support (NCES 2000 2000-153). The selection issue is then modeled in a way similar to that employed by Boyes, Hoffman and Low (1989) regarding loan defaults given the granting of a loan and Greene (1992) on consumer loan default and credit card expenditures. Our two-equation model for the i^{th} student enrolling given he or she is college-qualified is then based on

y_i^p = 1, if student actually enrolled in a college, and 0 otherwise.

T_i = 1, if student is college-qualified, and 0 otherwise.

As with the standard Heckman selection model, this two-equation system involving bivariate choice and selection can be estimated in a program like LIMDEP.

[10] Linsenmeier, Rosen and Rouse (2003) report in a study based on a single institution that substituting grants for loans did not have a significant effect on the likelihood that low-income students actually start college at the school making the switch. However, the switch did have a larger effect on minorities than other like low-income students, suggesting that minority students' expectations of their post-college income are less certain, giving a bigger impact to the importance of the financial-aid mix.

REFERENCES

Betts, Julian R. 1996. "What Do Students Know About Wages*?" Journal of Human Resources*, 31: 27-56.

Boyes, W., D. Hoffman and S. Low. 1989. "An Econometric Analysis of the Bank Credit Scoring Problem." *Journal of Econometrics*, 40: 3-14.

Caudill, Steven B. 1988. "An Advantage of the Linear Probability Model over Probit and Logit." *Oxford Bulletin of Economics and Statistics*, 50: 425-427.

Dominitz, Jeff, and Charles F. Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources*, 31: 1-26.

Greene, William H, 1992. "A Statistical Model for Credit Scoring." Department of Economics, Stern School of Business, New York University, September 29.

Griliches, Zvi. 1957. "Specification Bias in Estimates of Production Functions." *Journal of Farm Economics*, 39: 8-20.

Heckman, James. 1979. "Sample Bias as a Specification Error." *Econometrica*, 47: 153-162.

Heckman, James. 1990. "Varieties of Selection Bias." *American Economic Review*, 80 (May): 313-318.

Heckman, James, and J. M. Snyder. 1997. "Linear Probability Models of the Demand for Attributes with an Empirical Application to Estimating the Preferences of Legislators." *Rand Journal of Economics*, 28: 142-189.

Heller, Donald E. 2003. "Review of NCES Research on Financial Aid and College Participation: Report prepared for the Advisory Committee on Student Financial Assistance." Draft: March.

Kane, Thomas. 2003. "A Quasi-Experimental Estimate of the Impact of Financial Aid on College-Going" NBER Working Paper No. W9703 May.

Klein, R. and R. Spady. 1993. "An Efficient Semiparametric Estimator for Discrete Choice Models." *Econometrica*, 61: 3870-3921.

Lee, Lung-Fei. 1982. "Specification Error in Multinomial Logit Models: Analysis of the Omitted Variable Bias." *Journal of Econometrics*, 20: 197-209.

Linsenmeier, David, Harvey Rosen and Cecilia Rouse. 2003. "Financial Aid Packages and College Enrollment Decisions: An Econometric Case Study." National Bureau of Economic Research, Working Paper No. 9228.

Manski, Charles. 1990. “Nonparametric Bounds for Treatment Effects.” *American Economic Review*, 80 (May): 319-323.

Newey, Whitney, James Powell and James Walker. 1990. “Semiparametric Estimation of Selection Models: Some Empirical Results.” *American Economic Review*, 80 (May): 324-328.

St. John, Edward. 2002. The Access Challenge: Rethinking the Causes of the New Inequality. Indiana Education Policy Center, School of Education Policy Issue Report #2002-1.

Theil, Henri. 1957. “Specification Errors and the Estimation of Economic Relationships.” *Review of the International Statistical Institute*, 2: 41-51.

van der Klaauw, Wilbert. 2002. “Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discounting Approach.” *International Economic Review,* November: 1249-1288.

Yatchew, Adonis, and Zvi Griliches. 1985. “Specification Error in Probit Models.” *Review of Economics and Statistics*, 67: 134-139.

APPENDIX A

ADVISORY COMMITTEE ON STUDENT FINANCIAL ASSISTANCE

AGENDA

COLLOQUIUM ON ACCESS RESEARCH
MACALESTER COLLEGE
(WEYERHAEUSER LOUNGE AND BOARD ROOM)
SAINT PAUL, MINNESOTA

JUNE 12, 2003

8:15 a.m. **Continental Breakfast**
(Weyerhaeuser Lounge and Board Room)

8:30 a.m. **Opening/Welcoming Remarks**

Dr. Michael S. McPherson, President
Macalester College

9:00 a.m. **Background, Purpose, and Importance of Colloquium**

Dr. Brian K. Fitzgerald, Staff Director
Advisory Committee on Student Financial Assistance

9:30 a.m. **Review of NCES Research on Financial Aid and College Participation**

Dr. Donald E. Heller, Associate Professor of Higher Education
Pennsylvania State University

10:00 a.m. **Statistical Consequences and Implications in NCES Research**

Dr. William E. Becker, Professor of Economics
Indiana University

10:30 a.m. **Response of Reviewers to Papers***

Moderator: Dr. Brian K. Fitzgerald, Staff Director
Advisory Committee on Student Financial Assistance

Dr. Sandy Baum, Professor of Economics
Skidmore College

Dr. David W. Breneman, University Professor and Dean
Curry School of Education, University of Virginia
Response of Reviewers to Papers*

Dr. Patrick M. Callan, President
The National Center for Public Policy and Higher Education

Dr. Michael S. McPherson, President
Macalester College

Dr. Edward St. John, Professor of Education
Indiana University

***Other reviewers, who could not attend, include:**

Dr. Anthony P. Carnevale, Vice President
Educational Testing Service

Dr. Bridget T. Long, Assistant Professor of Education
Harvard Graduate School of Education

Mr. Jamie P. Merisotis, President
Institute for Higher Education Policy

Dr. Morton Owen Schapiro, President
Williams College

Dr. William Trent, Professor of Educational Policy Analysis
University of Illinois, Urbana-Champaign

11:30 a.m. **Full Roundtable Discussion Among Presenters, Reviewers, and the following Invitees with Audience Participation**

Dr. David B. Laird, Jr., President and Chief Executive Officer
Minnesota Private College Council

Dr. Paul E. Lingenfelter, Executive Director
State Higher Education Executive Officers (SHEEO)

Dr. Charles S. Lenth, Vice President of Research and Policy Development
Minnesota Private College Council

Dr. Derek V. Price, Director, Higher Education Research
Lumina Foundation

12:30 p.m. **Informal Discussion continued over Lunch**
(Weyerhaeuser Board Room)

THE PARTICIPANTS

Dr. Sandy Baum is professor of economics at Skidmore College. She is the author and co-author of numerous publications including *"College Education: Who Can Afford It?" The Finance of Higher Education: Theory, Research, Policy, and Practice* (Agathon Press, 2001). She also worked on several book reviews, such as Tuition Rising: Why College Costs So Much (Ronald Ehrenberg. Eastern Economic Journal. Forthcoming) and Keeping College Affordable (Michael McPherson and Morton Owen Schapiro, Eastern Economic Journal. 19(1).Winter 1993). Dr. Baum earned her B.A. in sociology at Bryn Mawr College and her M.A. and Ph.D. degrees in economics at Columbia University.

Dr. William E. Becker is professor of economics at Indiana University, Bloomington, and an adjunct professor at the University of South Australia, where he was last in residence in 2000. His research appears in numerous journals such as the American Economic Review (Refereed and Proceedings). He is the author, co-author, and editor of numerous books and publications, such as, Statistics for Business and Economics (South-Western, International Thomson Publishing), and Econometric Modeling in Economic Education Research (Kluwer-Nijhoff), among others. He earned his bachelor's degree in mathematics from the College of St. Thomas, a master's degree in economics from the University of Wisconsin, and a doctorate in economics from the University of Pittsburgh.

Dr. David W. Breneman is a scholar specializing in the economics of higher education and public policy. He is currently university professor and dean of the Curry School of Education at the University of Virginia. He is the author of several books and publications on higher education including Liberal Arts Colleges: Thriving, Surviving, or Endangered?, published in 1994 by Brookings. Dr. Breneman received his B.A. in Philosophy from the University of Colorado, his Ph.D. in Economics from the University of California at Berkeley and in 1999 received an honorary Doctor of Education degree from Worcester State College.

Mr. Patrick M. Callan is a leader in higher education and president of the National Center for Public Policy and Higher Education. Mr. Callan has co-authored and authored numerous articles and papers on education, educational opportunity, public accountability, and leadership, such as, *Public and Private Financing of Higher Education: Shaping Public Policy for the Future* (1997) and recently collaborated with Gene Maeroff and Michael Usdan on *The Learning Connection, New Partnerships between Schools and Colleges*, published by Teachers College Press in 2001.

Dr. Anthony P. Carnevale is the vice president for Education and Careers at Educational Testing Service (ETS). From 1983-1993, he served as the president of the Institute for Workbased Learning of the American Society for Training and Development and was appointed by President Clinton as chair of the National Commission for Employment Policy. He is the author of numerous books, manuals and articles on diversity training, job skills, education and school reform. He holds a Ph.D. from the Maxwell School of Citizenship and Public Affairs of Syracuse University and a B.A. from Colby College.

Dr. Donald E. Heller is a scholar specializing in higher education finance, tuition pricing, financial aid, and student access. He is currently an associate professor and senior research associate at the Center for the Study of Higher Education at The Pennsylvania State University. Dr. Heller earned an Ed.D. in Higher Education from the Harvard Graduate School of Education and holds an Ed.M in Administration, Planning, and Social Policy from Harvard. He received his B.A. in Economics and Political Science from Tufts University. He is the editor of the books *The States and Public Higher Education Policy: Affordability, Access, and Accountability* (Johns Hopkins University Press, 2001) and *Condition of Access: Higher Education for Lower Income Students* (ACE/Praeger, 2002).

Dr. David B. Laird, Jr. is the president and chief executive officer of the Minnesota Private College Council (MNPRIVCO), an association of 17 four-year, private liberal arts colleges and universities. Dr. Laird is considered a national expert on higher education financing and public policy. Dr. Laird earned his B.A. degree in government and history and M.Ed. in education and history from St. Lawrence University. He earned his Ph.D. from the University of Michigan where he served as a teaching fellow.

Dr. Paul E. Lingenfelter is executive director of the State Higher Education Executive Officers (SHEEO). Prior to joining the MacArthur Foundation, Dr. Lingenfelter served as Deputy Director of Fiscal Affairs for the Illinois Board of Higher Education and has been retained as a consultant by the United States Corporation for National Service, the Laidlaw Foundation in Canada, the Education Commission of the States, the New York Borad of Regents, and the U.S. Office of Education. Dr. Lingenfelter received his A.B. in Literature from Wheaton College, his MA from Michigan State University, and his Ph.D. in higher education from the University of Michigan.

Dr. Charles S. Lenth is vice president for research and policy development for the Minnesota Private College Council, Fund and Research Foundation. As vice president, he undertakes collaborative demographic and institutional research, and contributes to the organization's functions in the areas of policy analysis, advocacy and evaluation relative to financial aid, among others. He received his Ph.D. in Political Science for the University of Chicago, specializing in South Asian politics. Effective July 1, 2003, Dr. Lenth will be senior associate with SHEEO in Denver.

Dr. Bridget T. Long is assistant professor of Education at the Harvard Graduate School of Education. Trained as an economist, Professor Long applies the theory and methods of economics in her work to examine various aspects of the market for higher education. Her research interests focus on college access and choice, the effects of financial aid policy, and the behavior of postsecondary institutions. She is the recipient of numerous awards and has authored a number of publications including *The Impact of Federal Tax Credits for Higher Education* (forthcoming). She earned her Ph.D. from Harvard University.

Dr. Michael S. McPherson, who began serving as president of Macalester College in 1996, is a nationally known economist, writer, and authority on the financing of higher education, as well as on philosophical dimensions of economics. Dr. McPherson is the co-author and editor of seven books, including *Keeping College Affordable: Government and Educational Opportunities* (Brookings, 1991) and *Paying the Piper: Productivity, Incentive and Financing in American Higher Education* (University of Michigan Press, 1993). He was also one of the founding editors of the journal, *Economics and Philosophy*, published by Cambridge University Press. He earned his B.A. in mathematics and his M.A. and Ph.D. in economics at the University of Chicago.

Mr. Jamie P. Merisotis is the founding president of the Institute for Higher Education Policy, a nonprofit, nonpartisan research and policy organization located in Washington, D.C. Mr. Merisotis serves as director of many of the Institute's major projects, including, most importantly, the "New Millennium" project, a multi-year study of college costs, pricing, and productivity funded by the Ford Foundation. He earned his B.A. degree at Bates College. He is the author of dozens of publications and articles on higher education and student aid, such as the Commission's heralded report, *Making College Affordable Again, Minority-Serving Institutions: Distinct Purposes, Common Goals.*

Dr. Derek V. Price is a researcher and the director of Higher Education Research for the Lumina Foundation for Education, where he is responsible for the Foundation's research and analyses on access and success in postsecondary education. Dr. Price has published original research in the *Journal of Student Financial Aid, Race, Gender & Class*, most recently, the *Journal of Poverty: Innovations on Social, Political and Economic Inequalities*. He holds a doctorate degree in sociology from American University, a master's degree from the University of Michigan in Ann Arbor, Michigan and a bachelor's degree from Duke University in Durham, North Carolina.

Dr. Morton Owen Schapiro became professor of Economics and the 16th president of Williams College on July 1, 2000. From 1980 to 1991, he served as professor of Economics and an assistant provost at Williams College. He is among the nation's premier authorities on the economics of higher education, with particular expertise in the area of college financing and affordability, and on trends in educational costs and student aid. Dr. Schapiro has written more than fifty articles and five books, including (with his long-term co-author Dr. Michael S. McPherson) *The Student Aid Game: Meeting Need and Rewarding Talent in American Higher Education* (Princeton University Press, 1998). He received his bachelor's degree in economics from Hofstra University in 1975 and his doctorate from the University of Pennsylvania in 1979.

Dr. Edward St. John is professor of Education at Indiana University. Dr. St. John had extensive experience with policy research on both higher education and school improvement issues prior to joining Indiana University. He is the author and co-author of numerous publications, including *Refinancing the College Dream: Access, Equal Opportunity, and Justice for Taxpayers* (John Hopkins University Press) and *Keeping public colleges affordable: A study of persistence in Indiana's public colleges and universities*. Dr. St. John earned his Ed.D. in Administration, Planning and Social Policy from Harvard University and his M.Ed. and B.S. degrees from the University of California, Davis.

Dr. William Trent is professor of Educational Policy Studies at the University of Illinois at Urbana-Champaign and professor, Department of Sociology. He has authored and co-authored several publications including Justice, equality of educational opportunity and affirmative action in higher education (2000) and Focus on equity: Race & gender differences in degree attainment (1976-76). Dr. Trent earned his Ph.D. in Sociology from the University of North Carolina at Chapel Hill, his M.S. in Sociology from George Washington University in Washington, D.C., and his B.S. in Sociology from Union College, Barbourville, Kentucky.