Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES
TASK FORCE REPORT

EMERGING ISSUES IN
POSTSECONDARY ACCESS AND CHOICE

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

EMERGING ISSUES IN POSTSECONDARY ACCESS AND CHOICE

EXECUTIVE SUMMARY

The *Postsecondary Choice Project* was conducted by the National Institute of Statistical Sciences (NISS) for the National Center for Education Statistics (NCES) to address existing and emerging issues of postsecondary access and choice. The project culminated in a workshop held on 26 January 2009 at NCES in Washington, DC, where the papers that make up this report were presented.

Postsecondary choice comprises a complex set of issues with strong scientific content and important policy implications. The processes are inherently complicated, involving a multiplicity of participants and stakeholders. Consequently there is a multiplicity of goals for collecting data, which has implications for NCES' postsecondary data collections.

The goal of the *Postsecondary Choice and Access Project* was to examine the possibility of creating a set of economic- and social-based constructs for describing and reasoning about postsecondary choice today and into the future. Once defined, these constructs could be mapped into ongoing and potential future data collections by NCES. To the extent feasible, data elements and instrument designs could be identified in order to operationalize and measure the constructs.

This report includes the five papers presented at the Workshop. The process of making choices was viewed in terms of what decisions are made, when, by whom, under whose influence, using what criteria, and on the basis of what information is obtained from which sources. Authors were asked to consider three aspects of the topic:

- What we know about the rapidly changing, increasingly heterogeneous processes of postsecondary choice?

- What components of postsecondary choice are possibly less completely understood than they should be? One example is the increasing multiplicity of agents - among them, students, parents, secondary schools, postsecondary institutions, the federal government, and lenders.

- How can extant and future NCES data collections better inform both policy and scientific understanding? Individual papers discuss background, criteria for defining constructs, and consequences for data collection.

In *A Conceptual Structure for Postsecondary Access and Choice,* Alan Karr sets the stage for discussing choice and access by providing a structure to be further defined by core constructs. The process is inherently complex, first because choice among alternative postsecondary paths is actually a series of choices, second because of the involvement of multiple agents beside the student, and third because processes themselves are evolving, reflecting diverse family characteristics, new student trajectories, new sources of information such as social media, and new forms of communication. Progress will depend upon collection additional data and on not restricting analyses to a limited set of model frameworks.

In *Emerging Issues in Postsecondary Access and Choice: Implications for the Conceptualization and Modeling of College Decisions*, Bridget Terry Long discusses theoretical and empirical tools that might improve understanding of higher education decisions. Economic models, such as the human capital (HC) model, could be used to better understand current trends and developments. Within Economics, theoretical advances especially in Behavioral Economics, might help researchers better conceptualize postsecondary access and choice. Then consideration is given to how both tools and theoretical frameworks might be utilized at NCES to map central constructs discussed onto existing NCES data collections to improve NCES surveys.

In *Data Collection for Econometric Analysis of Schooling Decisions*: Charles F. Manski considers how NCES data collections can contribute to effective econometric analysis of schooling decisions, with emphasis on postsecondary schooling. The credibility of econometric analysis depends on the realism of the model used to represent decision making. Since surveys rarely query respondents about their perception of the choice-set, it is not unclear what researchers really know about how decision makers perceive the available alternatives.

This paper offers a perspective on what modeling is feasible with existing data and explores how new data collection could enhance modeling capacity.

In *Improving NCES Data Collection* Eric Bettinger, focuses on three key points: the role of expectations in future data collection; the appropriateness of behavioral economics as a framework for future data collection; and the integration and use of state administrative data with NCES data collection. From a different perspective, alternative survey designs could be advantageous in expanding the scope of NCES and allowing the flexibility for future changes in NCES data collection.

In *NCES Data Collections: Information and Gaps* Jack Buckley notes that the tension to balance the needs of many diverse stakeholders results in data collections - such as those in education – that are "general purpose" products. Buckley focuses models and specific issues in measurement of salient interest to NCES. Recommendations to NCES include changes to existing data collection activities to directly address perceptions. That would allow expanding analyses, going beyond econometric models and beyond a purely decision-based framework.

In *Conceptualizing Postsecondary Access and Choice: The Role of NCES Datasets* Laura W. Perna identifies further issues in particular the desirability of a multi-disciplinary perspective. While useful, human capital models alone are insufficient for understanding inconsistencies in college enrollment decision-making. With current data, these models shed little light on how students develop perceptions or understandings of financial aid. Specific recommendations for NCES data collections include: collect data to understand how students acquire and use information about college; develop better measures of students' postsecondary options; measure non-traditional pathways to degree completion; and consider data collection and sampling issues.

NATIONAL INSTITUTE OF STATISTICAL SCIENCES TASK FORCE REPORT

PREFACE

The five papers and discussion in this report were presented as part of the *Postsecondary Choice Project* conducted by the National Institute of Statistical Sciences (NISS) for the National Center for Education Statistics (NCES) of the Department of Education.

The first goal of the project was to support creation of a set of economic- and social-based constructs for describing and reasoning about postsecondary choice. The second goal was to initiate a process of mapping the constructs into NCES data collections, and address data elements and instrument designs needed to operationalize and measure the constructs.

The project culminated in a workshop held on 26 January 2009 at NCES in Washington, DC, where the five papers included here were presented.

Project Objectives and Goals

Postsecondary choice comprises a complex set of issues with strong scientific content and important policy implications. The processes are inherently complicated, involving a multiplicity of participants and stakeholders. Consequently there is a multiplicity of goals for collecting data, which has implications for NCES' postsecondary data collections.

The first goal of the project was to support creation of a set of economic- and social-based constructs for describing and reasoning about postsecondary choice. Constructs are meant to encompass both student (demand) and institutional (supply) perspectives. The second goal was to initiate a process of mapping the constructs into NCES data collections, and to address data elements and instrument designs needed to operationalize and measure the constructs.

The process of making choices was examined in terms of what decisions are made, when, by whom, under whose influence, using what criteria, and on the basis of what information is obtained from which sources. Authors were asked to consider three aspects of the topic:

- What we know about the rapidly changing, increasingly heterogeneous processes of postsecondary choice?

- What components of postsecondary choice are possibly less completely understood than they should be? One example is the increasing multiplicity of agents - among them, students, parents, secondary schools, postsecondary institutions, the federal government, and lenders.

- How can extant and future NCES data collections better inform both policy and scientific understanding? Individual papers discuss background, criteria for defining constructs, and consequences for data collection.

Papers presented at the Postsecondary Choice Workshop focused on the creation of a set of economic- and social-based constructs for describing and reasoning about postsecondary choice today and into the future. Once defined, these constructs could be mapped into ongoing and potential future data collections by NCES. To the extent feasible, data elements and instrument designs could be identified in order to operationalize and measure the constructs. Individual papers developed background and criteria for establishing constructs primarily by focusing on postsecondary choice in terms of what decisions are made, when, by whom, under whose influence, using what criteria, and on the basis of what information is obtained from which sources.

NISS/NCES WORKSHOP AGENDA

December 1, 2010

A Conceptual Structure for Postsecondary Access and Choice
Alan F. Karr **–** National Institute of Statistical Sciences

Emerging Issues in Postsecondary Access and Choice:
Implications for the Conceptualization and Modeling of College Decisions
Bridget Terry Long **–** Harvard University

Data Collection for Econometric Analysis of Schooling Decisions
Charles F. Manski **–** Northwestern University

Improving NCES Data Collection: Response to Long and Manski
Eric Bettinger **–** Stanford University and NBER

NCES Data Collections: Information and Gaps
Jack Buckley **–** New York University

Conceptualizing Postsecondary Access and Choice: The Role of NCES Datasets
Laura W. Perna **–** University of Pennsylvania

Institute of Education Sciences
National Center for Education Statistics

EMERGING ISSUES IN POSTSECONDARY ACCESS AND CHOICE:

A CONCEPTUAL STRUCTURE FOR POSTSECONDARY ACCESS AND CHOICE

Alan F. Karr, PhD

A CONCEPTUAL STRUCTURE FOR POSTSECONDARY ACCESS AND CHOICE

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

A CONCEPTUAL STRUCTURE FOR POSTSECONDARY ACCESS AND CHOICE

I.   INTRODUCTION

Postsecondary choice processes are inherently complicated, involving a multiplicity of goals for collecting data, which have implications for NCES' postsecondary data collections. Further, the choice processes concerning postsecondary education by students and institutions take place in a rapidly evolving environment. The processes themselves are evolving, reflecting diverse family characteristics, new student trajectories, new sources of information such as social media, and new forms of communication.

The first goal of the project was to support creation of a set of economic- and social-based constructs for describing and reasoning about postsecondary choice today and into the future. The constructs are meant to encompass both student (demand) and institutional (supply) perspectives. The second goal was to initiate a process of mapping the constructs into ongoing and potential future data collections by NCES, and to the extent feasible, to address data elements and instrument designs needed to operationalize and measure the constructs.

Questions that authors were asked to consider included:

- What we know about the rapidly changing, increasingly heterogeneous processes of postsecondary choice?

- What components of postsecondary choice are possibly less completely understood than they should be? One example is the increasing multiplicity of agents - among them, students, parents, secondary schools, postsecondary institutions, the federal government, and lenders.

- How can extant and future NCES data collections better inform both policy and scientific understanding?

Papers and discussions were presented at a workshop, held in Washington, DC on January 26, 2009.

II.   CHOICE PROCESS

Postsecondary choice comprises a complex set of issues with strong scientific content and important policy implications. The processes are inherently complicated, involving a multiplicity of participants and stakeholders. There is a consequent multiplicity of goals for collecting data, which has implications for NCES' postsecondary data collections.

Moreover, choice processes concerning postsecondary education by students and institutions take place in a rapidly evolving environment. Examples include the increasing role of for-profit, degree-granting institutions, the widespread (but not universal) de- emphasis of early admission, effects of the current US financial situation, new financial aid policies at highly endowed universities and the ubiquity of distance learning. The processes themselves are evolving, reflecting diverse family characteristics, new student trajectories, new sources of information such as social media, and new forms of communication.

It appears to be central to describe and understand four characteristics of postsecondary choice:

- **Heterogeneity**. Postsecondary education is dramatically more heterogeneous than five, let alone 15-25[1], years ago. Examples include a broader range of student trajectories, new modes of instruction (especially, internet-based instruction, whether as "distance learning" or not), more kinds of institutions (such as national for-profit universities) and more sophisticated institutional behavior (for instance, use of predictive admissions models).

- **Dynamics**. For most students, postsecondary choice is not one decision, but instead a series of decisions taking place over a substantial portion of their adult lifetime.

- **Information**. The number, nature and quality of sources on information used by students and institutions have changed enormously. Traditional sources for students, such as family, secondary school personnel, institutionally-created print advertising and peers, have been supplemented or even supplanted by broadcast media, the worldwide web and forms of information exchange such as social networking web sites. Institutional "marketing" efforts employ a multitude of information sources.

- **Uncertainty and Risk**. Decisions made by students and institutions are rife with uncertainty. Some of this uncertainty is inherent: students cannot know whether they will complete their studies, how their personal or financial situations will change during the process, or what career path they will follow (and what they will earn) afterward. Other sources of uncertainty include incomplete or inaccurate knowledge of (seemingly objective) information, such as conditions for availability of financial aid. Risk perception and tolerance affect students' decisions. Institutions also face risks: admitting students who cannot complete degree requirements is costly both financially and in other ways, and there is intensifying pressure to measure institutional "value added."

NCES does not wish, of course, that a gap develops between what is actually happening and the data that it collects in order to understand what is happening. The workshop, papers and subsequent actions by NCES make clear its desire to collect data that support scientific understanding and can be used, if desired, to formulate, implement and evaluate interventions.

III. CORE CONSTRUCTS

Since this project and the workshop focus principally on students, there is a consensus that a conceptual structure (set of constructs) is needed that is *rooted in human capital models of postsecondary choice*.

With some oversimplification, "traditional" human capital models for postsecondary choice assumed that:

- Students follow a path directly from secondary to postsecondary study at a four-year, degree-granting institution, at which they remain until completing degree requirements.

- A monolithic decision maker selects the alternative with the highest (expected) net benefit, subject to a family-dependent budget constraint.

---

[1] Postsecondary choice was an object of intense academia scrutiny during the 1980's, but seems to have undergone a decline in attention since approximately 1990. Long, Manski, Bettinger, Buckley and Perna are, of course, leaders in directing renewed attention to the topic.

- There is a known choice set (of institutions, majors, financial aid packages, …).

- Decision makers possess accurate, complete, cost-free knowledge of (discounted) costs (principally, out-of-pocket expenses and foregone earnings) and benefits (future earnings) for each alternative (work, military, institution A with major X, institution B with major Y, …).

To these models would be added constructs that accommodate:

- *Dynamics*: Postsecondary choice is a series of decisions, each of which may enlarge or restrict the choice set of subsequent ones.

- *Multiple decisions*, including secondary school curriculum, how and where to seek information about postsecondary education, whether to undertake immediate and/or deferred postsecondary participation, changes of major or institution, and engagement in "life-long learning," as well as whether and to what extent to incur debt.

- A complex, time-varying *locus of decision making*, comprising multiple decision makers whose roles and characteristics change over time: students, families, postsecondary institutions and lenders.

- *Choice sets underlying decisions*, with attention to alternatives not chosen.

- The *diverse information on which decisions are based*, consisting of (conceptually) objective knowledge, which may, however, be known incompletely, inaccurately or at an inappropriate time[2]. An example of such information is availability and characteristics of financial aid.

- *The cost of information*, which may be substantial, and is only partly monetary[3].

- *Decision makers' expectations* on which decisions are based: inherently uncertain and subjective inferences about the consequences of those decisions. Examples of expectation are whether a course of study will be completed successfully and future earnings will result. The distinction between information and expectations is central.

- *Uncertainties* associated with expectations, which introduces the need to consider of risk perception and tolerance.

- *Default choices for decisions* - sociologically or psychologically determined choices that reflect a student's environment in complex ways.

- *Explanatory factors from behavioral economics*, such as bounded rationality, bounded willpower and bounded self-interest, as means of understanding decisions that appear to deviate from rationality as defined by human capital models. An example of phenomenon is debt aversion, which can appear to be inconsistent with students' self-interests in the sense that assuming the debt is economically rational.

---

[2] Completeness, accuracy and timeliness are three of many dimensions of data quality; see A.F. Karr, D.L. Banks, A.P. Sanil (2008). Data Quality: A Statistical Perspective. *Statistical Methodology* 3(2) 137-173. Perna described such information as "measures of knowledge."

[3] The 2010 Nobel Prize in Economic Sciences, awarded to Peter A. Diamond, Dale T. Mortensen and Christopher A. Pissarides, recognizes research on an analogous "friction" associated with job searches.

- *Sociological and psychological effects*, in addition to economic effects, including sociologically-based "default" choices, and decisions makers' response to uncertainty and risk.

The constructs can accommodate a number of important specific issues:

- Actual as opposed to perceived choice sets. In a sense, this can be seen as extending a concept of default choices to default choice sets.

- Information about unchosen alternatives. Such information is essential to some models of decision making, and of scientific value in any event.

- Formation of expectations, that is, the process - as opposed to the "content" of those expectations.

- Cost and budget constraints, whose nature and roles may differ qualitatively, for instance, in low-income as to compared to high-income settings.[4]

- The complex, dynamic role of debt, including as well phenomena as debt aversion.

- The role of context, of which default choices and habitus are concrete instantiations. This is a central point where the economic perspective is complemented by psychological and sociological perspectives.

This conceptual structure is a *student-centric* view of the roles of postsecondary institutions, lenders, and the federal government. It is likely to evolve over time, reflecting evolution of the processes it describes. It may not capture the full complexity of the postsecondary education market, to which some or perhaps many standard economic assumptions do not apply. In this market,

- Neither buyers (students) nor sellers (institutions) are homogeneous.

- Prices may be negotiable (and therefore invisible to others), for instance in terms of the amount and nature of financial aid. Prices may also reflect "exogenous" entities such as state legislatures that set levels of, or bounds on, tuition.

- Not all sellers and buyers can interact, and certainly not all do interact.[5]

- The cost of acquiring information (for example, by means of campus visits) can be substantial.

- Price is not the sole market clearing mechanism.

- There exist large externalities - in particular, governments and lenders.

- The conceptual structure also may not capture fully the non-monetary aspects of utility to either students or institutions. To illustrate, access to particular occupations of interest for other-than-financial reasons is for some students a principal justification for postsecondary participation. Home-to-institution distance has major effects on students' decisions. Some institutions admit children of alumni who may be "less qualified" than others denied admission.

In the end, can a single conceptual framework can capture all of the current and emerging heterogeneity? For instance, does the world in which high school followed by four-year college/university enrollment

---

[4] To illustrate, the role of students' own earnings differs.
[5] Although actions are taken and costs are incurred beforehand, the application is the concrete instantiation of interaction.

followed by graduate or professional study is the default choice differ qualitatively from the world in which employment, military service[6] or part-time enrollment in a community college is the default choice? If so, does the extent of that divergence require alternative abstractions (not just different models) to describe and understand them? Heterogeneity is also an issue in the small: what once seemed to be well-defined concepts, such as family, now have multiple variants.

IV. IMPLICATIONS FOR NCES

The following material sketches one way in which NCES can think about data collections that incorporate the conceptual structure articulated above as well recognizes budgetary realities[7].

General Discussion

NCES can construe its postsecondary data collections as supporting two principal set of activities:

1. Description: Creation of national estimates of the condition of postsecondary education and the ways and extent to which it is changing. Description also includes characterizing effects of previous policy interventions.

2. Scientific understanding that captures the full complexity of the conceptual structure. Some examples are:

   - For which subgroups does debt aversion exist, and how does it impact their decisions?[8]
   - What is the precise role of distance as a cost?
   - What, if any, is the role of economic discounting - for instance, of future earnings - in postsecondary choice?
   - How important is cost of acquiring information?
   - What is the real locus of postsecondary decision making?
   - How do postsecondary decision makers deal with external and self-imposed uncertainties?
   - How can the quality of information used in postsecondary choice be measured?
   - What roles do risk perception and tolerance play in postsecondary choice?

In an operational sense, the distinction between "description" and "scientific understanding" is that between finite-population and super-population analysis of survey data. The former, for instance, requires intricately constructed weights whereas the latter may not. Reflecting the interests and expertise of their authors, the five papers focused to a significant degree on "scientific understanding" goal.

These two merge in the context of policy-relevant counterfactual analysis. To illustrate, prospectively evaluating the effect of the federal government's assuming a dominant-to-exclusive role in making student loans, or revision of the FAFSA, entails predicting the actual effects of such interventions, but is not possible without a model of the processes by which the policies would operate.

---

[6] Somewhat surprisingly, given that it is a both a source of direct education and a means of access to further education (via benefits to military personnel), as well as the number of people who pursue it, military service received relatively little explicit attention in the papers and at the workshop.

[7] Neither the workshop nor this summary addresses costs in detail. Nevertheless, it seems clear that enormous expenditure would be required to address even every question raised at the workshop or in the papers, let alone the myriad of other questions that exists.

8 To illustrate, it was asserted at the workshop that currently available data do not even support knowing whether debt aversion actually exists.

Attempting to meet these goals within the context of one family (even one evolving family) of data collections may not be the most sensible or cost-effective strategy for NCES to follow, since the consequences could include complex, burdensome data collection instruments, costly data collection processes and consequent negative effects on data quality. Instead, NCES might consider a two-pronged approach, which emerged in discussion at the workshop.

The first prong is to modify existing general purpose data collections to address descriptive data needs associated with the conceptual structure, which NCES has been doing and continues to do. These data collections include NPSAS, B&B, BPS, and the HSLS:09 postsecondary follow-up. Doing so will entail modification of survey instruments, and may require that some items be removed in order accommodate new ones. The second prong is to address scientific questions by means of small, targeted data collections, which can be either standalone surveys or embedded in general purpose data collections as subsamples.[9]

Obviously the situation would not be static. In particular, based on insights from targeted data collections, some scientific questions may shift to the descriptive category. To illustrate, consider the (currently, scientific) questions of existence of debt aversion and differential prevalence of debt aversion among subgroups of students. Significant scientific insight about these questions could be obtained by means of statistical analysis of a relatively small sample from a targeted data collection, because ascertaining which factors are important does not require national estimates of their effects. At some point, however, debt aversion might move to the descriptive or policy analysis category.

In parallel with this approach, NCES may wish to consider the benefits from identifying and leveraging supplementary data sources, including other federal agencies and state administrative databases such as K-16+ state longitudinal data systems (SLDS). A persuasive argument can be made for the necessity of doing this with emerging opportunities. However, it is also true that the associated challenges are daunting. For instance, NCES may not be allowed to merge data from multiple SLDSs.[10] Even if merging is permitted, record linkage introduces hard-to-characterize uncertainties. Privacy and data quality are certain to be issues.

A Possible Entry Point

Building on the approach outlined above, this section describes a set of initial targets that support both description and scientific understanding. The following constructs seem to be the most immediately approachable from the descriptive perspective:

- *Dynamics*: What decisions were made, when, and how did they interact?
- *Locus of Decision Making*: Who participated in which decisions, and in what role?
- *Information*: What are the principal sources of information on which each decision was based?[11]
- *Default choices*: For each decision, was there an a priori default choice? If so, what was it, and how was it determined?

---

[9] Or some other means entirely, such as focus groups.

[10] There are, however, statistical technologies based on secure multi-party computation that allow principled analysis of distributed datasets without actually merging the datasets. See A. F. Karr, W. J. Fulp, X. Lin, J. P. Reiter, A. P. Sanil, F. Vera, S. S. Young, Secure, privacy-preserving analysis of distributed databases. Technometrics 49(3) (2007) 335-345.

[11] Some aspects of information, especially those relating to quality and cost of acquisition, do not seem promising as short-run emphases.

Collecting data about these constructs within extant data collections (especially, HSLS:09) seems feasible, and may be only minimally disruptive. It is possible that much can be learned from relatively few survey items. The results would help frame scientific questions more sharply before data are collected concerning them.

Over the short run, this path leaves in the "scientific question world" issues of:

- *Choice sets* underlying decisions, which are difficult to deal with in the absence of attention to alternatives not chosen.

- *Expectations* on which decisions are based. Despite significant cause for optimism expressed by Manski, there appear to remain unresolved issues ranging from the right abstractions to instrument design. Collecting reliable information about probabilistic expectations is especially challenging.

- *Uncertainties* associated with expectations. The underlying issues here are risk perception and tolerance, which are rather broadly seen as difficult to ascertain.

- *Other explanatory factors from behavioral economics*, which would help explain "irrational" decisions, but are elusive because the definition of rational is model - rather than data-based.

These are important constructs, but the workshop seemed to confirm that current understanding of what they are and how to measure them have not yet reached the point of justifying NCES' expending major resources to collect data about them.

A different interpretation is that this path places initial emphasis on the way people act, which is relatively easier to describe and understand, as opposed to the way people think. The latter, as is well known, is made even more difficult by self-reporting as well as by data subjects' memory or inability or unwillingness to articulate their thought processes.

An alternative perspective it to focus not on decisions, but rather on the flows of students that result from those decisions. These flows are among, inter alia, secondary study, employment, military service, enrollment in two-year institutions, enrollment in a four-year year institutions, graduate/professional study and continuing education, as well as combinations of more than one of these simultaneously. These flows are functions of time, student characteristics, and externalities, and they constitute a tangible manifestation of the heterogeneity that underlies the conceptual structure so far. Current data NCES collections, even augmented by other data collections, do not seem assured to capture fully the richness and dynamics of these flows.

Other Considerations

While t utility-based discrete choice models d are important, they are not the only class of models relevant to postsecondary choice. Others include loss-based models with explicit representations of uncertainty (statistical decision theory) and models with explicit dynamics (Markov decision processes).

Finally, NCES is uniquely positioned to integrate demand-side data (on students) with supply-side data (on institutions). Although such integration was not a focus of the papers or the workshop, it is a significant opportunity.

APPENDICES

Appendix A: Author

Appendix A: Author

Alan F. Karr, PhD

Director, National Institute of Statistical Sciences

Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE of STATISTICAL SCIENCES
TASK FORCE REPORT

EMERGING ISSUES in POSTSECONDARY ACCESS and CHOICE

IMPLICATIONS for CONCEPTUALIZATION and MODELING of COLLEGE DECISIONS

Bridget Terry Long, PhD

IMPLICATIONS FOR CONCEPTUALIZATION AND MODELING OF COLLEGE DECISIONS

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

IMPLICATIONS FOR CONCEPTUALIZATION AND MODELING OF COLLEGE DECISIONS

I. INTRODUCTION

Higher education is an increasingly complex industry encompassing a diverse set of students and institutions. Many recent trends underscore the dynamic nature of postsecondary education in the United States, from the increasing presence of new populations, growth of more varied educational pathways, development of new ways to deliver education, and evolution of institutional and government policy. For instance, demographic trends reflect that students are increasingly coming from communities of color; postsecondary enrollment for Black and Latino students is projected to increase 26 percent and 39 percent, respectively, between 2006 and 2017 (Hussar and Bailey, 2008). In many other ways beyond race and ethnicity, the profile of the typical college student has changed drastically in recent decades. Only 27 percent of students match the circumstance of what has been considered a "traditional" college student: an 18- to 22-year-old who attends full-time and is dependent on their parents (Choy, 2002). Many students now delay initial entry into college, attend part- time, and over three out of ten undergraduates is age 25 or older (*Digest of Education Statistics*, 2008). Students enter with intentions that can range from a bachelor's degree to a certificate to just a few courses. In 2002-03, approximately 68.5 million people took formal courses or training that was not part of a traditional degree, certificate, or apprenticeship program for reasons related to their job or career (O'Donnell, 2005). Students also increasingly stop out and then later return, attend multiple institutions, utilize distance education, and study outside of the traditional schedule, including summers, weekends, and evenings.

Current surveys of students and institutions, which serve as the basis of the statistics and research cited above, provide copious amounts of data to trace the backgrounds of students over time, their general progress through college, and the prices and finances of postsecondary institutions. However, given recent developments, there is a need to reevaluate what is understood about postsecondary access and choice and develop tools, both in terms of theory and empirical analysis, that might improve understanding of higher education decisions today and in the near future. This paper considers this mandate and discusses possible directions for future data collection and research. The process of preparing, applying, and choosing to attend college is influenced by a complex array of interrelated background factors, educational experiences, and social contexts. To understand these decisions, researchers have mainly focused on several major groupings of variables. An incomplete list of some of the most common variables includes: family background (e.g., demographic profile, income, parent's education), academic preparation (e.g., courses taken and test scores), expectations, college costs (e.g., tuition net financial aid), and benefits (e.g.,

earnings). With the growing diversity of students, goals, and pathways, more nuanced thinking about these variables is warranted. These factors have been measured with varying degrees of success, as it is difficult to get a complete picture of some aspects of the college decision-making process or factors that might influence it.[1]

However, in other cases, it may be possible to refine measures and/or consider ways to better define the concept research models are trying to capture. Moreover, there are factors that have received little attention but seem increasingly relevant given the growing diversity of students, institutions, and educational pathways.

This paper considers potential tools and frameworks that could provide helpful insight in the examination of contemporary college decision-making. First, it discusses how economic models could be used to better understand current trends and developments. This partly involves extending economics-based constructs such as the Human Capital Model (HC Model) in ways to better incorporate the current realities of college access and choice. As the basic assumptions used in much of research may no longer hold for most students, much could be gained from integrating factors that are increasingly prevalent among the diverse set of students now entering higher education into the approaches of researchers. Additionally, I consider how developments within Economics, such as the advancement of Behavioral Economics (BE) theories, might help researchers better conceptualize postsecondary access and choice today.

While the discussion of models focuses on ways to improve understanding of enrollment patterns and provide testable hypotheses about college decision-making, the paper also considers how to operationalize and measure these refined constructs with respect to the data collections of NCES. Building on some of the concepts discussed in relation to models, the paper highlights additional data constructs that could advance the study of college access and choice.

II.   THE NEW REALITIES OF COLLEGE ACCESS AND THE HUMAN CAPITAL MODEL

The primary economics framework for the study of college decisions is the HC Model first developed by Becker (1964). Education is thought to increase human capital, a set of skills that can be "rented out" to employers for income. When deciding whether to continue their education, individuals compare the benefits of human capital to the costs of obtaining it. In terms of higher education decisions, an individual will weigh the costs and benefits, both monetary and otherwise, to decide whether to prepare for college, enroll in a postsecondary institution, and continue until completing a college degree. Theory suggests, and many studies have shown, that college demand is related to the net benefit (benefits minus costs) of education, the prices of alternatives, and the preferences of the individual subject to a lifetime budget constraint. Among the costs most often included in research are tuition and foregone earnings, the income that an individual could have made had he or she decided to enter the labor market rather than attend school. On the other side, the benefits of higher education include increased earnings. Additional non-monetary costs and benefits have also been considered, such as the psychic costs of studying and the consumption value of college. From a societal point of view, reductions in crime and government dependency have also been taken into account. The HC Model can be applied to multiple steps of the

---

[1] For instance, though it would be ideal to have a sense of permanent or long-term family income when taking into account a student's background, such a measure would be costly and subject to is measurement error if collected by asking families to recall earnings from years ago. Expectations can also be difficult to measure accurately (see the discussion below).

college enrollment decision including the decisions of whether to prepare academically, to apply to certain colleges, to initially enroll, and to persist each year until graduation. It is also important to note that the HC Model is often used in comparative statics analysis. This refers to the consideration of how the change of one factor might influence educational decisions in a positive or negative manner. For example, even if one does not have good measures of all of the factors thought to influence the decision to attend college, the model is helpful in understanding that a change in a cost, for instance, is likely to reduce the likelihood of attending a postsecondary institution, holding all other factors constant (i.e., *ceteris paribus*).

While the formulation of the HC Model described above is how it is commonly applied to educational decisions, the flexibility and usefulness of the model goes far beyond this basic characterization. The variables most often used when applying the model could be measured better or reconceived to more closely approximate the circumstances of current students. Additionally, some aspects of the model are commonly overlooked, such as the role of expectations, uncertainty, and discount rates, deserve more attention as we learn more about the factors that influence the decisions of students today.

Costs and Budget Constraint

One area of possible improvement in the way researchers conceptualize the college enrollment decision is how costs are defined. As noted above, most often researchers include in the HC Model college tuition rates, perhaps net basic financial aid awards. While studies focused on earlier cohorts of students, such as high school graduates in 1972, found that differences in tuition costs were related to a student's decision on whether or not to attend college, the same analysis using a cohort who graduated high school in 1992 did not get the same result (Long, 2004).[2] The lack of statistical significance for the more recent cohort gives way to two possible theories: cost is no longer an important factor for all students and/or it is not being accurately measured. Although there may be some truth to the former for certain subgroups, let us first consider the validity of the latter.

Measurement error in cost variables could stem from several sources.[3] Most relevant for this discussion is how net price is measured. The most common definition of net price is list tuition price minus grants, aid that does not need to be repaid.[4] Is this method the most accurate way to incorporate financial aid into price? The most obvious critique is the fact that researchers often do not know the full slate of aid a student could receive and will only include the Pell Grant when accounting for grants. This is due to lack of data, particularly about college options not chosen, and it is difficult to get around such concerns without substantial data collections.[5]

---

2 The analysis used the National Longitudinal Survey (NLS) and the National Educational Longitudinal Survey (NELS). To have a better sense of net price, the amount of Federal Pell Grants that a person could expect to receive was subtracted from the list tuition price. Other types of scholarships and grants could not be accounted for since information on a student's specific need is unavailable, and much of grant aid beyond Pell Grants is very individualized.

3 For example, uncertainty is inherent in any model using projections of college costs. When high school freshmen are deciding whether to prepare for college, they are unable to accurately predict the costs they will face years later even if they have an accurate list of institutional prices at the time. Moreover, even when entering college, freshmen have little idea what price they will be charged as college seniors. This uncertainty leads to measurement error of a sort that one is unlikely to be able to fix.

4 If one is considering short term costs, sometimes researchers will also include loan aid as it helps with the immediate need to cover college expenses.

5 The individualized nature of most grant aid beyond Pell Grants makes determining the net price a student in the midst of deciding whether to apply or attend college could expect at a range of possible schools extremely difficult. After applying, students will receive information about this at some schools, but it is difficult to solicit complete information on these offers.

Two other issues worth considering are the timing of financial aid awards and the particular types of aid received. With regards to time, financial aid awards must be known (or suspected) at the moment of the decision if they are going to influence college access and choice. I elaborate on this point below with respective to concerns about information and the interaction of affordability and academic preparation. The second concern is that all aid is not the same, and increasing evidence suggests that students perceive different types of aid in vastly different ways. For example, research suggests that aid labeled as a "scholarship" is perceived differently than a "grant," which demonstrates that the particularly labeling and packaging of aid can influence college decisions (Avery and Hoxby, 2004). More relevant given recent trends is the increasing prominence of loans, tax credits, and other support, which may or may not influence student decisions. There are particular questions about how loans are perceived and incorporated into the enrollment considerations of a potential student. Although all students with college expenses are eligible for a federal loan, not all students will choose to obtain one, and there are differences in the propensity to take out a loan by background. Financial aid administrators report anecdotally that students from traditionally disadvantaged backgrounds are often unwilling to incur substantial debt to attend college, and Baum (2003) finds that the prospect of substantial borrowing discourages enrollment among some students, especially those from low-income and underrepresented groups. The negative reactions of some students to the prospect of loans suggest they are not perceived as a simple instrument to help with immediate costs, and these students consider future interest payments as harmful.

The HC Model presents several hypotheses for why some students may react differently to loans. The first possibility is that students are still actually liquidity constrained even in the face of current government loan programs. For instance, some community colleges do not participate in the Federal Loan programs, and so students may not have the opportunity to take out debt like we assume. This would affect the budget constraint. Moreover, students may not be aware of the availability of student loans or may lack clear information and understanding about debt as financial literacy differs by background. There are also psychic costs that accompanying debt as students must go through the process of applying for, receiving, and repaying the loan. Another hypothesis is that loans may reduce the option value of a college degree as debt may preclude a student from choosing a particular low-paying career. As illustrated by these hypotheses, the HC Model is a useful when considering the role of loans, but more research is needed to understand how students and their families consider whether to take on debt and whether these theories are valid. Given the substantial shift in federal aid policy from grants to loans during the last fifteen years, such differences in the propensity to take out student loans have important implications for college access and choice, and concerns about how students perceive loans should be incorporated into researchers conceptualization of net college costs.

Beyond net price, another cost that also is relevant in the college access and choice decisions is distance. Proximity is increasingly thought of as an important factor in enrollment, and several studies have included this variable (Long and Kurlaender, 2009; Long, 2004; Rouse, 1995). However, distance should be included more often as one of the many college costs and better measures of it would improve enrollment models. Previous research has had to use crude approximations of the number of miles "as the crow flies" between a student's home and a potential college. A more accurate measure would take into account the differences between miles that must be driven in a rural area versus public transportation availability in a city. The relevance of this factor is not just in terms of the length of distance; proximity also gives a sense of the exposure a student might have to a postsecondary option. Related to costs in the HC Model is a

student's budget constraint. The way the model is applied is often predicated on the view that the student is of traditional college age and reliant on their parents for most of their income.[6] This results in the assumption that a portion of family income will be used for the payment of college costs, and a student's earnings will all go to education-related expenses. While this framing may accurately reflect the perspective of many middle- and upper-income students who comprise what thought of as traditional college students, it fails to adequately capture the reality of increasing numbers of potential students who come from much more diverse family situations. Much of the college access literature focuses on students from traditionally-disadvantaged backgrounds, and this has several implications for how one might model the costs and possible resources available to pay for college. Numerous anecdotal accounts suggest that the income of low-income students is often a critical part of a family's income. Students are expected to contribute to family bills and expenses (McSwain and Davis, 2007). In this circumstance, family income is not accurately measured as the summation of the parents' earnings nor should one assume any income made by the student is available to cover college costs. Additional information is needed on what role the student has in contributing to the family's resources as this would affect the actual resources available and the perceived cost of attending college (i.e., the potential loss of income to the family). While it is difficult to understand how this information might be directly incorporated into a quantitative measure of costs in the HC Model, it certainly changes how one should define the budget constraint, and more research is needed to establish how prevalent this practice is. Moreover, the role of a student's earnings is relevant in the consideration of comparative statistics. Changes in parental employment and other dependents in the family will affect the necessity of student earnings for purposes other than college expenses.

Benefits

As a mirror image to costs, benefits such as increased earnings are routinely incorporated into the HC Model. Most often, researchers approximate this value using mean or median income levels by educational attainment. However, the evidence suggests that the return to college has become increasingly varied for those with some postsecondary training. Since 1970, there has been a clear widening of the income distribution among college-educated workers (Hoxby and Long, 1999). The implication is that using the average value of benefits it not entirely helpful when considering the postsecondary decisions of students. Ideally, one would use more localized and realistic benefit assumptions based on the student's profile (i.e., race, gender, age, major, GPA, school type, geographic area, etc.). This is not feasible without a substantial amount of data and computations. Still, it is worth considering whether the return to education varies in more simplistic and systematic ways that could be incorporated into estimates. For instance, there are strong patterns of difference by subject area, school selectivity, and state/region.

Related to the specific benefit estimates by individual profile is the great deal of uncertainty involved in projections of the future, and the level of uncertainly could differ by student background.[7] Many students do not complete the amount of postsecondary education they originally intend, and the likelihood of completing a degree differs by background. Only 36 percent of low-income students who were college qualified[8] completed a bachelor's degree within eight years, while 81 percent of high-income students did

---

[6] This conceptualization is perpetuated in the design of the federal financial aid system, which requires students to supply income information about their parents until age 24 unless they can prove independence.

[7] Uncertainty is also relevant when considering college costs. Students are unsure about future college prices along with what financial aid will be available and if they will be awarded merit-based aid.

[8] The definition of "college-qualified" is from Berkner and Chavez (1997). Students were judged to be college qualified if they met

so (Adelman, 2006). This difference is partly due to the correlation between background and academic preparation. However, given the persistence gap, one should consider modeling the benefits of higher education taking into account that all students do not have the same chance of completion. In fact, based on growing data on college outcomes, researchers could incorporate graduation probabilities by demographic profile. One might imagine a probabilistic equation with different likelihoods assigned to different outcomes and their related benefits. What is most relevant is the acknowledgement that students sometimes understand that the likelihood of graduation is far less than one, and so assuming otherwise in models of their decisions is not accurate in many cases.[9] With any amount of uncertainty about benefits, an individual's own risk preference (e.g., the degree a person is risk loving versus risk averse) must also be considered, as differences in preferences could cause two individuals facing the same conditions to make very different enrollment decisions. What is less clear is how to measure an individual's risk preference, though one could look at indicators of risky behavior as has been done in other research literatures.

An individual's discount rate is also an important factor when considering the benefits of a college education. A discount rate is a measure of how an individual perceives the value of money in the future. In theory, differences in discount rates could explain why one person might choose to attend college while the other would not; a higher discount rate means a person views the future benefits of education at a lower value than a person with a lower discount rate, and this has implications for the total perceived worth of a degree. For instance, some theorize that differences in the discount rates of men versus women could help explain the gender gap in education. One could also consider students who have strong, immediate needs for resources in the present as having high discount rates.[10] However, like risk preferences, discount rates are difficult to measure. Although a single discount rate is often arbitrarily chosen when completing cost-benefit analysis, there has been little research to establish appropriate rates for different kinds of students. Past research has used the propensity to engage in risky behavior as a proxy for a person's discount rate, but more direct measures of how students trade off between present day and future benefits would be helpful.

The Role of Information

Central to the HC Model, as well as many other Economic models, is information. To truly maximize their benefits (or utility), individuals must have an accurate sense of costs, benefits, and their budget constraint. The puzzles of why some students do not choose to prepare academically, apply, or enroll in college could be partly be explained by bad information or lack of information altogether. In fact, differences in awareness across groups may also provide some answers as to why enrollment rates differ by background. To improve understanding of attendance patterns, careful consideration must be taken of what students really do know about postsecondary preparation and options.

Research suggests misinformation or a lack of information about college is a serious problem. Researchers

---

any of five criteria that would place them among the top 75 percent of four-year college students for that criterion. The minimum values for qualified were a class rank of the 46th percentile, an academic GPA of 2.7, an SAT combined score of 820, an ACT composite score of 19, or a NELS-88 test score of the 56th percentile.

[9] Uncertainty is also relevant when specifying benefits due to the fact that the monetary return to education is determined by the interaction of supply and demand factors in the labor market. Obviously, one cannot predict what will happen for the rest of an individual's working life to accurately determine the present discounted benefits (PDV) of increased earnings in the future. This is true for all potential students, though the amount of uncertainty involved may vary across different kinds of career paths.

[10] For example, students who must contribute their earnings to their families may have a strong need for earnings in the present (perhaps more than their desire for higher future earnings).

have found a significant lack of information among prospective college students in general (e.g. Ikenberry and Hartle, 1998). There is also a lot of misinformation about financial aid among parents and students. Most studies also find that students and their parents greatly overestimate the costs of college (Horn, Chen, and Chapman, 2003). Furthermore, a poll commissioned by the Sallie Mae Fund found that two- thirds of all parents and young adults planning to go to college did not name grants as a possible source of funds when asked about types of financial aid (Sallie Mae Fund, 2003). Many believed all financial aid has a merit component. Even those aware of need-based sources incorrectly estimate the income levels that are eligible for aid. Awareness about aid and college costs appears to be especially small among low-income students. The Sallie Mae Fund study found that low-income families had the least amount of information about how to pay for college. Research by Kane and Avery (2004) also demonstrates that low-income high school students have very little understanding and information about actual college tuition levels and financial aid opportunities. There are also differences by race. For example, Grodsky and Jones (2004) find that parents of color are less likely to be able to estimate the cost of tuition.

It is vitally important that researchers find a way to examine and incorporate the role of information in college decisions. Similar conclusions have been made about other social programs. With respect to government support programs like TANF, Medicaid, and the Earned Income Tax Credit, Currie (2004) notes that economists have historically paid little attention to how information is made available to eligible persons, the specifics of eligibility rules, or how these rules are enforced. More details on these factors with regard to postsecondary enrollment would increase our understanding college access and choice. They would also have strong implications for policy. Knowing the exact nature of students' and families' understanding of college prices and aid would help authorities tailor efforts to increase attendance in the most effective way possible. For instance, what do students know, when do they know it, and is this information correct? If the complexity of programs and application procedures is a real deterrent, policymakers could have a significant impact by addressing these factors. Beyond the enrollment decision itself, the quantity and quality information students have is likely to influence earlier behavior. For example, it is probable that expectations are related to information. What is unclear is whether high levels of aspirations and expectations cause students to seek out college information or having information helps form students' and their family's beliefs. If students and their parents believe college is unaffordable, they will likely negatively adjust their expectations about whether they can attend college. With the reality that college is a lot less expensive than what most potential students believe, better information earlier in the process could raise the expectations of some families. Much more research is needed to understand the interaction between these variables, but regardless, information, and especially incorrect information, could have negative effects on aspirations.

The amount and accuracy of information about postsecondary options could also affect whether a student decides to prepare for college. Again, if college-going is perceived as unaffordable by students, parents, and counselors, then individuals may not choose to prepare academically for college-level work. In this case, the timing of the information is vitally important. Because personalized estimates of financial aid awards are largely not revealed until late in the senior year of high school, families often do not form accurate views of affordability until a few months before actually attending college, a point at which it is too late to take the necessary high school courses.[11] Some students never even reach the stage of getting accurate aid and

---

[11] With the creation of the FAFSA 4caster and other similar tools, personalized information is more readily available. However, internet access is limited among low-income students, especially high-speed internet access, which is advisable when trying to

price information because their misperceptions convince them that applying is not worthwhile. However, early information about costs is not enough. Because proper preparation for postsecondary study involves more than meeting high school graduation requirements, students need details on specific college entry requirements in order to get ready for that next step. At selective institutions, knowledge about application procedures is also necessary. Unfortunately, research suggests that low-income high school students know little about how to navigate the admissions process (Kane and Avery, 2004).

III. LESSONS FROM BEHAVIORAL ECONOMICS

While the HC Model provides useful insight, Behavioral Economics (BE), a growing field within the discipline, may also contribute to the understanding of college decisions. Core to the field is the observation that individuals often do not act in ways that are completely rational. Mullainathan and Thaler (2000) summarize three ways in which individuals deviate from standard economic models. First, they note that problem solving can be constrained by human's limited cognitive abilities (i.e., "bounded rationality"). Second, individuals sometimes make choices that are not in their long-term interest (i.e., "bounded willpower"). Third, they observe that individuals are sometimes willing to sacrifice their own interests to help others (i.e., "bounded self-interest"). There are examples of each point, mostly in the finance literature, and they may provide helpful ways to reconsider how we conceptualize education decisions.

Several examples point to the importance of the default. Individuals will often react in totally different ways depending on whether the default decision is defined in one way versus another. One example in the literature involves 401k retirement accounts. Madrian and Shea (2001) studied the savings behavior of employees at a large American corporation. Originally, employees had to deliberately elect to participate in the company's 401(k) plan. However, after a policy change, employees were automatically enrolled unless they opted out. The researchers found that although none of the economic features of the plan were altered, the change in the 401(k)-enrollment policy resulted in a substantial change in savings behavior. According to what they term "default behavior," the employees had a high likelihood of sticking to the default conditions of the saving plan.

The importance of the default may be especially important in cases in which the decision to be made is very complex. Beshears, Laibson, Choi, and Madrian (2006) investigate an instance in which those who are not experts in the relevant topic are especially daunted by the task of making a decision. As a result, they find individuals tend to procrastinate in making a decision. However, they find an intervention that greatly simplifies the decision by collapsing the set of options into a binary choice between the status quo and a pre-selective alternative prompted individuals to make a decision.

These findings about how the default can affect decision-making sheds light on multiple aspects of college access and enrollment. Although college attendance has become the norm among most students, and some would even term it as a right or necessity, the system is still designed as if it is a privilege that students must jump through hoops to receive. In some ways, this is warranted because there are large private returns to higher education, and it commands significant social investment; students should be required to meet certain criteria and invest their own time and resources in order to participate. If, however, one believes educational attainment is vitally important and has significant social return, then

---

get aid information. There is also the problem of lack of awareness about such tools among target families.

perhaps there should be some thought to how current defaults unnecessarily discourage students. Low savings rates coupled with the importance of retirement saving is what has spurred changes to the enrollment default for 401(k) plans, and perhaps the same thing needs to be considered in higher education.

Although the word "default" is not mentioned in the education literature, past research already suggests that the framing of a program can be influential in college access. Dynarski (2002) examines the Social Security Student Benefit (SSSB) Program, which gave 18- to 22-year-old children of dead, disabled, or retired Social Security beneficiaries monthly support while they were enrolled full-time in college. In contrast to other aid, awareness among potential beneficiaries was high due to the default automatic notification of eligible students by the government and the extremely simple application process that followed. Dynarski found a huge response when the program was discontinued in 1982, larger than found for other aid programs, thereby suggesting that the default design of the program was probably influential.

Taking this lesson and applying it to current financial aid policy could have important repercussions. Analysis by Dynarski and Scott-Clayton (2006) demonstrate that the current system of awarding aid could be closely approximated by only using a few pieces of information on the family. For instance, over three-quarters of variation in Pell Grant awards could be explained using parents adjusted gross income, marital status, family size, and the number of family members in college. In addition, an experiment by Bettinger, Long, and Oreopoulos (2008) has demonstrated that tax data could be used to quickly populate the FAFSA with the remaining questions taking less than 10 minutes to complete. These points attest to the feasibility of changing the way financial aid is awarded without losing much in the targeting power of our need-based programs. If IRS data were used to automatically calculate aid eligibility and send this information to potential students thereby changing the aid application default, it is likely that we would see an enrollment response.

There are other stages in the college enrollment pathway that could be altered with the importance of the default in mind. For instance, some schools and states now require students to take college entrance examinations (i.e., the SAT and/or ACT). Rather than making the student seek out information to sign up and prepare for the tests as well as secure transportation to the testing site, the default in these places is to have everyone do it unless they deliberately choose to opt out (which is not possible in all places). One might also consider the tradeoffs of making the default high school curriculum be what is appropriate for college preparation or assigning the submission of college and financial aid applications as graduation requirements. With regards to data collections, more information about the defaults a student faces (i.e., in high school) would be helpful to investigations of this issue.

As also noted in the savings literature, procrastination is an important concern in higher education. Students often think they can put off doing the things to prepare for college. Some take college examinations just before applying, rather than during junior year when there is time to retake the test if necessary. Research also suggests that many submit financial aid applications well past the recommended deadline (ACE, 2004). BE may provide some theory behind this phenomenon. O'Donoghue and Rabin (1999) suggest people do not fully participate in 401(k) plans due to procrastination and hyperbolic discounting, which suggests that individuals typically are impatient about relatively short time horizons but are much more patient about the long run (in contrast, the standard assumption of exponential discounting assumes patience is independent of horizon). The implication for education is that although students may want to

eventually attend college, in the short run there are constant distractions offering greater immediate rewards. Consequently, they continue to delay the college decision. As noted in the research of Beshears, Laibson, Choi, and Madrian (2006), simplifying and changing the default option could deal with this problem.

Another anomaly that researchers have found is the fact that individuals will sometimes make suboptimal investments. As noted above, one reason for such decisions could be incomplete or incorrect information. However, research in BE documents that suboptimal investments may persist even after providing accurate information about benefits that have not been collected. For instance, Choi, Laibson, and Madrian (2005) identify a case in which individuals have strong incentives to invest and get the maximum matching contribution from their employer (i.e., these employees can make penalty-free withdrawals for any reason so are not required to keep the matched funds in the account). However, they find that many contribute below the threshold, thereby violating the no-arbitrage condition and losing as much as 6 percent of their salary. Providing the individuals with information did not change their behavior, thereby begging the question of why they would forego the "free lunch" available to them. This is a case in which the optimal investment strategy is clearly being ignored.

This pattern mirrors much of what we see in education. Given the extraordinary benefits of higher education, especially relative to the salaries of those with only a high school degree, it is puzzling that some students still choose not to attend. As noted by Choi, Laibson, and Madrian, their example underscores the failure of monetary incentives alone to cause an individual to make a certain investment. While BE has not solved this puzzle, it is a reminder that researchers should acknowledge the importance of factors beyond monetary benefits in college decisions.

Decision anomalies are also relevant to college choice. Avery and Hoxby (2004) set out to test whether high-ability students respond to their menus of colleges and financial aid offers like rational human capital investors. They find that the typical high-aptitude student chooses his college and responds to aid in a manner that is broadly consistent with rational investment. However, they also find some serious anomalies. First, student had an excessive negative response to loans and work-study. Meanwhile, they responded to an institutional grant's share of a college costs rather than its amount of the grant itself.

Finally, students had a strong response to superficial aspects of a grant, such as whether it has a name. This example again underscores the fact that students view various kinds of aid in different, perhaps non-rational, ways. Moreover, given the very selective and affluent nature of the sample, this is likely an upper bound on the sophistication of students and their families in weighing college options. More research is needed to document these types of anomalies as feedback between BE and Education research might provide insight.

IV. THE DATA IMPLICATIONS

The above sections outline ways in which college decisions might be better understood using Economics models. The discussion details ways to improve the conceptualization of old variables and suggestions about new factors to include in models of college decision-making. This section, however, considers how the NCES data collections might be used to address some of the suggestions. I map some of the central constructs discussed above onto existing NCES data collections and consider how surveys could be improved to the benefit of research on major issues related to college access and choice. The discussion is

not a fully inclusive list of the topics above, but instead moves from considering what would be ideal (the above sections) to what might be feasible (this section).

The Role of Information and its Relationship to Expectations and Preparation

The information and perceptions students have about college prices, financial aid, and preparation requirements could also have significant implications for college decisions. There are multiple dimensions of these types of information that deserve attention. The first is the quantity of information – what does the student know? Second is the quality of that information – it is accurate? And then, from where does the student get that information? Finally, and in some ways most critical, *when* does the student know certain things? While the first three questions have been measured to varying degrees in various NCES surveys, the last question about timing needs further exploration. Often, questions about information are only asked once. Therefore, even in longitudinal surveys, researchers are unable to examine how changes in college perceptions or information affect aspirations, preparation, or enrollment decisions. Additionally, most of the questions about student awareness are asked late in a student's secondary school career. This prevents analysis of how information and perceptions during middle school or earlier in high school affect subsequent behavior. A beneficial change would be to ask students at the earliest possible time (e.g., in tenth grade in ELS) and then periodically until senior high school. Follow-up questions while in college would allow for an interesting study on how students update their misperceptions and lack of information.

With regards to the other questions regarding information, it is still worth considering whether the current surveys maximize the use of such data. For example, questions about the source of information could aid to get more detail on key potential suppliers. With regards to guidance counselors, who have been the foci of many critiques about poor advising in urban high schools, it would be useful to get information about how often students meet with them and what types of information they provide. The specifics of peer groups and their aspirations would also be a contribution as well as details on parents and siblings' experiences with higher education, which might be an avenue for information. Finally, whether the student has participated in campus visits or visiting college websites might underscore gaps between different types of students.

As noted above, there may also be a connection between information and expectations, but the literature does not have clear answers about the direction of this relationship. Expectations could be proxies for a student's level of motivation, which would in turn prompt the student to seek out college information. In contrast, expectations might be amendable to policies that improve student information about college. As summarized by Chelimsky: "It is not possible to determine from the available studies whether knowing of financial aid availability is a precursor to the desire to pursue post-secondary or whether the desire to continue education explains the different in awareness of financial aid" (1991, p. 9). For example, the HSB survey included a great deal of information on students' perceptions of college prices and whether individuals plan to use a particular financial aid program. More recently, the 1999 NHES provides information on awareness about college price and aid. Unfortunately, neither NELS nor ELS includes information about college price perceptions. Therefore, while these more recent data do contain information on college expectations and aspirations, one cannot tie these factors to perceptions or awareness. They also have questions on where and how students get information about college.

Multiple measures of information and expectations over time are needed for attempts to establish the true relationship, which might also differ by individual. Moreover, it is worth noting that questions concerning

college information should be asked of *all* students. In some surveys, questions about college information are only asked of parents who expect their kids to go to college. This prohibits studies about how information impacts college expectations because it may be the case that parents who do not expect their kids to go to college greatly overestimate the cost of attendance.

With regard to academic preparation, information is essential to success as the requirements for colleges can be very different than what high school demand.[12] Interestingly, recent surveys suggest an overwhelming percentage of students expect to go to a postsecondary institution even though they are not taking active steps toward the goal in terms of academic preparation. While there is a great deal of data on exactly what students do to prepare (i.e., what courses they take), it would be interesting to discover what they perceive to be the requirements for their state flagship university. The response of students also begs the question whether their responses reflect true expectations, or whether students respond to questions about their expectations in the way they feel is socially acceptable due to pressure. Research on survey methods and the potential biases created by the framing of questions and presence of the interviewer might help with thinking around this problem.

The Role of Loans

As noted above, there are questions about how loans are perceived and incorporated into the enrollment considerations of a potential student. Given the substantial growth in the use of loans, these are topics of growing importance. On one hand, anecdotal reports suggest some groups of students have an aversion to debt. On the other hand, there is a debate about whether other students take out debt out of convenience, i.e., to fund recreation or luxury items rather than education (King, 1999). Little to no data exist to document either of these phenomena, and it would be a great service to have carefully designed survey questions to ascertain the motives of borrowers and non-borrowers. This relates to the longitudinal surveys, like ELS, as well as postsecondary surveys, like NPSAS.

In terms of examining students who may be averse to loans, it would be valuable to learn more about their and their families' experience with all kinds of debt. Researchers have used home ownership as a proxy for this, but a more complete picture, including car, credit card, and student loan debt, would be more instructive. Additionally, for the longitudinal surveys that begin with high school students, what are the parents' beliefs about debt, and do they have access to other forms of debt? This might help us to better understand the role of parents, and whether some students who choose not to take federal loans instead use other forms of capital. For those who take out significant debt, questions about the uses of the debt (tuition, room & board, food, car, etc.) might provide insight.

Another loan-related topic that is poorly understood is the long-term effects of this type of aid. As a precursor to that question, it would be valuable to learn what students believe to be their responsibilities in terms of debt. Do they have a good understanding of their total debt (as compared to external records)? Who do they expect will make the payments? How much do they think their monthly payment will be? And then finally, how does the amount of debt they have influence their educational decisions such as major choice? Of course, this last question necessitates careful survey design to make sure it captures students' true perceptions. Information on the earlier questions, however, might enable researchers to also examine

---

[12] Often students are placed into remedial or developmental courses once reaching college, and it would also be beneficial to learn of their expectations and perceptions of such courses before and during college.

the issue of whether debt has actually influenced such decisions. Finally, data on the budgets of recent college graduates would help provide perspective on what should be considered an excessive debt burden.

Family Resources: A More Accurate Picture of the Budget Constraint

Earlier in the paper, I called into question the common assumption that a portion of family income will go to pay for higher education and that a student's earnings can be devoted to that goal. Anecdotal evidence instead suggests that low-income students often contribute an important amount of their family's living expenses. If true, then family income, often a proxy for the student's budget constraint, is not accurately measured as the summation of the parents' earnings. One should also not assume that income made by the student is available to meet college expenses. To determine the prevalence of this type of family situation, it would be valuable to have surveys ask for more information on the role the student has in contributing to the family's resources. Questions about the use of summer and school-year earnings might shed some light on the issue. Additionally, this line of questioning could provide a clearer understanding of the family's living expenses compared to measures of income. Even after the student has left high school, information on changes in the income and employment of their parents would contribute to our understanding of a student's true budget constraint. Related to these issues, it is worth noting a strength of ELS: it asks what activities parents have undertaken to prepare financially for sending their child to college.

Other Issues

Finally, it is also worth noting several general aspects of the survey collections that are important and must be maintained. To truly understand the impact of a particular factor or policy on the decision to enroll, one must develop an appropriate strategy to isolate that effect from other possible influences as well as control for differences between the groups of students being compared. With respect to education research, often studies utilize samples of students across schools, states, or regions to accomplish this goal. This is due to the variation that frequently exists geographically. For instance, graduation requirements can differ by high school and tuition levels differ by state. The reliance on residential and school information makes knowing the exact residence, high school, and/or postsecondary institution attended essential. Though often only available if one has a restricted data license, this student-level information makes many innovative research designs possible. It also creates the possibility of sometimes linking between multiple datasets. Therefore, while the DAS and other data analysis tools give accessibility to much of the data, it is not enough for the types of causal analyses economists do.

V. CONCLUSIONS

Higher education is constantly evolving with increasingly diverse students seeking to attend a wide array of postsecondary institutions. The range of students' goals and educational pathways also underscore the dynamic nature of college decisions. With these trends as motivation, this paper attempts to consider new frameworks and tools that might provide helpful insight into contemporary college access and choice.

With regards to the HC Model, researchers could do more to incorporate the current realities of postsecondary trends. This would include more careful thought about how students perceive costs, the family resources that are (or are not) available to them, and the likely benefits they will receive given high college attrition among some groups. I also emphasize the importance of taking into account what students really know about postsecondary preparation and options. The quantity and quality of information they

have could greatly alter our view of the factors measured in the HC Model.

Additionally, I consider how contributions from Behavioral Economics might help researchers to reflect on postsecondary decisions. This field provides many examples of anomalies in decision-making, especially with regards to savings behavior, and its theories could help us to understand why some students choose not to attend college even in the face of overwhelming benefits. The framing of the default option appears to be especially important, and examples of this are also available in higher education. Policymakers, practitioners, and researchers should consider how redesigning the stepping stones to college (e.g., the financial aid application) might impact access.

Finally, I consider how the NCES data collections might be used to address some of the suggestions of the paper. Among the most important and feasible changes would be to ask more questions about students' information and perceptions. To be most beneficial, these questions should be asked repeatedly during a student's progression through and out of school. Such questions also have implications for how one thinks about the role of expectations and whether a student decides to prepare academically. Additionally, more information on perceptions of loans and their use would contribute to puzzles about loan aversion, convenience, and the longer-term effects of debt.

APPENDICES

Appendix A: References

Appendix B: Author

Appendix A: References

Adelman, C. (2006, August). *Internal Analysis*. Washington, DC: U.S. Department of Education, Office of Vocational and Adult Education.

Advisory Committee on Student Financial Assistance. (2005). The Student Aid Gauntlet: Making Access to College Simple and Certain. Final Report of the Special Study of Simplification of Need Analysis and Application for Title IV Aid. Washington, D.C.: Department of Education.

Avery, Chris and Caroline Hoxby. (2004). Do and should financial aid packages affect students' college choices? In Hoxby, C. (Ed.), *College Choices: The Economics of Which College, When College, and How to Pay for It*. Chicago: University of Chicago Press.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis*. New York: Columbia University Press.

Beshears, John, David Laibson, James J. Choi, and Brigitte C. Madrian, (2006). "Simplification and Saving." NBER Working Paper No. 12659.

Bettinger, Eric, Bridget Terry Long, and Philip Oreopoulos. (2008) Increasing Postsecondary Enrollment among Low-Income Families: A Project to Improve Access to College Information and Financial Aid. Project Description and Implementation Report. Harvard University mimeo.

Baum, S. (2003). *The Financial Aid Partnership: Strengthening the Federal Government's Leadership Role*. National Dialogue on Student Financial Aid. College Board.

Chelimsky, Eleanor. (1991). *Student Aid Information and Private Tuition-Guarantee Programs*. Testimony of the Assistant Comptroller General, Program Evaluation and Methodology Division of the General Accounting Office before the Subcommittee on Post-Secondary Education, U.S. House of Representatives, May 15, 1991.

Choi, James J., David Laibson, and Brigitte C. Madrian. (2005) "$100 Bills on the Sidewalk: Suboptimal Investment in 401(k) Plans." NBER Working Paper No. 11554.

Choy, Susan (2002) *Nontraditional Undergraduates: Findings from the Condition of Education 2002*. Washington: National Center for Education Statistics. Retrieved January 14, 2005, from http://nces.ed.gov/pubs2002/2002012.pdf.

Currie, Janet. (2004) "The Take Up of Social Benefits." National Bureau of Economic Research Working Paper No. W10488.

Dynarski, Susan. (2002) The Behavioral and Distributional Implications of Subsidies for College. *American Economic Review*. 92(2): 279-285.

Dynarski, Susan and Jill Scott-Clayton (2006) Cost of Complexity in Federal Student Aid: Lessons from Optimal Tax Theory and Behavioral Economics. *National Tax Journal*, LVIV(2).

Grodsky, Eric and Melanie T. Jones (2004). Real and Imagined Barriers to College Entry: Perceptions of Cost. Paper presented at the 2004 AERA Annual Conference.

Horn, Laura J., Xianglei Chen, and Chris Chapman. (2003). Getting Ready to Pay for College: What Students and Their Parents Know About the Cost of College Tuition and What They Are Doing to Find Out. National Center for Education Statistics Report No. 2003030. Washington, D.C.

Hoxby, Caroline. (1997) "The Changing Market Structure of U.S. Higher Education." Unpublished Harvard University mimeo.

Hoxby, Caroline and Bridget Terry Long. (1999). "Explaining Rising Income and Wage Inequality among the College-Educated." *National Bureau of Economic Research Working Paper No. 6873*, January.

Hussar, W.J., and Bailey, T.M. (2008). *Projections of Education Statistics to 2017* (NCES 2008078). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Ikenberry, S. O. and T. W. Hartle. (1998). *Too Little Knowledge is a Dangerous Thing: What the Public Thinks About Paying for College*. Washington, DC, American Council on Education.

Kane, Thomas J. and Christopher Avery. (2004) Student Perceptions of College Opportunities: The Boston COACH Program in Caroline Hoxby (ed.) *College Decisions: The New Economics of Choosing, Attending and Completing College*. Chicago: University of Chicago Press.

King, Jacqueline E. (1999) "Crisis or Convenience: Why are students borrowing more?" In Jacqueline King, ed. *Financing a College Education*. Phoenix: American Council on Education, Oryx Press.

Long, B. T. (2004) "How Have College Decisions Changed Overtime? An Application of the Conditional Logistic Choice Model." *Journal of Econometrics*, vol. 121, no. 1-2: pp. 271 -296.

Long, B. T. and Michal Kurlaender. (2009) "Do Community Colleges provide a Viable Pathway to a Baccalaureate Degree?" *Educational Evaluation and Policy Analysis, vol. 31, no. 1*: pp. 30-53

Madrian, Brigitte C. and Dennis F. Shea. (2001) "The Power of Suggestion: Inertia In 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics*, 116(4): pp. 1149-1187.

McSwain, C. and R. Davis. (2007). *College Access for the Working Poor: Overcoming Burdens to Succeed in Higher Education*. Washington, D.C.: Institute for Higher Education Policy.

Mullainathan, Sendhil and Richard Thaler. (2000) "Behavioral Economics." *International Encyclopedia of Social Sciences*, Pergamon Press, 1st edition: 1094-1100.

O'Donnell, K. (2005). *Tabular Summary of Adult Education for Work-Related Reasons: 20022003* (NCES 2005-044). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

O'Donoghue, Ted and Matthew Rabin. (1999) "Procrastination in Preparing for Retirement." In *Behavioral Dimensions of Retirement Economics*, Henry Aaron, Ed. Washington, D.C.: The Brookings Institution.

Rouse, Cecilia. (1995). "Democratization or Diversion? The Effect of Community Colleges on Educational Attainment." *Journal of Business Economics and Statistics* 13 (2), pp. 217-24.

Sallie Mae Fund (2003) Press release January 20, 2003.

Department of Education. (2008). *Digest of Education Statistics: 2007*. Washington, D.C.: IES National Center for Education Statistics.

General Accounting Office. (1990). *Higher Education: Gaps in Parents' and Students' Knowledge of School Costs and Federal Aid*. Report No. GAO/PEMD-90-20BR. Washington, D.C.: United States General Accounting Office.

Appendix B: Author

Bridget Terry Long, PhD

Professor, Education & Economics, Harvard Graduate School of Education

Research Associate, National Bureau of Economic Research (NBER)

Research Affiliate, National Center of Postsecondary Research (NCPR)

Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE of STATISTICAL SCIENCES
TASK FORCE REPORT

EMERGING ISSUES in POSTSECONDARY ACCESS and CHOICE:

DATA COLLECTION for ECONOMETRIC ANALYSIS of SCHOOLING DECISIONS

Charles F. Manski, PhD

DATA COLLECTION FOR ECONOMETRIC ANALYSIS OF SCHOOLING DECISIONS

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

DATA COLLECTION FOR ECONOMETRIC ANALYSIS OF SCHOOLING DECISIONS

I.   INTRODUCTION

This paper considers how NCES data collection can contribute to effective econometric analysis of schooling decisions, with emphasis on postsecondary schooling. We may, for example, want to forecast the impacts of revisions to financial aid programs, the introduction of new schooling alternatives, or changes to the returns to schooling. The credibility of a forecast depends on the realism of the model used to represent decision making. The model should embody a suitable conceptualization of decision making, informed as much as possible by data that sheds light on actual decision processes.

Past innovations in modeling have been closely associated with the emergence of new data. Modern discrete choice analysis of postsecondary schooling began in the 1970s when national surveys with data on individual choice behavior first became available. Perhaps the earliest such survey was the 1966 *School to College: Opportunities for Postsecondary Education* (SCOPE). Radner and Miller (1975) and Kohn, Manski, and Mundel (1976) used SCOPE data to study college entry decisions at completion of high school.

Subsequent NCES and Department of Labor (DOL) surveys have provided the empirical foundation for numerous studies of secondary and postsecondary schooling decisions. The NCES surveys include the *National Longitudinal Study of the High School Class of 1972* (NLS72), *High School and Beyond* (HSB) initiated in 1980, the National Educational Longitudinal Study (NELS) begun in 1988, and the Educational Longitudinal Study (ELS) begun in 2002. DOL has sponsored the 1979 and 1997 *National Longitudinal Surveys of Youth* (NLSY79 and NLSY97). Some of the many empirical studies that have used data from the NCES and DOL surveys to perform form econometric analysis of schooling decisions include Arcidiacono (2004), Cameron and Heckman (2001), Eckstein and Wolpin (1999), Keane and Wolpin (1997), Long (2004), and Manski and Wise (1983).

NCES should continue to play a major role in provision of data that enables econometric modeling of the complex of schooling decisions that contribute to educational outcomes. It is important to know what modeling is feasible with existing data and to explore how new data collection could enhance our modeling capacity. This paper offers my perspective on the matter.

To provide background, Section 2 describes the basic features of discrete choice analysis of schooling decisions, which seeks to infer decision processes from data on choice behavior and on covariates that describe the choice environment. I briefly consider discrete choice analysis in generality and then focus on the college entry decision for specificity. An Appendix provides a more detailed technical introduction to discrete choice analysis.

Section 3 considers how new data collection could enable more realistic econometric analysis of schooling decisions. As described in Section 2, econometric analysis presumes that the researcher observes, or somehow otherwise knows, what choices decision makers make and how they perceive their choice sets;

that is, the alternatives available to them. Existing surveys provide only a small part of the requisite data, so researchers routinely augment the available data with assumptions. Unfortunately, the assumptions made regularly lack empirical foundation. This diminishes the credibility of the analysis performed.

Existing NCES and DOL surveys ask respondents to describe their schooling-related choices, a sine qua non for choice analysis. However, existing surveys seek little if any corresponding description of non-chosen alternatives, which also is essential to choice analysis. In the absence of data, researchers make often heroic assumptions about how decision makers perceive their choice sets.

Section 3 calls for collection of data on choice-set perceptions. When discussing perceptions of choice sets, it is useful to distinguish respondent information and expectations. By information, I mean respondent knowledge of objectively verifiable facts that may be relevant to decision making. By expectations, I mean respondent beliefs about the outcomes they would experience if they were to make alternative choices. Collection of expectations data in a form that is useable for choice analysis requires a major departure from past NCES survey practice. Hence, I focus most of Section 3 on this matter.

Whereas Section 3 discusses new data collection in general terms, the concluding Section 4 addresses the practicalities of collection of data on choice-set perceptions. I first consider what has been done and what is possible to accomplish within large-scale surveys of the type that NCES has traditionally performed. The high cost and the respondent burden of lengthy interviews constrains what is feasible in large-scale surveys. Hence, I recommend intensive interviewing of smaller samples of respondents as a complement to the usual NCES survey approach.

II.   DISCRETE CHOICE ANALYSIS OF SCHOOLING DECISIONS

I first briefly sketch the general principles of discrete choice analysis. I then use the Manski and Wise (1983) study of college entry to illustrate application to schooling decisions. An Appendix provides a more detailed technical introduction to discrete choice analysis. The material in this section and the Appendix draws on Manski (1995, Chapter 5) and Manski (2007a, Chapter 13).

Basic Principles of Discrete Choice Analysis

Discrete choice analysis provides a framework for empirical study of choice among finite sets of alternatives. The framework has five essential features.

1. *Random Utility Model Representation of Behavior*: Discrete choice analysis describes the behavior of a population of heterogeneous decision makers, each assumed to make the best choice from a finite set of alternatives. A random utility model formally expresses the idea that the fraction of persons who choose a particular alternative equals the fraction of persons who prefer this alternative to all others.

2. *Representation of Alternatives and Actors by Attributes*: Discrete choice analysis enables thought experiments asking how behavior would change if new alternatives were to become available, existing ones were to become unavailable, or new decision makers were to face choice problems. This is achieved by characterizing alternatives and actors as attribute bundles.

3. *Analysis with Incomplete Attribute Data:* Empirical researchers typically possess incomplete attribute data. Unobserved attributes are treated as random variables or as parameters.

4. *Practicality:* Discrete choice analysis aims to be computationally tractable. McFadden (1974) introduced the conditional logit model. Subsequent work has broadened the set of models.

5. *Extrapolation*: An estimated model may be used to predict choice behavior in scenarios where members of a population face counterfactual choice sets.

College Choice in America

I shall use a highly simplified version of my own early analysis of college-going behavior, performed with David Wise, to illustrate application of the conditional logit model. Manski and Wise (1983, Chapters 6 and 7) used observations from the NLS72 to estimate a random utility model of college enrollment; see also Fuller, Manski, and Wise (1982). We then used the estimated model to predict the enrollment impacts of the Pell Grant program, the major federal college scholarship program.

The starting point for our analysis was to assume that the patterns of college enrollment and labor force participation observed among the NLS72 respondents are the consequence of decisions made by these students, by colleges, and by employers. Colleges and employers make admissions decisions and job offers that determine the options available to each high school senior upon graduation. Each senior selects among the available options.

An Idealized Binary Choice Setting

What do the NLS72 data reveal about the decision processes generating postsecondary activities? If we assume that a student chooses the most preferred alternative from the available options, observations of chosen activities partially reveal student preferences. For simplicity, assume that after high school graduation a student has two alternatives: d = 1 is college enrollment and d = 0 is work. (The model actually estimated posed multiple schooling and work alternatives.) If we observe that NLS72 respondent j chose to go to college, we may infer that the utility of college enrollment exceeds that of working. Formally, $u_j(1) \geq u_j(0)$. If person j chose to work, then $u_j(0) \geq u_j(1)$. The NLS72 data provide a large set of these inequalities, one for each respondent.

The preference inequalities implied by observation of these activity choices do not provide sufficient information to allow us to predict how a student not in the sample would select between college and work, or how a student in the sample would have behaved if conditions had differed. To extrapolate behavior, we must combine the NLS72 data with assumptions restricting the form of preferences.

For example, we might assume that the utility of college enrollment to student j depends on his ability A_j and his parents' income I_j, on the quality Q_j and net tuition T_j of his best college option, and an unobserved variable $\varepsilon_j 1$. Similarly, the utility of working might depend on his best potential wage W_j and an unobserved variable $\varepsilon_j 0$. For example, the utilities might have the linear form

$$uj(1) = \beta_1 v_1(Q_j, A_j) + \beta_2 v_2(T_j, I_j) + \varepsilon_j 1$$

and

$$u_j(0) = \beta_3 v_3(W_j) + \varepsilon_j 0.$$

Here $v_1(Q_j, A_j)$ measures the perceived benefit of college enrollment, which may vary with college quality and student ability, and $v_1(Q_j, A_j)$ measures the perceived cost of enrollment, which may vary with net

tuition and parental resources. The quantities $(\beta_1, \beta_2, \beta_3)$ are parameters to be estimated.

Assume that we observe or otherwise somehow know the choice y_j made by student j and the values of the attributes $(A_j, I_j, Q_j, T_j, W_j)$. Then the preference inequalities revealed by the NLS72 data are

$$y_j = 1 \Rightarrow \beta_1 v_1(Q_j, A_j) + \beta_2 v_2(T_j, I_j) - \beta_3 v_3(W_j) + \varepsilon_j \geq 0,$$

$$yj = 0 \Rightarrow \beta_1 v_1(Q_j, A_j) + \beta_2 v_2(T_j, I_j) - \beta_3 v_3(W_j) + \varepsilon_j \leq 0,$$

where $\varepsilon_j \equiv \varepsilon_j 1 - \varepsilon_j 0$. Assuming that the distribution of ε is continuous, the probability that a student with observed choice-set and personal characteristics v chooses to enroll in college is

$$P(y = 1 \mid v) = P(v\beta + \varepsilon \geq 0 \mid v),$$

where $v \equiv (v_1, v_2, -v_3)$ and $\beta \equiv (\beta_1, \beta_2, \beta_3)$. Assuming that ε_1 and ε_0 are independent and identically distributed with the Type I extreme value distribution, ε has the standard logistic distribution and the choice probabilities have the binary logit form

$$P(y = 1 \mid v) = \frac{e^{v\beta}}{1 + e^{v\beta}}.$$

Predicting the Enrollment Effects of Student Aid Policy

Manski and Wise (1983) estimated a random utility model that is considerably more complex than the one described above, but not qualitatively different. The estimated model was used to study the impact on freshman college enrollments of the Basic Educational Opportunity Grant program, later renamed the Pell Grant program. This federal scholarship program was initiated in 1973, so the NLS72 respondents were not eligible at the time of their initial postsecondary schooling decisions.

In the context of our random utility model, the Pell Grant program influences behavior by changing the college net tuitions that students face. Given knowledge of the program eligibility criteria and award formula, we can estimate the net tuition of college to any given student in the presence of the program. This done, we can predict how students would behave in the presence and absence of the program. We can then aggregate these predictions to generate predictions of aggregate freshman college enrollments in the United States.

Table 1 presents some of the findings concerning the version of the program that was in effect in 1979. The predictions indicate that the Pell Grant program was responsible for a substantial increase (59 percent) in the college enrollment rate of low-income students, a moderate increase (12 percent) in middle-income enrollments, and a minor increase (3 percent) in the rate for upper-income students.

Table 1: Predicted Enrollments in 1979, with and without the Pell Grant Program (thousands of students)								
	All Schools		Four-year College		Two-year College		Voc-Tech School	
Income Group	With Pell	Without Pell	With Pell	Without Pell	With Pell	Without Pell	With Pell	Without Pell
Lower	90	370	128	137	49	210	13	23
Middle	98	354	162	164	02	168	4	22
Upper	15	600	377	378	10	198	8	24
Total	603	1324	668	679	61	576	74	69
Lower income = below $16,900. Upper income = above $21,700. Source: Manski and Wise (1983), Table 7.4.								

Overall, we predicted that 1,603,000 of the 3,300,000 persons who were high school seniors in 1979 would enroll in full-time postsecondary education in 1979-80. In contrast, only 1,324,000 would have enrolled had the Pell Grant program not been in operation. The table indicates that the enrollment increases induced by the existence of the program were totally concentrated at two-year colleges and vocational-technical schools. Enrollments at four-year schools were essentially unaffected.

Power and Price of the Analysis

Federal scholarship programs with varying eligibility criteria and award formulae have been proposed, but only a few programs have actually been implemented. Revealed preference analysis of college enrollments makes it possible to predict the impacts of a wide variety of proposed and actual programs. This ability to extrapolate is very powerful.

The price of extrapolation is the set of assumptions imposed. The assumption of rational choice alone yields very limited ability to extrapolate. The power of revealed preference analysis emerges only when restrictions of the type discussed above are placed on preferences and when the researcher assumes that he can suitably describe the choices made and the choice sets faced by a sample of decision makers. This analysis of college choice illustrates the kinds of assumptions that have typically been imposed in empirical studies.

III. USING SURVEYS TO LEARN HOW DECISION MAKERS PERCEIVE THEIR CHOICE SETS

Perceptions and Choice

Rational choice is a subjective concept. Decision makers are assumed to choose the most preferred alternative, given their perceptions of the options available. There is no requirement that these perceptions be "correct" or "objective" in any sense. Applications of discrete choice analysis have traditionally assumed that researchers know how decision makers perceive their choice sets. For example, Manski and Wise (1983) assumed that their measurements of student ability, family income, college quality and net tuition, and potential wage correspond to the way youth and their families perceive these factors.

There is often much reason to question whether researchers really know how decision makers perceive their alternatives. Unless surveys directly ask respondents how they perceive their choice sets, researchers performing choice analysis must make assumptions about respondent perceptions. Existing surveys rarely

query respondents about choice-set perceptions. Hence, researchers have routinely made assumptions. Data collection would enable more realistic choice analysis.

Information and Expectations

When discussing perceptions of choice sets, it is useful to distinguish respondent information and expectations. By information, I mean respondent knowledge of objectively verifiable facts that may be relevant to decision making. When considering college application and entry decisions, these facts may include the current admissions and financial aid policies, course offerings, student-body compositions, and graduation rates of various institutions. Some youth and their families may not even be aware of the existence of colleges that may potentially be good matches. When aware of their existence, they may not know or may misperceive the objective attributes of these institutions.

In the absence of survey data on the information that respondents possess, researchers usually assume that youth and their families know the facts of their choice situations. When these assumptions are incorrect, choice analysis may mis-interpret preferences. For example, a researcher who observes that students with specified characteristics rarely enroll in a given type of college might infer that such students dislike this type of college. The actual reason for non-enrollment may be misperception or lack of knowledge of the attributes of the institutions.

Whereas information refers to knowledge of facts, expectations means respondent beliefs about the outcomes that may occur if they make alternative choices. Expectations are fundamental to schooling decisions, as persons evaluate their options by the outcomes they believe will result from them. Expectations differ from information because they are inherently subjective. There are no facts with which they can be compared.

Schooling decisions may depend on expectations of many events. Consider, for example, the decision to enroll in a particular college major. A student's choice of major may depend on his beliefs about the likelihood of successful completion of the required program of courses. It may also depend on his beliefs about the earnings he will receive if he should or should not successfully complete the program.

In what follows, I focus primarily on collection of data on expectations, because this requires substantial innovation relative to past NCES survey practice. For specificity, I first consider expectations of the monetary returns to schooling, which has long been a central concern of economists who study schooling decisions. Section 3.2 critiques the current practice of making assumptions about these expectations, in the absence of data. Section 3.3 describes an early exploratory study that used a survey to measure expectations of the returns to schooling, and cites some subsequent work of this type. Section 3.4 discusses in general terms the history and current status of expectations measurement in surveys. Section 3.4 calls attention to our very limited understanding of how persons form their expectations. Finally, Section 3.5 addresses collection of data on information. Sections 3.2 and 3.3 draw in part on Manski (1993) and Dominitz and Manski (1996). Sections 3.4 through 3.6 draw in part on Manski (2004).

Inquiring About the Locus of Decision Making

Before proceeding, I think it important to observe that existing surveys of youth and their parents do not directly ask respondents to state who makes schooling decisions in their families. It could be the youth alone, the parents alone, or some family decision process. In the absence of data, the standard econometric practice has been to assume that observed decisions result from optimizations by a single decision maker, sometimes thought of as the student and sometimes as the family.

This practice may be flawed. Some decisions, particularly those made during the early high school years, might better be conceptualized as the result of an intense interaction between students and their parents, who may have different preferences and beliefs. Later on, at the time of high school completion, parents may play a less comprehensive role in decision making but may still have a major say when comparing postsecondary schooling alternatives that differ in cost.

There is a need for surveys to collect data that sheds light on how schooling decisions are actually made in American households. This done, the participants in the decision processes should be questioned about their perceptions of choice sets.

Choice Analysis Using Assumptions on the Expected Returns to Schooling

Economists analyzing schooling decisions assume that youth compare the expected outcomes from schooling and other activities and choose the best available option. Viewing education as an investment in human capital, economists use the term *returns to schooling* to refer to the outcomes from schooling relative to non-schooling choices. Although this term can encompass both the monetary and non-monetary returns to schooling, economics have usually focused on earnings outcomes.

Given the centrality of the expected returns to schooling in economic thinking on schooling behavior, it might be anticipated that economists would make substantial efforts to learn how youth and their families perceive the returns to schooling. However, hardly any data have been collected. In the absence of data, the norm in analysis of schooling decisions has been to assume that expectations of the returns to schooling are formed in specific ways.

Researchers have generally supposed that all decision makers condition their beliefs on the same variables and process information in the same way. However, the hypothesized conditioning variables and information processing rule vary considerably across studies. For example,

a. Freeman (1971) analyzed the major field decisions of male college students. He assumed that each person chooses the field offering the highest expected lifetime income. Moreover, he assumed that expectations formation is myopic. Each person believes that, should he select a given college major, he would obtain the mean income realized by the members of a specified earlier cohort who did make that choice.

b. Willis and Rosen (1979) analyzed the college enrollment decisions of male veterans of World War II. They assumed that a person chooses to enroll if the expected lifetime income associated with enrollment exceeds that of not enrolling. They assumed that youth condition their income expectations on their ability. They also assumed that youth have *rational expectations*; that is, youth know the actual process generating lifetime incomes conditional on their ability and schooling. Willis and Rosen hypothesized that each youth applies his knowledge of the process generating lifetimes incomes to predict his own future income should he enroll or not enroll in college.

Several more recent studies have followed Willis and Rosen in assuming that youth have rational expectations. These include Arcidiacono (2004), Cameron and Heckman (2001), Eckstein and Wolpin (1999), Keane and Wolpin (1997), and Manski (1987), among others. The practice in these studies is to use longitudinal survey data on earnings to infer the objective returns to schooling and to assume that subjective expectations are the same as the objective returns.

c. In the Manski and Wise (1983) analysis of college choice, we did not assume that youth know the lifetime incomes realized by earlier cohorts nor the process generating their own future incomes. Instead, we assumed youth believe that the benefits of enrollment in a given college depend on their own SAT score and on the average score of students enrolled at the college. Regarding the potential wages from work, we used data on the realized wages of NLS72 respondents who chose to work to estimate a wage regression predicting wage conditional on various observed respondent attributes. We then assumed that respondents use this regression to predict their own wages.

One of the many ways that expectations assumptions differ across studies concerns how youth use ability information to form expectations. Whereas Freeman (1971) assumed that youth do not condition their expectations on ability, Willis and Rosen (1979) assumed that they do condition on ability and that they have rational expectations. Manski and Wise (1983) assumed that they use SAT score, which may be associated with ability, to form expectations that may or may not be rational.

Unfortunately, there is no evidence supporting any of the assumption's economists have made about expectations of the returns to schooling. In Manski (1993), I showed that these assumptions are consequential for choice analysis. I used a simple human capital model to demonstrate that interpretation of data on schooling choices can depend critically on how youth actually use ability information to form expectations of the returns to schooling. A particularly striking finding was that, if youth do not actually condition their expectations on ability, a researcher who assumes they do so may mistakenly conclude from observed choices that youth are unconcerned with the returns to schooling.

Eliciting Student Expectations of the Returns to Schooling

Knowledge of how students perceive the returns to schooling is essential to informed analysis of schooling decisions. However, only the barest empirical evidence has been available. A few studies have asked youth to report point beliefs about earnings unconditional on future schooling. Freeman (1971) and Betts (1996) asked college undergraduates about the average earnings of persons in various occupations or major fields - they did not ask respondents about the earnings they themselves would expect to receive if they were in these fields. Smith and Powell (1990) asked a sample of college seniors for point predictions of their "anticipated annual income in 10 years" and their "expected earnings" in the first year of their first job. Blau and Ferber (1991) asked a sample of college seniors to forecast "how much they would expect to earn initially and after 10 and 20 years if they were to be *continuously* employed in their preferred occupation after leaving school."

The above studies did not ask respondents for their expectations of the returns to schooling, which requires comparing alternative schooling choices. Nor did they permit respondents to express uncertainty in their beliefs. Uncertainty can be very important to schooling decisions. Students may be uncertain how well they would perform if enrolled in alternative institutions and programs of study. Moreover, they may be uncertain about the monetary and non-monetary outcomes that they would experience given different schooling choices.

Only a few studies have elicited returns to schooling and enabled respondents to express uncertainty about their expectations. The first of these was the exploratory effort of Dominitz and Manski (1996), who elicited probabilistic expectations of the returns to schooling from high school and university students in Madison, Wisconsin. We designed an interactive computer-assisted self- administered interview (CASI) survey eliciting from high school students and college undergraduates their expectations of the income they would earn if they were to complete different levels of schooling. We also elicited respondents' beliefs about current earnings distributions. We found that respondents were willing and able to respond meaningfully to questions eliciting their earnings expectations in probabilistic form. The 110 respondents varied considerably in their earnings expectations but there was a common belief that the returns to a college education are positive and that earnings rise between ages 30 and 40. There was a common belief that one's own future earnings are rather uncertain. Moreover, respondents tended to overestimate the current degree of earnings inequality in American society.

I describe this work in some detail because it provides a useful starting point for thinking about future data collection. Our substantive findings are of some interest, but the most innovative aspects of this exploratory study were the design and implementation of the survey. I focus on these aspects here. I cite subsequent research of this type at the end of the section.

Survey Medium and Structure

The survey was implemented on personal computers in student classrooms. The CASI software enabled us to design a survey that appears straightforward to respondents but that actually incorporated an extensive question-branching algorithm. CASI also enabled us to incorporate several tools intended to aid respondents in expressing meaningful expectations. *Training screens* explained basic probabilistic ideas through examples before the respondent begins the actual survey. *Help screens* could be accessed by the respondent at any time. *Error checks* informed the respondent if a response to a probabilistic question was not a proper probability or if the response was logically inconsistent with earlier responses. *Review-and-revise* screens showed the respondent his or her responses to each completed sequence of questions and allowed the respondent to revise these responses if desired.

Following the introductory and training screens, the CASI software led the respondent through the various parts of the survey. We sequentially elicited

1. *Unconditional Earnings Expectations*: Each respondent was queried about unconditional expected earnings (i.e., his or her subjective probability distribution of real earnings) at ages 30 and 40.

2. *Expected Earnings Under Alternative Schooling Scenarios*: Each respondent was asked about his or her expected earnings at ages 30 and 40 under the hypothetical scenario that the respondent continues in school at least through receipt of a bachelor's degree. The respondent was also asked about expected earnings at ages 30 and 40 under an alternative scenario assuming that less schooling is completed. In the version of the survey administered to high school students, the alternative scenario supposed that the respondent attains a high school diploma but no further schooling. In the version administered to college undergraduates, the alternative supposed that the respondent completes the current semester and then leaves school permanently.

3. *Schooling Expectations*: High school respondents were asked to state the probabilities that they will (a) attend college before age twenty-one, (b) receive a bachelor's degree before age 30, and (c) still be in school at age 30. College undergraduate respondents were asked to respond to (b) and (c).

4. *Beliefs About Current Earnings Distributions*: Male (female) respondents were asked about the current distributions of earnings among American men (women) of age 30 who have attained at least a bachelor's degree and among those who have attained a high school diploma but no further schooling.

I describe below how we elicited subjective earnings distributions and how we specified schooling scenarios.

Eliciting Subjective Earnings Distributions

We considered eliciting earnings expectations by asking respondents to report quantiles of their subjective distribution of future earnings or, alternatively, probabilities that earnings would exceed various thresholds. We decided to use a hybrid approach. First, we elicited the median of a subjective distribution and then the probabilities that earnings would exceed a sequence of thresholds, posed in increasing order. The thresholds about which a given respondent was queried were determined by the respondent's elicited median.

The present discussion uses our questions about unconditional earnings expectations to illustrate ideas. We used the same approach to elicit expectations under alternative schooling scenarios and beliefs about current earnings distributions.

Eliciting Median Earnings: We first asked the respondent to state the median of his or her subjective distribution for future earnings. Eliciting the median determines the probabilistic center of a respondent's expectations and provides a natural starting point for subsequent selection of threshold values in probability elicitation. At the beginning of the survey, respondents received this on-screen instruction defining the median of a distribution:

"The first question will ask you about the median amount of money that you think you will earn at some time in the future. The median is the amount of money for which there is a 50 percent chance that you will earn more than it and a 50 percent chance that you will earn less than it. So, to answer this question and others like it, you should try to pick the amount of money that you think there is just as good a chance you will earn more than it as less than it."

The question we used to elicit the median of the distribution of unconditional earnings had this format:

"Look ahead to when you will be X years old. Think about the kinds of jobs that will be available for you and that you will accept. What is the median amount of money that you think you will earn per year by the time you are X years old?"

Eliciting Probabilities That Earnings Exceed Thresholds: It is easy to pose readily understood questions eliciting probabilities that earnings would exceed specified thresholds. Here is the format we used to elicit unconditional earnings expectations:

"When you are X years old, what do you think is the percent chance that you will earn more than Y thousand dollars per year? That is, what are the chances out of 100 that you will earn more than Y thousand dollars per year?"

Each respondent answered this question with X taking the values 30 and 40 and Y taking three values that depend on the respondent's elicited median. Elicitation of probabilities for three thresholds, rather than

more or fewer, was our resolution of the tension between our desire to collect as refined data as possible and our concern that too much attempt at refinement would sap respondents' attention.

Effective selection of the threshold values Y is a non-trivial matter. Asking questions about a range of thresholds spanning the support of a respondent's subjective distribution should yield more information about the shape of the distribution than would the same number of questions asked about a narrower or wider range of thresholds. Morgan and Henrion (1990) warn that the specific thresholds chosen may influence respondents' beliefs, a phenomenon called *anchoring* by psychologists. Suppose, for example, that a respondent expects his earnings to be no greater than $30,000. Psychologists fear that, if the first question asked concerns the probability that earnings will be greater than $50,000, the respondent may be influenced to think that this amount is objectively reasonable and so may report a higher probability than believed a priori.

One of the reasons that we first asked about median earnings was that the reported median provides a natural self-anchor for later selection of thresholds. Our question-branching algorithm used the reported subjective median to select three thresholds out of this set of six candidate values:

{$10,000, $30,000, $50,000, $75,000, $100,000, $150,000}. The first three values were selected if the reported median was less than 40 thousand dollars. The next set - 30, 50, and 75 thousand - were chosen for medians from 40 - 65 thousand dollars. Successively higher threshold values were chosen for medians from 65 - 90 and above 90 thousand dollars.

Posing Schooling Scenarios

Finally, I describe our questions eliciting earnings expectations under alternative schooling scenarios. I focus on the version of the survey administered to high school students. The same considerations arise in the version administered to college undergraduates.

We first gave respondents this on-screen introduction:

"The next sets of questions ask you to put yourself in one of two hypothetical situations. In the first situation, you assume that you continue in school until you finish your senior year of high school and obtain your diploma, and you do not continue in school after that. In the second hypothetical situation, you assume that you continue in school at least until you finish your senior year of college and obtain your college diploma (a bachelor's degree). When responding to these questions, please attempt to fully place yourself in the hypothetical situation as it is described."

After this, respondents receive these more detailed instructions concerning the first scenario:

"In the first hypothetical situation, assume that you continue in school until you complete your senior year of high school and obtain your high school diploma. Please respond under the assumption that you do not return to school at any time after high school. Remember, this is a hypothetical situation. Just think about the kinds of jobs that would be available for you and that you would accept. Think about the amount of money you would make on these jobs. Again, you should ignore the effects of price inflation on earnings."

Following this, respondents were asked to state their median earnings at age 30 and the probabilities that earnings would exceed three thresholds.

Observe that the wording used to describe the two schooling scenarios offered no reason why one or the other scenario might be realized. This was intentional. We did not want respondents to draw from the descriptions of the scenarios information that might influence their expectations.

Subsequent Studies

Following the Dominitz-Manski (1996) exploratory study, several other small-scale surveys have undertaken probabilistic measurement of expectations of the returns to schooling and of other expectations relevant to schooling decisions. Wolter (2000) performed an exploratory study in Switzerland that is similar to the Dominitz-Manski work and compared the Swiss and American findings. Rouse (2004) elicited subjective median earnings expectations with and without a college degree from a sample of high school seniors in Baltimore. Zafar (2008) elicited probabilistic expectations regarding the monetary and non-monetary returns to different college majors from a sample of American college undergraduates. Stinebrickner and Stinebrickner (2008) elicited probabilistic expectations of academic performance (grade-point average) from a sample of American college undergraduates and related the expectations to their subsequent college completion outcomes.

Thus far, the only large-scale survey to provide probabilistic data on the expectations of youth is the NLSY97. The NLSY97 has not elicited expectations of the returns to schooling. However, the base year questionnaire, administered in 1997, asked about 3500 youth aged fourteen to sixteen a module of questions asking them to state the unconditional probabilities of various schooling, work, parenthood, crime, and mortality outcomes. Fischhoff *et al*.(2000), Dominitz, Manski, and Fischhoff (2001), and Walker (2001) report analyses of the responses. Similar questions were administered in the fourth and fifth annual survey rounds.

Measuring Probabilistic Expectations: Ideas and Applications

The exploratory study described in Section 3.3 is an early example of the general idea of probabilistic measurement of expectations, which has been applied increasingly often in surveys from the 1990s on. The idea has many potential applications to schooling surveys, so I think it important to discuss it at some length. This section, which draws on Manski (2004), describes some of the history of research on expectations, beginning with the verbal tradition of attitudinal research and then turning to elicitation of probabilistic expectations in cognitive psychology and economics.

Attitudinal Research

Attitudinal researchers have long used verbal questions to measure expectations. When asked to predict some outcome, respondents may be asked to report whether they "think" or "expect" that the event will occur. Sometimes they are asked to report the strength of this belief by reporting whether it is "very likely," "fairly likely," "not too likely," or "not at all likely" that the event will occur. A prominent example is this Michigan Survey of Consumers question on business conditions, the responses to which have been used since the 1950s to measure consumer confidence (Curtin, 1982):

Survey of Consumers Business-Conditions Question: Now turning to business conditions in the country as a whole - do you think that during the next 12 months we'll have good times financially, or bad times, or what?

Another is this question on job loss in the General Social Survey, a pivotal sociological survey in the United States (Davis and Smith, 1994):

General Social Survey Job-Loss Question: Thinking about the next twelve months, how likely do you think it is that you will lose your job or be laid off - very likely, fairly likely, not too likely, or not at all likely?

These questions illustrate a persistent problem that researchers face in interpreting verbal expectations data-assessment of the interpersonal comparability of responses. How do respondents to the Michigan survey interpret the phrases "business conditions" and "good times financially?" Do different respondents to the General Social Survey interpret the phrases "very likely, fairly likely, not too likely, or not at all likely" in the same way? Cognitive research does not give reason to think that responses should be or are comparable. Indeed, the available empirical evidence indicates that interpretation of verbal expectations questions varies substantially between persons (Lichtenstein and Newman, 1967; Beyth-Marom, 1982; Wallsten *et al.*, 1986). One may also question whether responses are intra-personally comparable; that is, a given respondent may interpret verbal phrases in different ways when asked about different events.

A second persistent problem is that the coarseness of the response options limits the information contained in the responses. Consider, for example, the fertility question asked female respondents in the annual June Supplement to the Current Population Survey of the U.S. Bureau of the Census:

Current Population Survey Fertility Question: Looking ahead, do you expect to have any (more) children?

           _____ YES   _____ NO   _____ UNCERTAIN

The three response options do not enable respondents to express the degree of uncertainty they perceive about their future childbearing.

NCES survey questions on student plans and expectations similarly have not permitted respondents to express uncertainty about their future behavior. For example, the 2004 follow-up to the Educational Longitudinal survey posed this question to high school seniors:

Education Longitudinal Survey Expectations Question: As things stand now, how far in school do you think you will get? (Mark One Response): Less than high school graduation; GED or other equivalency only; High school graduation only; Attend or complete a 1- or 2-year program in a community college or vocational school; Attend college, but not complete a 4- or 5-year degree; Graduate from college (4- or 5- year degree); Obtain a Master's degree or equivalent; Obtain a Ph.D., M.D., or other advanced degree; Don't know.

The influential Fishbein-Ajzen model of intentions illustrates the loose manner in which attitudinal researchers have thought about expectations. Fishbein and Ajzen (1975) propose that "intention" is a mental state that causally precedes behavior and that can be elicited through questionnaires or interviews. According to Ajzen and Fishbein (1980), a person's "behavioral intention" is his subjective probability that the behavior of interest will occur. (They refer to the response to a yes/no intentions question as "choice intention.") It seems, however, that social psychologists do not use the term "subjective probability" as a Bayesian statistician would. Ajzen and Fishbein (1980, page 50) state: "we are claiming that intentions should always predict behavior, provided that the measure of intention corresponds to the behavioral criterion and that the intention has not changed prior to performance of the behavior." In a review of attitudinal research, Schuman and Johnson (1976, page 172) write that the Fishbein-Ajzen model implies

that "the correlation between behavioral intention and behavior should approach 1.0, provided that the focal behavior is the same in both cases and that nothing intervenes to alter the intention." It is difficult to reconcile these statements with the idea that behavioral intention is a subjective probability, unless that probability is always zero or one.

Probabilistic Expectations in Cognitive Psychology

If persons can express their expectations in probabilistic form, elicitation of subjective probability distributions should have compelling advantages relative to verbal questioning. Perhaps the most basic attraction is that probability provides a well-defined absolute numerical scale for responses; hence, there is reason to think that responses may be interpersonally comparable. Another attraction is that empirical assessment of the internal consistency of respondents' expectations is possible. A researcher can use the algebra of probability (Bayes Theorem, the Law of Total Probability, etc.) to examine the internal consistency of a respondent's expectations about different events.

When probability has a frequentist interpretation, a researcher can compare elicited subjective probabilities with known event frequencies and reach conclusions about the correspondence between subjective beliefs and frequentist realities. Such *calibration studies* have a long history in cognitive psychology. Lichtenstein, Fischhoff, and Phillips (1982) review findings from 1906 on, and McClelland and Bolger (1994) update the review with findings from 1980 through 1994. Whereas the older studies mostly examined the accuracy of experts (e. g., weather forecasters' reported probabilities of precipitation), much recent research analyzes the expectations of non- experts, especially students in a cognitive laboratory.

Within cognitive psychology, there has been controversy about the way in which humans internally represent their beliefs, and their ability and willingness to express their beliefs as numerical probabilities. Koriat, Lichtenstein, and Fischhoff (1980) and Ferrell and McGoey (1980) posed models in which individuals may have some difficulty expressing beliefs as numerical probabilities, but nevertheless concluded that elicitation of numerical subjective probabilities is feasible.

However, Zimmer (1983, 1984) argued that humans process information using verbal rather than numerical modes of thinking, and he concluded that expectations should be elicited in verbal rather than numerical forms.

Erev and Cohen (1990) and Wallsten *et al.* (1993) have reported that a majority of respondents prefer to communicate their own beliefs verbally and to receive the beliefs of others in the form of numerical probabilities. This asymmetry is intriguing but only marginally relevant to the design of expectations questions. The relevant question is not what communication mode respondents prefer to use, but rather what modes they are willing and able to use. Wallsten *et al.* (1993) report that virtually all of their respondents were willing to communicate their beliefs numerically, should the situation warrant it.

Probabilistic Expectations in Economics

Among economists, the idea that survey measurement of probabilistic expectations might improve on the verbal approaches of attitudinal research appears to have originated with Juster (1966). Considering the case in which the behavior of interest is a binary purchase decision (buy or not buy), Juster considered how responses to traditional yes/no buying intentions questions should properly be interpreted. He wrote (Juster, 1966, page 664):

> "Consumers reporting that they 'intend to buy A within X months' can be thought of as saying that the probability of their purchasing A within X months is high enough so that some form of 'yes' answer is more accurate than a 'no' answer."

Thus, he hypothesized that a consumer facing a yes/no intentions question responds as would a decision maker asked to make a best point prediction of a future random event. Working from this hypothesis, Juster concluded that it would be more informative to ask consumers for their purchase probabilities than for their buying intentions. In particular, he proposed questions that associate verbal expressions of likelihood with numerical probabilities to elicit purchase expectations for automobiles and other household appliances:

Juster Purchase Probability Questions: Taking everything into account, what are the prospects that some member of your family will buy a _____ sometime during the next _____ months, between now and _____?

> Certainly, Practically Certain (99 in 100); Almost Sure (9 in 10); Very Probably (8 in 10); Probably (7 in 10); Good Possibility (6 in 10); Fairly Good Possibility (5 in 10); Fair Possibility (4 in 10); Some Possibility (3 in 10); Slight Possibility (2 in 10); Very Slight Possibility (1 in 10); No Chance, Almost No Chance (1 in 100).

He went on to collect data and concluded that elicited purchase probabilities are better predictors of subsequent individual purchase behavior than are yes/no intentions data.

Market researchers were attracted to Juster's proposal (e.g., Morrison, 1979). The idea that expectations might be elicited probabilistically from survey respondents did not, however, draw the immediate attention of economists. By the time Juster's article was published, economists were preaching that empirical research on decision making should be based on choice data alone. A quarter century passed before economists began to systematically collect and analyze probabilistic expectations data.

The conventional economic wisdom finally unraveled in the 1990s. Several large-scale surveys now use probabilistic formats to elicit expectations, and a new field of empirical research on expectations has emerged. Expectations have been elicited for macroeconomic events (stock market returns), for risks that a person faces (job loss, crime victimization, mortality), for future income (earnings and Social Security benefits), and for choices that persons make (durable purchases and voting choices).

The major American platforms for methodological exploration and substantive research have included the Health and Retirement Study (Juster and Suzman, 1995; Hurd and McGarry, 1995), the Survey of Economic Expectations (Dominitz and Manski, 1997a, 1997b), the NLSY97 (Fischhoff *et al.*, 2000; Dominitz, Manski, and Fischhoff, 2001; Walker, 2001), and the recently begun RAND American Life Panel internet survey. Even the venerable Michigan Survey of Consumers has sometimes included probabilistic questions along with its traditional verbal questions (Dominitz and Manski, 2004).

Manski (2004) and Hurd (2009) review the new literature on probabilistic measurement of expectations in surveys. These review articles describe what has been learned methodologically about effective question design, as well as what has been learned substantively about a large variety of expectations relevant to family decision making. They also discuss still-open questions about the cognitive processes at work when persons form expectations and respond to survey questions.

Extrapolation and Knowledge of Expectations Formation

Sections 3.2 through 3.4 focused entirely on the measurement of expectations. I have not described research that aims to understand how persons form expectations and revise them with receipt of new information. In fact, to estimate random utility models of decision making, researchers do not need to understand expectations formation. They need only to measure the expectations that persons held when they made their observed choices.

The situation may change when one wants to extrapolate from observed behavior, using an econometric decision model to predict choice behavior in a new scenario. Measurement of the expectations associated with observed choices suffices if it is plausible to assume that the new scenario does not affect expectations or that it changes them in some obvious way. However, econometric decision models often are used to predict behavior following policy interventions or other events that may alter expectations in non-obvious ways. Then credible prediction requires an understanding of expectations formation, a large subject about which little is known.

Experimental psychologists and economists have studied how persons update objective probabilities following receipt of random sample data in highly structured settings similar to those presented in textbook statistics exercises. A particular concern has been to test adherence to and characterize departures from application of Bayes Theorem; see, for example, Tversky and Kahneman (1974) and El-Gamal and Grether (1995). However, it is difficult to draw lessons from this work for expectations formation in real life, where the information that persons receive rarely maps cleanly into a textbook exercise in probability updating. Bayesian updating, which expresses new information through the likelihood function, presumes that data are generated by a well-understood sampling process. Expectations formation in real life requires persons to assimilate government announcements and media reports, communication from friends and family, as well as personal experiences and observations of the experiences of others. The sampling process generating these forms of information is obscure.

One can learn something about updating in real life by eliciting expectations longitudinally or from repeated cross sections of the population. Dominitz (1998) elicited earning expectations at six- month intervals from a sample of respondents. He examined the association between revisions to expectations and the earnings that respondents realized between interviews.

Research that measures revisions to expectations and associates them with observed event realizations can be informative. However, I think that understanding expectations formation will also require intensive probing of persons to learn how they perceive their environments and how they process such new information as they may receive. Large-scale population surveys are not amenable to investigations of this type - the time available to query respondents is too limited and the standardized question-response format of interviews is too confining. Researchers may need to engage small samples of respondents in lengthy, semi-structured "conversations."

Querying Survey Respondents about Facts

Having discussed collection of expectations data at some length, finally consider asking respondents about the information they possess; that is, their knowledge of objectively verifiable facts. Survey researchers routinely ask respondents to report facts. For example, respondents to NCES and DOL surveys are asked about their schooling and work histories. While this has not been done to date, respondents could be asked about various facts that are relevant to schooling decisions. For example, high school seniors contemplating college entry could be asked about the tuition of various postsecondary institutions, their eligibility for Pell grants and student loans, the available magnitudes of such awards, and the repayment options for loans. Asking such questions would not require innovation in NCES survey practice and could shed light on how students perceive their choice sets.

Although it is useful to distinguish information from expectations when considering the verifiability of responses, the two concepts need not be distinct from the perspective of respondents. Respondents often have only partial knowledge about facts. For example, high school students and their parents may have only a loose idea of how the Pell Grant and student loan programs work. Respondents who are uncertain about these matters may choose not to respond to conventional questions asking about facts, or they may give responses that are not much more than guesses.

Modifying traditional survey research practices to permit respondents to express uncertainty about facts can potentially shed light on the information they possess. Constructive steps in this direction have been taken by the Health and Retirement Study (HRS), which has used *unfolding bracket* questions to enable respondents to flexibly provide interval data on their income and assets. Respondents who are willing to provide point responses can do so. Those who are unwilling to respond to questions eliciting point responses are asked whether the quantity of interest lies above or below a sequence of specified thresholds. See Juster and Suzman (1995) and Hurd (1999).

Probabilistic elicitation of facts offers another route for improvement of survey research practice. When the fact is categorical, respondents can be asked to report a subjective probability. For example, consider eligibility for a Pell Grant. A student who knows that he is eligible can state that he has a one- hundred percent chance of eligibility while someone who is unsure can state a lower percent chance. When the fact of interest is real-valued, such as the amount of a Pell Grant, the method used by Dominitz and Manski (1996) to elicit expectations of the returns to schooling may be applied.

Although probabilistic elicitation of facts appears not to have previously been proposed as a tool of survey research, it is of interest to note that the idea has long had proponents in educational testing. Shuford, Albert, and Massengill (1966) argued that requiring a student to choose one answer to a true- false, multiple-choice, or fill-in-the-blank question reveals (page 125) "only a very small fraction of the information potentially available from each query." They proposed that students instead be asked to state subjective probabilities for the correctness of alternative answers to a question.

IV. THE STATE OF DATA COLLECTION AND RECOMMENDATIONS FOR FUTURE NCES SURVEYS

The absence of data on choice-set perceptions has seriously afflicted empirical analysis of schooling decisions from the earliest research through the present. Over thirty years ago, when Radner and Miller (1975) and Kohn, Manski, and Mundel (1976) used SCOPE data to perform the first discrete choice analyses of college entry, it was necessary to "impute" choice sets. The SCOPE questionnaire asked respondents to state their postsecondary choices but did not ask them to identify non-chosen alternatives.

The NLS72 was a major advance in data collection because its first follow-up survey asked respondents to state all postsecondary institutions to which they had applied and been admitted. However, the NLS72 did not ask respondents to characterize non-chosen alternatives to schooling and did not ask them about their expected returns to schooling. Manski and Wise (1983, Chapter 6) remarked on this in their conclusion, writing that an open question (p.117)

> "concerns the relation between the student's postsecondary choice and his future. One aspect of this question . . . . . asks how students form expectations about the future consequences - particularly earnings consequences - of their postsecondary decisions. Since expectations are themselves unobserved, our choice model must implicitly incorporate the expectations formation process. Empirical analysis then constitutes a joint test of the choice model and the model of expectation formation."

The NCES surveys that followed NLS72 have not gone any further towards provision of data on choice-set perceptions. In the mid-1990s, collection of data on the expected returns to schooling was proposed in the NLSY97 design effort, in which I participated. However, collection of expectations data was ultimately scaled back to the small module described in Section 3.3, which asked only a few simple questions about unconditional schooling expectations.

Looking back over the past forty years of NCES and DOL schooling surveys, I can see several possible reasons why these surveys have not done more to inquire about choice-set perceptions. First, the main motivation for collection of such data is econometric analysis of schooling decisions. Choice analysis has been highly important to economists, but it has not engaged the attention of the broader community of social scientists and education researchers who have designed the NCES and DOL surveys. Second, the economists who should want such data to be a high priority in questionnaire design have not pushed for it. Instead, most economists have been content to make assumptions about choice-set perceptions, despite the absence of empirical foundations for their assumptions. Third, a comprehensive effort to inquire about choice-set perceptions requires substantial interview time. Even with the best of intentions, this is difficult to obtain within the large-scale NCES and DOL surveys, which aim to provide a broad spectrum of data for diverse research and policy purposes.

I believe that the first two reasons for inaction cited above can be overcome through serious discourse. First, it should be possible to make a compelling case that effective econometric analysis of schooling decisions is important both to educational research and to evaluation of educational policy. Second, economists are already showing diminished faith in traditional rational-expectations assumptions about choice-set perceptions and, correspondingly, are showing increased interest in empirical measurement of

information and expectations. Indeed, the rationale for questioning survey respondents about their choice-set perceptions has strengthened over time, as the introduction of non-traditional secondary (e.g., home schooling) and postsecondary (e.g., distance-learning) options has broadened the set of schooling alternatives available to students.

I think that NCES can overcome the third possible reason for inaction, interview length, in two ways. First, it should be possible to place relatively short, say three to five minute, modules of questions on choice-set perceptions in future large-scale surveys of the NLS72 type. To make space for these modules, I would eliminate the traditional attitudinal questions on plans and expectations that have been included in past surveys. As explained in Section 3.4, these traditional questions are not especially informative.

Second, to enable comprehensive questioning on choice-set perceptions, NCES could administer substantially longer, say thirty to sixty minute, modules of questions to smaller samples of respondents. The Health and Retirement Study, funded by the National Institute on Aging (NIA) provides a good model that shows how such intensive interviewing may be accomplished in practice. From its inception in 1992 through the present, the HRS has combined its core questionnaire, which is administered to all respondents, with modules of additional questions that are administered to randomly drawn sub-samples of respondents. The HRS has also drawn dedicated samples of respondents, separate from the main sampling frame, for lengthy interviews on topics considered to be important to the NIA mission. For example, a special survey on the cognitive functioning of older persons has recently been completed. I believe that NCES could fruitfully adapt both of these HRS mechanisms to enhance the value of its surveys, not only to learn about respondent choice-set perceptions but to collect new data of other types as well.

Yet another potentially valuable mechanism for intensive interviewing is development of an internet panel of respondents who may be queried repeatedly about subjects of interest. The ongoing RAND American Life Panel demonstrates the possibilities.

APPENDICES

Appendix A: Discrete Choice Analysis

Appendix B: References

Appendix C: Author

Appendix A: Discrete Choice Analysis

Econometric analysis of discrete choice, pioneered by McFadden (1974), seeks to predict the distribution of choices that would be made by the members of a population in specified choice scenarios. Suppose that a researcher observes the decisions made by a population of persons, each of whom faces a choice problem. McFadden showed that these data, combined with assumptions on the population distribution of preferences, enable estimation of parametric probabilistic choice models. He then showed how probabilistic choice models may be used to predict population choice behavior in other settings. There are five steps, which I describe here. I conclude with a brief discussion of developments subsequent to McFadden's early contribution.

Random Utility Representation of Choice Behavior

Applying basic utility theory to the problem of discrete choice, assume that each member of a study population faces a finite choice set and selects an action that maximizes utility. I will suppose here that all members of the population face the same choice set, denoted C, which is a finite subset of a universe of potential alternatives, denoted A. The assumption of a common choice set simplifies the exposition, but it is not critical to what follows.

Formally, let person j have a utility function $u_j(\cdot)$ that assigns a utility value to each alternative in A. Suppose that this person is observed to choose some action $c \in C$. Then the standard revealed preference argument of economic theory holds that person j ranks c highest among the feasible alternatives. Thus, $u_j(c) \geq u_j(d)$, all $d \in C$.

Assume that indifference between alternatives occurs with probability zero. Then the fraction of the population who choose c equals the fraction of the population for whom c is the utility maximizing action. Thus, the *choice probability* for alternative c is (1) $P(y = c) = P[u(\cdot): u(c) \geq u(d), \text{ all } d \in C]$.

Equation (1) is a *random-utility-model* description of observed choices. In psychology, random utility models date back to Thurstone (1927) as a way of conceptualizing semi-rational behavior. The psychological interpretation, exposited in Luce and Suppes (1965), assumes that each decision maker carries a distribution of utility functions internally and selects one at random whenever a decision must be made. McFadden re-interpreted the randomness as arising from variation in utility functions across the population rather than variation within an individual.

Representation of Alternatives and Actors by Attributes

Suppose that one could learn the population distribution of utility functions. Then a random utility model would not only describe observed population behavior but would enable prediction of behavior in counterfactual scenarios where members of the population face choice sets other than C. Suppose, for example, that all persons were to face choice set D. The choice probability for each $c \in D$ is

(2) $P[y(D) = c] = P[u(\square): u(c) \; \square \; u(d), \text{ all } d \in D]$.

Learning the distribution of utility functions from choice data is no small problem. Observation of the study population only yields equation (1), or the analogous equation conditioning on observed covariates of decision makers. Clearly, many distributions of utility functions can solve this equation.

To resolve this identification problem, the researcher brings to bear assumptions that restrict the distribution of utility functions. McFadden's central idea was to transform the hitherto qualitative distinctions among alternatives and decision makers into quantitative differences in their attributes. Thus, the second step in the development of discrete choice analysis is to assume that each alternative d can be characterized through a vector of attributes, say w_d, and that each decision maker j can be characterized through a vector of attributes, say x_j. Then the utility of alternative d to person j has the form (3) $u_j(d) = u(w_d, x_j)$, where $u(\cdot, \cdot)$ maps the attributes of alternatives and decision makers into utility values. With knowledge of these attributes and the form of $u(\cdot, \cdot)$, a researcher can determine the utility of any alternative to any decision maker and, hence, can predict the decision maker's choice behavior. For example, a transportation researcher can predict how commuters with specified income and job attributes would choose among travel modes with specified travel-time and travel-cost attributes.

Analysis with Incomplete Attribute Data

One more conceptual step is needed to yield a viable approach to discrete choice analysis. It is not realistic to think that an empirical researcher will have complete data on the attributes of alternatives and decision makers. Nor is it realistic to think that the researcher fully knows the form of the function $u(\cdot, \cdot)$ mapping attributes into utility values.

A viable approach emerges if the researcher has enough partial knowledge of $u(\cdot, \cdot)$ and of the population distribution of unobserved attributes. In particular, McFadden assumed that $u(\cdot, \cdot)$ has the linear form

(4) $$u(w_d, x_j) = v(w_{do}, x_{jo})\beta + \varepsilon_{jd}.$$

Here $v(\cdot, \cdot)$ is a known vector-valued function of the components of w_d and x_j that the researcher can observe, labeled w_{do} and x_{jo}, while β is a commensurate parameter vector. The unobserved variables $\varepsilon_j \equiv (\varepsilon_{jd}, d \in C)$ express the contribution of unobserved attributes to utility.

With this setup, McFadden derived choice probabilities conditional on the observed attributes. Let $W_{Co} \equiv (w_{do}, d \in C)$. Consider the sub-population of persons with observed covariates x_o. Assume that the researcher knows the distribution $P(\varepsilon|x_o, W_{Co})$ up to the value of a parameter vector α. Let this parametrized distribution be denoted F_α. Then the random utility model implies that the choice probability for alternative c conditional on (x_o, W_{Co}) is

(5) $$P(y = c \mid x_o, W_{Co}) = F_\alpha[\varepsilon: v(w_{co}, x_o)\beta + \varepsilon_c \geq v(w_{do}, x_o)\beta + \varepsilon_d, \text{ all } d \in C \mid x_o, W_{Co}].$$

These conditional choice probabilities provide the basis for inference on the parameters (β, α) using data on the choices and attributes of a random sample of decision makers. Observation of the study population reveals the conditional choice probabilities on the left-hand-side of equation (5). Hence, the identification region for (β, α) is the set of parameter values that make (5) hold for all alternatives c and for all values of x_o in the support of $P(x_o)$. The parameters are point-identified if these equations have a unique solution. If no parameter value solves (5) for all values of c and x_o, the model does not correctly portray choice behavior.

Practicality

*A*t this point the conceptual development of modern discrete choice analysis was complete. The remaining question was practicality, a matter whose interpretation must depend on the computational power of the day.

McFadden was concerned that estimation of the parameters (β, α) be tractable. He judged that, with the computational technology available at the time of his work, discrete choice analysis would be practical only if the conditional choice probabilities have a quite simple form. With this in mind, he searched for a convenient assumption to impose on the distribution of the unobserved variables ε.

He found what he was looking for in earlier work on psychological random utility models. Adapting results proved by Marschak (1960) and Holman and Marley (cited in Luce and Suppes, 1965), he assumed that ε is statistically independent of (x_o, W_{Co}) and that the components of $(\varepsilon_d, d \in C)$ are independent and identically distributed, each according to the Type I extreme-value distribution. These assumptions imply that the conditional choice probabilities have the *multinomial logit* form

$$(6) \quad P(y = c \mid x_o, W_{Co}) = \frac{\exp[v(w_{co}, x_o)\beta]}{\sum_{d \in C} \exp[v(w_{do}, x_o)\beta]} .$$

The distribution of ε having been fully specified, there are no α parameters to be estimated here, only the β parameters.

Extrapolation

A primary objective of discrete choice analysis is to enable prediction of choice behavior in scenarios where members of the population face counterfactual choice sets. How is this done?

For concreteness, suppose that all members of the study population were to face choice set D rather than C. Let $W_{Do} \equiv (w_{do}, d \in D)$ denote the observable attributes of the elements of D. Then, for each $c \in D$, repetition of the derivation leading to (5) yields the counterfactual conditional choice probability

$$(7) \quad P[y(D) = c \mid x_o, W_{Do}] = P[\varepsilon: v(w_{co}, x_o)\beta + \varepsilon_c \geq v(w_{do}, x_o)\beta + \varepsilon_d, \text{ all } d \in D \mid x_o, W_{Do}].$$

If β and the distribution $P[\varepsilon_d, d \in D \mid x_O, W_O]$ are known, the right-hand-side of (7) is computable. If β and $P[\varepsilon_d, d \in D \mid x_O, W_O]$ are partially known, one may be able to deduce bounds on the counterfactual choice probability; see Manski (1988, 2007b).

Extrapolation using the conditional logit model is particularly simple. Extending the distributional assumption made for $(\varepsilon_d, d \in C)$, assume that the unobserved variables $P(\varepsilon_d, d \in D)$ are independent and identically distributed, each with the Type I extreme-value distribution. Then the counterfactual choice probability (7) has the form

$$(8) \quad P[y(D) = c \mid x_o, W_{Do}] = \frac{\exp[v(w_{co}, x_o)\beta]}{\sum_{d \in D} \exp[v(w_{do}, x_o)\beta]} .$$

Subsequent Development of the Literature

Since the 1970s, the econometric analysis of discrete choice has retained the basic aspects of McFadden's original work - the random utility model framework, the attribute characterization of alternatives and

decision makers, and the resulting conditional choice probabilities. However, the particular distributional assumptions of the conditional logit model have been subjected to close examination and models making many alternative assumptions have been studied. The idea has been to make distributional assumptions that yield computationally tractable conditional choice probabilities, retain point identification of model parameters, and yet are reasonably flexible in their behavioral implications.

Many researchers have studied *multinomial probit* models, which assume that $(\varepsilon_d, d \in C)$ are distributed multivariate normal (e. g., Daganzo, Bouthelier, and Sheffi, 1977; Hausman and Wise, 1978; Lerman and Manski, 1981; Berry, Levinsohn, and Pakes, 1994). Some such models have a *random-coefficients* feature, allowing the parameters β to vary randomly across persons rather than be fixed across the population. And McFadden has generalized the conditional logit model in various ways. In one stream of research, he retained the assumption that each ε_d has an extreme-value distribution but relaxed the assumption that the variables $(\varepsilon_d, d \in C)$ are mutually independent. This yielded the *generalized-extreme-value* model introduced in McFadden (1978) and studied in more depth in McFadden (1981). In another stream, he studied random-coefficient versions of the logit model (McFadden and Train, 2000).

The models described above aim to weaken the assumptions of the conditional logit model. However, they all assume that the distribution of ε is a priori known up to specified parameters, which may be estimated from the available data. Concern with the restrictiveness of any parametric model has led to the development of a branch of discrete choice analysis that imposes nonparametric assumptions that enable partial identification and estimation of β and the distribution of ε. Manski (1988) and Horowitz (1999) review much of this literature.

Appendix B: References

Arcidiacono, P. (2004), "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 121, 343-375.

Ajzen, I. and M. Fishbein (1980), *Understanding Attitudes and Predicting Social Behavior*, Englewood Cliffs: Prentice-Hall.

Berry, S., J. Levinsohn, and A. Pakes (1994), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Betts, J. (1994), "What do Students Know about Wages? Evidence from a Survey of Undergraduates," *Journal of Human Resources*, 31, 27-56.

Beyth-Marom, R. (1982), "How Probable is Probable? A Numerical Translation of Verbal Probability Expressions," *Journal of Forecasting*, 1, 257-269.

Blau F. and M. Ferber (1991), "Career Plans and Expectations of Young Women and Men," *Journal of Human Resources*, 26, 581-607.

Cameron, S. and J. Heckman (2001), "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109, 455-499.

Curtin, R. (1982), "Indicators of Consumer Behavior: The University of Michigan Surveys of Consumers," *Public Opinion Quarterly*, 46, 340-352.

Daganzo, C., F. Bouthelier, and Y. Sheffi (1977), "Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm," *Transportation Science*, 11, 338-358.

Davis J. and T. Smith (1994), *The General Social Surveys, 1972-1994, Cumulative File*, National Opinion Research Center, Chicago.

Dominitz, J. (1998), "Earnings Expectations, Revisions, and Realizations," *Review of Economics and Statistics*, 80, 374-388.

Dominitz, J. and C. Manski (1996), "Eliciting Student Expectations of the Returns to Schooling," *Journal of Human Resources*, 31, 1-26.

Dominitz, J. and C. Manski (1997a), "Using Expectations Data to Study Subjective Income Expectations," *Journal of the American Statistical Association*, 92, 855-867.

Dominitz, J. and C. Manski (1997b), "Perceptions of Economic Insecurity: Evidence from the Survey of Economic Expectations," *Public Opinion Quarterly*, 61, 261-287.

Dominitz, J. and C. Manski (2004), "How Should We Measure Consumer Confidence?" *Journal of Economic Perspectives*, 18

Dominitz, J, C. Manski, and B. Fischhoff (2001), "Who are Youth *At-Risk*?: Expectations Evidence in the NLSY-97," in R. Michael (editor), *Social Awakenings: Adolescents' Behavior as Adulthood Approaches*, New York: Russell Sage Foundation, 230-257.

Eckstein, Z. and K. Wolpin (1999), "Why Youths Drop out of High School: the Impact of Preferences, Opportunities, and Abilities," *Econometrica*, 67, 1295–1339.

El-Gamal, M. and D. Grether (1995), "Are People Bayesian? Uncovering Behavioral Strategies," *Journal of the American Statistical Association*," 90, 1137-1145.

Erev, I. and B. Cohen (1990), "Verbal versus Numerical Probabilities: Efficiency, Biases, and the Preference Paradox," *Organizational Behavior and Human Decision Processes*, 45, 1-18.

Ferrell, W. and P. McGoey (1980), "A Model of Calibration for Subjective Probabilities," *Organizational Behavior and Human Performance*, 26, 32-53.

Fischhoff, B., A. Parker, W. Bruine de Bruin, J. Downs, C. Palmgren, R. Dawes, and C. Manski (2000), "Teen Expectations for Significant Life Events," *Public Opinion Quarterly*, 64, 189-205.

Fishbein, M. and I. Ajzen (1975), Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research, Reading: Addison-Wesley.

Freeman, R. (1971), *The Market for College-Trained Manpower*, Cambridge, Mass.: Harvard University Press.

Fuller, W., C. Manski, and D. Wise (1982), "New Evidence on the Economic Determinants of Postsecondary Schooling Choices," *Journal of Human Resources,* 17, 477-498.

Hausman, J. and D. Wise (1978), "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," *Econometrica*, 46, 403-426.

Horowitz, J. (1998), *Semiparametric Methods in Econometrics*, New York: Springer-Verlag.

Hurd, M. (1999), "Anchoring and Acquiescence Bias in Measuring Assets in Household Surveys," *Journal of Risk and Uncertainty*, 19, 111-136.

Hurd, M. (2009), "Subjective Probabilities in Household Surveys," *Annual Review of Economics*, forthcoming.

Hurd, M. and K. McGarry (1995), "Evaluation of the Subjective Probabilities of Survival in the Health and Retirement Study," *Journal of Human Resources*, 30, S268

Juster, T. (1966), "Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design," *Journal of the American Statistical Association*, 61, 658-696.

Juster, T. and R. Suzman (1995), "An Overview of the Health and Retirement Study," *Journal of Human Resources*, 30, S7-S56.

Keane, M. and K. Wolpin (1997), "The Career Decisions of Young Men," *Journal of Political Economy*, 105, 473-522.

Kohn, M., C. Manski, and D. Mundel (1976), "An Empirical Investigation of Factors Which Influence College Going Behavior," *Annals of Economic and Social Measurement*, 5, 391-419.

Koriat, A., S. Lichtenstein, and B. Fischhoff (1980), "Reasons for Confidence," *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107-118.

Lerman, S. and C. Manski (1981), "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in C. Manski and D. McFadden eds. *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, Mass.:M.I.T. Press.

Lichtenstein, S. and R. Newman (1967), "Empirical Scaling of Common Verbal Phrases Associated With Numerical Probabilities," *Psychonomic Science*, 9, 563-564.

Lichtenstein, S., B. Fischhoff, and L. Phillips (1982), "Calibration of Probabilities: The State of the Art to 1980," in D. Kahneman, P. Slovic, and A. Tversky (editors), *Judgment Under Uncertainty: Heuristics and Biases*, New York: Cambridge University Press.

Long. B. (2004), "How have college decisions changed over time? An application of the conditional logistic choice model," *Journal of Econometrics*, 121, 271-296.

Luce, R. and P. Suppes (1965), "Preference, Utility, and Subjective Probability," in R. Luce, R. Bush, and E. Galanter, eds. *Handbook of Mathematical Psychology*, Vol. 3, New York: Wiley.

Manski, C. (1987), "Academic Ability, Earnings, and the Decision to Become a Teacher: Evidence from the National Longitudinal Study of the High School Class of 1972," in D. Wise (editor) *Public Sector Payrolls*, Chicago: University of Chicago Press, 291-312.

Manski. C. (1988), "Identification of Binary Response Models," *Journal of the American Statistical Association*, 83, 729-738.

Manski, C. (1993), "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" in C. Clotfelter and M. Rothschild, eds., *Studies of Supply and Demand in Higher Education*, Chicago: University of Chicago Press.

Manski, C. (1995), *Identification Problems in the Social Sciences*, Cambridge, Mass.: Harvard University Press.

Manski, C. (2004), "Measuring Expectations," *Econometrica*, 72, 1329-1376.

Manski, C. (2007a), *Identification for Prediction and Decision*, Cambridge, Mass.: Harvard University Press.

Manski, C. (2007b), "Partial Identification of Counterfactual Choice Probabilities," *International Economic Review*, 48, 1393-1410.

Manski, C., and D. Wise (1983), *College Choice in America,* Cambridge, Mass.: Harvard University Press

McClelland, A. and F. Bolger (1994), "The Calibration of Subjective Probabilities: Theories and Models 1980-94," in G. Wright and P. Ayton (editors), *Subjective Probability*, New York: Wiley.

McFadden, D. (1974), "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka ed. *Frontiers in Econometrics*, New York: Academic Press.

McFadden, D. (1978), "Modelling the Choice of Residential Location," in A. Karlqvist, L. Lundqvist, F. Snickars, and J. Weibull eds. *Spatial Interaction Theory & Planning Models*, Amsterdam: North-Holland.

McFadden, D. (1981), "Econometric Models of Probabilistic Choice," in C. Manski and D. McFadden eds. *Structural Analysis of Discrete Data with Econometric Applications*, Cambridge, Mass.:M.I.T. Press.

McFadden, D. and K. Train (2000), "Mixed MNL Models for Discrete Response," *Journal of Applied Econometrics*, 15, 447-470.

Morrison, D. (1979), "Purchase Intentions and Purchase Behavior," *Journal of Marketing*, 43, 65-74.

Morgan, G. and M. Henrion (1990), *Uncertainty: A Guide to Dealing With Uncertainty in Quantitative Risk and Policy Analysis*, New York: Cambridge University Press.

Radner, R. and L. Miller (1975), *Demand and Supply in U. S. Higher Education*, Carnegie Commission on Higher Education.

Rouse, C. (2004), "Low Income Students and College Attendance: An Exploration of Income Expectations," Social Science Quarterly, 85, 1299-1317.

Schuman, H. and M. Johnson (1976), "Attitudes and Behavior," *Annual Review of Sociology*, 2, 161-207.

Shuford, E., A. Albert, and H. Massengill (1966), "Admissible Probability Measurement Procedures," *Psychometrika*, 31, 125-145.

Smith, H. and B. Powell (1990), "Great Expectations: Variations in Income Expectations Among College Seniors," *Sociology of Education*, 63, 194-207.

Stinebrickner, T. and R. Stinebrickner (2008), "Learning about Academic Ability and the College Dropout Decision," Department of Economics, University of Western Ontario.

Thurstone, L. (1927), "A Law of Comparative Judgment," *Psychological Review*, 34: 273-286.

Tversky, A. and D. Kahneman (1974), "Judgement Under Uncertainty: Heuristics and Biases," *Science*, 185, 1124– 1131.

Walker, J. (2001), "Adolescents' Expectations Regarding Birth Outcomes: A Comparison of the NLSY79 and NLSY97 Cohorts," in R. Michael (editor), *Social Awakenings: Adolescents' Behavior as Adulthood Approaches*, New York: Russell Sage Foundation, 201-229.

Wallsten, T., D. Budescu, A. Rapoport, R. Zwick, and B. Forsyth (1986), "Measuring the Vague Meanings of Probability Terms," *Journal of Experimental Psychology: General*, 115: 348-363

Wallsten T., D. Budescu, R. Zwick, and S. Kemp (1993), "Preferences and Reasons for Communicating Probabilistic Information in Verbal or Numerical Terms," *Bulletin of the Psychonomic Society*, 31, 135-138.

Willis, R. and Rosen, S. (1979), "Education and Self-Selection," *Journal of Political Economy*, 87, S7- S36.

Wolter, S. (2000), "Wage Expectations: A Comparison of Swiss and U. S. Students," *Kyklos*, 53, 51-69.

Zafar, B. (2008), "College Major Choice and the Gender Gap," Federal Reserve Bank of New York.

Zimmer, A. (1983), "Verbal vs. Numerical Processing of Subjective Probabilities," in R. Scholz (editor) *Decision Making Under Uncertainty*, North-Holland, Amsterdam.

Zimmer, A. (1984), "A Model for the Interpretation of Verbal Predictions," *International Journal of Man- Machine Studies*, 20, 121-134.

Appendix C: Author

Charles F. Manski

Department of Economics and Institute for Policy Research

Northwestern University

Prepared for the NCES Postsecondary Choice Workshop, January 26, 2009, Washington, DC

Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES
TASK FORCE REPORT

EMERGING ISSUES IN POSTSECONDARY ACCESS AND CHOICE

IMPROVING NCES DATA COLLECTION: RESPONSE TO LONG AND MANSKI

Eric Bettinger, PhD

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

IMPROVING NCES DATA COLLECTION:
RESPONSE TO LONG AND MANSKI

INTRODUCTION

NCES data collection is one of the cornerstones of empirical education research throughout the United States. Educational researchers produce dozens and dozens of empirical papers each year utilizing these valuable data. The comments by Manski and Long prevent valuable frameworks which may enhance and guide future data collection efforts.

My response to their thoughtful essays focuses on three key points. First, I emphasize the role of expectations in future data collection. Second, I briefly discuss the appropriateness of behavioral economics as a framework for future data collection. Third, I discuss the integration and use of state administrative data with NCES data collection. I also discuss alternative survey designs which can expand the scope of NCES and allow the flexibility for future changes in NCES data collection.

Expectations

At the heart of any decision to acquire additional education is the comparison of the expected stream of future benefits from the additional education and the costs associated with acquiring it. The key is that individuals project forward and create expectations as to what the future would be like with the new education. This forward projection or expectation remains somewhat of a puzzle to educational researchers.

The expectation of future earnings is conceptually easy for an academic to define, but empirically measuring the expectation is quite difficult. Too often researchers ignore the expectation and simply make comparisons of actual earnings. However, as Manski elaborates, assigning actual earnings assumes an informational base and an accuracy in forecasting that few high school or college students have or are capable of. For example, many students do not know what the future earnings associated with a specific major in college is (Zafar 2008). Additionally, many studies document that the level of information that students have about the college application process and its benefits are deficient, especially in low-income populations (Avery and Kane 2004).

Both Manski and Long argue that gathering data on expectations would greatly improve our ability to understand college access, choice, persistence, and completion. I concur that data on expectations would represent a significant improvement to NCES data on college access and subsequent experiences in college. There are a few considerations that must be taken into account in gathering data on expectations.

First, which expectations matter most? Clearly, Long and Manski argue that expectations of absolute and relative future earnings matters, but there are other expectations that matter. For example, prior work by Manski shows that expectations of individual ability may influence college enrollment decisions. Other areas where expectations could matter include institutional selectivity, affordability, the non-pecuniary

benefits of a college degree, and the costs of college including its non-monetary costs. Additionally, it may not be only the child's expectation that matters but also the parents.

Second, when should we measure expectations? There are several key times for students making decisions about college – registering for college entrance exams, choosing and submitting applications to colleges, weighing competing financial aid packages, completing the FAFSA, and actually attending classes for the first time. At each of these points, students weigh different options and because of the sequential nature of these actions, information and thus expectations may evolve over this period. NCES can either attempt to prioritize which of these expectations is most important, or alternatively NCES can explore alternative methods of data collection where rotating subsamples of students are interviewed more frequently.

Third, economists are just starting to understand how individuals form expectations. Information is certainly important, and both authors accentuate the need to gather better data on the quality of information that students and parents have when they are formulating expectations. Additionally, formulating expectations requires an ability to imagine the future and make trade-offs in the future. However, we know that many populations make inter-temporal decisions that appear irrational or inconsistent. One study showed that almost 30 percent of low-income adults make such inconsistent decisions (Bettinger and Slonim 2007).

Finally, most of the examples that I have used as well as the examples in Long and Manski focus on college entrance, yet the human capital model can also potentially explain the decisions to attend one or two years of college. Over the last 30 years, access to higher education has increased dramatically, yet degree completion has remained stagnant (e.g. Turner 2004). If NCES is to use the human capital as a framework for future data collection, then it must also extend the model into decisions affecting college persistence. Just as expectations play a part in access, they also play a part in decisions to persist. Hence, gathering data on expectations which sheds light on persistence can help inform public policy and potentially improve the college completion rate.

Behavioral Economics

Behavioral economics has become increasingly popular. There are several situations in which individuals appear to behave non-rationally or behave in a way which seems incongruent with the simplest human capital model. Ultimately, many behaviors in education may be explained by these models. For example, research in behavioral economics has shown why individuals may choose either to under-invest or to invest in efficient portfolios (Choi, Laison, and Madrian 2005). The same type of procrastination might explain why many families do not save for their children's college educations.

While behavioral economics may yet shed light on the nature of individual decision-making in education, it is not clear what role NCES should play in identifying or even changing individual's behaviors. For example, NCES might be helpful in identifying "default" behaviors of individuals (e.g. expectation of 12th grade completion only, decisions to start saving for college); however, most of the interventions aimed at changing the default (e.g. K-14 versus K-12) occur at the local or state levels. NCES can gather information about "default" behavior, but I expect that it will have little variation at the national level. The more interesting variation will occur from state to state and locality to locality, and NCES's sampling strategies may make it difficult to identify such variation.

Behavioral economics might provide some motivation to understanding discount rates and risk preferences; however, measuring discount rates and risk preferences is a developing science. The methods for measuring discount rates and risk preferences rely on a battery of questions, and while these measures may be important, it may be difficult to allocate time to these measures in a student interview. The length of time that interviewers have with students is quite small, and additional questions will crowd out other topics. Even if we can measure discount rates through a small number of questions, it is not clear that one measure of discount rates accurately predicts individual's time preferences in every setting.

Behavioral economics may help shed light on multiple phenomena in education; and lessons from behavioral economics may ultimately suggest new policies which can improve educational outcomes. However, NCES's role in identifying these lessons may be minimal given the local nature of educational interventions aimed at changing behavioral "defaults."

Changes in Data Collection

Over the last decade, the ability of school districts and states to track student-level data has dramatically increased. Many states (e.g. Florida, Texas) have the capacity to track students from pre-school through the attainment of advanced college degrees. Additionally, the National Student Clearinghouse currently tracks 90 percent of all higher education enrollments and degree completions throughout the country, and they continue to expand their coverage.

These administrative datasets may provide additional opportunities to expand NCES databases. For example, NCES currently collects substantial data on individual's college enrollments, college transcripts, time to degree, and degree completion. In the future, administrative records might be able to provide much of these data. Removing these items would allow surveyors to focus more extensively on the types of variables that may not be available through administrative sources. Administrative records might provide a wealth of data on prospective college student backgrounds. We could conceivably identify all of the classes, grades, peers, and teachers with whom students interacted throughout their primary and secondary school careers. The sample sizes available in these datasets dwarf the samples that we can gather in their absence.

Clearly there are limitations to state administrative data. Currently, there is insufficient representation across states, and oftentimes, state administrative data bases lack significant variables which would aid in any empirical analysis. However, over time these databases will improve and may provide opportunities to partner with states and to build capacity and scale NCES has already started to take advantage of some of these collaborative opportunities and is well placed to continue using these data in the future.

Additionally, there are other ways in which NCES might be able to expand scale. One way to expand the number of questions is to gather data more frequently for subsamples of the survey sample. For example, NCES could administer the primary survey instruments to the overall survey frame. NCES could then select random subsamples to have specific follow-up questions. These groups could be on a rotating basis so that the individual burden on a specific student is not too large. Yet the rotation across subsamples would allow data collection at more frequent intervals for a wider scope of data. One rotation group could even receive an "experimental" module which would allow NCES to identify and to test new questions and metrics.

Finally, traditionally NCES has pursued a school-based approach to drawing the sample. For many NCES instruments, this might be the optimum way to continue to operate. However, one growing trend in higher

education is the participation of non-traditional students. These students are typically older and have significant lifetime experiences relative to traditional college students. NCES current sampling makes it difficult to track and identify these students. Given their increased representation in higher education, serious consideration is warranted to identify ways to track these students' higher educational experiences and outcomes.

Conclusion

NCES data collection has long informed educational policy. As NCES evaluates how to improve data collection, lessons from the human capital model, especially the roles of expectation and information deserve increased attention. Additionally, cooperation and collaboration with states may enable NCES to test new models of data collection and to greatly expand the scope of NCES data collection through the integration of new state administrative databases.

APPENDICES

Appendix A: References

Appendix B: Author

Appendix A: References

Avery, Christopher and Thomas J. Kane (2004) "Student Perceptions of College Opportunities: The Boston COACH Program" in *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, ed. Caroline Hoxby. U of Chicago Press: Chicago. Pp. 355-391.

Bettinger, Eric and Robert Slonim (2007), "Patience in Children: Evidence from Experimental Economics." *Journal of Public Economics* 91(1-2): 343-363.

Choi, James J., David Laibson, and Brigitte C. Madrian. (2005) "$100 Bills on the Sidewalk: Suboptimal Investment in 401(k) Plans." NBER Working Paper No. 11554.

Turner, Sarah (2004), "Going to College and Finishing College: Explaining Different Educational Outcomes" in *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, ed. Caroline Hoxby. U of Chicago Press: Chicago. Pp. 13-56.

Zafar, Basit, "College Major Choice: Experimentation and Revisions to Expectations." Northwestern University Mimeo.

Appendix B: Author

Eric Bettinger, PhD

Associate Professor of Education, Stanford University School of Education

Faculty Research Fellow, National Bureau of Economic Research

Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES TASK FORCE REPORT

EMERGING ISSUES IN POSTSECONDARY ACCESS & CHOICE

NCES DATA COLLECTIONS: INFORMATION AND GAPS

Jack Buckley, PhD

NCES DATA COLLECTIONS: INFORMATION AND GAPS

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

NCES DATA COLLECTIONS: INFORMATION AND GAPS

INTRODUCTION

As the Federal statistical agency for education, the National Center for Education Statistics (NCES) plays a major role in the collection and analysis of data on all aspects of postsecondary education. One area of recurring interest to the agency is the collection of data to facilitate better understanding postsecondary choice and answer policy relevant questions both descriptively - how do both traditional and non-traditional college students choose whether and where to attend? - and counterfactually - what are the likely consequences of a change in Federal loan or grant policy?

As part of a process of improving and modernizing its data collection strategies around postsecondary decision making, NCES commissioned the National Institute for Statistical Sciences (NISS) to assemble a small panel on the subject. Two papers were commissioned - "Data Collection for Econometric Analysis of Schooling Decisions," by Charles F. Manski, and "Emerging Issues in Postsecondary Access and Choice," by Bridget Terry Long. Both authors are prominent economists who study, *inter alia*, issues of post-secondary choice and access in the U.S. and their papers each provide several specific suggestions that can help NCES improve collection and analysis in this area.

In this brief review, I will examine these two papers with the goal of highlighting points that I believe are particularly salient to NCES. One of the tensions felt by all agencies in the Federal statistical community is the requirement to balance the needs of many diverse stakeholders in the data collection and analysis processes, including policy makers, academics, interest groups, and the general public. As a result many Federal data collections strive to be "general purpose" products that seek to provide data that are not wedded to any particular academic discipline or theoretical perspective. This seems to be particularly the case in education, where the research questions and approaches of psychologists, sociologists, economists, psychometricians, educators, educational institutions, government officials, and others often compete for scarce space in questionnaires and assessments. Accordingly, my review is not limited to the discussion of the papers in the context of economic theory and econometrics, but instead seeks to put the suggestions of the authors in a broader context.

I begin the review by briefly discussing the various sources of data collected by NCES relevant to the postsecondary choice process. I then turn to a review of the theoretical frameworks the two authors explicitly or implicitly use as the foundation of their discussion. Following this, I discuss the empirical concerns, including modeling and measurement, raised by Manski and Long. I conclude with some recommendations for NCES based on the ideas put forward in the two papers.

Relevant Data at NCES

As both authors acknowledge, postsecondary choice is a complex decision problem where decision makers students and families - are potentially constrained and influenced by a variety of actors, including teachers, peers, lenders, several levels of government, the media, and the postsecondary institutions themselves (Figure 1). Any theoretical framework must address some set of these actors, but most theoretical and empirical examinations of postsecondary choice examine only a limited subset.

Since historically NCES has not collected data motivated by any one theoretical perspective, the various datasets that address the issue of postsecondary choice tend to measure factors that cut across several of the actors illustrated in Figure 1. For example, the High School and Beyond (HS&B) longitudinal survey that tracked 10th graders for ten years beginning in 1980 asked students about their own expectations for postsecondary attendance in 10th and 12th grade (as well as retrospectively in grades 6-9) but also asked them what was expected of them from parents, teachers, guidance counselors, and peers. Additionally, HS&B collected some data on the importance of various criteria used in their choice of postsecondary institution such as selectivity, reputation, job placement, religious environment, sector, and crime rate. The students were also asked about the availability of financial aid.

In subsequent secondary longitudinal studies - the National Educational Longitudinal Study (NELS) of 1988 and the Educational Longitudinal Study (ELS) of 2002, NCES expanded measurement of the various factors thought to be linked to postsecondary choice. The NELS added items to attempt to measure the extent of parental involvement in the college search process as well as items to expand the attributes of the alternatives considered by students and the families. Questions were also added to measure more precisely the availability of financial aid, the amount of information parents had about the aid process, and some details about parent's aid information search strategies. The ELS study kept most of the preceding items and, importantly, expanded the amount of detail on Federal financial aid (Stafford, Perkins, and Plus loans and Pell grants) by matching student records to the National Student Loan Data System (NSLDS).

NCES also collects data relevant for understanding choice of postsecondary institution in the Beginning Postsecondary Students (BPS) longitudinal studies, begun in 1990, 1996, and 2003. Because the sampling frame for these studies is the National Postsecondary Students Aid Study (NPSAS) conducted in the preceding year, the BPS data files have extensive data on aid, including funds from non-Federal sources. Moreover, unlike the secondary longitudinal studies, the BPS series samples from *all* beginning postsecondary students - including the nontraditional students who are not members of the immediately entering high school cohort. The BPS data are limited with respect to the secondary studies, however, in that the sample is restricted to successful college entrants.

There are some other NCES data collections that contain information relevant to postsecondary choice (e.g. the adult education modules of the National Household Educational Survey measures awareness of education tax credits and adult enrollment, and the Baccalaureate and Beyond studies collect data on graduate education), but the high school longitudinals and the BPS are the most important sources of data within NCES's portfolio.
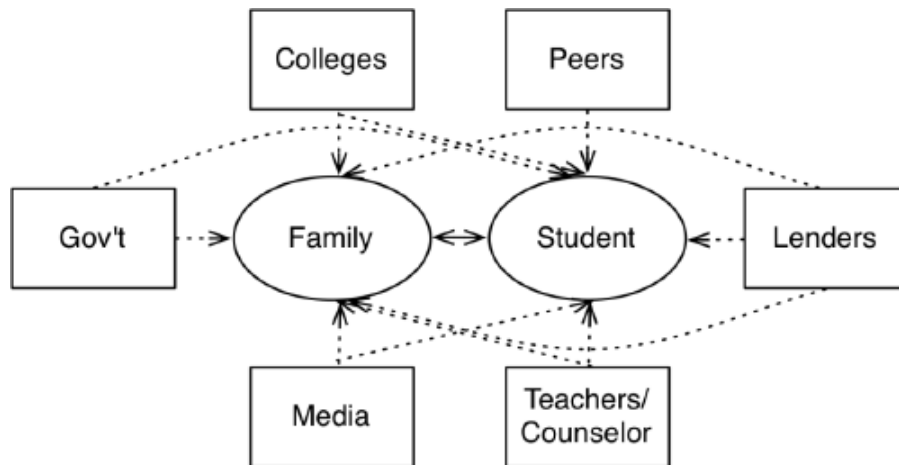
Figure 1: A stylized representation of the actors influencing the decision makers (students and families) in the postsecondary choice process. Dashed lines denote financial or informational influence.

How People Choose

Before discussing specific gaps or measurement problems in the NCES data, Long and Manski both introduce a theoretical foundation for discussion. Long situates the postsecondary education choice problem in two theoretical frameworks: the human capital model of economics and, to a lesser extent, the growing subfield of behavioral economics. Manski appears generally more concerned with empirical issues, but also frames his observations in the human capital literature and, to a much larger extent than Long, in the deep interdisciplinary research in decision theory that includes behavioral economics as a subset. As I shall discuss below, it is difficult to disentangle elements of each framework. More sophisticated formulations of the human capital model, for example, include many elements of behavioral decision theory.

As an aside, I should point out that both Manski and Long characterize the postsecondary choice question as a *decision* problem: a student or family (as I discuss below, Manski makes an interesting point about the limitations placed on analyses due to the standard assumption of a unitary decision maker in the economics of college choice literature) is assumed to be the locus of a decision, with budget constraints, payoffs, and information either assumed to be exogenous or else partially incorporated into the model. Neither author discusses the college choice problem as a *game*, however, in which the decisions of other student/families and of postsecondary institutions as strategic actors must be considered. One way in which the college choice process has been considered in the game theory literature is as a "two-sided matching problem" (Mumcu and Saglam 2007; Gale and Shapley 1962). Others consider the college choice as a "signaling" game between students and firms in the labor market (Spence 1973). Either theoretical approach would be likely to have consequences for data collection. However even when considered solely as a decision problem, the choice of postsecondary institution remains a rich and complex area of study.

In the past fifty years there has been a great deal of interdisciplinary scholarship concerned with the processes of judgment and decision making, producing well over a dozen distinct theoretical approaches. Beach and Mitchell (1998) divide these competing theories of decision analysis into three major categories: normative models, behavioral decision theory, and naturalistic decision theory.

Normative models, such as expected utility decision theory (Keeney and Raiffa 1976; Von Neumann and Morgenstern 1943), subjective expected utility decision theory (Edwards 1954) and the analytic hierarchy process (Saaty 1986), are prescriptive, focusing primarily on how decisions *should* be made. Grounded in mathematical theories of probability and economic theories of utility, the normative models are often described as "elegant" and are frequently used by formal theorists.

Despite the attraction of these normative models and their widespread application in fields as diverse as sociology, anthropology, political science and, of course, economics, many researchers have taken issue with their cognitive demands and their Herculean assumptions. Perhaps the most cogent criticisms are built on the seminal work of political scientist Herbert Simon (Simon 1955; Simon 1957; Simon 1978; Simon 1985). In particular, Simon's concept of "bounded rationality" is widely recognized as providing the foundation for what is now termed behavioral decision theory.

In general terms, behavioral decision theory seeks to apply empirical psychological findings concerning cognitive limits, shortcuts and sub-optimal judgment to the normative models discussed above. Prominent approaches include judgment analysis (Cooksey 1996; Hammond et al. 1975); heuristics (Tversky and Kahneman 1974); prospect theory, (Kahneman and Tversky 1979; Kahneman et al. 1982); search for dominance structure (SDS) theory (Montgomery 1983); fuzzy decision theory (Zadeh et al. 1975; Smithson 1987) and differentiation and consolidation theory (Svenson 1999). While these theories modify, sometimes extensively, the assumptions of the normative models, they often retain at least a partially normative focus on prescriptions for "good" decision making.

Beach and Mitchell's third category, naturalistic decision theory, consists of theories of judgment and decision making derived from observed decision behavior, not normative ideas about what such behavior ought to be. These naturalistic theories are rooted in cognitive science; they are relatively "messy" and often intractable by simple formal analysis. The category includes image theory (Beach and Mitchell 1987; Beach and Mitchell 1998); recognition-primed decision theory (Klein 1993); and conflict/constraint theory (Janis 1989; Janis and Mann 1977).

The core of the human capital model, as summarized by Long, is a cost-benefit calculation: decision makers weigh the costs of postsecondary education (both directly and in terms of foregone wages) against the benefits - the returns to schooling - adjusted by a discount rate. Thus, especially in the most traditional formulations that both Long and Manski review, the human capital model fits squarely in the normative tradition of research on decision making: decision makers are assumed to have full and accurate information on costs and benefits and to use this information in a utility maximizing framework.

As Manski notes, it is likely that this classical human capital model, as with any model that assumes full information, sets the standard too high for most student and their families (and indeed for most actors) to meet. As March observes:

> Time and capabilities for attention are limited. Not everything can be attended to at once. Too many signals are received. Too many things are relevant to a decision. Because of those limitations, theories of decision making are often better described as theories of attention or search than as theories of choice. They are concerned with the way in which scarce attention is allocated. (March 1994, 10).

March's position is congruent with many of the empirical findings of behavioral decision theory and the recent growth of research in behavioral economics noted by Long. In particular, the work of Gabaix and

Laibson (2005; Gabaix et al. 2006) on directed cognition theory demonstrates experimentally that the process of decision making is better explained by a theoretical model that explicitly includes an information search phase and/or a "meta-rational" model in which decision makers first choose the degree of rationality to be employed in the decision process based on a separate cost/benefit calculation (see also Payne 1976; Payne et al. 1993; Kahneman and Tversky 1979; Beach 1990; Svenson 1999; and Tversky 1972).

As I note above, the various data relevant to postsecondary choice that are currently collected by NCES do not appear intended to test any particular set of hypotheses generated within either a classical utility maximization formulation of the human capital model or a more behavioral, search-driven model of decision making. To be sure, there are data elements that fit into either framework: net price, family financial information, student levels of preparation, student and family levels of information about postsecondary options, to name a few. However, as both Long and Manski point out, even many of these measures are flawed or limited (more on this below). The challenge for NCES is to produce data that is congenial to several theoretical frameworks without alienating any particular set of stakeholders. I concur with Manski and Long that economics and the broader behavioral decision theory literature are excellent choices to consider as a foundation for this task.

Measurement, Modeling, and Data Considerations

Manksi and Long each provide a sizeable list of specific issues in measurement or data collection and, to a lesser extent, in modeling. As many of these provide significant opportunity for NCES to improve its data collection in this area, in this section I will review several of their points in detail.

Manski

- *Locus of Decision*. As noted above, and as pointed out by Manski, the postsecondary choice literature in economics generally assumes a unitary decision maker. Anecdotal evidence suggests that this is unlikely to be true and it is possible that the assumption, when mistaken, has consequences for the validity of inference about choice. NCES's secondary longitudinal collections could provide a relatively easy way to measure to assess the relative influence of the actors in Figure 1 (above) on the decision process. As Manski notes, it would also be useful to measure the locus of decision over time, and the secondary longitudinal studies would facilitate this as well. For non-traditional students, retrospective items could be added to the BPS (or perhaps in the NHES), but the data would be more limited.

- *Expectations and Elicitation*. Manski points out that there is little empirical evidence about how students (and families) perceive the returns to schooling, a crucial input in modeling choice under the human capital framework. Based on his own work, he recommends that NCES measure these expected returns and that the uncertainty in these expectations be assessed as well (following the example of the 1997 NLSY). Moreover, he suggests a specific elicitation strategy for expectations of both schooling and income, both unconditionally and under certain education scenarios. He also suggests the elicitation of probabilistic beliefs about key facts that are relevant to the decision problem. I believe his proposed approach is feasible for NCES and would be worth the cost in terms of development and questionnaire time. Space permitting (or perhaps with a subsample), I would also recommend extending the elicitation of probabilistic expectations and knowledge to parents as well. These innovations would keep NCES data relevant and useful in this area in light of advances in the survey literature and in the behavioral decision theory framework discussed above.

- *Interpersonal Comparability*. Although he raises the issue only as an aside during his brief review of research on attitudes, Manski's concern over whether responses to attitude items can be naively compared between persons is an important area that has been largely overlooked by NCES and I believe that this issue warrants closer consideration. Many empirical studies of choice (including existing work in postsecondary choice and other educational decision making) attempt to identify important differences in attitudes and perceptions and to link these differences to choices, search behavior, choice set construction or other proximal outcomes to the decision. Unfortunately there is an increasingly large body of evidence that suggests that many observed differences at the interpersonal or subgroup levels are, in fact, often the result of artifacts of measurement (Javaras and Ripley, 2007; King et al., 2004; Rossi et al., 2001; Baumgartner and Steenkamp, 2001; Heine et al., 2000; de Vijver and Leung, 1997; Chen et al., 1995; Mullen, 1995; Greenleaf, 1992; Poortinga, 1989). Much of this research focuses particularly on cross-cultural differences in the usage of Likert scales or individual categorical items drawn from such scales. Baumgartner and Steenkamp (2001) provide a useful summary of the various *response styles* or differences in response scale usage that can lead to bias in cross-cultural attitude research. Although individual respondents' idiosyncratic usage of different response styles adds noise to attitude survey data, systematic differences in response style across subgroups can introduce far more serious biases in both descriptive statistics and inferential results from more complex models. For example, Chen et al. (1995) report that Chinese and Japanese secondary students are more likely to use the midpoint of a seven-point Likert-type item while U.S. students exhibit a greater tendency toward extreme responses than the Asian students or their Canadian counterparts, although they find little effect on cross-national comparisons of item means. Watkins and Cheung (1995) examine response styles of high school students from five countries and report substantial variation in the tendency to exhibit several response styles, including extreme response style and noncontingent (or random) response, on academic self-esteem items. Marin et al. (1992) compare U.S. Hispanics to non-Hispanic Whites and find a greater incidence of several response styles (e.g. acquiescence response bias) among the Hispanic population, particularly the less educated and less acculturated. Bachman and O'Malley (1984) find similar results comparing Black to White respondents. These findings, particularly the research on secondary school populations, suggest that heterogeneity in response style could be a potential source of bias in the secondary analysis of NCES data on postsecondary choice. While there is a growing literature introducing technical "fixes" for this problem (Javaras and Ripley, 2007; King et al., 2004; Rossi et al., 2001) and NCES's international studies are investigating several of these methods, bias can often be reduced through simple measures such as including reverse-scaled items to construct balanced scales and the removal of items indicating susceptibility to scale-usage heterogeneity or differential item functioning. NCES should continue to conduct research in both areas.

- *Econometric Modeling*. Manski provides a helpful summary of both the alternative specific conditional logit, or McFadden's choice model, frequently used in applied econometric studies of discrete choice and some of the extensions and generalizations of this model. While I am aware of few applications of McFadden's choice model in NCES reporting (although there have been some official reports published based on estimation of multinomial logit models with attributes of the choosers but not the choices), NCES should strongly consider applying more advanced methodology in future reports on postsecondary choice, keeping in mind the need for published reports to be based on models considering survey design, weighting, and complex variance estimation issues.

Long

- *Measuring Cost*. As discussed above, cost (and the perception of cost) is a central input to the human capital model of educational choice. Long, building on her own work and others', notes that problems with the measurement of the cost of postsecondary education to families may account for recent findings that decision makers appear less sensitive to cost. In particular, she points out the need for better understanding of how students and their parents think about loans and about the disutility of accruing debt to finance education. She also points out the importance of distance (correctly measured) as a cost factor. I would hypothesize that distance appears particularly important for non-traditional students and students in non-selective two- and four-year institutions. Long also points out that, for low-income families, student income (including perhaps even education loans and grants) may not be safely assumed to be spent on education, and she recommends better measurement of the budget constraints of the traditionally disadvantaged.

- *Measuring Benefits*. Echoing Manski, Long argues that postsecondary decision makers consider the benefits of continuing education probabilistically - and, thus, data on expectations should elicit this uncertainty. Long also notes that the current datasets provide very little information regarding either the risk preferences or the discount rates of students and families regarding postsecondary education, although measurement may be problematic.

- *Timing of Information*. Long notes that, despite their longitudinal nature, the NCES secondary surveys do not generally ask questions about students' information and perception of postsecondary options until late in their secondary education (usually senior year) - thus researchers are unable to measure change over time at the individual level. Truly longitudinal data on levels and sources of postsecondary information would enable researchers to disentangle the cross-sectional endogeneity between information and motivation. Long also makes a case for expanded measurement of student interaction with guidance counselors (something that NCES considered for the forthcoming High School Longitudinal Study [HSLS] of 2009) and of peer group interaction and aspirations to continuing education. She also notes that it would be useful to determine via student surveys the accuracy of student perceptions of requirements for postsecondary acceptance.

Recommendations to NCES

I conclude with a series of recommendations for NCES drawn from the authors' papers and my discussion above. The first set of recommendations is for changes to data collection activities in existing surveys or families of surveys. Although it is almost certainly too late for any broad changes to the first wave of the HSLS survey instruments or design, NCES will have an opportunity to consider adopting some of these recommendations in later waves, subsequent secondary longitudinal surveys, or other data collection activities. I have attempted to order my recommendations by ease of implementation (easier to more difficult).

Collection Recommendations

1. Add items to the student and parent survey instruments of secondary or postsecondary longitudinal studies to enable better measurement of the locus of decision either prospectively (secondary studies) or retrospectively (postsecondary). This would require a modest amount of questionnaire time and any items added would be straightforward.

2. Measure the amount, quality, and source of information about postsecondary options in a truly longitudinal way (i.e. repeated measures) in the secondary studies. This would most probably involve only a slight expansion/alteration of existing measures, but would obviously require repetition over panel waves. Another alternative is Long's proposal to take advantage of variation in the timing of completed responses using the survey operations paradata.

3. Improve the measurement of expectations by adding survey items designed to explicitly measure the education and income expectations of students (and parents for their students) at several time points of a secondary longitudinal study. These measures could also be truly longitudinal (per 2, above).

4. Improve measurement of the cost/perception of cost to students and families by including more nuanced measures of distance to alternatives (these would likely be derived variables) and the budget constraints facing families, particularly low-income families.

5. Pilot the large-scale elicitation of probabilistic beliefs over both expectations and key facts influencing the postsecondary decision problem. This recommendation is related to item 7, below.

I also suggest the following recommendations pertaining to analysis, modeling, or research (again in ascending order of estimated complexity):

Research/Modeling Recommendations

1. Investigate potential biases in descriptive or inferential statistics introduced by interpersonal or subgroup variation in response styles or differential item functioning for items relevant to postsecondary choice.

2. To support the elicitation of probabilistic information on facts and expectations, conduct small- scale research (cognitive labs, small pilot surveys) to develop and test items and strategies. This research should also include an examination of the process of expectation formation. Begin with the methodology proposed in the Manski paper.

3. Build internal capacity for discrete choice modeling via the alternative specific conditional logit model and related methods, and also work on modifying methods to allow for the consideration of design and poststratification weights and complex variance estimation.

APPENDICES

Appendix A: References

Appendix B: Author

Appendix A: References

Bachman, J. G. and O'Malley, P. M. 1984. "Yea-Saying, Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles." *Public Opinion Quarterly*, 48(2):491–509.

Baumgartner, H. and Steenkamp, J. E. M. (2001). "Response styles in Marketing Research: A Cross-National Investigation." *Journal of Marketing Research*, 38(2):143–156.

Beach, Lee Roy. 1990. *Image Theory: Decision Making in Personal and Organizational Contexts*. Chichester, England: Wiley.

Beach, Lee Roy and Terence R. Mitchell. 1987. "Image Theory: Principles, Goals and Plans in Decision Making." *Acta Psychologica* 66:201-20.

Beach, Lee Roy and Terrence R. Mitchell. 1998. "The Basics of Image Theory." in *Image Theory: Theoretical and Empirical Foundations*, edited by L. R. Beach. Mahwah, NJ: Lawrence Erlbaum Assoc.

Chen, C., Lee, S., and Stevenson, H. W. 1995. "Response Style and Cross-Cultural Comparisons of Rating Scales Among East Asian and North American Students." *Psychological Science*, 6(3):170–175.

Cooksey, Ray W. 1996. *Judgment Analysis: Theory, Methods and Applications*. San Diego, Calif.: Academic Press.

de Vijver, F. J. R. V. and Leung, K. 1997. "Methods and Data Analysis of Comparative Research." In Berry, J. W., Poortinga, Y. H., and Pandey, J., editors, *Handbook of Cross-Cultural Psychology, Volume 1: Theory and Method*, pages 257–300. Allyn and Bacon, Boston, Mass.

Edwards, Ward. 1954. "The Theory of Decision Making." *Psychological Bulletin* 51:380-417.

Gabaix, Xavier and David Laibson. 2005. "Bounded Rationality and Directed Cognition." Unpublished manuscript.

Gabaix, Xavier, David Laibson, Guillermo Moloche, and Stephen Weinberg. 2006. "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model." *American Economic Review* 96(4): 1043-68.

Gale, D. and Shapley, L.S. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69(1): 9-15.

Greenleaf, Eric A. 1992. "Measuring Extreme Response Style." *Public Opinion Quarterly*, 56(Fall): 328–361.

Hammond, Kenneth R., Thomas R. Stewart, Berndt Brehmer, and Do Steinmann. 1975. "Social Judgment Theory." *In Human Judgment and Decision Processes*, ed. M. Kaplan and S. Schwartz. New York: Academic Press.

Janis, Irving L. 1989. *Crucial Decisions: Leadership in Policymaking and Crisis Management*. New York: Free Press.

Janis, Irving L., and Leon Mann. 1977. *Decision Making: A Psychological Theory of Conflict, Choice and Commitment*. New York: Free Press.

Javaras, K. N. and Ripley, B. D. 2007. "An 'Unfolding' Latent Variable Model for Likert Attitude Data: Drawing Inferences Adjusted for Response Style." *Journal of the American Statistical Association*, 102(478): 454–463.

Kahneman, Daniel, Paul Slovic and Amos Tversky. 1982. *Judgment under Uncertainty: Heuristics and Biases*. New York: Cambridge University Press.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47:263- 91.

Keeney, Ralph L., and Howard Raiffa. 1976. *Decisions with Multiple Objectives: Preferences and Values Tradeoffs*. New York: John Wiley.

King, G., Murray, C. J. L., Salomon, J. A., and Tandon, A. 2004. "Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research." *American Political Science Review*, 44(April): 341–355.

Klein, Gary A. 1993. "A Recognition-Primed Decision (RPD) Model of Rapid Decision Making." *In Decision Making in Action: Models and Methods*, ed. G. A. Klein, R. Calderwood, J. Orasanu, and C. E. Zsambok. Norwood, N.J.: Ablex.

March, James G. 1994. *A Primer on Decision Making: How Decisions Happen*. New York: Free Press.

Montgomery, H. 1983. "Decision Rules and the Search for Dominance Structure: Towards a Process Model of Decision Making." In *Analyzing and Aiding Decision Processes*, ed. P. C. Humphreys, O. Svenson, and A. Vari. Amsterdam: North-Holland.

Marin, G., Gamba, R. J., and Marin, B. V. 1992. "Extreme Response Style and Acquiescence Among Hispanics: The Role of Acculturation and Education." *Journal of Cross-Cultural Psychology*, 23(4): 498–509.

Mullen, M. (1995). "Diagnosing Measurement Equivalence in Cross-National Research." *Journal of International Business Studies*, 26: 573–596.

Mumcu, Ayse, and Ismail Saglam. 2007. "College Decisions Under Early Admission." Munich Personal RePEc Archive, Paper no. 1906. http://mpra.ub.uni-muenchen.de/1906/

Payne, John. 1976. "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis." *Organizational Behavior and Human Performance* 16:366-87.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. New York: Cambridge University Press.

Poortinga, Y. P. 1989. "Equivalence of Cross-Cultural Data: An Overview of Basic Issues." *International Journal of Psychology*, 24:737–756.

Rossi, P. E., Gilula, Z., and Allenby, G. M. 2001. "Overcoming Scale Usage Heterogeneity: A Bayesian Hierarchical Approach." *Journal of the American Statistical Association*, 96(453): 20–31.

Saaty, Thomas L. 1986. "Axiomatic Foundation of the Analytic Hierarchy Process." *Management Scienc*e 32:841-53.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics* 69:99-118.

Simon, Herbert A. 1957. Models of Man. New York: John Wiley.

Simon, Herbert A. 1978. "Rationality as Process and as Product of Thought." *American Economics Review* 68:1-16.

Simon, Herbert A. 1985. "Human Nature in Politics: the Dialogue of Psychology and Political Science." *American Political Science Review* 79:293-304.

Smithson, Michael. 1987. *Fuzzy Set Analysis for Behavioral and Social Sciences*. New York: Springer-Verlag.

Spence, A. M. 1973. "Job Market Signaling." *Quarterly Journal of Economics*, 87:355-374.

Svenson, Ola. 1999. "Differentiation and Consolidation Theory: Decision Making Processes Before and After a Choice." In *Judgment and Decision Making: Neo-Brunswikian and Process-Tracing Approaches*, ed. P. Juslin and H. Montgomery. Mahwah, N.J.: Lawrence Erlbaum.

Tversky, Amos. 1972. "Elimination by Aspects: A Theory of Choice." *Psychological Review* 79:281-99.

Tversky, Amos and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185:1124-31.

Von Neumann, John, and Oskar Morgenstern. 1943. *Theory of Games and Economic Behavior*. New York: John Wiley.

Watkins, D. and Cheung, S. 1995. "Culture, Gender, and Response Bias: An Analysis of Responses to the Self-Description Questionnaire." *Journal of Cross-Cultural Psychology*, 26(5): 490.

Zadeh, Lofti A., King-Sun Fu, Kazau Tanaka, and Masamichi Shimura. 1975. *Fuzzy Sets and their Applications to Cognitive and Decision Processes*. New York: Academic Press.

Appendix B: Author

Jack Buckley, PhD

Research Associate Professor, Applied Statistics, Social Science, and Humanities, Steinhardt School of Culture, Education, and Human Development, New York University

Institute Fellow, American Institutes for Research

Deputy Commissioner, National Center for Education Statistics, Apr 2006 – Jun 2008

NATIONAL INSTITUTE OF STATISTICAL SCIENCES TASK FORCE REPORT

EMERGING ISSUES IN POSTSECONDARY ACCESS AND CHOICE

CONCEPTUALIZING POSTSECONDARY ACCESS AND CHOICE: THE ROLE OF NCES DATASETS

Laura W. Perna, PhD

CONCEPTUALIZING POSTSECONDARY ACCESS AND CHOICE: THE ROLE OF NCES DATASETS

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

CONCEPTUALIZING POSTSECONDARY ACCESS AND CHOICE: THE ROLE OF NCES DATASETS

INTRODUCTION

The NCES datasets have undoubtedly played a critical role in advancing knowledge of postsecondary access and choice. These datasets are especially important for understanding two persisting problems: (1) the need to improve postsecondary educational attainment all students; and (2) the need to eliminate gaps in attainment across groups of students, particularly gaps by race/ethnicity and family income.

Among the well-known strengths of these datasets are the careful sampling strategies, large sample sizes, high response rates, and large number of available variables. Despite the strengths, however, the datasets are not without limitations. In their papers, Bridget Terry Long and Charles Manski offer thoughtful and insightful suggestions for adapting NCES data collections to improve economic analyses of postsecondary access and choice.

In this paper, I offer comments on the two papers and identify a few other issues for consideration. These comments focus on addressing two goals of this workshop:

1. Address key current and emerging issues in postsecondary choice process from multidisciplinary perspective; and
2. Construct conceptual framework that informs NCES' secondary and postsecondary data collections, especially longitudinal studies.

Contributions from Long and Manski

As might be expected given the expertise of the authors, both Long and Manski conceptualize postsecondary access and choice from an economic perspective. Manski emphasizes discrete choice analysis, while Long focuses on human capital theory and behavioral economics. A substantial body of research uses an economic approach to college-enrollment decision-making (see Perna, 2006 for one review).

The papers describe the relevant frameworks and identify challenges associated with using these frameworks to understand postsecondary access and choice. Both human capital models and discrete choice analysis emphasize that college-related decisions are based on a comparison of the expected benefits with the expected costs. The expected benefits include both monetary and non-monetary benefits, while the expected costs include the costs of attendance and foregone earnings. Human capital theory predicts that calculations of expected benefits and costs are influenced by an individual's academic preparation for college and availability of resources to pay the costs of attendance.

Together, the papers identify important limitations in available measures for key constructs of the human capital model. Long argues that improvements are needed in the measurement of expected costs, particularly college prices and financial aid, willingness to use loans to pay college prices, distance to a

higher education institution, and availability of family resources to pay college prices. Both Long and Manski discuss the need for better measures of expected returns to schooling and the need to consider students' perceptions of academic requirements, college prices, and financial aid.

As the authors note, improving measures of expectations requires attention to several issues. One is types of expectations. Manski focuses on developing better measures of expected returns to schooling. Other potentially important expectations are the expectation that loans are required to pay college prices, as well as the expectation that a student will complete a degree program and realize the earnings needed to repay the loans.

The two papers also offer insights for considering ways to collect data on expectations. One challenge is to frame questions in ways that avoid the tendency of students to report the answers that they believe are socially appropriate or valued (Avery & Kane, 2004; Long). Manski raises another challenge, namely the need to consider certainty of expectations, and discusses several approaches to measuring expectations.

Both Long and Manski also describe the need to know more about students' college related knowledge and perceptions. "Information" is one potentially useful lever for promoting college access and choice. The vast number of websites, public awareness campaigns (e.g., KnowHow2Go), publications, books, etc. suggest the perceived utility of improving students' knowledge of college prices, financial aid, and academic requirements for college access and success. Nonetheless, despite substantial investments in information, numerous reports document that students and their parents typically lack accurate knowledge and understanding of college prices and financial aid (see Perna, 2004 for a review).

A Multi-Disciplinary Perspective

In addition to identifying better measures of the components of a human capital model, more attention should also be given to the forces that contribute to these components. For example, in order to understand how to change students' expectations, we need to know more about how different groups of students develop these expectations. Developing a better understanding of these forces will also likely improve our understanding of the "anomalies" in students' college-related decisions identified by Long and others (e.g., Avery & Hoxby, 2004; Heller, 1997).

A rational human capital investment model assumes that, even when the expected benefits and costs are the same, two individuals may make different college choices because of differences in their preferences, tolerance for risk, and uncertainty (DesJardins & Toutkoushian, 2005). But, human capital theory does not explore why preferences, tastes, and expectations differ across groups (DesJardins & Toutkoushian, 2005).

Therefore, although useful for conceptualizing the effects of costs and benefits on students' college enrollment behaviors, traditional human capital approaches alone are insufficient for understanding inconsistencies in college-enrollment decision-making or how students develop perceptions or understandings of financial aid. Drawing on other frameworks provides additional insights.

For example, both behavioral economics and psychological approaches to decision making assume that, because of informational and computational constraints on information processing capacities, individuals adopt such strategies as satisficing or bounded rationality (Hogarth, 1987; Long, 2009). Thus, because of cognitive limitations, individuals make decisions that are bounded by time and resource constraints.

Drawing on a sociological perspective, McDonough (1997) uses Bourdieu's notion of habitus to illustrate the ways that bounded rationality influences college-related decision making. This perspective assumes that college-related decisions reflect an individual's habitus, or the internalized system of thoughts, beliefs, and perceptions that is acquired from the immediate environment. Rather than consider all possible alternatives, habitus defines and limits the alternatives that are considered, how different alternatives are perceived and valued, and the choices that are made (McDonough, 1997; Paulsen & St. John, 2002). These alternatives are defined and delimited by the social, organizational, and cultural contexts in which individuals are embedded, including class-based values and the organizational habitus of high school attended (McDonough, 1997). According to McDonough, the notion of organizational habitus considers both the availability of resources and structures within a high school to promote college enrollment, as well as the underlying norms and expectations for college enrollment at a school, norms and expectations that reflect class-based values about college-going.

Among other conclusions, these theories, and the research that tests these theories, suggest the ways that various aspects of "context" frame and define postsecondary access and choice decisions. More specifically, as suggested in Figure 1, students make postsecondary access and choice decisions that are largely determined by the norms and values embedded within a student's family, peers, and high school, as well as influences from the higher education and broader economic, social, and policy context (Perna, 2006). Greater attention to context may help identify the structures and processes that define and limit students' college-related behaviors. These understandings, in turn, may suggest ways to improve the construction, implementation, and/or marketing of programs so as to more effectively promote postsecondary enrollment and attainment for all students.

One application of this framework pertains to students' perceptions of loans and willingness to use loans to pay college prices. An economic perspective predicts that students consider loans in terms of a benefit-cost analysis. Specifically, students are expected to consider the costs of borrowing in the context of their expectations that they will complete their program and realize some expected earnings premium. But, economic theories alone are insufficient for understanding differences in students' perceptions of loans and willingness to borrow. Sociological frameworks, with their attention to social, cultural, and contextual differences, offer additional insights (e.g., Cunningham & Santiago, 2008; Perna, 2008). Among the potentially relevant aspects of context are a family's prior experience with loans (mentioned by Long), as well as culturally-specific perceptions of debt and availability of school-based financial aid counseling to provide information about loans. Perceptions of loans may also vary based on students' perceptions of other resources to pay college prices, particularly work (Gladieux & Perna, 2005). In addition, willingness to borrow may vary based on the state context, particularly a student's expectation that state grant aid will be available to help pay college prices.

Recommendations for NCES Data Collections

Based on my review of the Long and Manski papers and other perspectives, I offer four recommendations for NCES data collections: (1) collect data to understand how students acquire and use information about college; (2) develop better measures of students' postsecondary options; (3) measure non-traditional pathways to degree completion; and (4) consider data collection and sampling issues.

Collect Data to Understand How Students Acquire and Use Information about College

As both Long and Manski suggest, future NCES data collections should focus on generating a better understanding of how students acquire and use information about college. The underlying question that such data collection should address is: What types of knowledge "matter" (i.e., promote various college-enrollment behaviors), at what points in time, for which groups of students? If "information" is to serve as an effective policy lever for increasing college access and choice, then we need to know more about whether information can promote changes in student behavior and, if so, what kinds of information different groups of students need to have at different points in the pathway to and through college (Mundel & Coles, 2004).

Attention to the timing of college-related information is particularly important, as we currently know very little about whether information causes students to engage in college-related behavior or whether students who engage in college-related behaviors acquire information (Long; Perna, 2004). Knowing more about the timing of information will enable analyses of how information about college may promote a range of college-related outcomes, including not only enrollment but also other outcomes including aspirations and academic preparation.

In order to identify potential vehicles for improving knowledge, NCES data collections should also provide more complete measures of the sources of information that students use. The conceptual framework (Figure 1) suggests that potential sources of information may include other family members (e.g., older siblings and extended family members), high school staff, peers and other students attending the same high school, pre- college outreach programs (e.g., Upward Bound), colleges and universities, and federal and state sponsors of student financial aid. Available research includes limited attention to only some of these potential sources. Moreover, understanding the roles of these different sources requires attention not only to whether a student used a particular source (yes or no), but also to the quality and quantity of information that each source provides.

One source of information is a student's family. Both Long and Manski note the need for better measures of the relative roles of students and parents in college-enrollment decisions (including payment of college prices). Measures of "parental involvement" should also be refined to better reflect the growing diversity of family structures as well as the role of older siblings and members of the extended family (Tierney & Auerbach, 2005).

More complete measures of the quality and quantity of information available from a student's school are also required. Little is now known about high school counselors' college related knowledge, given the absence of relevant national data (McDonough, 2004). Additional measures are also required to understand how, and with what consequences, other school staff may be providing financial aid counseling to students. In the absence of sufficient financial aid counseling from school counselors, at least some students may be relying on information from teachers and coaches to learn about financial aid (Perna, Rowan-Kenyon, Thomas, Anderson, & Li, 2008).

Measures of other sources of college-related information are also needed. In particular, little is known about the ways that colleges and universities influence students' understanding and use of financial aid. More information is required about the full range of postsecondary educational options that students face, including programs offered by for-profit providers, community colleges, and other colleges and universities. Little is known about the ways that postsecondary educational institutions contribute to the structuring of

college opportunity, by "marketing" financial aid to students or working with high schools to provide information about college and financial aid (Mundel, 2008; Perna, 2004; Perna, Rowan-Kenyon, Thomas et al., 2008). Better measures are also required to understand how college-related programs that are sponsored by federal, state, and local providers affect students' college-related knowledge and behaviors.

Develop Better Measures of Postsecondary Options

As described above, various theoretical perspectives predict that students consider only a sub-set of potential postsecondary options, and that these options are constrained and defined by the context in which students are situated. Therefore, in order to know how to encourage more students to pursue and attain postsecondary education, better data are needed about the options faced by students, especially students from groups that are underrepresented in higher education. These options include working, working and attending college, delaying college entry, and entering the military. As suggested by the conceptual model, additional measures of the context in which students are embedded are also needed to understand the postsecondary choices that students actually consider.

Measure Pathways to Degree Completion, Including Non-Traditional Pathways

Policymakers are interested in improving not only postsecondary access and choice but also persistence to degree completion. NCES data collections should be designed to recognize that access is just one step toward the longer-term goal of program or degree completion, and that, increasingly, students are engaging in "non- traditional" college-related behaviors en route to program and degree completion. Modeling the educational pathways of individuals who delay entry, enroll part-time, and stop out is certainly complex. Nonetheless, few sources of longitudinal data exist to understand the many policy-relevant questions about non-traditional educational pathways.

Consider Other Data Collection and Sampling Issues

By definition, identifying ways to improve educational attainment for underrepresented groups of students requires sufficient numbers of students for analysis. In order to understand the role of the school context for college-going, NCES data collections should ensure sufficient numbers of students attending each school. Small numbers of students per school may produce samples of students that are not representative of the population of students attending the school and reduce the statistical power to detect variations across schools. Sampling strategies should also ensure sufficient numbers of students of color at each school so as to model variations in the ways that school structures and characteristics contribute to racial/ethnic differences in enrollment across schools (e.g., Perna & Titus, 2005).

NCES data collections are also an important resource for considering how differences in the state context for college-going influence the postsecondary options that students consider, as well as other aspects of students' college-related behaviors (e.g., Perna & Titus, 2004). Therefore, future data collections should include students attending schools in all 50 states, in order to maximize the number of cases at the state level (for multi-leveling modeling).

In addition, NCES data collections should continue efforts to link NCES data collections with other sources that describe various aspects of context (e.g., administrative data, state datasets, other NCES datasets). Providing opportunities to link with existing data sources reduces the data collection burden on students and other individuals and increases the quality and quantity of available measures.

Finally, NCES should consider timing various data collection efforts to overlap with the collection of other data by both NCES and other sources (e.g., Census). This overlap would enable researchers to draw on other sources of data to complement data from any particular survey.

Conclusion

These papers offer useful insights and appropriate suggestions for enhancing the NCES data collections. As Long and Manski suggest, the NCES datasets can and should provide data that generate more complete understandings of students' college-related knowledge and expectations. The datasets should also include more attention to measuring the forces that may contribute to the formation of knowledge and expectations, as well as more complete measurement of the postsecondary options that students consider and the context that delimits these options. Although human capital models offer important insights into college-going decisions, other theoretical perspectives also inform how expectations and knowledge are formed and help understand "anomalies" in college-related decision-making.

APPENDICES

Appendix A: References

Appendix B: Figure 1 - Proposed Conceptual Model

Appendix C: Author

Appendix A: References

Avery, C., & Kane, T. J. (2004). Student perceptions of college opportunities: The Boston COACH program. In C. M. Hoxby (Ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It* (pp. 355-394). Chicago: University of Chicago Press.

Cunningham, A. F., & Santiago, D. A. (2008). *Student Aversion to Borrowing: Who Borrows and Who Doesn't*. Washington, DC: Institute for Higher Education Policy.

DesJardins, S.L., & Toutkoushian, R. K. (2005). Are students really rational? The development of rational thought and its application to student choice. . In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research*, Volume 20 (pp. 191-240). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Gladieux, L., & Perna, L. (2005). *Borrowers Who Drop Out: A Neglected Aspect of the College Student Loan Trend*. San Jose, CA: National Center for Public Policy and Higher Education.

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

Hogarth, R. M. (1987). *Judgement and Choice: The Psychology of Decision* (2nd Edition). Chicago: John Wiley & Sons.

McDonough, P. M. (1997). *Choosing Colleges: How Social Class and Schools Structure Opportunity*. Albany: State University of New York Press.

McDonough, P. M. (2004). *The Impact of Advice on Price: Evidence from Research*. Boston, MA: TERI.

Mundel, D. (2008). What do we know about the impact of grants to college students? In S. Baum, M. McPherson, & P. Steele (Eds.), *The Effectiveness of Student Aid Policies: What the Research Tells Us* (pp. 9-38). Washington, DC: The College Board.

Mundel, D. S., & Coles, A. S. (2004). *Summary Project Report: An Exploration of What We Know About the Formation and Impact of Perceptions of College Prices, Student Aid, and the Affordability of College-Going and a Prospectus for Future Research*. Boston, MA: TERI.

Paulsen, M., & St. John, E. P. (2002). Social class and college costs: Examining the financial nexus between college choice and persistence. *Journal of Higher Education*, 73, 189236.

Perna, L. W. (2004). *Impact of Student Aid Program Design, Operations, and Marketing on the Formation of Family College-Going Plans and Resulting College-Going Behaviors of Potential Students*. Boston, MA: The Education Resources Institute, Inc. (TERI).

Perna, L. W. (2006). Studying college choice: A proposed conceptual model. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research*, Vol. XXI (pp. 99-157). Springer.

Perna, L. W. (2008). Understanding high school students' willingness to borrow to pay college prices. *Research in Higher Education*, 49, 589-606.

Perna, L. W., Rowan-Kenyon, H., Thomas, S. L., Bell, A., Anderson, R., & Li, C. (2008). The role of college counseling in shaping college opportunity: Variations across high schools. *Review of Higher Education*, 31, 131-160.

Perna, L.W., & Titus, M. (2004). Understanding differences in the choice of college attended: The role of state public policies. *Review of Higher Education*. 27(4) 501-525.

Perna, L.W., & Titus, M. (2005). The relationship between parental involvement as social capital and college enrollment: An examination of racial/ethnic group differences. *Journal of Higher Education*, 76, 485-518.

Tierney, W. G., & Auerbach, S. (2005). Toward developing an untapped resource: The role of families in college preparation. In W. G. Tierney, Z. Corwin, & J. E. Colyar (Eds.), *Preparing for College: Nine Elements of Effective Outreach*. Albany, NY: State University of New York Press.
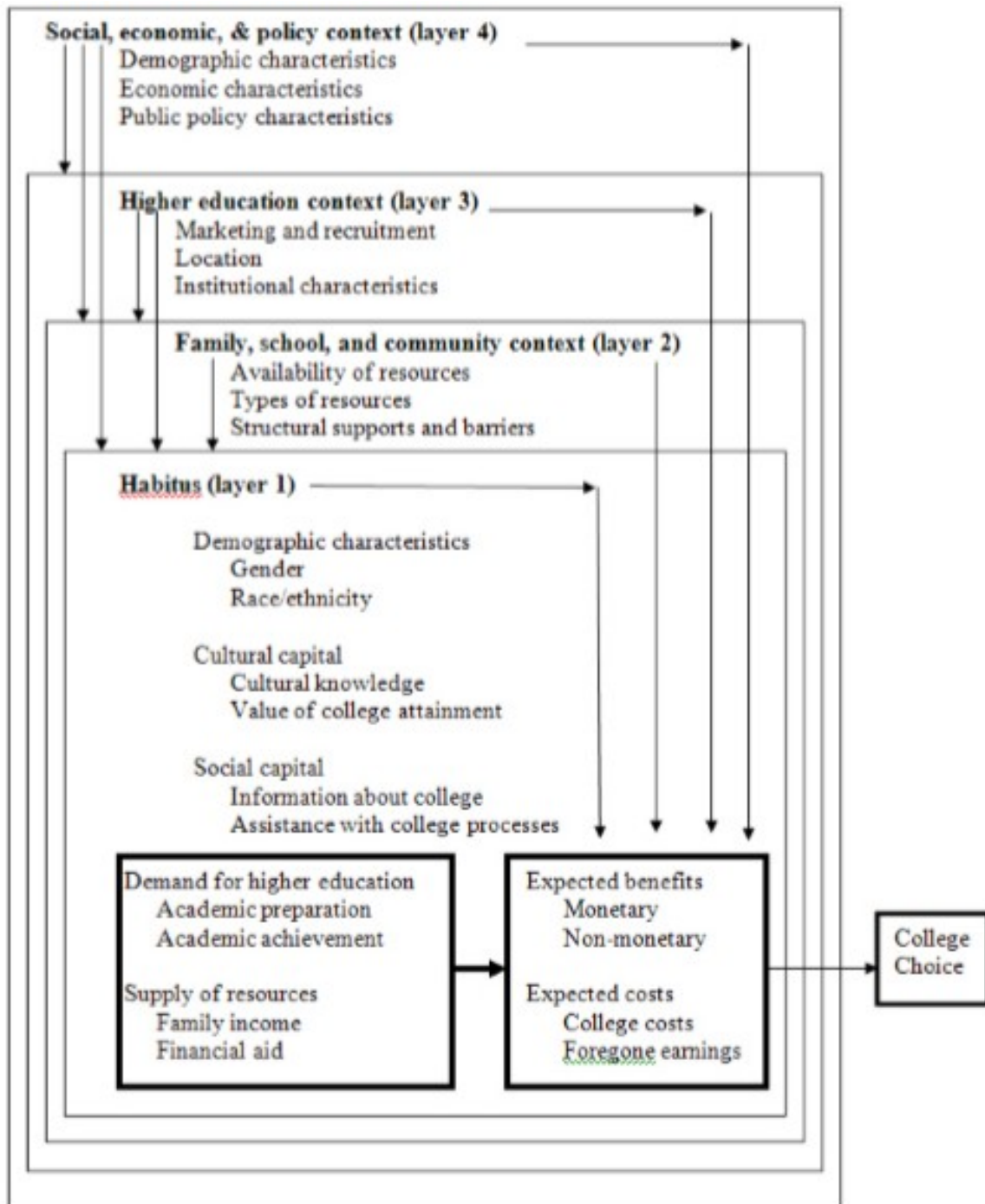
Appendix B: Figure 1 - Proposed Conceptual Model

Figure 1. Proposed Conceptual Model of Student College Choice
Source: Perna (2006).

Appendix C: Author

Laura W. Perna, PhD

Professor, University of Pennsylvania