AERA Open
April-June 2016, Vol. 2, No. 2, pp. 1–13
DOI: 10.1177/2332858416634686
© The Authors 2016. http://ero.sagepub.com

A Future for the National Education Longitudinal Program

Barbara Schneider
Guan Saw
Michigan State University
Michael Broda
Virginia Commonwealth University

Forty years ago, the National Center for Education Statistics initiated the national longitudinal studies program in response to congressional concern for policy-relevant information on school-to-work transitions. This program has grown substantially, and not unexpectedly, questions have arisen about its usefulness and present operation. This essay briefly describes the scope of the program and its benefits for informing knowledge and policy, including (a) racial and socioeconomic inequalities and their influence on the academic and social development of student subgroups, (b) differential schooling experiences and their impact on children's later educational and occupational outcomes, and (c) new methods for estimating statistical inferences. Despite these benefits, the program needs serious modification especially with respect to its organization and design features. Several substantive and methodological recommendations are made for improving the program's effectiveness, including widening technical expertise; reframing sample populations; adjusting intervals for data collections; embedding experiments and conducting field studies; linking longitudinal panels with other federal, state, and local collections; and developing new measures and forms of delivery.

Keywords: *National Center for Education Statistics, generalizable samples, measuring schooling transitions*

Questions on how to improve the federal government's national longitudinal studies program in the National Center for Education Statistics (NCES) have been the topic of several policy reports (e.g., Grissmer & Ross, 2000; National Academy of Education, 2014). The major concern has been how to increase the effectiveness of the longitudinal program for monitoring the achievement and socioeconomic gaps in student performance and career choices from adolescence into adulthood. Given the time and labor costs of conducting longitudinal studies and the possibilities of harnessing data from other multiple collections, recently the issue of improvement has escalated to one of relevance and usefulness (Warren, 2015). Some justifiably assert that the benefits of the longitudinal program for producing causal claims regarding policies and practices are methodologically limited (Schneider et al., 2007). Despite these criticisms, the value of these data sets should not be underestimated. The longitudinal panels are particularly suitable for studying the associations among family, school, and community contexts on education achievement and occupational outcomes, for which other collections provide only partial coverage. This, however, does not constitute a blanket endorsement of the program. For the program to retain its effectiveness, several organization and design features require major modifications.

The National Education Longitudinal Studies Program: A Brief Synopsis

Nearly half a century ago, NCES began conducting a series of large-scale longitudinal studies in response to federal requests for policy-relevant research on the schooling experiences and school-to-work transitions of a nationally representative sample of young people (Ingels, 2004). The first federal national longitudinal program was the National Longitudinal Study of 1972 (NLS-72), which included a representative sample of U.S. high school seniors who were followed as they transitioned from graduation into the labor market and postsecondary education (Ingels, 2002). Since then, the program has grown considerably by expanding the age of the cohorts and incorporating new survey respondents, constructs, and supplemental data sets that could be linked with each respective collection.

Today, these various age cohorts can be categorized by their school-level transitions, including (a) birth to kindergarten entry; (b) kindergarten through middle school; (c) middle school through high school, postsecondary school, and the labor market; and (d) postsecondary education and the transition to work and careers. Table A1 (see appendix) presents some basic information about each of 12 national student

Creative Commons CC-BY: This article is distributed under the terms of the Creative Commons Attribution 3.0 License (http://www.creativecommons.org/licenses/by/3.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

longitudinal studies. In all of these data sets, there are unique problems with each of them, and there are the problems common to all of them. In the postsecondary data sets, for example, it is difficult to trace back to the students' experiences in high school, and we know that the quality of one's earlier education affects later outcomes. In the kindergarten studies, there are no plans to follow these children into adulthood. This is an issue because researchers and policymakers could use such follow-ups to learn about the longer impact of early childhood education on educational attainment and labor market. There are trade-offs as finding individuals over long periods of time and launching new waves can be costly, yet these two examples, it seems, are ones that should receive a high priority.

Most recently, the program expanded its focus from children by adding a longitudinal study of teachers, the Beginning Teacher Longitudinal Study (BTLS), which follows teachers originally interviewed in an earlier cross-sectional survey (the 2007-2008 Schools and Staffing Survey [SASS]). Each of these longitudinal cohorts has scheduled collections within them, typically every 2 years over an initial 6-year period with 10-year follow-ups (see http://nces.ed.gov/surveys/ for exact dates on reinterviews for each of the studies).

The budget for the longitudinal programs is relatively modest compared to the management and administration of other cross-sectional and administrative data collections (e.g., the yearly budget for the administration of the National Assessment of Educational Progress [NAEP] is $137 million (including the governing board). A single round of a longitudinal sample can cost anywhere from approximately $15 million to $25 million over a 3- to 5-year cycle (NCES, 2015), including the pilot, survey administration, and subsequent data cleaning, coding, and descriptive reports (policy reporting is not included in NCES's mission). Overall, the federal funding for the NCES currently stands at about $280 million annually, most of which is administered via competitive contracts to subcontractors (U.S. Department of Education [USED], 2015). This represents approximately 49% of the overall 2015 discretionary funding for the Institute of Education Sciences (IES), which houses NCES, and about 0.6% of the 2015 non-Pell Grant discretionary budget for the USED. For the 2016 fiscal year, the USED has requested an additional $21 million for NCES statistical activities (USED, 2015). Below is a brief description of several student longitudinal studies and the types of policy-relevant questions they have been designed to examine.

Original Longitudinal Collections: High School Through Postsecondary Education and the Labor Market

There are now five high school longitudinal studies, the most recent the High School Longitudinal Study of 2009 (HSLS:09). Earlier initiatives include the Education Longitudinal Study of 2002 (ELS:2002), the National Education Longitudinal Study of 1988 (NELS:88), High School and Beyond (HS&B), and the NLS-72. All five of these studies collect information on student background characteristics, aspirations, behaviors, attitudes, and peer group characteristics. Additionally, they all contain school questionnaires, parent surveys, cognitive assessments, high school transcripts, postsecondary attendance and completion, and labor market information. Collections beginning with HS&B were expanded to include teacher and counselor surveys and data links to a variety of administrative records, including but not limited to Census (e.g., geocode data), Common Core of Data, School District Data Book, Quality Education Data, Integrated Postsecondary Education Data System (IPEDS), and student postsecondary application and loan sources (e.g., National Student Loan Data System and Free Application for Federal Student Aid [FAFSA]).

Each of the high school studies focuses on a particular theme, reflecting changes in the nature of education itself or federal interests at the time (Ingels, 2004). NLS-72 examined the transition into the labor market, whereas HS&B looked at the performance of different groups of students in different types of schools (e.g., public versus private and religious schools). NELS:88 focused on the transition from eighth grade through high school into college and the labor market, targeting at what age students were most likely to leave school (i.e., developing new terminology to account for "stop outs" and other variations on school leaving and returning). ELS:2002 covered topics of educational technology, responding to the widespread introduction of computers into the teaching and learning environments, although the survey itself was administered with pencil and paper. HSLS:09 began with ninth graders, followed them through 11th grade, and administered an exit survey at the end of 12th grade centering on their interest and career preferences for work in science, technology, engineering, and mathematics (STEM).

Several researchers (Warren, 2015; Phillips, 2000) have raised concerns that the longitudinal studies do not follow students in short enough intervals for capturing test score and demographic information, but this is not accurate. Follow-up collections to base-year surveys are typically repeated every 2 to 3 years until postsecondary education, where there have been longer lapses. The early childhood studies have also tracked students over both semesters and years. Most recently, there has been a new initiative to follow up with the HS&B sophomore and senior cohorts who are now over 50 years old (Muller, Grodsky, Warren, & Black, 2015). These extended collections tend to be dependent on federal interests, availability of funds, and investigator initiatives.

Early Childhood: Birth Through Middle School

The early childhood longitudinal studies are relatively new and were initiated in the late 1990s. The first was the Early Childhood Longitudinal Study, Kindergarten Class of

1998-99 (ECLS-K), which surveyed students, teachers, and parents regarding the transitions from kindergarten through eighth grade. The Birth Cohort (ECLS-B) collection includes those students born in 2001 who were followed through kindergarten entry. The ECLS-B is one of the richest data sets we have on a longitudinal sample of young children transitioning into formal schooling. The scope of this data set is exceptional even when comparing it to other data sets, such as the National Longitudinal Study of Youth 1979 (NLSY79) Child and Young Adult data. The Kindergarten Class of 2010-2011 (ECLS-K:2011) is ongoing and second in the series of longitudinal studies of young children. These studies comprise direct assessments of children's competencies and skills as well as indirect measurements of their learning process/products and socioemotional status (as rated by their parents/caregivers and teachers) at multiple time points throughout the survey periods. Developmental and social psychologists view many of the measures as incomparable with respect to their construct and criterion validity for studying life course trajectories in contrast to those in other longitudinal studies (Roisman & Fraley, 2006; Willoughby, Blair, Wirth, & Greenberg, 2012).

Postsecondary Experiences Into the Labor Market

The first of this series, the Beginning Postsecondary Student Longitudinal Study (BPS), surveys first-time undergraduate students at the end of Years 1, 3, and 6 after being enrolled in a postsecondary program. BPS collects information on student demographic characteristics, school and work experiences, and college persistence, transfer, and completion. A second set of postsecondary longitudinal studies, the Baccalaureate and Beyond Longitudinal Study (B&B), examines students' education and work experiences after they obtain a bachelor's degree, with a special focus on the experiences of graduates who become teachers in elementary and secondary schools. Both BPS and B&B draw their initial cohorts from the National Postsecondary Student Aid Study (NPSAS), which uses a nationally representative sample of postsecondary students and institutions to examine how students finance their postsecondary education. The three postsecondary data sets now have three cohorts of first-time undergraduates who started their college career in 1989-1990, 1995-1996, and 2003-2004, respectively. These data sets are of major interest to economists, university personnel, and policymakers—all of whom are concerned about the rising costs of college, increases in student debt burden, and labor market projections for different types of postsecondary degrees.

Why We Use NCES Longitudinal Data Sets

There are several compelling reasons why the NCES data are especially useful for studying the conditions that influence students' lives in and out of school. First, NCES data sets provide nationally representative samples of students at multiple schooling periods, with oversamples of particular minority populations, such as Hispanic and Asian American and Pacific Islander students. These oversamples of different groups have produced multiple studies on the educational trajectories of Hispanics (Callahan & Muller, 2013), Asians (Han, 2008), and first-generation students (Chen & Carroll, 2005). Second, NCES data sets collectively offer standardized test scores that are comparable within and across cohorts. For example, the mathematics test battery of NELS:88 shared sufficient common items both across and within grade-level forms with the test battery of HS&B, enabling procedures for vertical equating as well as cross-sectional equating with prior and later student cohorts of NCES longitudinal studies. The achievement tests have also been especially valuable in assessing sex differences in subjects, such as science and math, across decades (Hedges & Nowell, 1995). Third, for those researchers who study socioeconomic disparities in academic achievement, NCES longitudinal data sets are specifically valuable as they typically collect a comprehensive set of family background variables, such as parental education, occupation, and household economic resources. Other national data sets, such as NAEP, or international studies, such as the Trends in International Math and Science Study (TIMSS) and the Program of International Student Assessment (PISA), contain only limited measures on family resources. And fourth, these data sets have captured new national trends in student transitions, such as the numbers and student characteristics of those entering community college and their subsequent degree attainment and entrance into the labor market, compared to those students who leave high school without a diploma or enter 4-year institutions (Calcagno, Crosta, Bailey, & Jenkins, 2007; Kalogrides & Grodsky, 2011).

Sample Selection

All of the NCES national longitudinal studies are based on similar sample designs that can provide nationally representative data across race/ethnicity groups and, in some cases, school types (e.g., public vs. private high schools). It is important to underscore the exact collection periods of these longitudinal surveys as they often occur when it is possible to compare student performance with other national cross-sectional data sets, such as NAEP. The period of these collections can also be linked to statewide collections, such as the Statewide Longitudinal Data Systems (SLDS), and urban school district collections, like those administered by the Consortium on Chicago School Research.

One other type of analysis that has been overlooked is in comparing national estimates on specific variables of interest with results obtained from purposive samples of a limited number of schools or geographical areas. More specifically, with a national probability sample, researchers are able to

determine how representative a local purposive sample is in comparison to national longitudinal studies designed to obtain inferences about the U.S. student population. For example, Mulligan, Schneider, and Wolfe (2005) compared the Alfred P. Sloan Study of Youth and Social Development with NELS:88 and Current Population Study; Arum and Roksa (2011) compared their sample of college students with BPS.

Data Measures and Comparability

It has been suggested that the SLDS data sets might be a good alternative to the longitudinal studies program as they can overcome selective age and grade issues, since information on students and their teachers is obtained every year. However, the quality of each state's data system is highly variable, and the student information tends to be strictly behavioral (Conaway, Keesler, & Schwartz, 2015). Although SLDS may be especially useful for monitoring student achievement over time, enrollment patterns, and teacher and course histories, it does not help us understand individual student aspirations and their connection with perceptions of self and subsequent actions. Topics such as participation in extracurricular activities, time allocations on homework, and even course designations—such as those taken online, as part of dual-enrollment status, and/or as credit recovery—are often omitted from state data sets. Additionally, most states do not track the students who move from the state, making it difficult to monitor the schooling progression of elementary or secondary students residing in different geographic locations. If the SLDS were the only source for studying many of the questions being explored with NCES longitudinal data sets, major item revisions would have to be undertaken with these collections (which are costly, totaling $34 million in 2014 and 2015).

There are other potential political problems if we move entirely to using the SDLS for studying longitudinal questions. Education in the United States operates under a state-based system. Legislatively, states have the prerogative to add or withdraw additional items beyond what is required for federal funding. Comparability of measures across states would be quite difficult without a federal mandate for participation and standardization. This is highly unlikely to occur given state control and national politics.

Depending on localized collections for longitudinal information also poses many problems, especially in giving access and critical information about the design and instruments to researchers who are not affiliated with initial and ongoing collections. Similar to states, local districts can refuse participation in new and additional data collections. Implementation of survey administration can also be a hurdle locally, where community members, teachers, and administrators may be unwilling or unable to comply with data collection procedures (particularly as survey collection now relies heavily on the local technological infrastructure).

This lack of transparency and refusal for participation makes it difficult to monitor the quality of data collection at both the state and local levels. This is not to say that the national longitudinal surveys are free from full disclosure of difficulty in securing participation of respondents and nonresponse of particular items. Missing data on critical items are estimated with statistical imputation techniques, and weighting procedures are used to correct for key groups underrepresented in the original sample frame. Where the difference lies is that contractors who collect and code the data are required to identify the methods and items used for data and sample replacement, which can be verified independently. Furthermore, there are specific guidelines for participant nonresponse that need to be addressed and remedied before a data file can be released. Additionally, the national longitudinal data sets are public and can be accessed on the web. There are also detailed procedures should an investigator choose to conduct more specialized analysis with restricted data sets.

Contributions to Research and Policy Discussions

Several state and local longitudinal data sets have produced significant insights into the impact of particular interventions and population trends. These results, however, are not generalizable to the nation as a whole, which is a major advantage of the findings from the national longitudinal studies. Findings from NCES longitudinal studies are often at the center of national, state, and local debates on educational inequalities, preschool effects, college readiness reforms, school organization, and trajectories into college and labor market. Examples of these types of studies are discussed below.

1. Early beginnings and problems of increasing racial and socioeconomic inequality. Two influential early childhood longitudinal studies focusing on relationships between racial and socioeconomic disparities among young students and their subsequent school achievement have been conducted by Fryer and Levitt (2004, 2006) and Reardon (2011, 2013). These studies, highly cited in the research literature and on social media, not only exemplify rigorous scholarship but also support the work of other leading economists, sociologists, and developmental psychologists. Analyzing multiple waves of the ECLS-K data, Fryer and Levitt (2004, 2006) found that the Black-White test score gap for children entering kindergarten can be almost entirely explained by family factors. This gap continues to grow, on average increasing by 0.10 standard deviations per school year. Over time these family factors become less determinant, suggesting that schools play a critical role in the education achievement of ethnic-minority students.

Whereas Fryer and Levitt's work focused on changes in racial test score gaps for a single cohort over time, Reardon (2011, 2013) examined the evolving trends in socioeconomic

achievement gaps across student cohorts for the past four decades using a collection of national data sets (six out of 19 are NCES longitudinal studies). His findings suggest that the socioeconomic achievement gap favoring wealthy children has been substantially growing over the years. In particular, the test score gap between children from high- and low-income families (i.e., 90th percentile vs. 10th percentile of the family income distribution) is roughly 30% to 40% larger among children born in 2001 than among those born 20 to 25 years before 2001.

2. College readiness: Variations in preparation and course taking. Studying curricular experiences across high school student subgroups and cohorts has relevant implications with respect to improving student college readiness and ensuring equality in opportunity to learn. Each year, about half of high school seniors graduate without the minimal requirements needed to apply to a 4-year college (American College Testing, 2010; Greene & Foster, 2003). Graduates who are low income, Black, and Hispanic are particularly less likely to be academically prepared for postsecondary education (Adelman, 2004; Long, Iatarola, & Conger, 2009). Domina and Saldana (2012), for example, analyzed HS&B, NELS:88, and ELS:2002 simultaneously and showed that race, class, and skills inequalities in academic math course completion (i.e., algebra, geometry, algebra II, trigonometry, and precalculus courses) have narrowed over the past three decades, yet unequal distribution of attainment in calculus, the most advanced math course in high school and gatekeeper for competitive 4-year institutions, has persisted.

Having access to more rigorous academic course work among traditionally disadvantaged students, however, has not translated into reduced achievement gaps when students take such courses in segregated schools. Using data from ELS:2002, Riegle-Crumb and Grodsky (2010) found that Hispanic students from low-income households and Black students enrolled in segregated schools fall further behind the math achievement of their White counterparts even when taking more advanced courses. Such results suggest that course titles may mean little with respect to actual course content, and these differences are more likely to occur in racially segregated and lower-income schools.

Large-scale national longitudinal data sets have also played a key role in developing a body of knowledge around the phenomenon of student school leavers. For decades, researchers have advocated for more rigorous and generalizable evidence on student and school characteristics associated with school dropout (Lee & Burkam, 2003; Rumberger, 2011). Initially, this evidence was developed from analysis of HS&B and NLSY79 and focused primarily on student factors, such as race/ethnicity, gender, and family background. However, research using NELS:88 broadened the scope of analysis to include school and community factors (Rumberger & Palardy, 2005).

Large-scale longitudinal data sets, such as NELS:88 and ELS:2002, that tracked individuals who have dropped out of school over time have also facilitated the development of alternative conceptions of dropping out, moving from the traditional notion of leaving school and not returning to more nuanced comparisons, such as leavers versus stayers and returners, and dropping out versus stopping out, as well as a deeper understanding of both the individual and societal consequences of leaving school. Here, the longitudinal perspective is essential, as students who appear to have left during one data collection may often return in later years and can be followed many years beyond their high school graduation.

3. Differential effectiveness of school organizations. In addition to student characteristics and their experiences in school, perhaps one of the most significant studies on the social context of education occurred in HS&B, when Coleman and Hoffer (1987; earlier analyses by Coleman, Hoffer, & Kilgore, 1981, and additional later work by Bryk, Lee, & Holland, 1993) examined the relationship between student academic success and the type of school they attended (i.e., public, private independent, and private religious [Catholic]). Results showed that students in private religious schools had higher standardized test scores than similar students in public schools, and these effects were more pronounced for students from disadvantaged backgrounds. Students in Catholic schools were more likely to complete high school than were similar students in public high schools. Examining data from HS&B 2 years after high school graduation, Coleman and Hoffer also found that Catholic and private school students were more likely to attend postsecondary school than were students of similar backgrounds in public schools. For example, lower-performing students who graduated from private schools were more likely to attend college than were similar students in public schools, suggesting that parents who pay for their children to attend private school are also more willing to pay again for them to attend college.

The school effect difference has been more recently studied with a focus on student achievement in schools of varying economic and social contexts. Benner and Crosnoe (2011), for example, found with ECLS-K that students attending more racially diverse schools perform better on standardized tests than their counterparts with similar achievement levels who enrolled in schools with lower levels of racial/ethnic diversity. Additionally, moving to a deeper level, they found that attending a school with like-race and like-ethnicity peers boosted the positive social and emotional development for all students. One of the major takeaways from these studies is that the value of the national longitudinal data sets rests not just in tracing the lives of the students but in recognizing the variations in context that are strongly related to student academic performance, educational attainment, and social and emotional well-being.

4. Postsecondary educational opportunities. Obtaining a college degree continues to be the single most significant investment young people can make in their future. The number of students now entering postsecondary school continues to increase. However, the high school longitudinal program shows that students from middle- and high-income families are much more likely to enroll in college than students from lower income families. Analyses of NLS-72 data and other sources show, in addition to increasing gaps in postsecondary entry, that gaps can be found for college persistence and completion between children from high- and low-income families (Bailey & Dynarski, 2011). What makes this finding particularly problematic is that individuals who are least likely to expect to enroll in postsecondary school are those that are most likely to benefit from college compared with those who expect to enroll in college (Brand & Xie, 2010). With respect to gender, the once male advantage in college enrollment has reversed. Using longitudinal data, researchers show that this advantage is being driven by larger numbers of females completing high school and immediately attending college after high school graduation (DiPrete & Buchmann, 2013).

Over the past 30 years, the average tuition at a public 4-year college has more than tripled (see Domestic Policy Council & Council of Economic Advisors, 2014). To be able to meet rising tuition and other college costs, 70% of students in public universities take out loans to complete their bachelor's degree. Recent studies showed that the percentage of borrowers who default on their loans has increased over the past 10 years and that those who drop out of college are more likely to default than those who do not (BPS data reported in Domestic Policy Council & Council of Economic Advisors, 2014). For those who default, the consequences can be severe, and the federal government has launched a new program to help students and their families learn about and manage their loan obligations. The postsecondary longitudinal program coupled with high school information poses serious questions regarding educational access and job security, especially for many minority and low-income students. Although much of the most cited postsecondary longitudinal work is descriptive in nature, policymakers, university administrators, and families turn to it to inform their postsecondary decisions and programmatic initiatives, especially during periods of economic uncertainty.

5. Additional benefits of NCES longitudinal data sets. The widespread use of large-scale national longitudinal data sets has also facilitated the development of advanced statistical methods that have wide applicability to other research areas in education and the social sciences. For example, to account for the nested nature of social or developmental settings as captured by most of NCES's education longitudinal data sets, in which students or teachers are nested within classrooms in schools, new statistical approaches were needed. One such approach is hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002), which allows for variance to be partitioned by level (student, school, etc.), thus producing more accurately estimated standard errors. This approach has since been extended to a wide variety of research paradigms, including those focused on measuring individual change, such as cognitive growth (Bryk & Raudenbush, 1987; Hong & Raudenbush, 2005), school and organizational effectiveness (Bryk et al., 1993; Lee & Bryk, 1989), and research synthesis and meta-analysis (Borenstein, Hedges, Higgins, & Rothstein, 2009; Kalaian & Raudenbush, 1994).

The national longitudinal data sets also serve as valuable teaching resources, with several statistics books (Murnane & Willet, 2011; Raudenbush & Bryk, 2002) using the data sets as examples for undergraduate and graduate students. The longitudinal data sets are released with a series of technical reports on sampling and other design components, including descriptive statistics, that facilitate opportunities to replicate analyses. Frequent reports, such as the yearly publication of *The Condition of Education* (http://nces.ed.gov/programs/coe/), serve as a valuable source for researchers interested in checking their baseline descriptive computations.

To encourage the use of these data for advancing knowledge and policy research and to enhance the capacity of scholars working with them, the American Educational Research Association (AERA), in conjunction with the National Science Foundation (NSF) and NCES, has for several decades operated a program that provides competitive small grants, fellowships, and training workshops to scholars in a variety of disciplines to conduct research using quantitative methods with data from longitudinal and other federal data sets sponsored by NSF, NCES, and other federal research agencies. This program has produced hundreds of dissertations and publications in top-tier refereed journals. Some have argued that these studies are not necessarily of high quality, but that does not seem to be the case when compared with studies reported in other social and behavioral science journals (see Schneider, 2008). Certainly in any field there are studies that could be improved. If we are sincerely interested in the value of replication especially for policy-relevant work, secondary analyses of these observational data sets, which often include multiple time points, are a valued resource.

In addition to the AERA programs, NCES also annually hosts free conferences and occasionally offers webinars that provide training and instruction on how to access and use longitudinal data sets. The National Center for Education Research at IES also annually provides funding for research on longitudinal data sets through its Education Research Grants program.

Recommendations for the Future of the National Longitudinal Program

While it may be useful and timely to reaffirm the purpose of the national longitudinal program, it is also important to recognize that the program could be redesigned to

accommodate other topics and that the administration could be improved to increase the value of the program. As others have suggested, there have been recommendations regarding the inclusion of (a) more refined measures, particularly in the area of social and emotional learning (Moore, Lippman, & Ryberg, 2015); (b) social network information and other biological measures (as in the National Longitudinal Study of Adolescent Health; Muller, 2014); (c) linkages to other data sets (Dynarski, 2014); and (d) other forms of survey administration, not only with computerized survey tools but also with other technology that incorporates new types of measurement (Hofmann & Patel, 2015). We agree with these suggestions, which we discuss below, and highlight some other organization and design issues.

Organization Considerations

1. Widen technical expertise. When some of the early longitudinal studies were being conceived, NCES invited researchers from different fields to design the sample, including how participants and schools should be selected, what questions should be asked, and what were the most important policies that the survey could inform. These reports were written before the request for proposals was issued. This seems to be a missing initial step in the design of more recent data programs. Today, once a contract is set, the designated firm, in conjunction with NCES, seeks advice from a group of scholars and some practitioners. This has sometimes resulted in keeping the same advisors on board through different collections for similar age cohorts. If one reviews the external experts listed on NCES technical reports, the overlap among experts involved in advising HS&B follow-ups, NELS:88, ELS:2002, and HSLS:09 item selection and sample design is considerable. In our rapidly changing world, trying to solicit new ideas early on in the process could spur creativity and investment in the surveys by prominent researchers, state administrators familiar with the SLDS, and other technological advisors.

Similarly, developing items for the questionnaires was previously considered a major task and subject to considerable scrutiny within the wider research community—this does not seem to be so much a priority anymore. In the past, a group of educational and sociological researchers were able to successfully negotiate taking the lead on a redesign of the school and teacher questions in NELS:88, even though the contract had already been awarded to a firm. Over the years, significant debates among scholars have been held on which questions would be retained for future trend analyses. While trend analyses are valuable, as shown above, especially in assessing inequality of opportunities over time, it may be time for more than limited modifications. For example, even when the surveys tackle new problems, there are relatively few special reports on specific topics. Trend reports and technical materials appear to follow a predictable routine format. It would seem that with the possibilities now of linking data, the reports could be more inventive in their substance, methodological techniques, and dissemination venues for different audiences.

2. Revisit intervals of data collections. When data collections should occur and who should be surveyed have been a long-standing topic of debate within the research community, with some scholars advocating for more extensive early childhood surveys with frequent test administrations (Phillips, 2000). If the question is solely about measuring change in test scores over specific periods, for individuals in particular groups, in distinctive situations, then this is not the type of collection that fits that purpose (Hauser, 2000). Others have argued for larger samples surveyed more often and followed less frequently, as described in a recent article by Warren (2015). However, a dramatic increase in the number of observations and times of collection could pose serious challenges for data collection, as schools are already overburdened with surveys. Frequently surveying students, who are already subject to multiple testing regimes, tends to aggravate a school's willingness to undergo additional data collections regardless of the purpose or incentives to participate.

This is not to say that these issues are absent from the national longitudinal program; they are not. One of the major questions this program continues to grapple with is how to increase school/family participation and reduce respondent burden, especially among specific populations and social systems of interest in low-resourced and urban schools serving underrepresented populations. We recommend that researchers examine alternative types of incentives for increasing cooperation, including individualized school reports, which provide useful information for school-level decision making, and other bonus programs, such as computers, tablets, and other instructional materials, especially when data collection periods occur within short intervals.

Design Considerations

3. Reframe sample population. The sampling frame for nearly all of the longitudinal studies is based on what is termed a two-stage national probability sampling process. This involves achieving generalizability of the population under investigation and, in some instances, the schools they attend. The difficulty with this situation is that our population shifts are quite dramatic, with rising numbers of minority students now attending U.S. public schools and an increasing number of children who are considered multiracial. Additionally, with the recession and other natural shocks, some of the population centers have shifted, and new urban centers are emerging while others have been significantly decreasing. For the national longitudinal studies to remain representative of our changing population, they may require more frequent information on population shifts than in the past, when the country

as a whole was less transient. But it is not just the issue of generalizability that is problematic.

One of the major shifts in policy has been the widening of public school choice, along with the more systematic identification of schools that are making more of a positive influence on student outcomes than one might have expected. This may warrant a redesign of the traditional two-stage sampling frame, not only depending on information being generated by yearly state samples taken from state longitudinal studies but also oversampling particular topics of interest, such as charter schools, homeschoolers, and so on. The sample design could be more efficient, drawing in additional state school districts that may maximize shifts in urban centers. The most recent high school longitudinal study (i.e., HSLS:09) drew generalizable student samples from 10 states that could be added to the national longitudinal study. This idea of sample augmentation has considerable value, especially in estimating population shifts within states and school districts and specifically within school districts implementing major reforms, such as Chicago, New Orleans, and New York City. While districts could obtain information on all their students, replicating the items used in the national study increases the possibility of standardization for replication and generalizability.

4. Embed experiments and the conduct of field studies. Some have argued that the national longitudinal data are useful for external validity but not internal validity. The argument is that observational data obtained in the national surveys are not the most efficient means of assessing treatment effects of programs or policies (Murnane & Willet, 2011; Shadish, Cook, & Campbell, 2002). This point has been a major concern of statisticians, and there have been considerable advances in statistical procedures to lessen spurious confounds, such as selection effects, by using procedures such as fixed effects, instrumental variables, and others (see Hong, 2015; Murnane & Willet, 2011). Certainly we need to encourage and support work that continues to model alternative methods for achieving more robust indicators of causal effects with observational data.

Another path that should also be taken with respect to determining causal effects could be to embed randomized control trials (RCTs) within the national longitudinal data sets. One could envision randomly assigning schools, or the students within them, to different conditions. Given the number of schools and students in a longitudinal cohort, it seems reasonable to identify enough subjects and sites to be able to detect a difference with adequate power between those assigned to treatment and control groups.

Embedding RCTs within existing national surveys does not quite go far enough. Several years ago, a group of researchers from multiple institutions wanted to conduct an intensive field study on some of the schools in one of the NCES longitudinal samples, but several confidentiality concerns prevented this from occurring. One concern was releasing the names and locations of the schools, which could potentially increase the possibility of identifying particular students. Since improved statistical methods for removing potential identifying information and procedures for accessing restricted data sets have become more prevalent, this no longer seems to be a compelling deterrent. This was a missed opportunity. Intensive field studies could help in more deeply understanding student experiences in different school and community contexts. In both the instance of the RCTs and field studies, one would expect that they would be conducted with the high quality standards of performance that are typically instituted in other national studies (NCES, 2014). For instance, NCES contractors are required to achieve high response rates of well over 80% and limit the number of missing items. We might adopt and strengthen standards for these types of embedded studies in the NCES longitudinal program. In the case of RCTs, for example, there could be precise procedures for estimating effects for NCES longitudinal samples (What Works Clearinghouse, 2014), or for intensive field studies, standard and rigorous procedures for data collection, coding, and verification (AERA, 2006).

5. Data linkages. National longitudinal data sets are designed in part to show trends over time, and although not as extensive as the state longitudinal data systems that collect student information every year, the scope and depth of the variables contained in these data sets are extensive and provide for several key analyses. While national longitudinal data do not have the same time coverage as the new state longitudinal data systems, the two types of data sets could be linked, thereby increasing the utility of both collections (see Dynarski, 2014, on this point).

The state longitudinal data systems track large numbers of students from kindergarten through high school into college and, in some instances, into the labor force. Not only are these systems tracking students, but they also are tracking teachers and changes in school composition. Presently 41 states are involved in developing state longitudinal data systems, but it is important to remember that these data sets are state, not national, so there is limited generalizability to the U.S. population. Dynarski (2014) suggests continuing to follow up NCES data sets into adulthood, and that is currently happening with HS&B and several other data sets being considered for future surveys (Muller et al., 2015). By moving to longer periods, it will be possible to analyze the effects of various student loan policies as well as estimate the long-term effects of college attendance.

Previously we discussed how investigators have already linked administrative data (e.g., IPEDS, FAFSA) to the national longitudinal databases. However, this information is not widely distributed, and researchers are unaware of the potential opportunities to obtain other measures and extend their investigations. We strongly recommend that NCES provide more information on present data linkages and ones that could be explored in the future.

Another point made by Warren (2015) and others has been the usefulness of linking the longitudinal measures and items with other international data sets. Clearly, administrators of longitudinal data sets should consider adding items being asked in TIMSS and PISA especially when the age cohorts overlap. Additionally, many other countries (e.g., Australia, Chile, Finland, Germany, Taiwan, and the United Kingdom) have initiated longitudinal surveys that include survey items that are also asked in the United States. Too often, researchers are unaware of these other international longitudinal data sets and the organizations and associations that support such collections. Bringing together countries that collect items similar to ours will help us gain a more global perspective on student transitions from preschool through the labor market.

6. Employ technology. We live in a world that is becoming increasingly technological. In some schools, students are encouraged to carry phones to call parents or 911 in emergencies; in some places, many of the lessons are given on tablets with interactive software. If we are interested in understanding the lives of children as they progress through their constantly changing environments, we need to employ some of our technology companies in helping us to find better ways of obtaining samples of student work. This includes (a) creating computerized systems of data collection that can quickly process Likert and other item formats; (b) devising new systems for obtaining information, such as digitized records of visual material; and (c) capturing student schooling experiences—such as videos, perceptions of their social worlds, and interactions with parents, teachers, and peers—with newer social network designs. This may mean changing not only the medium of data collection but also the approach. For example, smartphones loaded with experience-sampling method measures can be used to instantaneously record students' activities and feelings in the moment, over the course of a few days (see Hektner, Schmidt, & Csikszentmihalyi, 2007). These momentary data could add additional perspective on how students' experiences differ across a school day that a one-time survey (even with follow-ups) may not be able to provide.

7. Refine and develop new measures. One of the advantages of the national longitudinal surveys is that they have large samples of students, their families, teachers, schools, and communities. With such large samples, it would seem useful to conduct some health-related studies, such as monitoring exercise activities or instituting new meal programs, within a subsample of the population. The equipment for measuring heart rates, stress, and other hormonal samples is becoming less invasive, lower cost, and digitally linkable with other data sets. One could envision these types of studies being conducted descriptively to learn about patterns of behavior or as RCTs to measure the effect of a new program. These samples could also be beneficial for developing a more robust and generalizable set of measures of social and emotional learning, such as students' self-efficacy, self-control, or growth mind-set. Despite marked increases in attention to such measures and their relationship to students' academic performance and identity formation (Duckworth & Yeager, 2015), they could benefit from further development and application in a large-scale longitudinal context. Inclusion of these measures in future student data collections could lead to stronger generalizability between school and student populations, a comparison point that may be highly useful to policymakers and researchers seeking to understand how students' social and emotional learning may interact within a variety of organizational contexts.

Conclusion

The strength and purpose of the longitudinal program is its ability to examine life course developmental patterns among different groups of students and address problems of inequality of access and opportunity from pre-K through postsecondary school and the labor market. Presently there is no other national program that tackles schooling transitions for different groups of students so expansively, beginning with birth and continuing up through one's schooling career, supported by a set of behavioral and subjective measures not captured in other surveys or the SLDS. Specifically, we have highlighted studies that use these data that have had major impacts on local, state, and national conversations regarding (a) racial and socioeconomic inequalities and their influence on the academic trajectories of student subgroups and (b) different schooling experiences and their impact on children's schooling careers and occupational outcomes. These data sets have informed the development and applications of new methods for drawing inferences and provided training opportunities for the next generation of quantitative analysts in education and social and behavioral research.

In times of fiscal constraint and changing administrations, the breadth and scope of the NCES program will be an annual subject of debate and negotiation. Although the costs of operating the longitudinal program are real, the data have provided unique benefits for informing both knowledge development and policy. We do not advocate that the program should proceed as business as usual. If this program is to remain substantively relevant and methodologically rigorous both today and in the future, it needs a serious review of its present organization and design components. Future efforts need to prioritize the following two areas: (a) encouraging innovative approaches to survey design, data collection, coding, and linking; and (b) developing a community of diverse, well-trained researchers (such as those young scholars supported by IES predoctoral and other postdoctoral training programs) who can use national longitudinal data sets to produce rigorous empirical evidence that can inform educational policy and practice.

Appendix

TABLE A1
Basic Features of Selected National Education Longitudinal Studies by the National Center for Education Statistics

Survey	Purpose (special focus)	Base-year sample (sample size, *N*)[a]	Survey years[b]	Types of data/instrument	Key dimensions of measures
Early childhood					
Early Childhood Longitudinal Study, Birth Cohort (ECLS-B)	Children's health, development, care, and education during the formative years from birth through kindergarten entry	9-month-olds (*N* = 10,700)	2002, 2004, 2006, 2007, 2008	Questionnaire (parent, caregiver, teacher), cognitive test, child observation, and physical measurement	Children's cognitive (literacy), social, emotional, and physical development across multiple settings
ECLS, Kindergarten Class of 1998-99 (ECLS-K)	Children's development and experiences in the early grades, and their progression through eighth grade	Kindergarteners (*N* = 21,260)	1999 (fall and spring), 2000 (fall and spring), 2002 (spring), 2004 (spring), 2007 (spring)	Interview (parent), questionnaire (student, caregiver, teacher, school administrator), cognitive test, and physical measurement	Children's cognitive, social, emotional, and physical development across multiple settings
ECLS, Kindergarten Class of 2010-11 (ECLS-K:2011)	Children's development and experiences in the early grades and later development, learning, and experiences in school	Kindergarteners (*N* = 18,170)	2011 (fall and spring), 2012 (fall and spring), 2013 (fall and spring), 2014 (spring), 2015 (spring), 2016 (spring, planned)	Interview (parent), questionnaire (student, caregiver, teacher, school administrator), cognitive test, and physical measurement	Children's cognitive, social, emotional, and physical development across multiple settings
Secondary education					
National Longitudinal Study of the High School Class of 1972 (NLS-72)	High school, postsecondary education, and transition to work	12th graders (*N* = 19,001)	1972, 1973, 1974, 1976, 1979, 1986	Questionnaire (student, school), cognitive test, high school records, and transcript (postsecondary)	Academic achievement, student educational and vocational experiences with later outcomes
High School and Beyond (HS&B)	High school, postsecondary education, and transition to work (school sectors: public vs. private)	Two cohorts: sophomore class (*n* = 30,030), senior class (*n* = 28,240)	1980, 1982, 1984, 1986, 1992 (only for the sophomore class)[c]	Questionnaire (student, parent, and school administrator), cognitive test, and transcript (high school and postsecondary)	Academic achievement, socioemotional skills, family and school context, educational attainment, and employment status
National Education Longitudinal Study of 1988 (NELS:88)	Middle school, high school, and postsecondary education and transition to work (dropout)	Eighth graders (*N* = 24,599)	1988, 1991, 1992, 1994, 2000	Questionnaire (student, parent, teacher, school administrator), cognitive test, and transcript (high school and postsecondary)	Academic achievement; school, work, and home experiences; parent, peer, and neighborhood; educational and occupational aspirations
Education Longitudinal Study of 2002 (ELS:2002)	High school, postsecondary education, and transition to work (educational technology)	10th graders (*N* = 15,362)	2002, 2004, 2005, 2006, 2012	Questionnaire (student, parent, teacher, librarian, and school administrator), cognitive test, and transcript (high school and postsecondary)	Academic achievement; school, work, and home experiences; parent, peer, and neighborhood characteristics; computer use; transition into college and work

(continued)

TABLE A1 (CONTINUED)

Survey	Purpose (special focus)	Base-year sample (sample size, *N*)[a]	Survey years[b]	Types of data/instrument	Key dimensions of measures
High School Longitudinal Study of 2009 (HSLS:09)	High school, postsecondary education, and transition to work (STEM education)	Ninth graders (*N* = 21,444)	2009, 2012, 2014, 2016 (planned)	Questionnaire (student, parent, teacher, counselor, and school administrator), cognitive test, and transcript (high school and postsecondary, planned)	Academic achievement; school, work, home, neighborhood experiences; transition into postsecondary education and work; student majors and career choices (STEM)
Middle Grades Longitudinal Study of 2016-17 (MGLS:2017)	Factors relating to student academic success, high school readiness, and positive life development, such as high school graduation, college and career readiness, healthy lifestyles of all students	Planned: Sixth graders (*N* = about 15,000–20,000)	Planned: 2017, 2018, 2019	Planned: Interview (parent), questionnaire (student, teacher, and school administrator), cognitive test, student observation, and school records	Academic achievement, school and home experiences, transition from middle school into high school
Postsecondary education					
National Postsecondary Student Aid Study (NPSAS)	Student-level financial information at the postsecondary level; serves as the base-year data collection for two postsecondary longitudinal studies (BPS and B&B)	First-time beginning undergraduates (*N*, 1987 = 59,886; 1990 = 70,000; 1993 = 72,000; 1996 = 48,000; 2000 = 70,200; 2004 = 101,010; 2008 = 132,800; 2012 = 128,120)	1987, 1990, 1993, 1996, 2000, 2004, 2008, 2012	Interview (student), institution records, and administrative databases	Demographics, family circumstances, education and work experiences, student expectation, financial aid, and institutional characteristics
Beginning Postsecondary Students Longitudinal Study (BPS)	Postsecondary education issues such as access, choice, enrollment, persistence, progress, curriculum, attainment	First-time beginning undergraduates (*N*, 1990 = 70,000; 1996 = 10,300; 2004 = 16,680)	BPS:90/94 cohort: 1990, 1992, 1994; BPS:96/01 cohort: 1996, 1998, 2001; BPS:04/09 cohort: 2004, 2006, 2009	Interview (student), transcript, administrative databases	Demographics, school and work experiences, college persistence, transfer, and degree attainment
Baccalaureate and Beyond Longitudinal Study (B&B)	Transition from college to work and access to graduate and professional school, and rates of return on investment in education	Bachelor's degree recipients (*N*, 1993 = 11,192; 2000 = 11,700; 2008 = 17,160)	B&B:93/03 cohort: 1993, 1994, 1997, 2003; B&B:2000/01 cohort: 2000, 2001; B&B:08/12 cohort: 2008, 2009, 2012	Interview (student), transcript, administrative databases	Educational attainment, access to graduate and professional schools, the rate of return on educational investment, patterns of preparation and engagement in teaching (teacher sample)

Note. STEM = science, technology, engineering, and mathematics.
[a]The number of participants surveyed in the base year for a given study. The sample sizes for the following waves of data collection in each study varied (not reported here).
[b]Year in which spring semester of academic year occurred; for example, 2002 refers to 2001–2002 school year.
[c]One new follow-up survey is currently being undertaken to reinterview approximately 15,000 sophomores and 12,000 seniors from the original sample in 1980. The purpose of the new follow-up survey is to explore how high school and early adult experiences affect people's lives in their 50s and beyond (see Muller, Grodsky, Warren, & Black, 2015).

References

Adelman, C. (2004). *Principal indicators of student academic histories in postsecondary education, 1972–2000*. Washington, DC: U.S. Department of Education.

American College Testing. (2010). *A first look at the Common Core and college and career readiness*. Iowa City, IA: Author.

American Educational Research Association. (2006). *Standards for reporting on empirical social science research in AERA publications*. Washington, DC: Author.

Arum, R., & Roska, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Bailey, M., & Dynarski, S. (2011). Inequality in postsecondary education. In G. Duncan & R. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 117–132). New York, NY: Russell Sage.

Benner, A. D., & Crosnoe, R. (2011). The racial/ethnic composition of elementary schools and young children's academic and socioemotional functioning. *American Educational Research Journal*, *48*(3), 621–646.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, *1*(2), 97–111.

Brand, J. E., & Xie, Y. (2010). Who benefits most from college? Evidence for negative selection in heterogeneous economic returns to higher education. *American Sociological Review*, *75*(2), 273–302.

Bryk, A. S., Lee, V. E., & Holland, P. B. (1993). *Catholic schools and the common good*. Cambridge, MA: Harvard University Press.

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*, *101*(1), 147–158.

Calcagno, J. C., Crosta, P., Bailey, T., & Jenkins, D. (2007). Does age of entrance affect community college completion probabilities? Evidence from a discrete-time hazard model. *Educational Evaluation and Policy Analysis*, *29*(3), 218–235.

Callahan, R. M., & Muller, C. (2013). *Coming of political age: American schools and the civic development of immigrant youth*. New York, NY: Russell Sage.

Chen, X., & Carroll, C. D. (2005). *First-generation students in postsecondary education: A look at their college transcripts* (NCES 2005-171). Washington, DC: U.S. Department of Education.

Coleman, J. S., & Hoffer, T. (1987). *Public and private high schools: The impact of communities*. New York, NY: Basic Books.

Coleman, J. S., Hoffer, T., & Kilgore, S. (1981). *Public and private schools*. Report to the National Center for Education Statistics. Chicago, IL: National Opinion Research Center.

Conaway, C., Keesler, V., & Schwartz, N. (2015). What research do state education agencies really need? The promise and limitations of state longitudinal data systems. *Educational Evaluation and Policy Analysis*, *37*(1), 16S–28S.

DiPrete, T. A., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. New York, NY: Russell Sage.

Domestic Policy Council & Council of Economic Advisers. (2014). *Taking action: Higher education and student debt*. Retrieved from https://www.whitehouse.gov/sites/default/files/docs/student_debt_report_final.pdf

Domina, T., & Saldana, J. (2012). Does raising the bar level the playing field? Mathematics curricular intensification and inequality in American high schools, 1982–2004. *American Educational Research Journal*, *49*(4), 685–708.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251.

Dynarski, S. (2014). *Building better longitudinal surveys (on the cheap) through links to administrative data*. Retrieved from http://www.naeducation.org/cs/groups/naedsite/documents/webpage/naed_160695.pdf

Fryer, R. G., & Levitt, S. D. (2004). Understanding the Black–White test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447–468.

Fryer, R. G., & Levitt, S. D. (2006). The Black–White test score gap through third grade. *American Law and Economics Review*, *8*(2), 249–281.

Greene, J. P., & Forster, G. (2003). *Public high school graduation and college readiness rates in the United States* (Report No. 3). New York, NY: Manhattan Institute for Policy Research.

Grissmer, D. W., & Ross, J. M. (Eds.) (2000). *Analytic issues in the assessment of student achievement*. Washington, DC: U.S. Department of Education.

Han, W. J. (2008). The academic trajectories of children of immigrants and their school environments. *Developmental Psychology*, *44*(6), 1572–1590.

Hauser, R. M. (2000). Response: Two studies of academic achievement. In D. W. Grissmer & J. M. Ross (Eds.), *Analytic issues in the assessment of student achievement* (pp. 157–166). Washington, DC: U.S. Department of Education.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science*, *269*, 41–45.

Hektner, J. M., Schmidt, J. A., & Csikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.

Hofmann, W., & Patel, P. V. (2015). SurveySignal: A convenient solution for experience sampling research using participants' own smartphones. *Social Science Computer Review*, *33*(2), 235–253.

Hong, G. (2015). *Causality in a social world: Moderation, mediation, and spill-over*. West Sussex, UK: Wiley-Blackwell.

Hong, G., & Raudenbush, S. W. (2005). Effects of kindergarten retention policy on children's cognitive growth in reading and mathematics. *Educational Evaluation and Policy Analysis*, *27*(3), 205–224.

Ingels, S. J. (2002, June). *Longitudinal studies of youth: Recent American experience*. Paper presented at the International Seminar of the Korea Institute for Youth Development, Seoul, Korea.

Ingels, S. J. (2004, April). *Creating time series data sets: Reconciling the conflicting imperatives of continuity and change*. Paper presented at the American Educational Research Association Annual Meeting, San Diego, CA.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, *1*(3), 227–235.

Kalogrides, D., & Grodsky, E. (2011). Something to fall back on: Community colleges as a safety net. *Social Forces*, *89*(3), 853–878.

Lee, V. E., & Bryk, A. S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, *62*(3), 172–192.

Lee, V. E., & Burkam, D. T. (2003). Dropping out of high school: The role of school organization and structure. *American Educational Research Journal*, *40*(2), 353–393.

Long, M. C., Iatarola, P., & Conger, D. (2009). Explaining gaps in readiness for college-level math: The role of high school courses. *Education Finance and Policy*, *4*(1), 1–33.

Moore, K. A., Lippman, L. H., & Ryberg, R. (2015). Improving outcomes measures other than achievement. *AERA Open*, *1*(2), doi:10.1177/2332858415579676

Muller, C. (2014). *New tools for measuring context*. Retrieved from http://www.naeducation.org/cs/groups/naedsite/documents/webpage/naed_160701.pdf

Muller, C., Grodsky, E., Warren, R., & Black, S. (2015). *High school and beyond: Human capital over the life cycle*. Retrieved from https://hsandb.norc.org/

Mulligan, C., Schneider, B., & Wolfe, R. (2005). Non-response and population representation in studies of adolescent time use. *Electronic International Journal of Time Use Research*, *2*(1), 33–53.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research*. New York, NY: Oxford University Press.

National Academy of Education. (2014). *Workshop to examine current and potential uses of NCES longitudinal surveys by the education research community*. Retrieved from http://www.naeducation.org/NAED_081267.htm

National Center for Education Statistics. (2014). *NCES statistical standards*. Retrieved from http://nces.ed.gov/statprog/2012/

National Center for Education Statistics. (2015). *National Center for Education Statistics contract awards: Fiscal years 2011 and 2012*. Retrieved from http://ies.ed.gov/pdf/NCESFY20112012.pdf

Phillips, M. (2000). Understanding ethnic differences in academic achievement: Empirical lessons from national data. In D. W. Grissmer & J. M. Ross (Eds.), *Analytic issues in the assessment of student achievement* (pp. 103–132). Washington, DC: U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Reardon, S. F. (2011). The widening academic-achievement gap between the rich and the poor: New evidence and possible explanation. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 91–116). New York, NY: Russell Sage.

Reardon, S. F. (2013). The widening income achievement gap. *Educational Leadership*, *70*(8), 10–16.

Riegle-Crumb, C., & Grodsky, E. (2010). Racial-ethnic differences at the intersection of math course-taking and achievement. *Sociology of Education*, *83*(3), 248–270.

Roisman, G. I., & Fraley, R. C. (2006). The limits of genetic influence: A behavior-genetic analysis of infant-caregiver relationship quality and temperament. *Child Development*, *77*(6), 1656–1667.

Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge, MA: Harvard University Press.

Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal*, *42*(1), 3–42.

Schneider, B. (2008). Assessing quality in educational journals. In P. B. Walters, A. Lareau, & S. Ranis (Eds.), *Education research on trial: Policy reform and the call for scientific rigor* (pp. 83–104). New York, NY: Routledge.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). *Estimating causal effects using experimental and observational designs*. Report from the Governing Board of the American Educational Research Association Grants Program. Washington, DC: American Educational Research Association.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.

U.S. Department of Education. (2015). *FY 2016 education budget summary and background information*. Retrieved from http://www2.ed.gov/about/overview/budget/budget16/index.html

Warren, J. R. (2015). The future of NCES's longitudinal student surveys. *AERA Open*, *1*(2). doi:10.1177/2332858415587910

What Works Clearinghouse. (2014). *What Works Clearinghouse procedures and standards handbook* (Version 3.0). Washington, DC: U.S. Department of Education.

Willoughby, M. T., Blair, C. B., Wirth, R. J., & Greenberg, M. (2012). The measurement of executive function at age 5: Psychometric properties and relationship to academic achievement. *Psychological Assessment*, *24*(1), 226–239.

Authors

BARBARA SCHNEIDER is the John A. Hannah Chair and University Distinguished Professor in the College of Education and Department of Sociology at Michigan State University. Dr. Schneider is the principle investigator of the College Ambition Program (CAP), a study that tests a model for promoting a STEM college-going culture in 15 high schools that encourages adolescents to pursue STEM majors in college and occupations in these fields. Most recently she is the recipient of a NSF international award to study how to increase science engagement and learning in chemistry and physics high school classrooms. Her research focuses on how the social contexts of schools and families influence the academic and social well being of adolescents as they move into adulthood. Professor Schneider has published 15 books and over 100 articles and reports on family, social context of schooling, and sociology of knowledge. She received her Ph.D. from Northwestern University. She was President of the American Educational Research Association and a fellow of the American Association for the Advancement of Science as well as the National Academy of Education.

GUAN SAW is a PhD candidate in measurement and quantitative methods in the College of Education at Michigan State University. His research focuses on developing and applying experimental and quasi-experimental designs in multilevel and longitudinal settings for examining educational theories, policies, and practices.

MICHAEL BRODA is an assistant professor in the Department of Foundations of Education in the School of Education at Virginia Commonwealth University. His research focuses on applying multilevel modeling and longitudinal data analysis to inform practical challenges faced by schools, teachers, and students.