Surveying the nation: longitudinal surveys and the construction of national solutions to educational inequity

Ethan L. Hutt

To cite this article: Ethan L. Hutt (2016): Surveying the nation: longitudinal surveys and the construction of national solutions to educational inequity, Ethics and Education, DOI: 10.1080/17449642.2016.1185834

To link to this article: http://dx.doi.org/10.1080/17449642.2016.1185834

Published online: 23 May 2016.
Surveying the nation: longitudinal surveys and the construction of national solutions to educational inequity

Ethan L. Hutt
Department of Teaching and Learning, Policy and Leadership, University of Maryland – College Park, College Park, Maryland, USA

ABSTRACT
This paper examines the origins and influences of the introduction of longitudinal student data-sets as a way of gaining insight into the operation of American schools and as a tool for policy-makers. The paper argues that the creation of this new form of data in the 1960s and 1970s represented a relatively new way of thinking about American schools that allowed policy-makers to view the American education system as relatively uniform and the goal of policy to optimize its function. The use of data in this way produced, somewhat paradoxically, a higher precision and more distorted view of the American schooling. The detailed picture provided by these data-sets offered policy-makers new and important insights into the pathways through schooling of individual students while at the same time presented scholars and policy-makers with an increasingly abstract and decontextualized view of the American education system as a whole. That is to say, to view the system uniform and national terms. The upshot of the unprecedented scale and detail of the data collection effort was a more stylized portrait of the American education “system” – one that appeared more amenable to federal intervention but no more likely to secure educational equity for its students.

Americans have grown accustomed to thinking about educational inequality in national terms. Since the 1970s, the yearly release of college admissions tests scores (SAT) and the biannual release of the ‘nation’s educational report card’ (NAEP) have become the education version of a national holiday. Indeed, the cycle of anticipatory commentary – ‘Who is to Blame if NAEP Reading and Math Scores Fall?’ (Klein 2015) – analysis – ‘NAEP Scores Extend Dismal Trend in US Education Productivity’ (Coulson 2013) – and counteranalysis – ‘When Bad Things Happen to Good NAEP Data’ (Sawchuk 2013) – has all the fireworks, nationalism, and selective history of the fourth of July just without the camaraderie and good cheer.

Beyond providing a perennial occasion for finger-pointing and recriminations, these scores have come to provide the most prominent evidence and potent language for discussing educational inequality in the US – the ‘Black-White Achievement Gap’ (e.g. Gamoran and Long 2007; Ladson-Billings 2006). Addressing this achievement gap has become a...
primary and explicit goal of federal educational policy, as the full title of the 2001 No Child Left Behind Act makes abundantly clear: ‘An Act to Close the Achievement Gap with Accountability, Flexibility, and Choice, so that No Child is Left Behind.’

Though we have grown accustomed to thinking about and describing inequality in national terms – NAEP did not even start collecting representative state level data until 2001, well after the ‘achievement gap’ had become a dominant policy frame – it is not at all obvious why we should do so. It seems natural in some respects, of course, given that despite the tradition of localism, the federal government has long used a variety of resources – not just money but land – to support education. The federal government has consistently used those resources to secure educational opportunity whether via the educational land grants of the Northwest Ordinance in 1787; the construction of public school buildings during the Great Depression; or the funding of vocational education programs in the early 1950s (Kaestle and Smith 1982). More recently, the federal government invested in science and math curriculum after the launching of Sputnik and compensatory educational funding became explicit tool for fighting President Johnson’s ‘War on Poverty’ and for building his ‘Great Society’ – a society that would transcend the regional legacies of inequality and racism (Cohen and Moffitt 2010; Phillips 2014).

That education became an issue of national concern does not necessarily mean that we would come to talk about a national system of education. Indeed, there are probably few countries in the world where it makes less sense to describe a school system in national terms than in the US. In 1963, the year NAEP was first conceived, there were nearly 32,000 school districts in the US containing more than 100,000 public schools and nearly 18,000 private schools (NCES 1993). Even today after some considerable consolidation there are still more than 13,000 school districts and 132,000 public and private schools in the US (NCES 2012). These schools are governed by 50 distinct state constitutions; regulated by a patchwork of tens of thousands of federal, state, and local laws and codes; actualized by more than 3 million school teachers certified via 20,000-plus teacher degree and certificate programs; and the entire calico patchwork is cloaked in the mantra of ‘local control.’ Despite the pervasiveness of the construction, it seems fair to ask: given the lack of centralized control, differing constitutional rights to education, or uniformity of schooling provision, what does it mean to talk about the average achievement of the American 9-year-old? Let alone to talk about, as the NAEP reports so often do (NCES 2008), how this achievement compares to the average 9-year-old student of different races or the average American student from four decades ago?

As a historian, I am less concerned with what we currently mean – or think about – when we talk about a national educational system than I am with tracing the shifts in our thinking about the American education system overtime. One important aspect of this story – the allure and value of quantification – has already begun to be addressed by education scholars in history and philosophy (Smeyers and Depaepe 2010). This article builds on this important earlier contribution by considering a specific kind of quantification – the nationally representative longitudinal data-set. These data-sets, which were first imagined during the 1950s when the power of systems analysis and the faith in social science research to solve social problems were at its peak, have provided the basis for an incredibly generative body of scholarship on the American education system – one that, from its inception, was explicitly aimed at securing greater educational opportunity. As the National Center of Education Statistics NCES explained at the outset, the aim was ‘to establish a factual basis for verifying
and refining federal policy concerned with maximizing individual access to educational and vocational opportunity’ (Peng 1977).

Though collection and dissemination of statistics had been the primary goal of the federal government since the nineteenth century, these longitudinal efforts distinguished themselves by promising to provide not just a snapshot of the system’s exterior but a time-lapse picture of the inner workings of the school system. As I will argue below, this shift represented a relatively new way of thinking about American schools that allowed policymakers to view the American education system as relatively uniform and the goal of policy to optimize its function. The use of data in this way produced, somewhat paradoxically, a higher precision and more distorted view of the American schooling. The detailed picture provided by these data-sets offered policymakers new and important insights about the pathways through schooling of individual students while at the same time presented scholars and policymakers with an increasingly abstract and decontextualized view of the American education system as a whole. That is to say, to view the system uniform and on national terms.

This development is significant, in no small part, because it ran against a growing body of evidence – from these same data – that pointed to the considerable limitations of ignoring contextual and individual differences between schools in favor of universalized comparisons across schools. Yet, in the hands of education researchers these data-sets – the first of which were heralded as producing ‘probably the richest archive ever assembled on a single generation of Americans’ with 1900 variables covering 22,000 individuals across 12 years – seemed to place a growing arsenal of simple fixes within the grasp of policymakers: curricular differentiation, Algebra by 8th grade, school vouchers, etc. In other words, the upshot of the unprecedented scale and detail of the data collection effort was a more stylized portrait of the American education ‘system’ – one that appeared more amenable to federal intervention but no more likely to secure educational equity for its students.

This article sketches the broad outline of this historical development and presents it in two parts. The first section provides an examination of the first attempts in the late 1950s to secure longitudinal system data. I argue that these efforts were intimately tied to the ambition to bring systems analysis to bear on questions of school reform – an ambition that created a demand for larger, more precise statistics on school operations – as well as to the Progressive Era view of educational equity as providing appropriately differentiated schooling. Though these early efforts failed to deliver on the promise to identify the precise relationship between identifiable school characteristics and school outcomes, they demonstrated the general viability of these techniques and pointed the way toward this new frontier of research.

In part two I consider the second generation of longitudinal systems in which the federal government took direct control of these efforts and examine how the failure of prior attempts to produce usable insights were recast as a justification for still more and greater data collection efforts. These efforts required researchers to construct an increasingly stylized view of the education system. Rather than consider how different systems might produce different results for different children, the emphasis was now on how education systems, despite their differences, could be made to produce the same results for all children. I also examine some of the ‘policy ready’ recommendations that have come from the analysis of these now built longitudinal data-sets and how they reflect and co-constructed the vision of a national school system in which de-contextualized data could be used to provide universal answers to the local problems of schooling.
In considering this history, I hope to make three contributions to literature on the history of educational research and social science more generally. First, in considering the historical origins of longitudinal data, this article seeks to draw attention to what has become a foundational part of the American educational research infrastructure and the influence that this infrastructure has had on education research. Second, I hope this article will contribute to a growing body of scholarship that has examined the ways in which – across fields ranging from meteorological modeling to poverty research (Edwards 2000; O’Connor 2001) – large-scale data collection and the decisions they make possible become profoundly interdependent and thus self-perpetuating. That is, they become a primary means for thinking about and acting on the underlying phenomenon. Thus, in the same way that the decision to view poverty in terms of an ‘income deficit’ led to a demand for longitudinal ‘microdata’ on the income of individuals and households and these data, in turn, defined our understanding of the ‘nature and causes of poverty,’ (O’Connor 2001, 182) so too did the federal desire to view the school system in national terms and view its deficiencies as rooted in that system led to data-sets that allowed it to see the problems and respond in those terms.

Thirdly, I hope to complicate the history of the federal involvement in American education policy and research. Normally, that story is dominated by the failure of federal efforts to spur ‘basic research’ or to produce a R&D infrastructure analogous to those developed in defense and medicine (e.g. Kiesler and Turner 1977; Vinovskis 1996). Though the hope of turning education into a hard science proved disappointingly ‘elusive’ in many respects, these failures should not prevent us from seeing the success of concurrent efforts to construct – at least at the level of data-sets and statistics – a national education system. Considering education research from this angle helps to make greater sense of the trajectory of education research in a field disillusioned, but not dissuaded, by the Coleman Report. Though the headline message of the Coleman report was that schools matter quite a bit less than society had hoped, and certainly less than non-school factors like social background, the secondary message was that school processes likely mattered more than the overall amount of school resources (Coleman et al. 1966).

While educational scholars, like the public at large, were dispositionally and, perhaps, culturally immune to the first message, the collection of longitudinal data concerning students’ movement through the system and into the world promised to speak – when translated into a certain statistical dialect – directly to the second. Indeed, it is no coincidence that Coleman himself would use the second federal longitudinal data-set (High School and Beyond) to argue that students who attend Catholic achieve more academically even when accounting for race and socioeconomic background – claims that perfectly reflect the triple move provided by longitudinal data: moving at once ‘inside’ the school, outside of any particular context, while at the same time laying claim to an expansive portion of an individual’s life course (Coleman, Hoffer, and Kilgore 1982; Hoffer and Coleman 1987).

**Part I – the search for talent and equity in the age of systems**

‘Schools are extremely complicated organisms,’ observed two RAND analysts J.A. Kershaw and R.N. McKean in a little read but widely searching memo – Memorandum RM-2473-FF – in October of 1959 (Kershaw and McKean 1959). The pair knew something about complicated systems. McKean was, with fellow analyst Charles Hirsch, in the process of developing the Planning-Programming-Budgeting System and writing the Economics
of Defense Spending in the Nuclear Age; Kershaw had already begun to distinguish himself as an analyst of social problems and would soon join the Johnson administration’s War on Poverty as head of the Office Research and Planning in the Office of Economic Opportunity.

‘The learning process is only dimly understood even by those who have spent their lives studying it,’ they explained before concluding ‘If we have anything to contribute, it is not as experts on education, but as interested outsiders who may be able to bring to bear a new methodology on old problems’ (Kershaw and McKean 1959, 6).

The old problem was the efficient organization of school systems and the new methodology was RAND’s in-house specialty: ‘systems analysis.’

The immediate impetus for considering the amenability of school systems to systems analysis was the concern for how America, locked in an Cold War arms race, could continue to supply ‘high quality education’ given that ‘with each passing year the flood of school-age children swells’ (Kershaw and McKean 1959, 56). To do so inefficiently was not only wasteful, they argued, but risked certain failure. The broader impetus was an effort – pursued by RAND analysts and a growing number of people inside the federal government – to ensure that the quantitative analytical techniques that were seen as instrumental in winning World War II were profitably applied to all sectors of American society ranging from the design of electrical grids to the management of paperwork inside the Social Security Administration (Akera 2000; Heyck 2015). Hunter Heyck has recently argued that these efforts should be understood as indicative of a much broader cultural and intellectual shift in the social sciences in the 1950s – a shift which he calls the ‘Age of System’.

The Age of System, according to Heyck, was characterized by both the re-imagining of the world as a complex, hierarchically structured bureaucratic system and by the redefinition of role of the social scientist as describer and analyst of these systems (Heyck 2015). A key driver of this shift was the ‘data explosion’ that had begun inside the federal government during the Great Depression and accelerated during World War II. The sheer volume of these data pointed both to the increased complexity and interconnectedness of systems in general but also to the difficulty of making the right choice without being overwhelmed by the available system information. In an increasingly complex, fast-changing, and dangerous world, decision makers’ choices were both more consequential and difficult than ever. The role of the social scientist, then, was to help rationalize the process of decision-making and to refashion the world as necessary in order to make it possible for such structured, rational choices to be made (Heyck 2015, 128–130). Indeed, Heyek notes, one hallmark of the Age of System was the re-framing of numerous fields around the science of decision-making and the exaltation of this science’s ability to produce optimal outcomes in the face of uncertainty – think Game Theory – and rational choices even in light of the limited cognition and potential irrationality of individual deciders.

The reverberations of the bold intellectual agenda of the age can be heard resounding throughout Kershaw and McKean’s exploration of the education and systems analysis. In their view, the ultimate upshot of systems analysis techniques would be the ability to make ‘formal quantitative comparisons of specific systems with variants of them in which changes and innovations are incorporated … that can help administrators and others choose improved educational systems’ (Kershaw and McKean 1959, iii, emphasis added) While the idea that individual school system variables could be isolated and manipulated to induce measurable educational changes reflected the unbridled analytic confidence of the day, the sheer size
of the analysis required to frame reasonable policy choices did give the analysts pause. ‘The collection of data would, of course be an enormous job’ they conceded (Kershaw and McKean 1959, 35). ‘For instance, if one used only eight variables and allowed each to have two levels, there would be 256 possible combinations. To speak realistically, we doubt if a very helpful analysis could be made with a sample that comprised less than 500 schools. And, it should be remembered that one would have to visit and examine a good many more school systems [to locate ones with suitable traits] than the number of combinations used in the final analysis’ (Kershaw and McKean 1959, 17).

Still, the pair took solace in the fact that the federal government had already funded two massive data collection efforts. The result of these efforts, the RAND men hoped, would provide the necessary raw data inputs for the analyses they envisioned. Their optimism on this account even led them to suggest that subsequent systems analyses be postponed until the data from the second of these studies were available – a delay that ‘might not ultimately cost anytime to speak of’ on account of the amount learned and the possibility that the study ‘will tell us what we need to know about the relationship between school characteristics and educational output’ (Kershaw and McKean 1959, 57).

Though the RAND memo perfectly captures the analysts’ abstract ambition to view American schools as a rational system with variables available for selection and manipulated from afar, the concrete experiences of the early data collection efforts they had pinned their hopes on revealed how much the reality of schooling diverged from the analysts’ vision and how much data infrastructure and statistical airbrushing was required in order to produce a school system available for analysis.

The first of the two federally funded efforts to secure longitudinal data on school operations referenced by the RAND analysts was the Quality Measurement Project in New York State. With money from the Office of Education, the Quality Measurement Project aimed to provide quantitative answers to four deeply pressing and interrelated questions: How much did the quality of school systems (as measured by pupil outcomes on standardized tests) vary within New York state?; Did variation exist even among schools systems with similar system contexts (e.g. urban/rural; socio-economic status of community)?; When accounting for system context, did systems produce similar results for students with similar academic potential (IQ-scores)?; Could any of these variations be accounted for by ‘factors that are within the control of school administration?’ (Goodman 1958, 3, 4). The elementary nature of the questions reflected both the immaturity of the field but also the particular educational goals of the moment. The aim was not to secure uniform achievement across school systems and students, but, in the words of the project’s lead investigator, ‘to capitalize maximally the potential of all types of pupils’ (Goodman 1958; Gamson 2007). Given that ‘a rural district cannot be transmuted into an urban district,’ the goal was to identify relative deficiencies by category of system and suboptimal outcomes for kinds of students within those systems (Goodman 1958, 34).

If these gaps, and the variables that predicted them, could be identified, the state could take a huge step forward in optimizing its education system and creating the necessary data infrastructure to allow for its continued monitoring and maintenance. In this respect, the goals of the project were unprecedented. Even the most ambitious longitudinal attempts to study school systems – notably the Eight Year Study (Aikin 1942) – had sought to demonstrate the possibilities of deliberately experimenting with the high school curriculum and the results of doing so. This was a far cry from monitoring variation in existing systems with
an eye toward exploiting that ‘natural’ variation for the sake of collective system optimization.

The researchers hoped that their role would primarily be as aggregators of available information rather than as creators of new information. But these hopes were quickly dashed. Even in a relatively centralized state whose system was overseen by a Board of Regents and had a uniform test for high school graduation, the researchers found that when it came to documenting student achievement, the system varied in every conceivable way: the standardized tests that were used, the grades that they were used in, and the frequencies of the administration (Goodman 1958, 10). Even their hope that they could generate standardized longitudinal data themselves from a direct observation of student records were dashed when the researchers found ‘only a small percentage of systems in the sample was found to have readily available longitudinal data on their pupils; and, among those that did have such data, variation in the span of years covered, in the tests used, and in the procedures for reporting was too great to permit use of the records by the Project’ (Goodman 1958, 10).

In order to subject the system to study they would have to generate their own standardized data first – a task they gave over to E.F. Lindquist’s Iowa Tests of Basic Skills, not because of its wide use in New York or its alignment with local curriculum, but in large part because of ‘the availability of an electronic scoring service, which had considerable appeal in an enterprise of the size of the Quality Measurement Project’ (Goodman 1958, 11). Standardization, in other words, not only had to occur but it had to occur apace.

Once in possession of their new data – covering roughly 70,000 students in grades 4, 7, and 10 – the researchers could go ahead with their planned analyses. The results they found were ‘an arresting fact.’ According to their data, the mean achievement levels varied by as much as four grade levels across the state systems (Goodman 1958, 17).

Faced with such massive differences in system quality, the Quality Measurement team performed what would become for education researchers in subsequent decades the ritual parsing of achievement gap responsibility. The researchers found that even with their limited statistical controls, half the variation could be accounted for by school location and parental occupation level. They also found, to their dismay, that the most obvious and readily observable school level variables – student expenditures and teacher experience – had very weak, and at times negative, correlations with student achievement particularly for those at the lowest levels (Goodman 1958, 38). Again foreshadowing decades of coming research on inequality, the scholars, undeterred, concluded ‘the finding about within-group differences means that variation in process, that is, in what schools in similar settings do with pupils, has a demonstrable effect upon pupil outcomes. This is the essential finding – the finding that supports the hope that substantial improvement of education in a state or in a nation can come from programs of quality assessment.’ Having just acknowledged that 50% of the variation was beyond school control, they chose to see the systems analysis glass as half-full.

Extracting this hope from the system, however, would require more data. In particular, the Quality Measurement Project noted the need for ‘the development of longitudinal data on system outcomes and patterns of pupil progress’ (Goodman 1958, 66). This data would allow them to peer both inside the system and into the future. Developing and standardizing more variables would provide for a better understanding of the ‘dynamics of school system effectiveness’ – information which, in turn, could be used to ‘develop better procedures for more precisely classifying systems in terms of institutional potential’ and ‘for classifying pupils.
in terms of educative potential’ (Goodman 1958, 67). The developing data loop would provide the means for maximum efficiency and optimal utilization of education potential given the inherent differences between pupils and school systems; it would also provide the basis for subsequent and on-going management of the school system.

Just about the time the researchers for the Quality Measurement Project were submitting their final report to the Office of Education, John Flanagan of the University of Pittsburgh and the President of the American Institute for Research, was in negotiations with the Office of Education to take up this call for longitudinal data with one of the most ambitious education research projects ever attempted. During World War II, Flanagan, a Harvard trained psychologist, had served in the Army Air Forces’ Aviation Psychology Program (Flanagan 1962). Flanagan spent the war devising instruments and procedures to better predict which aspiring pilots were most likely to complete their aviation training. The work was deemed a critical success – credited with saving hundreds of lives and millions of dollars and earned Flanagan a Legion of Merit award (Freeman 1996).

When the war ended, Flanagan believed that the same testing techniques could be used to solve the nation’s manpower problems by providing better guidance to high school students about their likely career trajectories. Just as a battery of tests could be used to determine who would succeed as a pilot and who was better suited to be the navigator, a much larger battery could be used to identify the unique talent signatures of each individual American student, which career would best suit them and, in turn, how their time in school might best be spent (Flanagan 1962, 5). Echoing the sentiments of the famous National Education Association of the United States (1918) and the researchers in the Quality Measurement Project, Flanagan believed that equitable treatment of students involved the individualization and differentiation of their curriculum. Talents, like aptitudes, were innate but they were also multi-dimensional and could be prosperously developed for the benefit of both the individual and the nation. Flanagan believed that embracing this view, as the military had in World War II, represented a fundamental shift in thinking about human development and schooling – drawing an analogy between the rejection of phlogiston theory for modern chemistry to his rejection of Spearman’s unitary intelligence for a multi-dimensional view of talent (Flanagan 1962, 21, 22; see also Kett 2012).

What was required, then, to pursue this vision was a massive, longitudinal ‘census’ of American talent. Only a large-scale survey carried out over an extended period of time would allow for the repeated sub-division of the sample by talent measures and the linking of those measures backward to particular features of the school in which they were developed and forward to success (or not) in the multitude of possible career paths. The result of this grand vision; funding from the Office of Education, National Institute of Mental Health, Office of Naval Research and National Science Foundation; assistance from 90 regional coordinators; and guidance from a technical panel of 31 prominent researchers – including luminaries Ralph Tyler, E. Franklin Frazer, Henry Chauncey, and Robert Thorndike – was a brand new two and a half day long academic and psychological test battery consisting of 25 academic and aptitude tests sub-tests, a student interest inventory, student activity inventory, a measure of student ‘preferences,’ and two short open-ended essays. The battery was given to a representative sample of 440,000 American high school students in 1353 high schools nation-wide. These students, representing 1 out of every 20 high schoolers in the country, would be followed into the working world and surveyed again 1, 5, and 11 years after graduation (Flanagan 1962).
The belief that this avalanche of a billion pieces of data – again helpfully generated by E.F. Lindquist’s computerized scoring machines at the University of Iowa – would unlock the secrets of the high school system is reflected in the astounding 93% cooperation rate that Flanagan received from the nation’s high schools (Flanagan 1962, 51). Town papers throughout the country even proudly announced the selection and participation of local schools in this unprecedented effort (e.g. Beatty 1960; Hartford Courant 1960). The reality of the initial data, however, proved somewhat less than the dream.

Flanagan had hoped – much like the researchers in New York – to use his data to create a taxonomy of 17 different kinds of high schools based on location, community demographics, etc. and to identify their unique strengths in developing certain forms of talent (Flanagan et al. 1962). Instead the research mostly offered support for the ‘small relation to the amount of student learning’ of such school factors as ‘size of school, size of classes, age of building, rural versus urban location, and dropout rate’ (Flanagan 1978, 17). He also found strong evidence of the limited influence of schooling on college going, finding instead a strong correlation between family socioeconomic status and achievement of the student: knowing a student’s academic ability was only a slightly better predictor of college attendance than knowing a student’s socioeconomic status (Flanagan 1978, 18, 19).

With respect to improving student guidance, Flanagan was able to demonstrate, in addition to the low college attendance of academically capable poor students, that many students’ possessed ‘unrealistic’ career plans – defined as a gap between vocational aspirations and current achievement and latent talent. Flanagan characterized these findings as evidence of his overriding concern in the ineffective guidance programs and talent waste in American high schools. But they did not come close to fulfilling the promise of being able to divine a student’s most promising occupation by charting their talent scores (Flanagan 1962). Indeed, Flanagan had fallen well short of the RAND analysts’ hopes that his study would ‘tell us what we need to know about the relationship between school characteristics and educational output’ (Kershaw and McKean 1959).

Though Flanagan was fond of likening standardized testing to the physical scientist using X-rays to study the crystalline structure of molecules (e.g. Flanagan 1962, 1), the seeming failure of his immense data-set to reveal the core structure of the school system – to unlock its secrets – has led many modern scholars and historians to ignore Flanagan’s effort outright or to dismiss it as ‘an exercise in overkill’ that was crushed by the weight of its own empiricism (Kett 2012, 157). Such dismissals however, fail to recognize how pathbreaking this work was and how much its results – and the reaction to them – would portend a defining trend of education research over the next four decades.

First, and perhaps most importantly, Flanagan’s massive Project Talent demonstrated to the federal government and other researchers that the technical tools – test batteries, survey instruments, and data processing machines – and analytic techniques necessary to conjure a stable, if fuzzy, image of the American school system. Flanagan’s portrait of the American High School and the American student provides a telling contrast when placed next to the other famous study of the American high school from this period: James B. Conant’s best-selling The American High School Today (1959). Completed only two years prior to the launch of Flanagan’s survey, Conant conducted his study by driving around the country visiting American high schools in person in order to understand the unique, subjective characteristics of each school. Indeed Conant doubted whether it was possible ‘to obtain information which one could generalize the success or failure of the American school in regard to the education
of any group of children’ and he characterized his evaluations of the schools he visited as ‘impressions’ (Conant 1959; see also, Hampel 1983).

The contrast with Flanagan’s survey design and published reports could not be more stark. Flanagan absolutely believed that such broad comparisons and generalizations were possible – a view reflected in his decision to make the data representative of the nation only – not of individual states or districts. He believed in the immutable traits of the schools as systems even if the optimal relationship between input and system characteristics might vary. As Flanagan explained in a brief section of one the Project Talent Reports entitled ‘Can Schools Really Be Compared?’, the answer had been ‘no’ but only because the proper variables had not been identified or the information available. Producing that information, however, necessarily imposed a level of abstraction on the rendering as academic test batteries reflecting no particular textbook or curriculum had to be created from scratch for the job. Flanagan’s study, like the New York study, loudly announced the ambition and intention of researchers to bring the American high school under study and to transform it, if necessary, in the process (Cooley and Lohnes 1962).

Second, the study also underscored the importance of tying school inputs and characteristics – longitudinally – to particular higher education and labor market outcomes. Though scholars normally associate the ‘shift to outcomes’ with the Coleman Report and the Nixon Administration, Flanagan’s explicit concern with these relationships casts these later efforts in slightly different light. What changed was not so much the focus on outcomes as the meaning of providing equity to students. Both the Quality Measurement Project and Project Talent took the reality of differential outcomes as a starting point and sought to optimize schools and, in the case of Flanagan, course offerings to local conditions recognizing that ‘a rural district cannot be transmuted into an urban district.’ This reflected more of the Cold War imperative to identify the gifted and talented and utilize the full manpower resources of the nation than the Supreme Court’s ‘all deliberate speed’ declaration in Brown v. Board.

The issue of shifting equity standards raises the third point about the contemporaneous interpretation of the Project Talent results. Foreshadowing decades of debates in education research, the fine grain-data of Project Talent – the most detailed picture of American education ever produced – became more of a mirror than a microscope – reflecting back the commitments of researchers. Christopher Jencks, for instance, would use the Project Talent data to argue, repeatedly, that it provided the ‘best currently available evidence’ of the complete inability of schools – regardless of their characteristics – to address inequity (Jencks 1972; Jencks and Brown 1975). For him, the data led to the inevitable conclusion that if Americans wanted to address inequality they should do so directly and leave the schools out of it. Others, however, like economist Alice Rivlin who would serve as Johnson’s Assistant Secretary of Planning and Evaluation in the HEW, explained away the Project Talent results by arguing that the data-set was large but not large enough. What was needed, in her view, was ‘a longitudinal data system for keeping track of individual students as they move through school’ including information about the class they take and the resources they are exposed to (Rivlin 1972, 64). The real value of Project Talent, Rivlin’s argued, was that it justified still ‘more complex and expensive longitudinal studies’ (Rivlin 1972, 65). Studies that, not coincidentally, the National Center for Education Statistics (NCES) would initiate while she was serving in the Johnson administration.

Even with these newly launched longitudinal studies, Rivlin conceded ‘the problem may be that the real world is not organized to generate information about production functions,
Part II – national longitudinal survey and beyond

It is impossible to tell the history of American education research in the second part of the twentieth century and not address the Coleman Report. The report, commissioned by the federal government as part of the passage of the Elementary and Secondary Education Act, produced some of the most controversial – and enduring (Gamoran and Long 2007) – findings in education research. The findings called into question the basic American assumption that schools could be an engine of upward mobility by providing evidence that race and socioeconomic status were strong predictors of school success than were in-school factors such as school facilities, teacher qualities, student body characteristics, etc. (Coleman, United States, and National Center for Education Statistics 1966). What is so striking – and strange – about the American response to these findings is how little they caused people to rethink the central role of schools in the pursuit of equality. As Leah Gordon has noted, this was, with few exceptions, no less true when one considers the responses to the Coleman Report emanating from inside the academy regardless of whether they appeared in white or black social scientific outlets (Gordon 2014).

If the report did not drive Americans to rethink their general faith in schooling, it did propel a narrative that the American education system was fundamentally broken. Whereas before, scholars imagined a mismatch between system design and its location and inputs, now scholars and the wider public suspected something more fundamental might be amiss. This narrative neatly dovetailed with the new brand of analyses that were being underwritten and undertaken by a newly emboldened federal government. The creation of the National Assessment of Education Progress (NAEP), which debuted in 1969 was intended to provide an ‘educational thermometer’ that its designers explicitly likened to the education version of the GDP. While NAEP could provide a running tally of the nation’s ‘achievement gap’ and provide a basis for thinking about the nation’s education system as a coherent whole (disaggregation of test scores by state was impossible), passive indicators of this kind could not provide the ‘solutions’ or the kind of actionable information that could serve as the basis for new policies.

One way that education researchers and those in the federal government hoped to fill this gap was through the newly designed National Longitudinal Study of the high school class of 1972 (NLS72). Coupling the empiricism of the old systems analysis and the post-Brown language of educational access and opportunity, those working in the NCES described the purpose of the NLS72 project as ‘establish[ing] a factual basis for verifying and refining federal policy concerned with maximizing individual access to educational and vocational opportunity, with improvement of the general education system as it impinges upon young people’ (Peng 1977, 1). The survey would accomplish this, like Project Talent before it, by repeatedly surveying a nationally representative sample of the high school class of 1972 as they made no matter how cleverly the statistics are collected. Perhaps the schools are too uniform, with too few important differences that are not correlated with the socioeconomic status of students’ (Rivlin 1972, 65). However plausible these concerns, they would hardly prevent educational researchers from endeavoring to try. If schools as they were found were not organized to divulge their inner secrets then they would need to be reimagined; if the schools were too uniform to the gaze of existing statistical controls, then new variables would need to be devised and new statistical ways of seeing would need to be constructed.
the transition from high school to higher education and/or the workforce. Implicit in this research design was that the ‘impinging’ factors of the education system were inherent to schools as systems rather than to particular locales as the findings from the data were only generalizable to the nation as a whole rather than to any school, district, or even state level system.

In many respects this design paralleled the shifts in the larger intellectual outlook within the discipline of economics. During this period there was fundamental reimagining of the ways in which big systems, like ‘the economy,’ should be studies. Rather than consider the larger forces at work in macroeconomic systems like national economies, they flipped the analysis around and began to ‘place macroeconomics on microeconomic foundations’ (Rodgers 2011, 47). That is to say, many economists came to believe that they could develop a picture of the strength and structure of the economy as a whole, simply by aggregating a large enough group of individual actions.

Even granting the ascendancy of economics – and microeconomics in particular – as a discipline, this post-Coleman Report design choice was particularly strange given that one of the other major findings of the Coleman report was that 80% of the variation in student achievement existed within rather than between schools (Coleman, United States, and National Center for Education Statistics 1966). This finding suggested that there was less to be learned by looking across schools than by looking within them – the kind of particularized analysis undertaken by Conant (and so many qualitative researchers) and also precisely the kind of study explicitly foreclosed by the survey design.

The implications of Coleman’s findings did not prevent scholars from utilizing national survey data to try and open up the black box of schooling in order to determine which features of the school organization – beyond resources – might contribute to differential school achievement when controlling, of course, for socioeconomic factors. One key parameter was curriculum, which has been a perennial source of complaint and presumed culprit for the presumed ineffectualness of American education.

Scholars had begun developing structural ‘school process’ models to account for the differential allocation of school resources to students within a school building particularly on account of tracking. Differentiating curricular tracks – college preparatory, commercial, and vocational – had long been a feature of the American high school particularly after Comprehensive High School movement of the 1950s. Anecdote and a growing body of largely qualitative evidence suggested, however, that tracking had been reappropriated as a strategy for addressing the challenges posed by school desegregation – middle class white advantage could be preserved in the college preparatory even as African-American students were given access to the lower curricular tracks at integrated schools (Gamoran and Berends 1987; Labaree 1992; Reed 2014).

What these qualitative studies had a more difficult time substantiating was whether the differences in achievement by track stemmed from the track placement itself or from differences in prior achievement levels. Venturing an answer to this question was a task well suited to the new crop of national representative longitudinal data-sets (Alexander, Cook, and McDill 1978). Not only could they attempt to disentangle prior achievement from track placement, but, in the case of NLS72, it could associate these curricular placements with life outcomes in an unprecedented way.

This line of research appeared to reap large dividends when researchers found that the large gaps in college degree attainment by race disappeared – though they remained for
socioeconomic status – when researchers controlled for the ‘academic resources’ of the student including, notably, curricular track (Alexander et al. 1982). The final lines of the article reporting these findings reflect the ascendance of the ‘can do’ attitude of education policy researchers:

Clearly, differences involving student background, especially SES, still exist and these fully warrant our concern. As a practical matter, though, the more pressing problem seems to be to assure that all youth who desire a college education acquire the sorts of academic resources that will enhance their prospects for doing so. We recognize that this may be no easy task, but the potential payoff is considerable. Improving student qualifications would serve much more to promote high levels of educational attainment than would even a wholly successful assault on the remaining background liabilities uncovered in our analysis. (Alexander et al. 1982, 330)

These conclusions place the potential and pitfalls of the longitudinal survey on full display: the ability to provide a comprehensive, longitudinal description of racial and socioeconomic educational inequality; the potential to identify specific past bottlenecks in the system; and the temptation to associate their mitigation with profound changes in an individual’s future life-course (i.e. graduating from college or not).

The elided caveat in findings of this sort was that their precise meaning was not as evident as it appeared. Because of a complete lack of standardization in American education curricula, the only way to measure a student’s track was by self-report of the student and the school. On account of the need to keep things standardized, these tracks could only be characterized in the broadest possible terms: in the case of the study above and NLS72 more generally ‘college track’ and ‘non-college track’ (Alexander et al. 1982, 318). What these designations meant in practice – in terms of the actual experiences of students residing in states from Massachusetts to Mississippi – with respect to actual course-taking, specific curricular content, educational experiences, or academic socialization was anyone’s guess.

The response to these concrete but indeterminate findings was a push for more precise and more chronologically expansive longitudinal data. Indeed, one major criticism of the NLS72 design was that it contained only one set of cognitive measures obtained from students when they were seniors. This design did not allow for measures of cognitive growth during high school or for much disentangling of track placement and academic growth (Rock 2012). These deficiencies were addressed in the second longitudinal study – High School and Beyond (HSB) – which followed the high school class of 1980 and obtained measures in both tenth and twelfth grade (and the follow up to HSB, NELS88, went one step further by following students beginning in the eighth grade).

Given the running 1970s controversy over the reality and cause of the declining SAT scores of America’s college bound seniors and the enduring concerns about academic rigor, the issue of curricular tracks and the differential course taking patterns they implied, clearly struck a nerve with lawmakers.

The National Commission on Educational Excellence commissioned a study of a representative historical sample of high school American transcripts by Clifford Adelman a researcher in National Institute of Education. A study of actual course-taking patterns – the ‘time on academic subject matter’ – would reveal, the theory went, the trends in academic quality of the American school and the relationship between those trends and higher education outcomes. The result of these efforts would feature prominently in the ‘Findings’ section of the Commission’s reform document par excellence, the infamous, A Nation at Risk: ‘Secondary school curricula have been homogenized, diluted, and diffused to the point that
they no longer have a central purpose'; ‘The proportion of students taking a general program of study has increased from 12 percent in 1964 to 42 percent in 1979’; ‘We offer intermediate algebra, but only 31 percent of our recent high school graduates complete it’ (Adelman 1983; National Commission on Excellence in Education 1983).

While the study’s inclusion in *A Nation at Risk* highlights the policy relevance of these findings and the growing comfort of policymakers and scholars of speaking generally about the American student, the process by which these findings were produced underscores an important point about the weight of the work and measure of abstraction required to allow us to speak in these terms. It took the National Center of Education Statistics more than two years of study to develop a classification system capable of comprehending the variety and idiosyncrasies of the American high school transcripts collected as part of the HSB study (Adelman 1983, 2).

Because of the specific parameters of the study, Adelman and his colleagues had to use a modified version of the NCES scheme which nonetheless involved a nested coding scheme involving grouping course titles under 131 course name categories, 13 general headings, and three curricular tracks (Adelman 1983, Appendix A, B; Ludwig 1982). And though the authors of *A Nation at Risk* discussed the findings in terms of students, the actual unit of analysis in the study was the course title/code not the curriculum or student. This was, as Adelman explained, because ‘the course is a standard ‘house’ within which students live for measured periods of time … it is thus a category that allows one to analyze aggregate behavior, and it was aggregate behavior in which we were interested’ (Adelman 1983, 6). It was on this basis that he was able to report that American students’ course-taking had become ‘diffuse,’ less academic, and that the value of a course credit had become ‘devalued’.

In subsequent years Adelman and fellow researchers would press further into this, now standardized, course-taking transcript data in order to better foreground the individual American student as the unit of analysis. As with the more generalized research on the mediating effects of academic resources on race and socioeconomic status, scholars were able to tie macro-life outcomes with even more micro elements of schooling. Notably, researchers found that a student’s highest level mathematics course turned out to be a very strong predictor of their likelihood to persist in college. Just as important, they found that the specific timing of math achievement was crucial: ‘early access’ to algebra was strongly associated with advanced high school math achievement and ‘socialization’ in math (Smith 1996).

Findings like these, gleaned from the analysis of longitudinal transcript data, lent credence to the massive push in the 1990s to make algebra by eighth grade a new ‘civil right’ – between 1990 and 2007 the percentage of eighth graders enrolled in math courses nearly doubled from 16 to 31% (Loveless 2008, 1). Not only did scholars identify early algebra taking as predictive of later math achievement but, the HSB transcript data revealed one variable in particular – HIGHMATH – was ‘an extremely powerful construct’ (Adelman 1999, 12; see, also Pallas and Alexander 1983) for predicting college level degree attainment.

Of course the origins of any particular policy development are numerous and complex, but the ease and authority with which scholars could move from research findings utilizing representative, longitudinal data to tidy policy prescriptions are well captured by the definitive Department of Education summary of the findings from NLS72 and HSB, appropriately
titled ‘Answers in the Toolbox’. In its pages one can almost hear the echoes of John Flanagan’s voice and ambition from Project Talent:

the story told by this voyage is clear … it helps us advise and guide students no matter what paths of attendance they follow through higher education. It tells us that if degree-completion lags for any student or group of students, the situation is fixable. We learn where to take the tool box, and what tools to use … One must acknowledge that SES has a continuing influence in life-course events. But the analysis here (and elsewhere) indicates how much education can mitigate those effects – and in both directions (downward mobility is not a chimera). If SES were an overpowering presence the tool box directions would be futile. Optimism is the preferred stance. (Adelman 1999, 83)

Putting aside the inappropriately causal language for conclusions based on a regression, what is striking about this statement is its imagining of an individual student and the capacity to speak to her amidst the babel of local system dialects. Not only that, but the findings also bear the other hallmark of research involving longitudinal data: a comfort with the elision of time. Though the data the study is based on reflects the experiences of the previous two decades of students, the author has no trouble projecting those same experiences onto future generations of students. The data, having provided secure ownership of the past, provides certain purchase of the future.

It is worth noting that the search for universal tools invited by longitudinal data-sets like NLS72 & HBS, was not limited to student level or curricular interventions. They proved just as useful for devising system level interventions as well. Though conservatives had long advocated for the introduction and expansion of school choice programs, these arguments received a big boost in the 1990s with the publication of John Chubb and Terry Moe’s Politics, Markets, and Schools (Chubb and Moe 1990b). The book, which purported to show the irredeemable flaws of public schools and inherent superiority of choice-based systems, was rooted in their analysis of the HSB data. In providing these conclusions, Chubb and Moe were building on the earlier work of James Coleman and his colleagues who had used the same HSB data-set to claim that Catholic schools, on account of their internal organization, produced superior academic outcomes to public schools even when controlling for student characteristics – controversial findings that sparked considerable debate in the academy (Coleman, Hoffer, and Kilgore 1982). In the ultimate illustration of how available data co-create both methods and findings, Coleman’s findings along with a quirk in the design of the HSB survey allowed for the application of a brand new statistical analysis – Hierarchical Linear Modeling – to be applied in the adjudication of school quality (Bryk and Raudenbush 1988; Raudenbush and Bryk 1986). As with the student level findings, the survey data allowed one to confidently ignore specific context to formulate a national reform and to easily jump from statistical association in the past to future policy solution: a universal choice program would be, the words of Chubb and Moe, a ‘panacea’ for the challenge of providing educational opportunity for all American children – and they had the long-term, nationally representative data to prove it (Chubb and Moe 1990a).

**Conclusion**

A 1997 New York Times editorial cartoon nicely captures the dangers inherent to research fields from too much data. The cartoon depicts a newscaster announcing ‘According to a report released today …’ while sitting next to three board game spinners – one each listing
a random behavior, a medical condition, and segment of the human/animal population – under the headline ‘Today’s Random Medical News from the New England Journal of Panic-Inducing Gobbledygook’ (Borgman 1997 as cited in Plantin et al. forthcoming). Though not stated in these terms, the underlying premise is that researchers armed with too much data on too many variables are bound to find and report some – likely spurious – correlation. Indeed in many fields – from epidemiology to meteorology – researchers have found that more data – ‘big data’ – is not always an unqualified good (Plantin et al. forthcoming).

In the case of education and education research our version of this story is clearly complicated. There is no question that some of our most important and enduring findings have been the result of the accumulation and analysis of very large data-sets. Given the American education traditions of federalism, inequality, and unbridled optimism, there is something unquestionably valuable and important about our ability to produce cross-cutting, statements that represent the experiences that are nationally representative for our student population. And, given how often our thinking in education policy is driven by visions of ‘other people’s children,’ there is something important, too, about being able to speak about the ‘average American student.’ But, there are also dangers inherent in this style of thinking. Too often there is a temptation to set aside our otherwise intimate knowledge of the historical legacies of ‘local control’ and of our immensely idiosyncratic education system and to imagine a coherent, uniform system available for our inspection and intervention. I think these dangers are particularly acute, as I have tried to argue in this paper, when we are dealing with data-sets that give us a window unto the lives of not only a nationally representative of students but unto a substantial portion of their life-course as well.

To be sure these efforts were always undertaken with the best of intentions – during the Cold War the belief was that systems analysis would ensure the maximum development of talent and optimal match between school system and its local context; when notions of equality shifted after Brown, researchers shifted from considering the maximum developed in light of local context, to searching for ways to produce equity in spite of local context. This meant using data collection – from Project Talent student profiles to NLS72 surveys to HSB student transcripts – to create a national system that, in turn, could be the subject of search and policy intervention; it also meant freezing the future by projecting onto it insights gleaned from the past. Though these developments well-served the needs of education researchers eager to demonstrate their utility in reforming a broken system – analyzing NCES data-sets has become more than a cottage industry – the broader outlook is less clear. Though we have reached the point where it has become common to speak about the educational and career trajectory of each child – ‘No Child Left Behind’; ‘K-16 pipeline’ – the irony of this kind of attention to individual needs is that in doing so we have found the need to efface that student’s individual context.

**Disclosure statement**

No potential conflict of interest was reported by the author.

**References**


