3-2016

The Question of School Resources and Student Achievement: A History and Reconsideration

Larry Hedges
Northwestern University

Terri D. Pigott
Loyola University Chicago, tpigott@luc.edu

Joshua Polanin
Development Services Group, Inc.

Ann Marie Ryan
Loyola University Chicago, aryan3@luc.edu

Charles Tocci
Loyola University Chicago, ctocci@luc.edu

Author Manuscript

This is a pre-publication author manuscript of the final, published article.

Recommended Citation

Hedges, L., Pigott, T., Polanin, J., Ryan, A., Tocci, C., & Williams, R.* (2016). The question of school resources and student achievement: A history and reconsideration. Review of Research in Education Centennial Edition, 40, p.143-168. (*Authors listed in alphabetical order.)

The Question of School Resources and Student Achievement: A History and Reconsideration[1]

Larry V. Hedges
Northwestern University

Terri D. Pigott
Loyola University Chicago

Joshua R. Polanin
Development Services Group

Ann Marie Ryan
Loyola University Chicago

Charles Tocci
Loyola University Chicago

Ryan T. Williams
American Institutes of Research

[1] Authors listed in alphabetical order

Author Note

The meta-analysis reported in this manuscript was partially supported by a grant from the National Science Foundation, NSF DRL-0723543

The Question of School Resources and Student Achievement: A History and Reconsideration

One of the most enduring questions in educational research is how, and sometimes even whether, the resources provided to schools relate to student achievement. This question has been summarized into the seemingly simple question: Does money matter to student outcomes? A close examination of the historical origins of this question as well as recent studies attempting to examine the influence of resources on student achievement highlights the tension between the competing priorities of efficiency and equity in U.S. public schooling. It also raises issues about the ways in which certain questions develop and become central to educational research.

This paper is organized into three sections. The first section presents an historical background on school resources and student achievement research. The second section entails a report on the results of a systematic review of quantitative studies examining the relationship between per-pupil expenditures and student achievement The results from our new systematic review updates work first conducted by Hanushek (1989) and later re-analyzed by Greenwald, Hedges & Laine (1996), thus incorporating relevant literature published from 1966 - 1993 along with the past twenty-three years. In the concluding section, we draw from both the historical narrative and the meta-analysis to discuss the limitations of contemporary research into the relationship between resources and student achievement as well as suggest ways that the field might develop better, more valuable questions to pursue.

Historical Background on Research into School Resources and Student Achievement

Like Hanushek (1989), we consider *Equality of Educational Opportunity* (Coleman et. al., 1966), also known as the "Coleman Report," to be a historical milestone. The report marked the start of the current era of research into the relationships between schooling inputs and outputs, a period characterized by an increasingly sophisticated use of inferential statistics with large scale datasets. Nearly 100 years of history lead to the publication of the "Coleman Report." The following section briefly reviews this history up through the "Coleman Report" to provide background on how statistical studies of educational resources and student achievement developed in relationship to contemporaneous scholarly and social concerns. The history is cleaved into three eras and organized around central questions of each of those eras: What is the state of US Schooling?: Collecting data on US Schools (1867-1891); Can the field of educational measurement assist in directing school resources in more efficient ways? (1892-1965); and Can educational measurement tell us if money matters? (1966-current). Each period describes the introduction of new statistical methods and threaded throughout are persistent debates about the nature of schooling in a socio-economically diverse, multilingual, and multi-racial society, the purposes of public education, and which academic fields were best equipped to answer questions about how resources relate to achievement.

What is the state of US Schooling? Collecting data on US Schools (1867-1891)

The United States "Department of Education" was formed in 1867 as part of the "Reconstruction Acts" passed by a Republican-controlled Congress. The Republicans, especially the "radical" faction lead by Charles Sumner and Thaddeus Stevens, held strong views on education, including the idea that widespread, publicly supported school systems were essential for the country. Indeed, Sumner noted the lack of public schools in the southern states and insinuated that this was a cause of their recent "rebelliousness" (Tyack, Thomas, & Benavot, 1987). While the "Radical Republicans" failed in the attempt to pass a federal law guaranteeing a public education to all citizens, they were able to establish the Department of Education to, in the words of Ignatius Donnelly, Republican Congressmen from Minnesota, "'enforce education, without regard to race or color, upon the population of all such States that fall below a standard to be established by Congress" (as quoted in Tyack et. al., 1987, p.141).

An immediate problem facing such an effort was that little was known about the state of American education nationally. To address this need, Congress provided the Department of Education with the purpose of

> "collecting such statistics and facts as shall show the condition and progress of education in the several States and Territories, and of diffusing such information respecting the organization and management of schools and school systems, and methods of teaching, as shall aid the people of the United States in the establishment and maintenance of efficient school

systems, and otherwise promote the cause of education throughout the country.'' (as quoted in Grant, 1993, p.1)

At this time the modern field of statistics was in its infancy, though advances in mathematics (Feinberg, 1992) and epidemiology (Freedman, 1999) were attracting wide interest for their ability to describe phenomena and make predictions. It is very likely that the notion that statistical information and facts about education are instrumental to supporting efficient and widespread schooling represents the influence of Horace Mann. This link is through Sumner, who was a self-professed friend and regular correspondent with Mann. In 1844, Sumner unsuccessfully ran for the Boston School Committee following Mann's encouragement (Reese, 2013). A year later, Mann introduced the nation's first system of standard examinations in an effort to gather objective information about the comparative quality of Boston schools and whether or not students were ready to graduate (Gallagher, 2003). Mann promoted this use of examinations and statistical information nationally, later connecting it to the abolition movement through advocacy for multi-racial common schools. The support for abolition helped garner Mann the Free Soil Party's 1852 nomination for Massachusetts governor. The short-lived political party was established in Massachusetts by Sumner, and by 1856, folded into the nascent Republican Party.

By the time the first Commissioner of Education was installed in 1869, Sumner had been marginalized within his party by President Grant and therefore unable to deter legislation downgrading the Department of Education to the "Office of Education" within the Department of the Interior and cutting the Offices' staff from three clerks to two. Nonetheless, the Office of Education developed and distributed its inaugural survey in 1870 to solicit information ranging from student enrollment totals to school expenditures to numbers of teacher to tallies of high school graduates to attendance figures. These efforts were hamstrung by sizable gaps in basic information, such as complete lists of schools and colleges for any state. But the Office persisted in its duty, hiring its first statistician in 1872 and publishing its first public report in 1875. While Donnelly's vision for a vigorous, forceful federal role in education never found sufficient political backing, the Office of Education was able to meet its informational mandate by progressively expanding its survey's scope and increasing the detail of published data. In 1890, the Office inquired about the subject areas taken by students, sources of public revenue, and the value of facilities and physical equipment from both public and private schools (Grant, 1993). A basic, yet robust statistical portrait of American education was emerging.

Can the field of educational measurement assist in directing school resources in more efficient ways? (1892-1965)

The late 19th century marked the developmental period of educational measurement characterized by trial and error, experimentation, and wide-ranging uses of this new field. Much like any new discipline, it allowed for the coexistence of seemingly contradictory perspectives within it. The late 19th and early 20th century also embodied the Progressive Era, which meant that the field of educational measurement grew as part of broader progressive efforts to develop and use scientific and social scientific methods to solve social problems (Feinberg, 1992; Freedman, 1999). As a result, some of these efforts embraced progressive education aimed at limiting opportunities (e.g. eugenics and IQ testing), and others attempted at expanding educational efforts to all students (e.g. early efforts to develop special education). In both cases, the issue at hand revolved around how schools could efficiently educate all children to become productive citizens in an era of compulsory mandates.

In the late 19th and early 20th century, district officials, researchers, and concerned citizens, including and almost exclusively businessmen, used or encouraged the use of descriptive statistics to investigate two central questions: 1) "How can we use statistics to understand what is happening in schools?" and 2) "How can we use the information gleaned from statistical analyses to best direct our resources?" Educational measurement did not yet, as an academic field, distinguish between the theory/practice divide. Educational researchers developed scientific methods for the explicit purpose of improving education and they worked diligently to have their ideas integrated into

formal educational policy with great immediacy. This urgency came about in large part from the perceived inefficiency of schools and the resultant need for reform.

The demands for school reform came from several sectors with business and industry pushing schools to be more efficient, as they engaged in scientific management, their own reform effort. This reform idea derived from concepts about industrial efficiency and scientific management put forth by Frederick Taylor, popularly known as Taylorism. The rise of the social efficiency movement in schools at the turn of the 20th century therefore resulted in large part from Taylorism (Kliebard, 2004). One of the more noted figures that gained wide attention at this time was Joseph Mayer Rice, a medical doctor with a keen interest in schools. Kliebard (2004) dubbed Rice "the father of comparative methodology" (p. 19) as a result of his surveys of American schools conducted in the 1890s. Rice started these in 1891 and published his findings in the educational journal *The Forum* beginning the next year. Although trained as a physician, Rice devoted his work to understanding the status of American education -- its curriculum, teaching, and the performance of students. He became interested in comparing student performance and educational conditions through administrative school surveys. Using his results, Rice advocated for better educational conditions for American students. According to Callahan (1962) Rice's use and application of statistics reflected limited knowledge and questionable results, nonetheless he was taken seriously at the time and considered a pioneer in the field of measurement (p. 100).

Rice published an expanded version of his work in a 1913 book, "Scientific Management in Education" where he made the case to hold administrators and teachers both accountable for defining educational goals and measuring the results of their efforts on meeting those goals through scientific measurement (Kliebard, 2004). He grounded these ideas in industrialism and the social efficiency culture that had begun to seep into American education. Taylorism made a large impression on educational reformers of the early 20th century. They saw its adherence to efficiency as a ripe solution to the challenges faced by school systems dealing with an expanding school population with a multitude of needs. Rice's surveys signaled the beginning of a broader trend. The school survey took hold in districts and found support not only from business interests, but from academics and professional education associations as well (Ryan, 2011).

In the 1910s, the *American School Board Journal* promoted the use of school surveys to examine the return of investments in schools, the efficiency and quality of teachers, and to some degree the efficiency of students (Callahan, 1962). Much of the work around efficiency stemmed from Taylorism, but also from the work of academics like Arthur C. Boyce of the Department of Education at the University of Chicago, a colleague of Franklin Bobbitt (Callahan, 1962). Teachers voiced concern with these rating systems, but had to accept them in most districts due to a lack of bargaining power. Callahan (1962) noted that there was little resistance to the movement to make

schools more efficient from professional circles. The move to conduct full school and district surveys required public support. To garner such support, school boards often enlisted the help of business groups or groups that represented "taxpayers" and appealed to the public's desire to use funds wisely to provide educational resources. George D. Strayer, a professor of educational administration at Teachers College Columbia, a key figure in the survey movement, played a large role in developing and conducting district-wide surveys that carried on well into the 1930s. These surveys left a lasting impression on the way district and school administrators approached their positions in schools, putting the data around administrative and management concerns at the forefront (Callahan, 1962).

The heightened focus on the efficiency of schools, teachers, and eventually students and their achievement led in part to the movement toward standardizing educational testing (Callahan, 1962). Statistics reflected how many students repeated grades or dropped out altogether, the chief concerns in larger districts, for example New York and Chicago, in the early 20th century (Tyack, 1974). The district's goal in collecting these statistics was to determine how schools would deal with "backward" children or, as they referred to them, the "feeble-minded". For example, in 1899 the Chicago Public Schools established a Department of Child Study, which, in 1911, tagged "educational research" on to the department's name. By 1918 the Chicago Public Schools had a department devoted to standards and statistics (Ryan, 2011).

Departments such as these coincided with the growing school population and compulsory school laws in order to manage their school populations and sort them. Simultaneously, calls for how to better differentiate the curriculum increased. Many educators sought to better meet the needs of their students and put their hopes in the use of IQ and other testing, as well as stratified curriculum to prepare children for what they might be best "suited for" in life.

Gould (1996) examined the introduction of intelligence testing in the United States and its European origins. Gould's seminal work, *The Mismeasure of Man*, addressed how key figures who introduced the field of measurement and testing to American education through the promotion of IQ testing and other standard forms of testing, rejected the cautions of French psychologist Alfred Binet who believed the "aim of his scale was to identify in order to help and improve, not to label in order to limit" (p. 182). Henry H. Goddard of the University of Chicago, Lewis M. Terman of Stanford University, and Robert M. Yerkes of Harvard University and then Yale University were early and renowned figures in the field of testing in the United States. All of these psychologists had a significant impact on the growth and use of measurement. Among them, Terman, more so than the others, was responsible for its growth in schools and across districts, with his development of the Stanford-Binet Scale. This instrument, while still focused on measuring the "intelligence" of individual children eventually broadened into other tests designed to assess all children by the late 1910s and early

1920s (Gould, 1996). According to Gould (1996), researchers like Terman took more interest in their "science" of hereditarianism (eugenics), than the burgeoning field of statistics. When confronted with information that contradicted his beliefs, e.g. a "correlation of 0.4 between social status and IQ", Terman advanced a multifaceted argument in support of nature over environment (p. 219). Ironically, Terman ended up backtracking some of his earlier arguments, but not until the late 1930s, after the Great Depression and eugenics had largely been discredited.

E.L. Thorndike, a professor at Teachers College Columbia and an influential psychologist in the early 20th century, adhered to eugenic beliefs of intelligence and had a heavy influence on ideas about the curriculum (Kliebard, 2004). Bobbitt and others who supported curriculum that would stratify American children and prepare them for their "station in life" based on IQ test results found support in Thorndike's conclusions from his published studies in 1924 (Kliebard, 2004). Lewis Terman and E.L. Thorndike believed intelligence was inherited and fixed, but other educators questioned that notion. Academics like Harold Rugg (1917) of Teachers College Columbia, a contemporary of Terman and Thorndike, published a textbook on statistics as early as 1917 for teachers with the hope that they would learn to use it as a tool of social science. Rugg believed teachers could make societal change through education and that students could learn and grow using the curriculum (Kliebard, 2004).

The educators in the Progressive Education Association (PEA) reflected a similar belief as academics like Rugg. We can attribute one of the more prominent uses of statistics during the World War II era to the PEA. The Eight-Year Study (1932-1940), directed by Ralph Tyler of the Ohio State University, examined 30 schools (the final tally was 29) with 15 progressive schools given curricular freedom and the remaining schools following their traditional curricula. At the close of the study almost 1500 students attended college from across the experimental and traditional schools with little difference in academic performance based on grade point average and other factors with the experimental students just edging out the others (Kliebard, 2004). This comparative study provided a good example of a large-scale investigation beyond the school survey where educational researchers began to employ an experimental design in the early era of educational measurement.

Although the use of educational statistics had primarily centered on how to better use resources and reduce waste in K-12 school districts, others began to see how statistics could also be used to address unequal conditions in education more broadly. Organizations like the National Association for the Advancement of Colored People (NAACP) pulled such statistics to address issues of school segregation and to equalize resources in graduate education. In 1935, Charles Hamilton Houston of the NAACP began efforts to desegregate law schools, arguing that separate but equal law schools for Black and White students would become prohibitively expensive for states. He also

saw this as a strategy to eventually call for equal schooling at other levels. In 1938, the Supreme Court heard the case of State of Missouri ex rel. Gaines v. Canada, and held that the state must furnish Gaines legal education equal to those for Whites. This case led to a series of cases brought by Thurgood Marshall, eventually leading to the landmark decision to desegregate schools with Brown v. Board of Education in 1953 and then again in 1954. This shift in thinking about how educational measurement, and in this instance statistics, could be marshaled to support the cause to equalize and perhaps even garner resources for those denied equal access would begin to shape the next period in the growth of the field of measurement in the 1960s.

Can educational measurement tell us if money matters? (1966 - present)

With the election of John F. Kennedy, two ideas were paired as central to federal social policy: a strong belief in the value of scholarly research to effectively design social policies was combined with a commitment to social welfare in the form of the expansion of civil rights and the alleviation of poverty (Featherman & Vinovskis, 2001, p. 49). Prominent academics from leading universities, particularly those with personal ties to members of the Kennedy and later the Johnson Administrations, were sometimes directly consulted and often solicited to prepare reports in support of key policy initiatives (Halberstram, 1993). These tendencies led to the emergence of two parallel approaches to education policy, traditions still present today at the nexus of scholarship and politics of public education. The first is the "compensatory" approach, codified by

Bloom, Davis, and Hess (1965), which primarily seeks to design and implement programs and policies that improve education for students in poverty and minority students. The second is the "efficiency" approach, modernized by the "Coleman Report" (Coleman et al., 1966), which primarily seeks to evaluate programs and policies in order promote the most effective and resource efficient among them.

The divergence began in the earliest weeks of the Johnson Administration as the President and his aides began pressing Congress to enact comprehensive civil rights legislation. The Civil Rights Act of 1964 had been proposed to Congress by the Kennedy Administration, but it was Johnson that saw the bill through to law in the wake of Kennedy's assassination. A small provision had been written into early drafts requiring the federal government to conduct a thorough national assessment of educational opportunities for children from all backgrounds. After a flurry of negotiations, Section 402 of the Civil Rights Act came to read,

> The Commissioner (of Education) shall conduct a survey and make a report to the President and the Congress within two years of the enactment of this title concerning the lack of availability of equal educational opportunity for individuals by reason of race, color, religion, or national origin in public educational institutions at all levels in the U.S., its territories, and possessions, and the District of Columbia.

The completed survey would come to be known as the "Coleman Report" (Grant, 1973).

After the Civil Rights Act's passage, the Johnson Administration began work on comprehensive education legislation independent from initial work on the Section 402 research survey. John W. Gardner, a psychologist by training and then president of the Carnegie Corporation, was tapped to form a commission to draft the new education bill. The "Gardner Commission" put forth a proposal to categorically direct federal education spending, with a significant entitlement program addressing the needs of children from poor families. This concept became the basic structure of the Elementary and Secondary Education Act (ESEA) of 1965, and the provision of aid directly to school districts educating children in poverty became Title I. Following ESEA's passage, Gardner was appointed Secretary of Health, Education, and Welfare (Thomas & Brady, 2005). In turn, he quickly contracted with the eminent educational psychologist Benjamin Bloom to organize a conference and publish its proceedings to make recommendations as to how Title I monies might be invested.

Bloom and his colleagues at the University of Chicago hosted the five day "Research Conference on Education and Cultural Deprivation" in June of 1965, recruiting thirty leading education scholars. The vast majority of these were psychologists, though several sociologists and two public schools officials were included. In the wake of the Brown v. Board of Education ruling, which drew heavily

on the Clarks' "doll tests" (Clark & Clark, 1947) to demonstrate the injury of segregated schooling, cognitive psychology took a central position in discussions about desegregation and education policy. These issues were typically framed in terms of "cultural deprivation;" as Bloom, Davis, and Hess (1965) explained in the introduction to the conference proceedings, the "cultural deprivation" discourse rejected the idea of natural born intelligence deficits among certain races in favor of emphasizing "homes which do not transmit the cultural patterns necessary for the types of learning characteristic of the school and the larger society" (p.4). These problems were to be addressed through "compensatory education," which sought to "prevent or overcome earlier deficiencies in the development of each individual" (p. 6). Frank Reisman (1963), the conference's opening speaker, explained that the goal was not "to train the disadvantaged to become 'good middle class' children" (p.345), but rather to change the way schools and teachers engaged culturally deprived students and families in order to better equip these children for success in broader society. This could be achieved through a variety of programs and curricular changes recommended by the conference participants. These policy suggestions ranged from providing free breakfasts and annual physical examinations to intentional efforts to increase contact between home and school to concerted efforts to identify appropriate curricula and pedagogies to effectively educate "disadvantaged" youths. This "compensatory" approach sought to

develop and install targeted programs aimed at improving the educational outcomes of students in poverty and minority students.

David Seeley, Assistant Commissioner of Education, was a listed observer to the conference. Seeley came to the Office of Education as a Yale-trained lawyer with a particular interest in modernizing the Office's historic data collection and publication functions. In 1964, Seeley successfully lobbied the Commissioner to hire Alexander Mood, a mathematician and former executive at the RAND Corporation, to bring expertise in inferential statistics and computers to the Office as Assistant Commissioner for Educational Statistics. One of Mood's first tasks was to contract a principal investigator for the Section 402 survey, and Mood's immediate recommendation was James Coleman (Grant, 1973). Mood was impressed by Coleman's 1961 book, *The Adolescent Society*, in which Coleman and a team of researchers surveyed over 4,000 students across nine Chicago-area high schools. The 175-item questionnaire at the heart of the study was paired with informal observation and interviews to present a portrait of the American teenager as overly-influenced by peers, being steered away from academic and mature social responsibilities and towards superficial entertainments and immature peer relationships.

Coleman agreed to lead the survey notwithstanding the short timeline to deliver a report (less than two years) and the Office's numerous contentious relationships with state and district leaders across the country stemming from attempts to enforce

desegregation orders. Despite these challenges, Coleman and his researchers were able to administer their survey to over 650,000 teachers and students across more than 3,000 schools over three days in October of 1965. Having defined "equality of educational opportunity" as the "equality of results, given the same individual input" (Coleman et al., 1966, p. 14), the survey generated data about individual students, their school contexts, and their academic performance. This massive dataset allowed the researchers to cultivate a variety of sample groups using results from the 1960 census, and to these sample groups apply a relatively new analytic method, input-output analysis, uncommon outside of economics.

Input-output analysis was initially developed by Soviet economists during the 1920s as a way to inform socialist economic planning. The codification and popularization of the method is attributed to Wasily Leontief, a Russian Jew who fled the USSR to Germany in 1925 at age 19, already with a masters degree in economics. After earning his doctorate in economics in Munich, Leontief fled rising anti-Semitism in Germany to take a position with the National Bureau of Economic Research (NBER) in 1931. From 1932 through 1975, Leontief also held a faculty appointment at Harvard, where he taught input-output analysis to successive generations of economists (Kaliadina, Pavlova, & Wittich, 2006; Kaliadina, Babaskina, & Wittich, 2006). Carl Christ was one of Leontief's early acolytes, publishing an influential paper on input-output analysis in 1955, and becoming a colleague of Coleman's at Johns Hopkins in 1959. The

use of input-output concepts as part of the report's regression analysis was groundbreaking and was soon picked up by other scholars (e.g., Entwisle & Conviser, 1969) to open new lines of inquiry. Yet, the regression methods employed by Coleman were poorly equipped to provide causal inferences (Hoxby, 2016) and were instead better suited to measure correlations between phenomena.

The "Coleman Report" was released on Friday, July 1, 1966 and ran over 700 pages. Its three major conclusions were that: racial segregation was widespread in public schools; there were distinct disparities in academic achievement between racial groups; and, that school effects on student achievement were much smaller than variation in individual background, particularly social class (Gamoran & Long, 2006). This final claim, that "schools are not acting as a strong stimulus independent of the child's background, or the level of the student body" (Coleman et al. , 1966, p. 311) was the result of regression analysis and became the report's most noteworthy argument. In the short term, the "Coleman Report" was generally ignored by the Johnson Administration, whose major foci were on school desegregation and alleviating poverty, questioned by other academics, and met with confusion by the news media, who found the report technical, dense, and difficult to effectively summarize (Grant, 1973). The report had no notable policy influences until 1968 when Daniel Patrick Moynihan wrote a laudatory review of in the *Harvard Educational Review*. Moynihan brought his enthusiasm for the report with him into the Nixon Administration, where

Moynihan arranged for Coleman to become advisor to the Cabinet Committee on Desegregation as well as a favored expert to testify for Congressional committees. The report's findings became central to the Emergency School Aid Act of 1970. This legislation initiated two key changes in federal education policy: a shift from punishing school districts that do not desegregate to rewarding districts that complied with desegregation mandates; and, targeted cuts in education spending under the rationale that school effects are comparatively small (Grant, 1973). Subsequent education policies passed under Nixon, Ford, and Reagan would adopt similar "efficiency" approaches, thereby establishing the "Coleman Report" as foundational to recent education policy and scholarship.

Introduction to Current Study

The "Coleman Report" conclusion that schools had a comparatively minor influence on student achievement spurred much research activity around examining the relationship between school outcomes such as achievement and school inputs using input-output analysis, or education production functions. Perhaps most reflective of the efficiency mindset in the post-Coleman era was the Reagan administration's report, "A Nation at Risk" (1983) and its ensuing budget cuts to the Department of Education. The report argued for a "back to basics" approach to education that focused on streamlined academic inputs in hopes of raising student achievement in core subject areas. Left unaddressed were any "compensatory" concerns about racial or socioeconomic

inequities and how these might be addressed through targeted program or differentiated curricular reforms.

Twenty years after the "Coleman Report" and during the renewed focus on efficiency in the 1980s, Eric Hanushek, an economist, reviewed the existing literature using educational production functions. Hanushek has a personal history with the "Coleman Report"; while a graduate student at Harvard, he participated in a year-long series of weekly seminar meetings among researchers from various backgrounds to closely parse the report's data, methods, and findings. Hanushek (2016) has written that this experience set him on path to researching education policy. In the following decades, Hanushek published a series of papers (1981, 1986, 1989, 1991) reviewing the educational production function literature, which typically uses ordinary least squares regression analysis to predict student achievement using a number of covariates including measures of school inputs such as per-pupil expenditure. The assumption in these analyses is that student background variables such as race, prior achievement and socioeconomic status can be adequately controlled so that we can infer a causal relationship among school resource inputs and student outcomes. Across these studies, he concluded that school resources did not have a consistent relationship with school achievement, essentially that money does not matter for student outcomes.

Hanushek used a method of research synthesis called vote counting. Vote counting categorizes each study into groups depending on the direction and

significance of the studies' conclusions. The analysis counted the numbers of studies that found a positive relationship between school resources and achievement, no relationship between resources and achievement, and a negative relationship between school resources and achievement. Hanushek found insufficient evidence that a majority of studies found a positive relationship between school resources and achievement. Since the time of Hanushek's research, methodological developments in meta-analysis provided more robust and statistically defensible alternatives to vote-counting. A series of papers by Larry Hedges, Robert Greenwald and Richard Laine (1994, 1995, 1996) used meta-analytic techniques to re-examine Hanushek's conclusions. Their analyses synthesized the actual values for measures of the relationship between school resources and achievement instead of characterizing the studies based on the direction of their results. Greenwald, Hedges and Laine (1996) found a small, but consistent positive relationship between school resources and student achievement.

While Greenwald, Hedges and Laine's (1996) methods followed the most current guidelines for research synthesis at the time, they encountered a number of difficulties in analyzing the education production function literature. Two of the major issues were the diversity of models used across the studies, and the number of models presented within each study. In the education production function literature, researchers do not have an agreed-upon set of covariates that should be included. Thus, when predicting academic achievement, researchers control for a wide range of student and school

characteristics such as gender, race, socioeconomic status, and prior achievement. Greenwald, Hedges and Laine included only those studies that controlled for either socioeconomic status or prior achievement in order to decrease the possibility that student background characteristics would confound the findings. Hanushek's (1989) vote-counting method did not account for the influence of other covariates in the studies' models.

The second issue concerns dependencies among the estimates of the relationship between achievement and per-pupil expenditures within studies. Studies included in the review typically reported more than one education production function model. Greenwald, Hedges and Laine used the median regression coefficient within each study to ensure that the coefficients used in the analysis were computed from independent samples.

Twenty years have passed since Greenwald, Hedges & Laine's (1996) work, and new meta-analytic techniques exist for handling some of the difficulties faced in the original work. The study reported here uses a subset of a larger work to provide an update of the synthesis of education production function studies. We focus here on the subset of studies measuring the impact of per-pupil expenditure on achievement. Hanushek's (1989) work included other resources such as teacher/pupil ratio, teacher education and teacher salary as this line of research flows directly from the historical concerns around efficiency.

Methods

Background

This study builds on the systematic review conducted by Greenwald, Hedges & Laine (1996) that expanded Hanushek's (1989) paper examining the relationship between school resources and student achievement. In addition to the studies used by Hanushek, Greenwald, Hedges & Laine (1996) conducted a search of electronic databases in economics, education and psychology, and examined the references from several narrative reviews of this literature. The final sample of studies in Greenwald, Hedges & Laine included 29 studies from Hanushek's (1989) review and an additional 31 studies found in their search process.

This work uses a subset of studies from a project designed to update the Greenwald, Hedges & Laine review. In the larger project, we conducted a search of studies published since 1993, the last year of the search in Greenwald, Hedges & Laine, in order to update the sample of studies focused on examining the relationship between school resources and student achievement. We utilized the same search terms as in the original study. The full dates of the search were 1993 – 2014. The search terms identified studies that were directly examining the relationships among school resources and student achievement. The search would not identify any studies where school expenditures are used as control variable in a study of another phenomenon. A list of search terms used in the updated review are provided in Appendix A.

Inclusion Criteria

In the larger study, we generally followed Greenwald, Hedges & Laine's inclusion criteria for the additional studies though we included unpublished research. We included all studies:

1. Conducted in the United States;
2. Where the outcome measure was some form of K-12 student academic achievement; and
3. That included a measure of educational expenditures such as per pupil expenditure, or teacher salary

Unlike Greenwald, Hedges & Laine, we included unpublished research given the changes in systematic review practice from 1996. Current guidelines for systematic reviews such as those in Cooper (2009) include both published and unpublished research. For the analysis discussed in this paper, we focus exclusively on studies that included a measure of per pupil expenditure in the models examining correlates of academic achievement. All studies included used independent samples. In some cases, studies used the same database; we used only the study that included the most complete model for the analysis.

Coding

All studies included in the analysis were coded by three of the authors. At the level of the study, coding categories included type of publication, year of publication,

and demographic characteristics of participants such as race, socioeconomic status, gender, and grade level. We coded every model fit within each study, recording descriptive statistics if provided, descriptions of each predictor variable and associated outcome variable, the estimated regression coefficients and their standard errors if provided, measures of the quality of the model such as R^2, and the level of the analysis, such as district or student level.

Analysis

Our focus in the analysis was the synthesis of the regression coefficient for per-pupil expenditure (PPE), a measure of the relationship among school expenditures and academic achievement. The studies included in the sample used some form of regression analysis to predict academic achievement from a set of covariates including per pupil expenditure. Studies typically reported more than one regression model, resulting in dependencies among the coefficients within the studies. Greenwald, Hedges & Laine (1996) computed the median value of the PPE regression coefficient for each study reporting more than one regression model. Since 1996, researchers have developed more sophisticated meta-analytic strategies for handling dependent effect sizes within studies.

Becker and Wu (2007) outline three key difficulties in combining multiple regression slope estimates. First, all model outcomes must be measured on a common scale. Second, the slope estimate of interest (focal slope) is measured on a common scale

across studies. Finally, each study estimates the partial relationship between the focal slope and the outcome using the model (i.e. includes an identical set of additional predictors). Maintaining these assumptions in any given synthesis will almost always be impossible.

An alternative approach that requires few assumptions and requires no additional information is robust variance estimation (Hedges, Tipton, and Johnson, 2010; Tipton, 2013). The authors identify three important features of this estimator. First, and most importantly, the covariance structure of effect size estimates is not needed. Second, parameter estimates converge on the target parameter as the number of studies, not the number of cases within studies, rises. The authors show that accurate standard errors are produced with as few as 10 to 20 studies, and Tipton (2013) provides a small sample correction for those cases with fewer than 10 studies. Third, the robust variance estimator is unbiased for any set of weights. Williams (2012) conducted a simulation study that examined using robust variance estimation in the case of synthesizing sample dependent focal slope estimates and as a means of synthesizing regression models across multiple samples. His results indicate that the robust variance estimator provides accurate standard errors across a wide range of circumstances. All analyses were conducted in R (R Core Team, 2008) using the *robumeta* package (Fisher & Tipton, 2014).

Several studies also used a log-transformation of the PPE variable in the model, potentially creating difficulties in synthesizing the PPE coefficient across studies. In order to correct for this problem, we divided the PPE regression coefficient by the mean PPE reported in the paper. All of the regression models that were included in the analysis reported on the mean PPE and could be included in the analysis.

Results

The analysis discussed in this paper focuses on the models that predict some measure of academic achievement, including a measure of per-pupil expenditure as a predictor, and control for race and either socioeconomic status or prior achievement in some manner. The meta-analysis was conducted separately for those studies conducted at the level of the district and those conducted at the level of the student.

Figure 1 provides the flowchart of the results of the search process for the studies included in this paper. We identified 2,641 new potential studies in the search conducted from 1993-2014. After screening titles and abstracts, we obtained 56 studies for full-text screening. We coded 35 studies from the full-text eligibility screening.

From the 95 eligible studies (60 from Greenwald, Hedges & Laine (1996) plus the 35 studies from our most recent search), 24 of these studies included a measure of PPE as a covariate in a regression model predicting some form of academic achievement. The remaining studies typically included some measure of teacher salary or administrative expenses rather than per-pupil expenditure. A majority of the 24 studies

were published in journals in the field of economics. To be eligible for the analysis, the regression model needed to include as a covariate a measure of students' race or the racial composition of the sample, and a measure of either prior achievement or the socioeconomic status of the participants in the sample. We included the racial composition of the sample as a necessary covariate in our analysis in addition to those required by Greenwald, Hedges & Laine (1996). As seen in Table 1, 13 of the 24 studies were missing the requisite control variables for inclusion in the meta-analysis. 12 of these 13 studies were missing any control variable for racial background or composition in the sample, and most of these studies were also missing a measure of prior achievement as a covariate.

A second inclusion criteria for the meta-analysis concerned the information needed to synthesize the PPE coefficients across studies. We used Greenwald, Hedges, and Laine's (1996) strategy to synthesize the PPE coefficients across studies, which requires the mean value of the achievement outcome in the study. We used the half-standardized partial regression coefficient for PPE as our measure of effect size where we divided the estimate of the regression coefficient for PPE by the standard deviation of the achievement outcome variable. The half-standardized partial regression coefficient measures the number of standard deviations of change in achievement associated with one dollar change in per-pupil expenditure. As seen in Table 1, three of

the eleven studies with the requisite control variables failed to provide the standard deviation of the achievement outcome variable.

A third inclusion criteria related to the level of analysis used in the study. Most of the remaining eight studies that reported on all requisite control variables and the standard deviation of the achievement outcome were studies collecting and analyzing data at either the level of the school district or the student. Two studies, however, were studies at the school or classroom level. We decided not to conduct a separate analysis of these two studies, leaving us with six studies meeting the following criteria: 1) A model that controls for race and either prior achievement or SES, 2) The reporting of the standard deviation of the outcome achievement measure, and 3) Data collected and analyzed either at the student or district level. A list of ineligible studies is provided in Appendix B.

A description of the six studies in the meta-analysis are given in Table 2. Three of these studies included data at the level of the district, and three studies included data at the level of the student. A list of the included studies are found in the references. All six studies focused on high school students, with one study including achievement measures from 7th graders. Two of the studies published in 1990 used the Test of Economic Literacy as an outcome, with the remaining studies using either achievement or measures of readiness for college such as the SAT or ACT. Four of the studies used national samples of students with two studies in single states (Virginia and Michigan).

We present the results of the robust variance meta-analytic model separately for the district and student-level dataset. For the three studies that included data at the district level, we could estimate thirteen effect sizes. The results yielded a very small, non-statistically significant but positive effect size (b = .00114, SE = .000287, t = 3.97, p = .13, 95% CI[-.00159, .00387]). To put the mean effect size in context, every $1000 increase in per-pupil expenditure would result in a 1.14 standard deviation increase in achievement. However, the confidence interval includes zero, indicating that, at the district level, PPE is not related to academic achievement. For the three studies that included data at the student level, we estimated eight effect sizes using the half-standardization procedures. The meta-analytic results again indicated a very small, non-statistically significant but positive effect size (b = .000067, SE = .000035, t = 1.91, p = .29, 95% CI[-.0003, .00043]). We can again conclude, based on this very limited dataset, that PPE may not be related to academic achievement. In comparison, Greenwald, Hedges and Laine (1996) found a median PPE effect of 0.0003.

Summary of Findings

Though new meta-analytic methods exist for synthesizing results from regression studies, we found that the models used in the education production function literature are diverse and limit our efforts at quantitative synthesis. Researchers focusing on the relationship between per-pupil expenditure and student achievement do not agree on a standard set of covariates, nor do they use similar measures of

achievement. Of the identified 24 studies that examine per-pupil expenditure, half of them do not include any control for race in the model, a critical omission given the "Coleman Report" findings that inspired this area of research. The studies eligible for the analysis are all focused on students at the high school level, and mostly focus on a single achievement measure such as economics or math. Generalizations from this set of studies to US schools is not warranted.

Our major finding of no-statistical relationship between per-pupil expenditure and academic achievement is based on a small set of studies at both the district and student levels. While we are confident that our meta-analytic results are a subset of studies representative of the education production function literature, they are, like all meta-analyses, not necessarily representative of the population of students or districts in the US. Our finding, while statistically consistent with Hanushek's original argument, is not based on a strong evidence base. These studies use narrow achievement measures, employ cross-sectional or short time frames, and use broad controls for race, socioeconomic status and prior achievement. Jackson, Johnson & Persico (2014) also note that this research base uses statistical methods (e.g. ordinary least squares) that cannot isolate the causal effects of per-pupil expenditure due to unresolved endogeneity biases.

Many research studies have been produced to examine school inputs and outputs, but the literature is too diverse and too inconsistent to employ meta-analysis to

estimate a reliable effect. Even if we were able to obtain a defensible estimate of the magnitude of the relationship between per-pupil expenditures and achievement, the studies in this literature would not help us to understand the underlying mechanisms of that relationship nor how to use PPE to increase achievement. A more important finding of this synthesis is that most of the studies identified do not control for basic student background differences, highlighting a major flaw in this literature.

In these ways, the recent literature fits squarely in the tradition set out by the "Coleman Report." It's a legacy that has been both enlightening and confounding. The "Coleman Report" found distinct disparities in academic achievement between racial groups, and yet the studies in our sample fail to account for race in their models. Since the "Coleman Report," the broader educational research field focused on student outcomes consistently recognizes the importance of race, socioeconomic status and prior achievement in understanding student performance. Further, policy makers and researchers worked for years under the assumption that schools had little influence on student achievement; numerous scholars sought to test this proposition despite the methodology used in "Coleman Report", which was inadequate to justify the claims put forward (Hoxby, 2016). The question of whether monetary resources directly translate into achievement gains has not been addressed adequately in the literature, and may be impossible to explore given the complexity of schools and school districts and the critical importance of student background in examining student performance. Instead,

we should re-frame the question into one about how school resources could influence student outcomes across a wide range of school contexts and student needs.

One productive line of research from economists centers on the impact of school finance reform. Prior to the 1970s, local property taxes funded most schools leading to large within-state differences among districts in per-pupil spending (Howell & Miller, 1997; Hoxby, 1996). Since 1971, many states have implemented school finance reform through court or legislative action (Jackson, Johnson & Persico, 2014). These efforts have been successful, to varying degrees, of equalizing school spending in low- and high-income districts. More importantly, Jackson, Johnson & Persico have also shown that low-income children born between 1955 and 1985 in districts that implemented school finance reform completed more years of education, earned higher incomes, and were less likely to experience poverty than poor children in districts that did not implement reform. These findings suggest room for new questions due to the broader set of outcomes this research examines. It also calls for research that can examine how resources can be deployed to support these student outcomes in a socio-economically diverse, multilingual and multi-racial society.

Conclusion

The question of how resources relate to achievement is an old and recurring one in American education. It dates as far back as the 1867 law establishing the federal Department of Education to promote the "establishment and maintenance of efficient

schools;" however, this question has also always been tied to debates about race, equity, and the purposes of schooling in American society. The way the question is asked and the methods used to answer it are a product of history, as well as a reflection of the scholarly, social, and political concerns at any given time. Considered from this perspective, there is no "best method" to answer the question unequivocally. Instead, an opportunity exists to bring together educators, researchers, policy makers, and other stakeholders to carefully consider what may be the best and most effective questions to ask in pursuit of shared goals in the interest of the educational welfare of children and and public education.

In examining the question of how resources have related to achievement over the last century and a half, it is clear that the responses have been driven by those from disciplines outside of education: Rice as a physician, Thorndike, Terman and later Bloom as psychologists, Coleman as a quantitative sociologist, and Hanushek as an economist. Most had little relationship or intimate knowledge of the inner workings of schools. Rice attempted to understand the work of schools and how they used resources, but did not have the perspective of a teacher or administrator. Psychologists focused on children and whether or not they *could* be taught, in other words they focused on whether or not intelligence was fixed or malleable. Quantitative sociologists and economists created models and functions to isolate the impact of particular resources in relation to achievement. The "Coleman Report" narrowed the definition of

equality of educational opportunity in just this way: "equality of results, given the same individual input" (Coleman et al., 1966, p. 14). The result has been to exclude or radically simplify the complex roles of social factors, such as race, class, and gender that are the inextricable context for schooling. The wide range of social sciences that focused on trying to ascertain whether or not one could tie student achievement to the resources devoted to a school or school district rarely included researchers from the field of education with substantial experience and familiarity with schools and school systems.

This lack of understanding of the problem and its context on the part of those researching the perceived problem at hand - a mismatch of resources and results - may have very well have set up a situation where the research question was flawed from the beginning. In the late 19th century, the focus remained simply on understanding what kinds of schooling were available and where, so that public education could be promoted nationally. The early 20th century brought with it the rise of efficiency and a new business model in the service of creating systems of public schools to educate all American youth. The modern "efficiency" approach, characterized by evaluating the inputs and outputs of schooling, is the vein in which resource-achievement research has been conducted. This has been countered by a "compensatory" approach (Bloom et al., 1965), which focuses on identifying and implementing interventions for more equitable schooling with a secondary concern for efficiency (Coleman et al., 1966). Our present scholarly and political debates about education are often caught between these two

approaches, whether aligning clearly with one side or attempting to argue an effective claim to both (e.g., Reading Recovery has been identified as a "what works" intervention, one reported to be highly efficient and highly effective in supporting literacy development for students in poverty and minority students (Institute of Education Sciences, 2013)).

A critically important point to note is that the statistical models used to examine the relationship among school inputs and student outcomes are not consistent across studies and do not support causal inferences. Policy has been made on the basis of these studies without appreciation for their limitations despite prescient warnings (Murnane, 1991). Moreover, some policy makers seek out research in support of their pre-existing views without acknowledging the implications of selecting research for ideological purposes (Plank, 2011). Taken together, we reiterate the call for clearer measurement of educational constructs, well-defined and articulated methods, and comprehensive results reporting. Without such efforts, the data we seek to use will be limited and the conclusions we draw will be suspect.

When we consider the new questions that can be asked about educational resources and student achievement, especially in this era of "big data," we must not confine our debates to a narrow sphere of experts, funders, and the various public and private entities that generate massive datasets. We must continuously and vigorously engage with the stakeholders that we intend to benefit from our work - policymakers,

school districts, communities, and families of all backgrounds - to better ensure that the questions we research are the questions that have shared value in the pursuit of better educational outcomes for all children. The history of this research, from the Radical Republicans of the late 1860s through to the present, illustrates the dangers of failing to do so.

The turn of the 21st century has instructed us on the value of asking more complex, sophisticated questions and considering a range of factors in our attempts to understand school systems and student achievement. These questions must be generative. How do we reimagine research on school resources and student achievement as part of a concerted, deliberate collaboration among scholars, practitioners, policy makers, and communities? What processes can help us develop questions reflecting shared goals for the educational welfare of children and in the best interest of school systems? Our scholarship must be critical, our research projects must be interdisciplinary, and our engagements must be with a diverse range of stakeholders in public education. We must endeavor to be rooted in the realities of those who understand schools at the ground level and those who work with students from all backgrounds and learning styles to even begin to secure a better future. We must build partnerships that allow us to ask better questions about education that best serves all children in our diverse society.

References

* Indicates studies included in the meta-analysis

Becker, B. J. & Wu, M.-J. (2007). The synthesis of regression slopes in meta-analysis. *Statistical Science, 22(3),* 414-429

Bloom, B., Davis, A., & Hess, R. (1965). *Compensatory education for cultural deprivation*. New York, NY: Holt, Rinehart, Winston, Inc.

Callahan, R. E. (1962). *Education and the cult of efficiency: A study of the social forces that have shaped the administration of the public schools.* Chicago: University of Chicago Press.

Christ, C. F. (1955). A review of input-output analysis. In *Input-output analysis: An appraisal,* NBER Book series studies in Income and Wealth. Princeton, NJ: Princeton University Press.

Clark, K. B. & Clark, M. P. (1947). Racial identification and preference among negro children. In E. L. Hartley (Ed.) *Readings in Social Psychology*. New York: Holt, Rinehart, and Winston.

Coleman, J. S. (1961). *The adolescent society: The social life of the teenager and its impact on education.* Glencoe, IL: The Free Press.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, F., Mood, A. M., Weinfeld, F. D., et al. (1966*). Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Cooper, H. (2009). *Research synthesis and meta-analysis (4th ed.)*. Thousand Oaks, CA: Sage Publications.

Entwisle, D. R. & Conviser, R. (1969). Input-output analysis in education. *The High School Journal, 52(4),* 192-198.

Featherman, D. L. & Vinovskis, M. A. (2001). The growth and use of social and behavioral science in the federal government since World War II. In D. L. Featherman & M. A. Vinovskis (Eds.), *Social science and policy-making: A search for relevance in the Twentieth Century* (pp.40-82). Ann Arbor, MI: University of Michigan Press.

Feinberg, S. E. (1992). A brief history of statistics in three and one-half chapters: A review essay. *Statistical Science 7(2)*, 208-225.

Fisher, Z., & Tipton, E. (2014). Robumeta (Version 1.1) [Software]. Available from: http://cran.r-project.org/web/packages/robumeta/robumeta.pdf

Freedman, D. (1999). From association to causation: Some remarks on the history of statistics. *Journal de societe francaise de statistique, 140(5),* 5-32.

Gallagher, C. J. (2003). Reconciling a tradition of testing with a new learning paradigm. *Educational Psychology Review*, 15 (1), 83-99.

Gamoran, A. & Long, D. (2006). *"Equality of Educational Opportunity": A 40-year retrospective.* WCER Working Paper 2006-9. Madison, WI: Wisconsin Center for Educational Research.

Gould, S. J. (1996). *The mismeasure of man*. New York: Norton.

Grant, G. (1973). Shaping social policy: The politics of the Coleman Report. *Teachers College Record, 75(1),* 18-54.

Grant, W. V. (1993). Statistics in the U.S. Department of Education: Highlights from the past 120 years. In T. Snyder (Ed.), *120 years of American education: A statistical portrait*, pp. 1-4. Washington, DC: National Center for Education Statistics.

Greenwald, R. Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, 66(3), 361-396.

Grimes, P. W., & Register, C. A. (1990). *Teachers' unions and student achievement in high school economics*. Journal of Economic Education, 21, 297-306. *

Halberstram, D. (1993). *The best and the brightest (20th anniversary edition)*. New York, NY: Ballentine Books.

Hanushek, E. A. (1981). Throwing money at schools. *Journal of Policy Analysis and Management, 1,* 19-41.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. *Journal of Economic Literature, 24,* 1141-1177.

Hanushek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18(4), 45-65.

Hanushek, E. A. (1991). When school finance "reform" may not be good policy. *Harvard Journal on Legislation, 28,* 423-456.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23(3),* 5-14.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39-65.

Institute of Education Sciences. (2013). *What works clearinghouse intervention report: Beginning reading: Reading Recovery.* Washington, D.C.: United States Department of Education. Retrieved from: http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/wwc_readrecovery_071613.pdf

Jackson, C. K., Johnson, R., Persico, C. (2014). The effect of school finance reforms on the distribution of spending, academic achievement, and adult outcomes. NBER Working Paper No. 20118. Cambridge, MA: National Bureau of Economic Research. http://www.wber.org/papers/w20118

Jones, J. T. & Zimmer, R. W. (2001). Examining the impact of capital on academic achievement. *Economics of Education Review*, 20, 577-588. *

Kaliadina, S. A., Pavlova, N., & Wittich, C. (2006). The family of W. W. Leontief in Russia. *Economic Systems Research, 18*(4), 335-345.

Kaliadina, S. A., Babaskina, T., Pavlova, N., & Wittich, C. (2006). W.W. Leontief and the repressions of the 1920s: An interview. *Economic Systems Research*, *18*(4), 347-355.

Kliebard, H. M. (2004). *The struggle for the American curriculum, 1893-1958*. New York, NY: RoutledgeFalmer.

Laine, R. D., Greenwald, R., & Hedges, L. V. (1995). Money does matter: A research synthesis of a new universe of education production function studies. In L. O. Picus (Ed.), *Where does the money go?: Resource allocation in elementary and secondary schools* (pp. 44-70). Newbury Park, CA: Corwin Press.

Lopus, J. S. (1990). Do additional expenditures increase achievement in the high school economics class? *Journal of Economic Education*, 21, 277-186.*

Moynihan, D. P. (1968). Sources of resistance to the Coleman Report. *Harvard Educational Review, 38*(1), 23-35.

Murnane, R. (1991). Interpreting the evidence on: `Does money matter?' *Harvard Journal on Legislation 28*(2): 457-464

Plank D. N. (2011). Minding the gap between research and policy making. In C.F. Conrad and R.S. Serlin (Eds.), *The Sage handbook for research in education (2nd edition).* Los Angeles, CA: Sage Publications, Inc.

R Development Core Team (2008). R: A language and environment for statistical computing (Version 3.2.3) [Software]. Available from: http://www.R-project.org.

Reese, W. J. (2013). *The testing wars in public schools: A forgotten history.* Cambridge, MA: Harvard University Press.

Register, C. A., & Grimes, P. W. (1991). Collective bargaining, teachers, and student achievement. *Journal of Labor Research*, 12, 99-110.*

Reisman, F. (1963, February). Research in review: The culturally deprived child: A new view. *Educational Leadership*, 337-347.

Ribich, T. I. & Murphy, J. L. (1975). The economic returns to increased educational spending. *Journal of Human Resources*, 10, 56-77.*

Ryan, A. M. (2011). From child study to efficiency: District administrators and the use of testing in the Chicago Public Schools, 1899 to 1928. *Paedagogica Historica* 47(3), 341-354.

Thomas, J.Y. & Brady, K.P. (2005). The Elementary and Secondary Education Act at 40: Equity, accountability, and the evolving federal role in public education. *Review of Research in Education, 29*, 51-67.

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods, 4(2),* 169-187.

Tyack, D. B. (1974). *The one best system: A history of American urban education*. Cambridge, Mass: Harvard University Press.

Tyack, D. B., James, T., & Benavot, A. (1987). *Law and the shaping of public education, 1785-1954.* Madison, WI: University of Wisconsin Press.

United States. National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform: A report to the Nation and the Secretary of Education, United States Department of Education*. Washington, D.C.: The Commission.

Unnever, J. D., Kerkhoff, A. C., & Robinson, T. J. (2000). District variations in educational resources and student outcomes. *Economics of Education Review*, 19, 245-259. *

Williams, R. T. (2012). *Using robust standard errors to combine multiple regression estimates with meta-analysis*. Unpublished doctoral dissertation, Loyola University Chicago.

Appendix A

Search Terms Used

Input-output and

ERIC

- Administrator qualifications
- Class size
- Cost effectiveness
- Educational assessment
- Educational facilities
- Educational finance
- Educational resources
- Expenditure per student
- Outcomes of education
- Productivity and education
- Resource allocation
- School effectiveness
- Teacher education
- Teaching experience
- Teacher salaries

PsychInfo

- Academic achievement
- Educational aspirations
- Educational objectives
- Income
- School learning
- Classroom environment
- School administrators
- School counseling
- Student characteristics
- School environment
- School facilities
- Teacher characteristics
- Teacher education

EconLit/EconPapers

- Analysis of education
- Economics of education and capital and value of human life
- Economics of education and economics of discrimination and economics of minorities

Appendix B

Studies Not Included In Analysis

Studies missing requisite covariates

Baum, D. N. (1986). A simultaneous equations model of the demand for and production of local public services: The case of education. *Public Finance Quarterly, 14,* 157-178.

Bieker, R. F., & Anschel, K. R. (1973). Estimating educational production functions for rural high schools: Some findings. *American Journal of Agricultural Economics, 55,* 515-519.

Boser, U. (2011). *Return on educational investment: A district-by-district evaluation of U.S. educational productivity.* Washington, DC: Center for American Progress.

Deller, S. C. & Rudnicki, E. (1993). Production efficiency in elementary education: The case of Maine Publish Schools. *Economics of Education Review, 12,* 45-57.

Gyimah-Brempong, K. & Gyapong, A.O. (1991). Characteristics of education production functions: An application of canonical regression analysis. *Economics of Education Review, 10,* 7-17.

Kiesling, H. J. (1967). Measuring a local school government: A study of school districts in New York State. *Review of Economics and Statistics, 49,* 356-67

Nyhan, R. C. (2001). *The influence of class size on student achievement in Florida's elementary school* (pp. 137-166). In C. D. Herrington & K. Kasten (Eds.)., *Florida 2001: Educational Policy Alternatives.* Tallahassee, FL: Florida Institute of Education.

Okpala, C. O., Okpala, A. O. & Smith, F. E. (2001). Parental involvement, instructional expenditures, family socioeconomic attributes, and student achievement. *The Journal of Educational Research, 95(2),* 110-115.

Perl, L. J. (1973). Family background, secondary school expenditure, and student ability. *Journal of Human Resources, 8,* 156-180.

Ritzen, J. M., & Winkler, D. R. (1977). The revealed preferences of a local government: Black/White disparities in scholastic achievement. *Journal of Urban Economics, 4,* 310-323.

Sander, W. & Krautman, A. C. (1991). Local taxes, schooling, and jobs in Illinois. *Economics of Education Review, 10,* 111-121.

Walberg, H., & Fowler, W. (1987). Expenditure and size efficiencies of public school districts. *Educational Researcher, 16(7),* 5-13.

Studies missing outcome standard deviation

Dobbs, R. A. (2012). *The No Child Left Behind Act (NCLB): A Texas economic analysis of accountability.* Unpublished doctoral dissertation, Texas Tech University.

Maynard, R., & Crawford, D. (1976). School performance. In D. L. Bawden & W. S. Harrar (Eds.)., *Rural income maintenance experiment: Final report* (Vol. VI, Part II, pp. 1-104). Madison, WI: University of Wisconsin, Institute for Research on Poverty.

Sebold, F. D., Dato, W. (1981). School funding and student achievement: An empirical analysis. *Public Finance Quarterly, 9,* 91-105.

Studies not analyzed at the student or district level

Dugan, D.J. (1976). Scholastic achievement: Its determinants and effects in the education industry. In J. T. Froomkin, D. T. Jamison, & R. Radner (Eds.), *Education as an industry* (pp.53-83). Cambridge, MA: national Bureau of Economic Research.

Gross, B., Rainey, L. & Goldhaber, D. (2006). *Seeing success: The impact of implementing model practices on student outcomes.* Seattle, WA: Center on Reinventing Public Education. CPRE working paper #2006-2.

Figure 1 – Results of Search

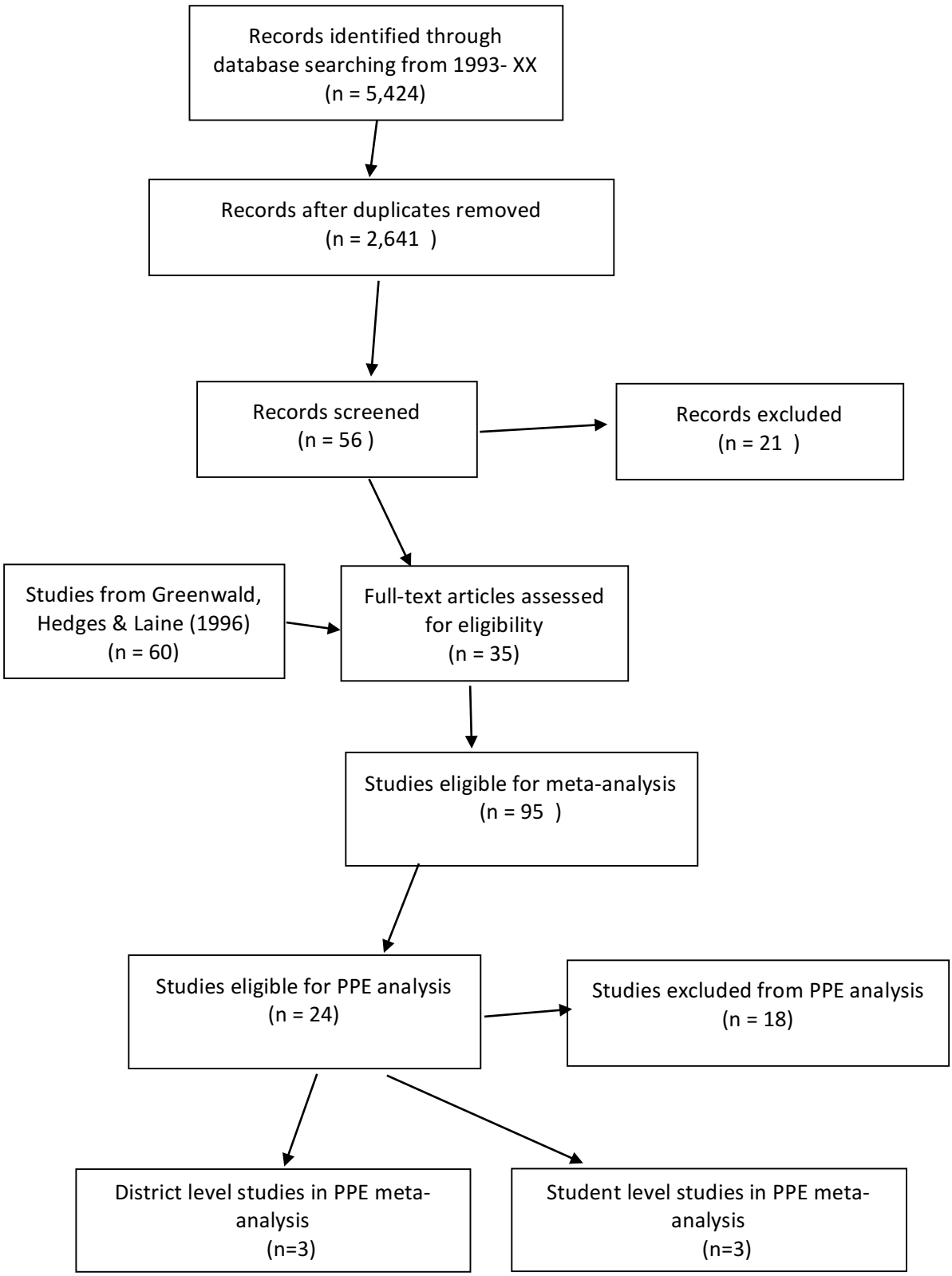

Table 1: Studies excluded from the meta-analysis

Study	Missing Control Variables			Missing Outcome SD	Analysis not at student or district level
	Race	Prior Achievement	SES		
Baum (1968)		√	√		
Bieker & Anschel (1973)	√				
Boser (2011)	√	√			
Deller & Rudnicki (1993)	√	√			
Gyimah-Brempong & Gyapong (1991)	√				
Kiesling (1967)	√		√		
Nyhan (2001)	√	√			
Okpala, Okpala & Smith (2001)	√	√			
Perl (1973)	√	√			
Ritzen & Winkler (1977)	√				
Sander & Krautman (1991)	√	√			
Walberg & Fowler (1987)	√	√			
Dobbs (2012)		√		√	
Maynard & Crawford (1976)				√	
Sebold & Dato (1981)				√	
Dugan (1976)		√			√
Gross, Rainey & Goldhaber (2006)		√			√

Table 2: Studies Included in the Analysis

Study	Level of analysis	Number of Models	Outcomes	Grade Levels	State	Control for SES	Control for Prior Achievement	Control for Race
Grimes & Register (1990)	Student	3	Test of Economic Literacy	High School	National sample	Mother & Father Education	SAT	Student is Black or White/Other
Lopus (1990)	Student	2	Test of Economic Literacy	High School	National sample	Parent Education	Pre-Test of Economic Literacy	Student is White or Non-White
Ribich & Murphy (1975)	Student	3	9th grade aptitude tests	High school	National sample	SES	9th grade tests	Student is Non-White or Other
Jones & Zimmer (2001)	District	8	7th and 10th grade test scores	Middle and High School	Michigan	Median district income	No	% of Black, Asian, Hispanic, American Indian
Register & Grimes (1991)	District	1	SAT and ACT	High School	National sample	Parents' occupation	Student grades	% of Non-white students
Unnever, Kerkhoff & Robinson (2000)	District	4	Math scores in grade 11, percent of 9th graders who graduate HS, percent of seniors aspiring to college	High School	Virginia	Average district income	Standardized 4th grade test scores	% of African-American students