

# Defining Productivity in Education: Issues and Illustrations

The American Economist  
2017, Vol. 62(2) 165–183  
© The Author(s) 2017  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/0569434516688207  
journals.sagepub.com/home/aex



Eric A. Hanushek<sup>1</sup> and Elizabeth Ettema<sup>2</sup>

## Abstract

To discuss how productivity in education might be improved, we must first measure where it stands. The U.S. Bureau of Labor Statistics reports productivity statistics for private sector industries, but not for education, leaving various researchers to attempt such calculations using various inputs and outputs. This article explores existing productivity measures in the private sector, and how these might be applied to education. It then uses examples of possible education productivity measures for the United States to explore education productivity trends over time and possible explanations for these trends. Across alternative measures, productivity in K-12 education has fallen for four decades.

*JEL Classifications:* J2, I21, E24

## Keywords

educational productivity, Baumol's Disease, education costs

## Introduction

Imagine a time traveler from a hundred years ago. About the only place that this person would feel comfortable today would be in our schools. Hospitals, the military, transportation, and of course private industry are all wondrously different. Schools, with the possible exception of whiteboards, offer refuge from the astounding changes that are seen ubiquitously in society.

A significant part of the observed change across the U.S. economy is not that we are just doing things differently but in fact are doing things better. In particular, we are able to do more in terms of things that are valued with any given inputs. Yet by most measures this is not the case in education.

The common summary measure for the overall changes—both in industries and in the economy as a whole—is productivity improvement. The underlying idea is that over time the economy learns to produce more goods with the same level of inputs. The added output from the resources available to society is in fact what makes societies richer over time. It is productivity improvement that translates into income gains.

When education discussions turn to productivity, however, there is often confusion about both how to define productivity and how to interpret productivity changes over time. As explained,

---

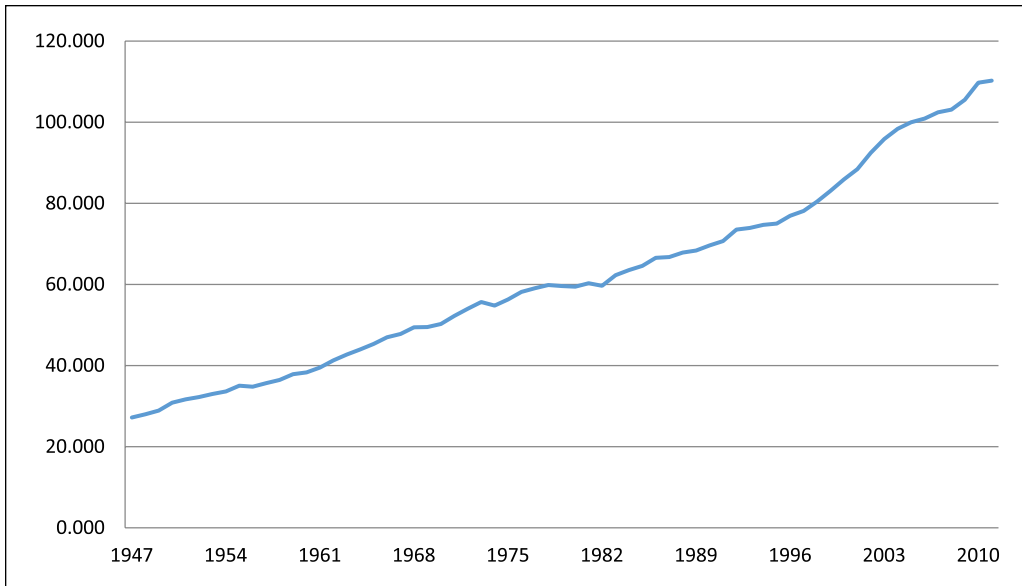
<sup>1</sup>Stanford University, CA, USA

<sup>2</sup>TNTP, Brooklyn, NY, USA

## Corresponding Author:

Eric A. Hanushek, Stanford University, 434 Galvez Mall, Stanford, CA 94305, USA.

Email: [hanushek@stanford.edu](mailto:hanushek@stanford.edu)



**Figure 1.** Labor productivity, private nonfarm business.

Source: Bureau of Labor Statistics, Major Sector Productivity and Costs, Labor Productivity (Output per Hour) Non-farm Business (Series ID PRS85006093).

part of this confusion arises from the results of attempting to apply concepts largely developed for private firms to government activities. Nonetheless, recent measures of outcomes in education make this extension plausible and useful.

This article examines how productivity is measured in the private sector, ways it could be measured in education, and trends in both. It also discusses factors that drive changes in productivity in the private sector and how these may apply to education.

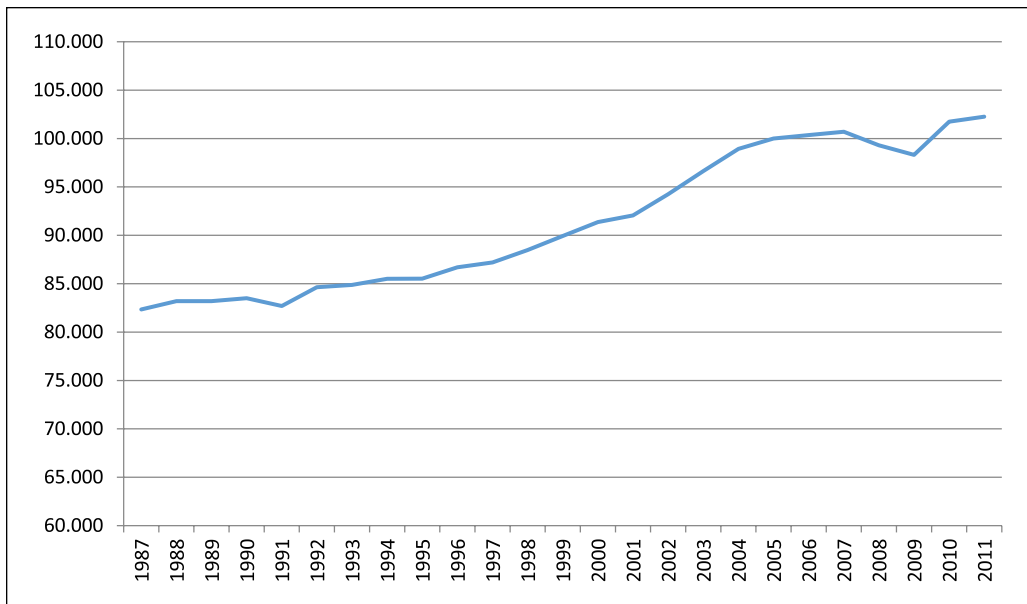
## Productivity in the Private Sector

Fundamentally, productivity is about how much output can be produced per unit of input. Calculating this for the economy, or for parts of the economy, is made feasible by having competitive markets such that outputs can be measured from market transactions where the dollar volume of products reflects how end users value them.

The U.S. Bureau of Labor Statistics (BLS) reports two types of productivity statistics: labor productivity and multifactor productivity. Labor productivity measures output per hour of labor, and multifactor productivity measures output per unit of “combined inputs,” which includes labor, capital, and sometimes intermediate inputs (e.g., fuel). Neither of these measures has much intuitive meaning on the surface, but changes over time—productivity growth—directly relate to aggregate improvements in economic well-being.

Figure 1 shows the pattern of labor productivity in the private sector from 1947 (when the BLS began gathering this information) to 2011. The BLS uses a base year of 2005 in this index, so that percent change in productivity is shown relative to 2005. Small periods of declining productivity coincide with recessions, but overall labor productivity in the private sector has increased steadily over the past 65 years.

Multifactor productivity in the private sector has also increased since the BLS began collecting these data (most recently, in 1987). Once again, 2005 is the base year for the index. Recessions



**Figure 2.** Multifactor productivity, private nonfarm business.

Source: Bureau of Labor Statistics Major Sector Productivity and Costs, Multifactor Productivity, Private Nonfarm Business Sector (Series ID MPU4910012).

are again apparent (and more pronounced), but an overall increase in multifactor productivity is clear (see Figure 2).

Productivity has increased over time in the private sector as new technologies and ways of doing business are developed (Kendrick & Grossman, 1980). Firms that are less productive than their competitors will go out of business, creating both an incentive for firms to be as productive as possible and a mechanism to weed out less productive firms and increase productivity in the aggregate. By definition, productivity can be increased by increasing outputs, decreasing inputs, or both.

## Measuring Productivity in Education

Decision makers in education have shied away from considering productivity. For the most part, resources for the sector have not been very constrained, and the whole focus of attention was simply on finding ways to increase outcomes (quality) without much attention to costs or the resources needed. Thus, discussions of productivity have almost never entered into the debates about educational policy.

Historically, the BLS does not report statistics for productivity in education (or any government function), because of the difficulty of measuring outputs. But this does not mean that education productivity cannot be meaningfully measured.

Various people have worked on this problem in the past, so we are not beginning from scratch. The underlying idea is conceptually clear even if its implementation is complicated. In simplest terms, productivity measurement in education should follow that in the private sector as best as possible.

To measure productivity in education, we must first define the inputs and the outputs. Not surprisingly, this is a much more difficult task in education than it is in manufacturing, or even in for-profit service industries such as banking.

In many ways the input side is easier, although there are some interpretative issues that arise below. As in the BLS statistics, we can measure inputs in terms of labor productivity (teachers employed) or multifactor productivity (total dollars spent). These data are regularly produced and provide the ability to measure trends over time.

The question of outputs is not as straightforward, however, as there is no consensus as to how to define success in education, let alone how to measure it. The ambiguity about outcomes, similar to that in other areas of public programs, is the fundamental reason why the BLS does not report productivity statistics for education. Indeed, government services have traditionally been measured by the dollar value of inputs. However, as interest in this subject has grown in recent years, several non-BLS researchers have attempted to do so.

In a report commissioned by the government of the United Kingdom, Tony Atkinson and colleagues review U.K. government output and productivity in multiple areas, including education (Atkinson, 2005). The goal of this report was to enable more sophisticated analyses of government productivity by setting guidelines for government output to be measured in a similar manner to output in the private sector, a departure from previous methods that treated the output of government services as equal to the inputs. In their words, “we start from Principle A: the measurement of government non-market output should, as far as possible, follow a procedure parallel to that adopted in national accounts for market output” (Atkinson, 2005, p. 36).

In calculating U.K. government productivity in education, the Atkinson Report uses total government expenditure as the input measure and the number of full-time equivalent students in the United Kingdom as an output measure. The students are classified and cost-weighted by school type (nursery, primary, secondary, and special schools), with total U.K. expenditure by school type used to construct the cost weights. In addition, a quality adjustment based on the results from the national examination system (General Certificate of Secondary Education, or GCSE) is used to account for the overall quality of education in the United Kingdom and any changes in that quality over time.

The authors note that this output measure could be improved by measuring actual school attendance rather than students enrolled in school, and that the quality adjustment could be improved by including interim exams as well as the end-of-course GCSE. They also discuss additional measures of school quality, including pupil attainment, quality of teaching, and class size.

Using the procedures outlined in the Atkinson Report, Baird, Haynes, Massey, and Wild (2010) calculate education productivity in the United Kingdom from 1996 to 2009. They report that over this period, inputs rose 36.8% while outputs rose 38.8%, causing productivity in the United Kingdom’s education system to remain virtually flat.

In their 2004 book *Productivity in the U.S. Services Sector*, Triplett and Bosworth examine productivity trends in 29 service sector industries, including education. They acknowledge that there is “very little agreement on how to develop strong quantifiable measures of either output or productivity” in the education sector (Triplett & Bosworth, 2004, p. 268) but go on to construct some basic measures nonetheless. They examine changes in enrollment, nominal and real expenditure, instructors, total staff, enrollment per instructor, and enrollment per staff member. The last two represent labor productivity in education, which they find has declined since 1970. However, they note that this should not be surprising because student–teacher and student–staff ratios were lowered with the goal of improving the quality of education provided, and they do not include any measure of education quality in these calculations. They discuss possible indicators of education quality such as the National Assessment of Educational Progress (NAEP)<sup>1</sup> scores for 17-year-olds or the change between 9-year-old and 17-year old NAEP scores (the former presumably measuring the end “product” of the education system and the latter measuring knowledge added by the system), but they do not incorporate these into any formal measures. They note, however, that “many of the available tests show very little change in the quality of elementary and secondary education at the national level in recent decades” (Triplett & Bosworth, 2004, p. 270). This

would lead to even more dramatically declining productivity in education if these measures of quality were incorporated into the calculations of labor productivity described above.

Finally, Boser (2011) constructs three measures of education productivity and applies them to more than 9,000 school districts in 46 states.<sup>2</sup> (The study was restricted to districts that enrolled at least 250 students and included all grades from K-12 and states with more than one district.) All three measures use the percent of fourth and eighth graders rated as proficient or above on state reading and math tests as the output, but the inputs are different. The Basic Return on Investment Index uses current per-pupil expenditures, adjusted for cost of living as well as weighted by percent low income, non-English speaking, and special education students. The Adjusted Return on Investment Index uses a regression analysis, rather than fixed weights, to account for differences in the cost of educating these student groups. Finally, the Predicted Efficiency Rating uses spending and demographic information to predict student achievement and measures how much better or worse districts do than expected.

The purpose of this study is different than the two described above. Although Baird et al. and Triplett and Bosworth examine changes in education productivity for an entire country over time, Boser focuses on differences within individual states at one point in time. Findings indicate that productivity varies greatly within states. Districts that spend the same amount of money per pupil in the same state have vastly different outputs as measured by student achievement. This study is useful to districts seeking to improve their productivity because it highlights similar districts that are getting greater outputs with the same (or less) input. Although the reasons for these disparities may be complicated to untangle and further analysis is needed, this study goes a step beyond documenting an overall trend of declining productivity in the national education system as a whole to documenting differences in inputs and outputs at the district level.

These studies demonstrate possible ways of measuring productivity in education, and utilize varying measures of education inputs, and especially education outputs. Data availability limits the ways in which education productivity can be calculated, but the examples above are by no means the only possibilities. Along the lines of the BLS figures for the private sector, both labor and multifactor productivity can be calculated for education, and a wide variety of output measures (or a composite measure constructed of multiple education outputs) can be used. As education data improve, even more possibilities for measuring education productivity arise.

## **Possible Simple Calculations of Education Productivity**

The prior work provides a range of possible ways to identify and measure changes in educational productivity. We now return to consider both some additional basic options and the variety of interpretive issues that arise with each. We start with the simplest approaches and then expand on these.

Consider for the moment measuring productivity if the quality of the product (students) is constant. The absolute number of students is largely driven by demographic forces, although the number of student-years varies also with the length of time each student remains in school.

### ***Pupil–Teacher Ratio***

The direct analogy to private sector labor productivity would simply be how much labor goes into each student-year. But indeed this is regularly calculated and presented. It is simply the pupil–teacher ratio or the pupil–staff ratio. We can think of each student as receiving an input of 1/20 (or 5%) of a teacher in a class of 20 children, or 1/16 (or 6.2%) of a teacher in a class of 16. Because classroom teachers are not the only adults employed in K-12 education, it is also natural to expand on this. Teaching aides, cafeteria staff, bus drivers, school counselors, and building- and district-level administrators are all education employees as well, and their labor could be

considered in labor productivity measures as well. If a more inclusive input measure is desired, pupil–staff ratios, which count not only teachers but all adults employed in K-12 education, could be used instead of pupil–teacher ratios.

### *Per-Pupil Spending*

Another way to define inputs for the purpose of measuring productivity in education is total spending. This accounts for money spent on nonlabor inputs, as do the BLS multifactor productivity statistics. Again, due to rising numbers of students we cannot just use the total dollars spent on K-12 education as an input: Rising spending may simply be due to rising enrollment. Instead, per-pupil spending could be used. Depending of course on the comparisons that are made, we would clearly like to recognize how input costs change over time. Per-pupil spending must be adjusted to reflect inflation and regional cost differences if comparisons are to be made over time or across states. Note that this issue does not arise when looking at labor productivity as defined above, because we compare real quantities—students and teachers or staff. (It also was relatively unimportant when calculating total factor productivity in the private sector when both output and inputs are measured in dollar aggregates.)

### *Cumulative Inputs*

One other issue is special to education. The outputs that we measure are the result of a lengthy schooling process. It would not make sense to attribute the 12th-grade knowledge of a student just to the schooling of the last year of high school. Indeed, it makes sense to consider the entire flow of inputs that went into a student at the point where outcomes are measured.

But once we think more broadly about the relevant inputs, we also realize that there are more than just schools that influence students and their outcomes. Families, other students, and neighborhoods are important as is the motivation and effort of the individual student. To the extent that we want to attribute any productivity changes to schools, it is important to ensure that changes in other inputs are not important—and confused with measures of productivity in the schools.

### *Output Issues*

These simple calculations for pupil–teacher ratios and for per-pupil expenditure rely, however, on a presumption that the relevant output is student-years of schooling. This presumption in turn requires that the output quality is not changing. Remember, for example, that there was a quality adjustment when productivity is considered by the U.K. Office for National Statistics. The U.K. Office for National Statistics (Atkinson, 2005; Baird et al., 2010) uses attendance-adjusted enrollment as their main output when measuring education productivity, but adds a quality adjustment (change in General Certificate of Secondary Education scores) to account for possible changes in the quality of education being provided.

Indeed, such considerations appear to be particularly important, because much of the decrease in pupil-teacher ratios and much of the increase in educational expenditure per pupil is motivated by an interest in improving educational quality. If, for example, smaller classes permit greater learning per student-year, it would be inappropriate simply to track pupil–teacher ratios and to interpret the declines in these as a fall in educational productivity. The same is true for interpreting the rise in per-pupil spending.

### *Possible Alternative Measures of Education Outputs*

If education were a factory, its product would be educated citizens. Children enter kindergarten “uneducated” and come out of 12th grade “educated.” Understandably, however, there is little

consensus regarding exactly what this entails. Most people agree that educated citizens should be able to read and do mathematics as well as know some information (though exactly what is certainly up for debate) about science, history, geography, and the arts. A more inclusive definition might also encompass physical education and health, citizenship, noncognitive skills, and other nonacademic outputs.

There are similarities in private industry. Counting student-years is quite similar to counting cars coming off an assembly line or insurance policies sold. And, with private industry and productivity, these quantity measures of production would suffer similar issues of quality changes and definitions of exactly what one unit of an output would be. The private sector approach relies on market information and uses the aggregate value of production—presuming that sales information gives a way of both aggregating across disparate products and allowing for quality changes. The problem is that education lacks such ready market information that can be directly related to the production of students.

A currently common way of summarizing the goal of K-12 education is preparing students for “college or career,” though this is also quite vague and varies widely depending on the college or career in question. States have quite broadly tried to define college and career readiness, but, because education is a responsibility charged to the U.S. states, there are 50 different sets of goals, curricula, standards, and assessments. Although this may be mitigated somewhat in the future through relying on the Common Core standards, not all states have adopted these standards and not all content areas are covered.<sup>3</sup> In addition, college admission and job requirements will likely continue to vary widely.

Even once the correct “outputs” of education are defined, measuring them is a challenge. Any number of existing measures (including graduation rates, college completion, employment statistics, or various tests) might be used to evaluate educational outputs, or a composite involving multiple measures could be constructed. Each option has its strengths and weaknesses, and the choice of which to use depends largely upon the purpose for which productivity is being measured. Several output measures (though by no means an exhaustive list) are discussed below, and from these, we can begin to develop a picture of educational productivity.

*High school graduation.* A high school graduate is perhaps the closest thing to a “product” that K-12 education produces—if a student successfully passes through those 13 years of education, he or she might be thought to come off the education assembly line as a finished product of the American education system. High school completion is frequently used as a goal of K-12 education. We could therefore use graduation rates as an output in education productivity measures. This builds upon the enrollment output measure above because a student only counts as “educated” if he or she graduates.

However, graduation standards and requirements vary by state and district, and simply passing through the system may not guarantee that a student is ready for work or college. Moreover, the requirements for graduation—at least in terms of course taking—have changed over time. Approximately half of the states also have required passing graduation examinations, and these exams vary widely across states and over time in terms of content and difficulty.

*College matriculation/graduation/remediation.* Because one goal of K-12 education is preparing students for college, it might make sense to use college enrollment as an output measure. If we are concerned that some students are not graduating from high school prepared for college, we could leave it to the colleges to determine which “products” of the education system are successful. However, many students who enroll in college do not earn a degree, and those who drop out may do so because they were not adequately prepared for college by their K-12 education. Furthermore, because college attendance is affected by tuition and cost, changes in these could alter attendance even without schools doing anything different.

College graduation could instead be used as an output measure, although many factors that play a role in this are completely unrelated to K-12 education, including the quality of college education itself. Students may also drop out for financial or personal reasons that have nothing to do with how well they were prepared for college. These are all factors that vary over time and across institutions.

A college-related output measure that may avoid some of the problems described above is the number of remedial classes required of college students. If college students need to take remedial courses, it could be argued that they were not successfully “educated” in high school, and certainly that they have not met the goal of being “college-ready.” Because remedial courses are an indicator of an unsuccessful education, it would make sense to use the percentage of college students not requiring remediation as an output measure. Again, however, if views about the requirements and about the utility of remediation vary over time, interpreting changes in remedial course taking will be complicated.

*Employment.* Ultimately, one of the most important goals of education at all levels is the preparation of students for employment. Obviously, jobs differ in the skills required, in their importance to society, and in the education that they need. Here, however, the analogy of the calculation of private sector productivity is closer, because competitive labor markets are often interpreted as paying different wages to reflect these different attributes of jobs. Thus, an extended measure of educational output may come from the wages that are commanded in the labor market.

Such considerations are especially important given the fact that not all high school graduates go on to college. For those who enter the workforce directly out of high school, employment data could be used. And comparisons across different levels of schooling give a way of comparing the output of different levels of output. (As with a number of these various factors, some interpretive issues will necessarily remain. For example, an issue considered in the labor economics literature is whether people completing college education differ from those completing only high school education in other ways that are unrelated to the output of schools—for example, ability, motivation, or family background.)

*Tests as proximate measures of outputs.* The measures of education output outlined above may be useful in analyzing productivity in education, particularly when combined, and when examined over time. If attendance, graduation rates, college attendance, and postgraduate employment were all declining while inputs were on the rise, for example, we would have strong evidence that education productivity was falling. However, these measures are not fine-grained enough to give policy makers or educators much useful information on what students are learning and not learning, information that could drive efforts to improve productivity in education. For more detailed insight, we might use tests as the output measure.

In advancing ideas of productivity for decision makers, it is also problematic if observations can only be made many years after the schooling is actually conducted—when students graduate from college or find later employment. Fortunately, we have found that achievement tests in fact measure many of the skills that are valued in the economy (Hanushek, Schwerdt, et al., 2015; Hanushek & Woessmann, 2008, 2015a). In simplest terms, individuals with more skill as identified by these tests systematically earn more than those with less skill. And, nations that have more skilled populations grow more rapidly than those with less skilled populations. Thus, while the tests themselves might not be what is valued, the skills that they capture are a good indicator of important elements and dimensions of the ultimate outcomes of schools.

There are several tests that are currently being administered to American students that could serve as useful output measures in education productivity analysis.

*NAEP.* One useful set of assessments to measure educational outputs is the NAEP. There are two different sets of NAEP tests. The Long Term Trend Assessment measures a nationally



representative sample of students in reading and math at ages 9, 13, and 17 years. The assessment has been conducted approximately every 4 years, and the content remains constant over time so that changes in student achievement can be measured. It was first conducted in 1971 for reading and 1973 for mathematics, allowing comparison of scores over time for the past 40 years. In addition to overall achievement, results are reported by subgroup, including race, gender, and family income. The NAEP Long Term Trend Assessment is the only test that measures nationally representative student achievement over time, and is therefore useful in measuring changes in education productivity at the national level. However, only reading and math are consistently tested, and these subjects are certainly not the only educational outputs that most people would consider important. (Science was tested until the last decade, but the assessments were subsequently altered so that consistent performance measures through today are unavailable).

The Main NAEP Assessment, now the centerpiece of the assessment program, differs from the Long Term Trend Assessment in several ways. First, the content of the assessments changes over time. Therefore, Main NAEP cannot be used to measure longitudinal changes in education outputs reliably. However, it does provide more detailed information than the Long Term Trend Assessment to measure education outputs at a given point in time. Main NAEP also assesses a nationally representative sample of students at three levels: Grades 4, 8, and 12. In addition to reading and math, assessments are conducted in science, writing, the arts, civics, economics, geography, and U.S. history at selected grade levels. This is useful in terms of measuring education outputs, as schools are charged with teaching more than just math and reading. Also useful is the fact that Main NAEP scores are not only nationally representative, but representative at the state for Grades 4 and 8, and in some cases the district level.<sup>4</sup> This allows for comparisons across states and large urban districts. Main NAEP scores could be used as outputs when comparing productivity in education at a given point in time across states or districts.

*State tests.* The No Child Left Behind Act<sup>5</sup> required states to test all students in reading and math every year in Grades 3 to 8 and once in Grades 10 to 12. It also requires that all students be tested in science once in Grades 3 to 5, once in Grades 6 to 9, and once in Grades 10 to 12. In addition to these requirements, many states test students more frequently, in additional subjects and in further grades. The specific content of these tests is left to the states, so standards across states may vary widely. States that adopt Common Core Standards may move toward one of two common assessments which would allow more cross-state comparisons, but currently such comparisons are difficult. In addition, states may change their assessments over time, limiting longitudinal comparisons within a state. As state tests are administered to every student in the tested grades, they are a useful output measure to compare education productivity in districts, or even schools, within a state. To the extent that the content is constant over time, they may also be used to track changes in productivity at the district or school level.

*College entrance tests: ACT/SAT.* The ACT and SAT are both standardized tests commonly used for college admissions. Typically, students who plan to attend college take one or both of these exams in their final 2 years of high school. Both exams are designed to measure high school achievement and to predict college success, so the scores on these exams could be a useful measure of education output. However, it is important to note that in most cases, students self-select to take these tests. If only students who plan to attend college take the exams, their scores do not accurately represent the achievement level of all students. Because a commonly accepted goal of K-12 education is to prepare students for college, however, ACT and SAT scores may be a useful indicator of how well they are doing so among the self-selected group of students that plan to attend college. These scores might be combined with other indicators to measure education outputs, but probably cannot paint a complete picture on their own. The exception to this is found within the limited number of states where students are required to take the ACT to graduate from

high school. ACT scores in these states could be used as outputs to compare education productivity across districts or schools, and within the state, district, and schools over time.

*International tests—PISA and TIMSS.* International comparisons and country-specific data on performance in mathematics and science can be obtained from two different sets of international tests.<sup>6</sup> Since 1965, what is currently called the Trends in International Mathematics and Science Study (TIMSS) has tested students in a number of voluntarily participating countries. This has been followed more recently by the Programme for International Student Assessment (PISA) of the Organisation for Economic Co-Operation and Development (OECD). These assessments provide performance data every 3 or 4 years that, for the past two decades, can be used to observe performance over time on a consistent basis. Although the number of tested students is notably smaller than available through the NAEP or SAT data, these assessments can be linked to changes around the world. These examinations also permit international comparisons (as seen below).

## Trends in Education Productivity

Although we have attempted to provide a somewhat nuanced view of various measures that could be used in one way or another to calculate productivity change in education, it is also useful simply to look at some of the data about education and to consider what has been behind the long-run trends in productivity. If this empirical work were done completely, we would undoubtedly want to think about some composite performance measures.

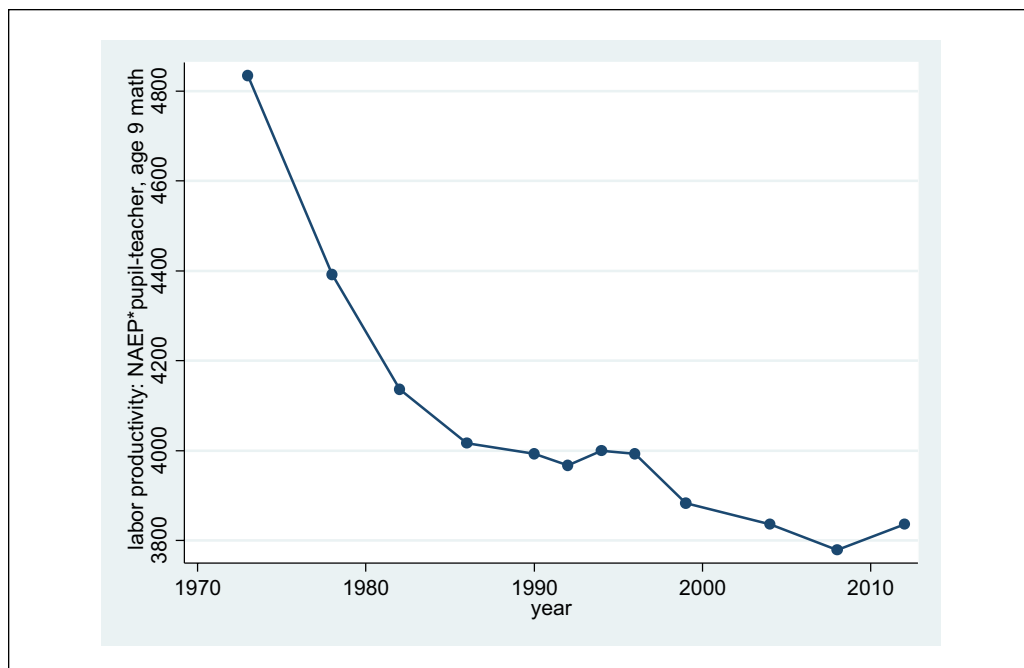
When presenting their data on education productivity, the U.K. Office for National Statistics points out that “it is unlikely that a single measure of productivity change will ever capture all the costs and benefits of education” (Baird et al., 2010). Various weaknesses in the possible output measures described above certainly lend weight to this assessment. It may make sense, therefore, to use a composite output measure consisting of several different outputs. We have little experience with determining which measures are used or how they should be weighted in constructing the composite.

At the same time, for the current illustrative purposes, the choice of output measures is not overly important. Why? Because many of the outcomes discussed, including high school graduation (Heckman & Lafontaine, 2010) and Long Term Trend performance on NAEP for 17-year-olds (U.S. Department of Education, 2005), have been virtually constant over the past three to four decades, there is not much distortion from selecting a few of the possibilities as examples.<sup>7</sup>

With the caveat that there are alternative possible measures for inputs and outputs, we next construct some basic measures of education productivity so that trends in education productivity can be compared to the trends in private sector productivity shown above. Figure 3 shows one example of labor productivity in education (outputs/labor input), with the input defined as teachers per student (the inverse of class size) and the output defined as score on the NAEP Long Term Trend Reading Assessment for 9-year-olds. For the sake of clarity, we show only one NAEP score, though graphs of labor productivity for the other tested areas of NAEP display a similar downward trend. Both input and output measures are national averages, and teacher–pupil ratios are averaged over the 5 years that a student at age 9 years will typically have spent in school.

The pattern of productivity change seen in this figure is quite clearly strongly falling. Although there is some variation and even a few periods where the productivity appears to have risen somewhat, there is no avoiding the conclusion that schools have exhibited a long-term decline in productivity by this measure.

Figure 4 is an example of Multifactor Productivity in education, with the input defined as cumulative current expenditure (in 2013 constant dollars) per pupil in average daily membership,



**Figure 3.** An example of labor productivity in education: Teachers per student and 9-year-old mathematics.

Source. Pupil-teacher ratios: U.S. Department of Education, National Center for Education Statistics, Digest of Education Statistics 2013, NAEP scores: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, NAEP Long-Term Trend Reading Assessments.

Note. NAEP = National Assessment of Educational Progress.

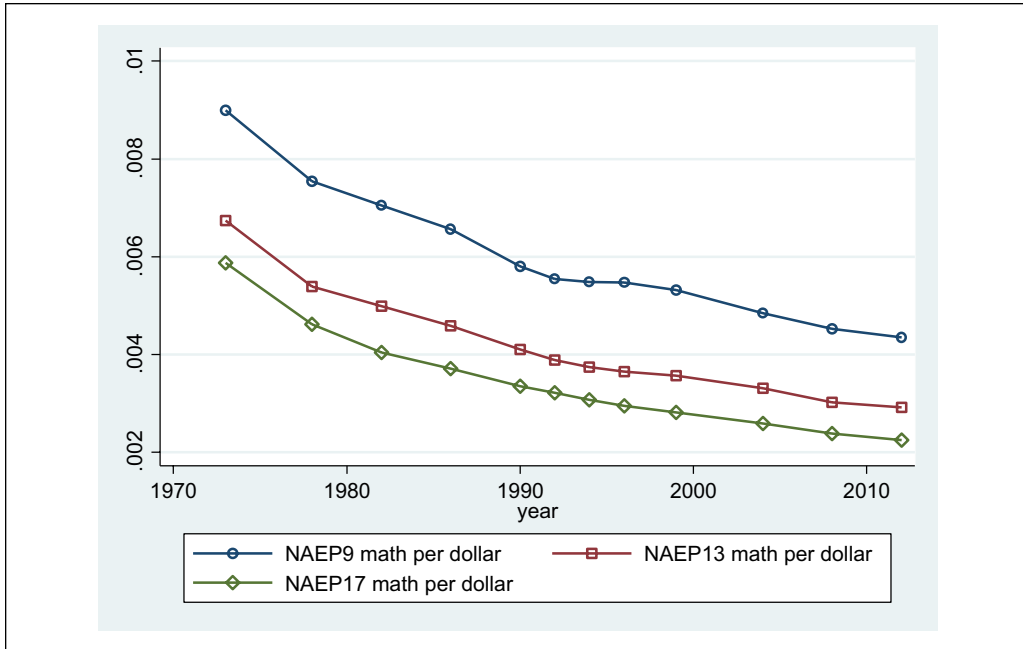
and the output defined as score on the NAEP Long Term Trend Math Assessment for 9-year-olds (as above), 13-year-olds, and 17-year-olds. Defined this way, there is a steady decline in productivity across all age groups.

Whether we look at labor productivity or multifactor productivity, it is clear that education productivity in the United States has declined over the past 40 years. The NAEP scores used in the calculations have increased over time for the youngest groups, while they have been quite flat for 17-year-olds. But even the increases at lower grades do not keep pace with the additional money spent or teachers hired to achieve them.

None of the NAEP measures or other potential outputs shows a sharp increase that matches the increases in labor and in spending. For example, while there have been some increases in college attendance rates, the rates of college completion have changed much less. Similarly, SAT and ACT scores have remained essentially flat for a long period of time, even if adjusted for participation on these tests.

The productivity decline displayed in the figures is caused by increased inputs (both in labor and in overall dollars) without a corresponding increase in outputs—a simple enough mathematical outcome. However, why increased inputs have not resulted in increased outputs is not as clear.

Interestingly, this picture of declining productivity in U.S. education also can be found in international evidence. Figure 5 uses the international achievement and international spending data to show that across countries, there is no consistent pattern of achievement gains accompanying increases in school resources. This figure plots the change in school spending per student between 2000 and 2010 against the change in international math and science achievement



**Figure 4.** An example of multifactor productivity in education: Cumulative PPE and NAEP math performance (NAEP points per dollar, age 9, 13, and 17 years).

Source: PPE: U.S. Department of Education, Digest of Education Statistics, 2013; NAEP scores: Data on NAEP Long-Term Trend Mathematics Assessments from data explorer at <http://nces.ed.gov/nationsreportcard/ltrdata/>.

Note: PPE = per-pupil expenditure; NAEP = National Assessment of Educational Progress.

according to the PISA examinations over the period 2000–2012. Although some countries have seen achievement gains with added spending (e.g., Poland), many have not.

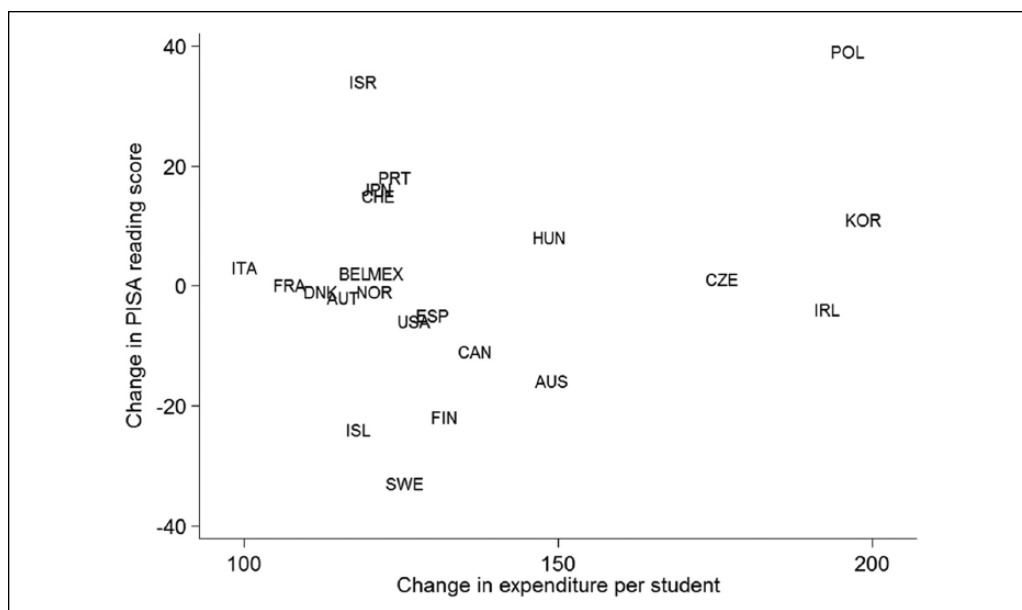
## Some Possible Explanations for Declining Productivity

We are not in a position to explain the decline in educational productivity fully. We can, however, address some of the issues that have been discussed and that potentially contribute to the overall outcome.

### Public Employee Unions

One of the major institutions involved in the American education system today is teacher unions. However, this was not always the case. The National Education Association (NEA) was founded in 1857, but operated as a professional organization run by administrators, not a labor union. Teacher unions, primarily the NEA and the American Federation of Teachers (AFT) did not begin collective bargaining until the 1960s, when many states adopted labor laws for the public sector that allowed this (Moe, 2011). Currently, approximately 65% of teachers are covered by collective bargaining, and 79% are union members (the difference occurs because some states, particularly in the south, have teacher unions but do not allow collective bargaining; Moe, 2011).

This rise in union membership and collective bargaining may be one factor in declining education productivity, especially because public sector unions do not have the fear of putting their employer out of business to temper their demands. If private sector unions demand salaries or working conditions that seriously damage the productivity of their employers, they risk putting



**Figure 5.** Changes in educational spending and in student achievement across countries.

Source: Hanushek and Woessmann (2015a).

Note. Scatter plot of the change in expenditure per student, 2000-2010 (constant prices, 2000 = 100) against change in PISA reading score, 2000-2012. PISA = Programme for International Student Assessment.

their employer out of business and losing their jobs completely. Teacher unions face no such risk when making demands, as public schools cannot go out of business, and therefore have no incentive to consider the productivity of schools when negotiating with their employers.

### *Baumol's Disease*

One common explanation for the productivity problems in education is the need for schools to compete with private industry to obtain teachers. The argument, developed by Baumol and Bowen (1965, 1966), was first developed for the arts. If there is an industry experiencing rapid productivity advances, it can afford to pay its workers more because they are producing more output for their work. If there is another industry that has no change in productivity but competing for the same labor inputs, it must adjust by either lowering quality or becoming more expensive. This simple argument is used to explain the rapid rise in education costs over the past century.

But this argument is neither a sufficient explanation for the observed performance of schools nor is it a statement of inevitability. The competitive story would offer some explanation for why teacher wages have risen over time (or teacher quality has gone down). But, since labor is the expensive input, a competitive story cannot explain why schools have in fact moved toward using more of the expensive input—through dramatic reductions in pupil–teacher ratios over time.

Hanushek and Rivkin (1997) find that the rising price of instructional staff, driven by growth in real wages in the overall economy, is a key factor in explaining growth in per-pupil spending in the past 100 years—but particularly over the past few decades, the reductions in pupil–teacher ratios are most important. Thus, the “competitive response” explanations of Baumol's Disease do not correspond with the observed actions of schools.

An argument sometimes added to this story is that the reductions in pupil–teacher ratios were made in an effort to improve outcomes—that these were investments that were needed to ensure

higher performance. Although that might be an explanation for the reasoning behind the decisions, it does not change the picture about productivity because student outcomes failed to improve in response to these reductions in pupil–teacher ratios.

The idea of Baumol’s Disease is that there is an inevitability to the problems facing sectors like education that have low productivity growth. It still takes a teacher just as long to teach a lesson or grade a paper as it did 100 years ago. Therefore, if we identify the output of education as “lessons taught” or “papers graded,” the declining productivity associated with Baumol’s Disease seems inevitable. However, education outputs may be (and arguably should be) defined differently, as some measure of what students have actually learned. If outputs are defined in this way, advances in technology can help education overcome Baumol’s Disease as has occurred in other sectors.

Triplett and Bosworth (2003) declare Baumol’s Disease “cured” in the U.S. private service industries, citing average labor productivity growth rates after 1995 that are equal to or higher than those of the economy as a whole. They attribute this in large part to investments in information technology. Hill and Roza (2010) discuss how similar gains could be made in education. They recommend that K-12 education look to other forms of organized learning, such as distance learning systems, franchise tutoring programs, military training, industry training, and the education systems of other countries for examples of how learning could take place differently (and more efficiently) than it currently does in traditional public schools.

### *Policy Focus on Distribution*

With the advent of No Child Left Behind, there has been increased attention to the achievement of various disadvantaged groups and to bringing up the bottom end of the distribution. As a result, even though both labor inputs and total spending have increased, if focused on particular subsets, the results might not show up in the overall average scores. We have seen that achievement for those aged 9 years has increased noticeably more than achievement for older students, consistent with the No Child Left Behind (NCLB) focus on earlier grades. (But, as seen above, the increases in the lower grades were insufficient to overcome the larger increases in resources.)

It is true that both the achievement gaps between whites and blacks and between whites and Hispanics have narrowed during the period after NCLB. Figures 6 and 7 show that the scores of both Black and Hispanic 9-year-olds have moved closer to those of Whites during the last decade, even if they remain very large.

The closing of racial and ethnic achievement gaps also shows up at age 17 years, but it is less pronounced since the introduction of NCLB (see Figures A1 and A2 in the Appendix). That is perhaps not so surprising, given that NCLB was more aimed at earlier grades.

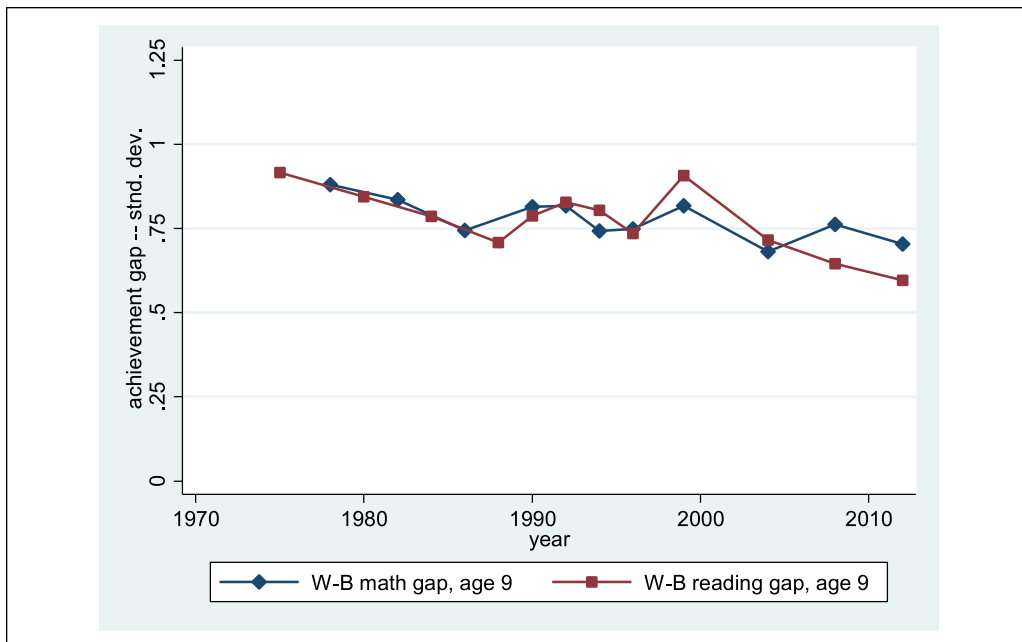
Nonetheless, the question that remains is whether this narrowing is sufficient to match the growth in inputs and to explain the fall in aggregate productivity.

### *Changes in the Quality of Labor*

Two major changes have occurred in the teacher labor market that may have affected the overall quality of labor in education and possibly contributed to the decline in education productivity. First, smaller classes require a greater number of teachers, meaning that administrators may have to hire less-skilled candidates to fill the additional jobs. Second, as a wider variety of employment opportunities have become available to women, the most qualified women may no longer choose to become teachers.<sup>8</sup>

### *Changes in the Quality of “Raw Materials”*

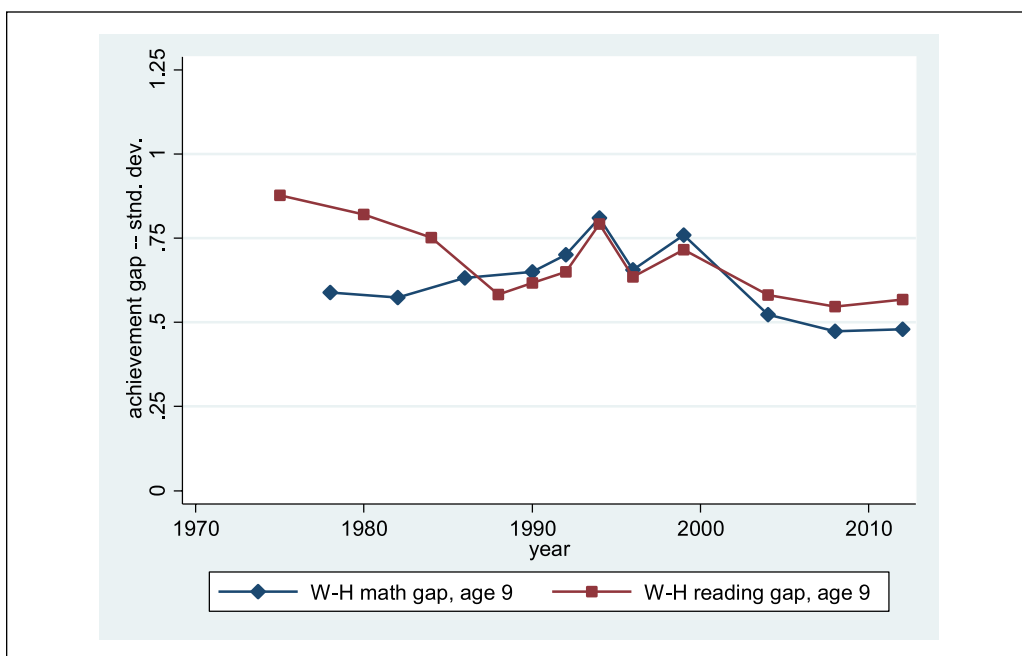
Students could be thought of as the “raw materials” in the education system. If students are coming to school less prepared to learn, this may cause declining productivity. A 2004



**Figure 6.** Black–White achievement gaps, 1975–2012.

Source. NAEP scores- Data on NAEP Long-Term Trend Mathematics Assessments from data explorer at <http://nces.ed.gov/nationsreportcard/ltrdata/>.

Note. NAEP = National Assessment of Educational Progress.



**Figure 7.** Hispanic–White achievement gaps, age 9 years.

Source. NAEP scores- Data on NAEP Long-Term Trend Mathematics Assessments from data explorer at <http://nces.ed.gov/nationsreportcard/ltrdata/>.

Note. NAEP = National Assessment of Educational Progress.

Manhattan Institute study by Greene and Forster investigates claims that higher spending has not led to increased achievement because today's students are harder to teach than those in years past. In the first analysis of its kind, Greene and Forster (2004) construct a Teachability Index composed of 16 social factors in the categories of school readiness, economics, community, health, race, and family and measure improvement or decline in each from 1970 to 2001. They find that declines in teachability on the race and family indicators are more than offset by increases in school readiness, economics, and community, resulting in an overall increase in teachability of 8.7%.

### **Resource Reallocations**

Major reallocations of resources have not occurred in education. Increased spending levels largely reflect increased labor costs, driven by both the rising price of teachers and falling class sizes (Hanushek & Rivkin, 1997). However, this need not be the case in the future—resources could be reallocated to promote efficiency in a variety of ways.

There has been extensive research into ways to improve school outcomes and also ways that have not proven to be generally successful. This research has been controversial, in part because it conflicts with many policy proposals that are being debated. Consistent with the discussion of productivity changes, this research has shown that simply providing more resources without changing the way that resources are used does not consistently raise student outcomes (Hanushek & Woessmann, 2015a). In particular, specifying inputs to the process—including spending, class size reduction, or the like—do not seem to raise outcomes consistently when they are not linked to outcomes and to incentives for the people in schools (Hanushek, 2003).

The most common view from the research is that a focus on teacher quality is essential, but the best ways of ensuring high teacher quality is not entirely clear. The policies that appear to be effective from the international evidence involve changing incentives in the system and include direct rewards for effective teachers, strong school accountability, greater local decision making, and parental choice of schools such as through charter schools in the U.S. (Hanushek & Woessmann, 2015a). Each of these policies appears to promote better resource allocation and greater achievement even if none by themselves present complete answers to the productivity problems in schools.

### **Conclusion**

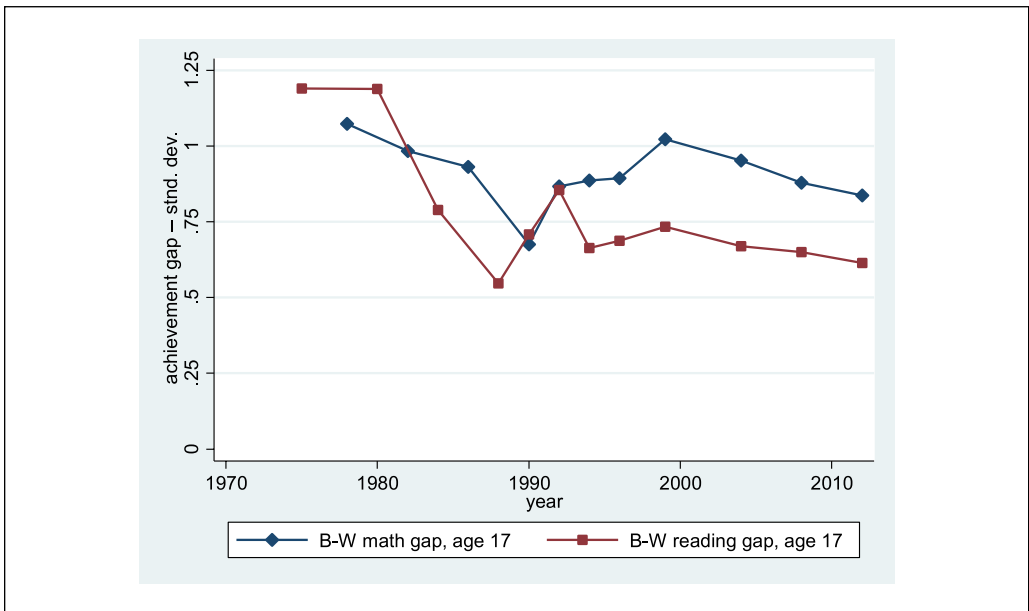
Private sector productivity in the United States has increased in recent years while productivity in education has declined. Although there are many ways that education productivity could be defined and measured, it is unlikely that any reasonable measure would show a great productivity increase, given a major escalation of inputs (whether measured by labor or dollars) with most output measures remaining relatively unchanged.

Education policy has traditionally ignored any consideration of productivity. Indeed, considering productivity has generally been viewed as something bad, something that can only have bad implications for educational quality and for school policy (Callahan, 1962). In today's world of fiscal imbalances and budgetary pressures, it seems impossible to continue ignoring productivity.

Given the disconnect between inputs and outputs and the fact that inputs are unlikely to continue to increase as quickly as they have in the past (Guthrie & Ettema, 2012), a close examination of the current practices that undermine productivity increases in education and a thoughtful exploration of alternative practices that might reverse this trend is in order.



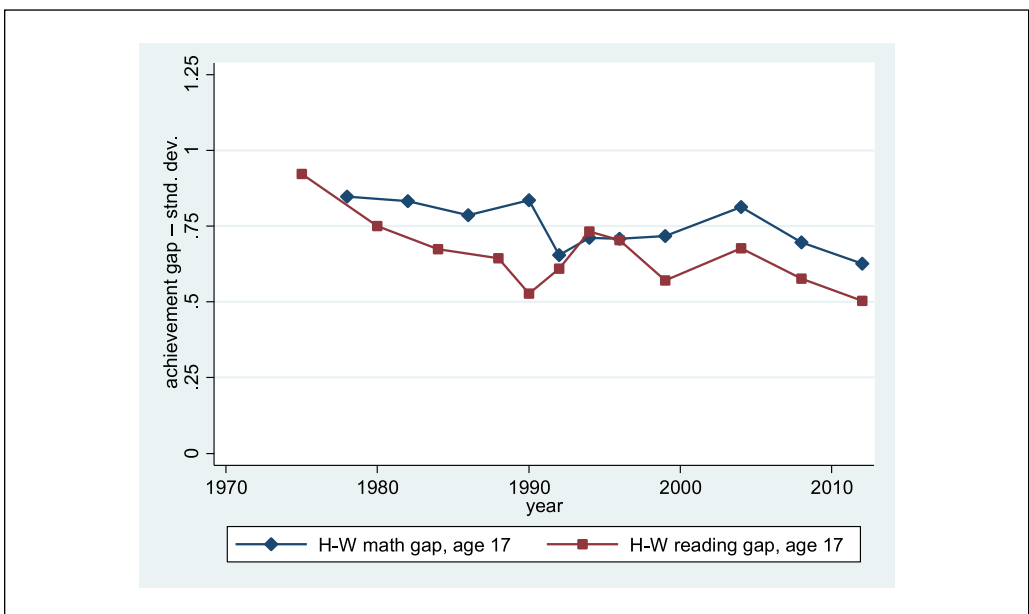
## Appendix



**Figure A1.** Black–White achievement gaps, 1975–2012.

Source. NAEP scores: Data on NAEP Long-Term Trend Mathematics Assessments from data explorer at <http://nces.ed.gov/nationsreportcard/ltrdata/>.

Note. NAEP = National Assessment of Educational Progress.



**Figure A2.** Hispanic–White achievement gaps, 1975–2012.

Source. NAEP scores- Data on NAEP Long-Term Trend Mathematics Assessments from data explorer at <http://nces.ed.gov/nationsreportcard/ltrdata/>.

Note. NAEP = National Assessment of Educational Progress.

## Authors' Note

The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of TNTP.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

## Notes

1. The National Assessment of Educational Progress (NAEP) is a nationally representative test in different subjects of students at different ages. It is often called "The Nation's Report Card" (see <http://nces.ed.gov/nationsreportcard/>).
2. This study is closely related to analysis in Texas (Combs, 2010).
3. The Common Core is a set of content standards for each grade and subject. It was designed to unify the standards across states, but it became controversial because it impinged on the right of each state to control its own schools. For a description of the effort, see <http://www.corestandards.org/>.
4. The focus on test measures of the quality of outputs matches other analysis that shows the importance of achievement and quality for growth—growth of states (Hanushek, Ruhose, & Woessmann, 2015) and of nations (Hanushek & Woessmann, 2015a, 2015b). There is strong evidence that the quality of human capital is by far the most important way to judge prior investments in skills.
5. The No Child Left Behind Act was federal legislation that governed accountability across all U.S. states. It was introduced in 2001 (after many states had already set their own systems of accountability). It was significantly modified in 2015, but state testing remained required in the follow-on Every Student Succeeds Act.
6. The history of the international assessments can be found in Hanushek and Woessmann (2011).
7. Murnane (2013) suggests, however, that there may have recently been increases in graduation rates.
8. Hanushek, Piopiunik, and Wiederhold (2014) find that expansion of women into other occupations has reduced the average skills (measured by test scores) of people entering teaching.

## References

- Atkinson, T. (2005). *Atkinson review: Final report. Measurement of government output and productivity for the national accounts*. New York, NY: Palgrave Macmillan.
- Baird, A., Haynes, J., Massey, F., & Wild, R. (2010). *Public service output, input and productivity: Education*. Newport: UK Centre for the Measurement of Government Activity.
- Baumol, W., & Bowen, W. (1965). On the performing arts: The anatomy of their economic problems. *The American Economic Review*, 55, 495-502.
- Baumol, W., & Bowen, W. (1966). *Performing arts: The economic dilemma*. New York, NY: The Twentieth Century Fund.
- Boser, U. (2011). *Return on educational investment: A district-by-district evaluation of U.S. educational productivity*. Washington, DC: Center for American Progress.
- Callahan, R. (1962). *Education and the cult of efficiency*. Chicago, IL: The University of Chicago Press.
- Combs, S. (2010). *Financial Allocation Study for Texas 2010*. Austin: Texas Comptroller of Public Accounts.
- Greene, J., & Forster, G. (2004). *The Teachability Index: Can disadvantaged students learn?* (Education Working Paper 6). New York, NY: Center for Civic Innovation, The Manhattan Institute.
- Guthrie, J., & Ettema, E. (2012). Public schools and money: Strategies for improving productivity in times of austerity. *Education Next*, 12(4), 18-23.
- Hanushek, E. (2003). The failure of input-based schooling policies. *Economic Journal*, 113(485), F64-F98.

- Hanushek, E., Piopiunik, M., & Wiederhold, S. (2014). *The value of smarter teachers: International evidence on teacher cognitive skills and student performance* (NBER Working Paper No. 20727). Cambridge, MA: National Bureau of Economic Research.
- Hanushek, E., & Rivkin, S. (1997). Understanding the 20th century growth in U.S. school spending. *Journal of Human Resources*, 31, 34-68.
- Hanushek, E., Ruhose, J., & Woessmann, L. (2015). It pays to improve school quality: States that boost student achievement could reap large economic gains. *EducationNext*, 16(3), 16-24.
- Hanushek, E., Schwerdt, G., Wiederhold, S., & Woessmann, L. (2015). Returns to skills around the world: Evidence from PIAAC. *European Economic Review*, 73, 103-150.
- Hanushek, E., & Woessmann, L. (2008). The role of cognitive skills in economic development. *Journal of Economic Literature*, 46, 607-668.
- Hanushek, E., & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 89-200). Amsterdam, The Netherlands: North Holland.
- Hanushek, E., & Woessmann, L. (2015a). *The knowledge capital of nations: Education and the economics of growth*. Cambridge, MA: MIT Press.
- Hanushek, E., & Woessmann, L. (2015b). *Universal basic skills: What countries stand to gain*. Paris, France: Organisation for Economic Co-Operation and Development.
- Heckman, J., & Lafontaine, P. (2010). The American high school graduation rate: Trends and levels. *The Review of Economics and Statistics*, 92, 244-262.
- Hill, P., & Roza, M. (2010). *Curing Baumol's disease: In search of productivity gains in K-12 schooling*. Seattle, WA: Center on Reinventing Public Education.
- Kendrick, J., & Grossman, E. (1980). *Productivity in the United States: Trends and cycles*. Baltimore, MD: Johns Hopkins Press.
- Moe, T. (2011). *Special interest: Teachers unions and America's public schools*. Washington, DC: The Brookings Institution.
- Murnane, R. (2013). U.S. high school graduation rates: Patterns and explanations. *Journal of Economic Literature*, 51, 370-422.
- Triplett, J., & Bosworth, B. (2003). Productivity measurement issues in services industries: "Baumol's disease" has been cured. *FRBNY Economic Policy Review*, 9, 23-32. <https://www.newyorkfed.org/medialibrary/media/research/epr/03v09n3/0309trip.pdf>.
- Triplett, J., & Bosworth, B. (2004). *Productivity in the U.S. services sector: New sources of economic growth*. Washington, DC: Brookings Institution Press.
- U.S. Department of Education. (2014). *Digest of Education Statistics, 2013* (Advanced Release ed.). Washington, DC: National Center for Education Statistics.

## Author Biographies

**Eric A. Hanushek**, senior fellow at the Hoover Institution of Stanford, has been a leader in the development of economic analysis of educational issues. His research spans such diverse areas as the impacts of teacher quality and the role of cognitive skills in international growth and development.

**Elizabeth Ettema** is an analyst at TNTP (formerly The New Teacher Project). She works on issues of school culture, including issues of how school culture impacts teacher retention and student outcomes.