2011

Getting Ahead or Left Behind? The Politics and Policy of Education Reform in the United States

Anna Elisabeth Brown
Oberlin College

Follow this and additional works at: https://digitalcommons.oberlin.edu/honors

Part of the Political Science Commons

Repository Citation
https://digitalcommons.oberlin.edu/honors/400

This Thesis is brought to you for free and open access by the Student Work at Digital Commons at Oberlin. It has been accepted for inclusion in Honors Papers by an authorized administrator of Digital Commons at Oberlin. For more information, please contact megan.mitchell@oberlin.edu.
Getting Ahead or Left Behind? The Politics and Policy of Education Reform in the United States

Anna Brown

American Politics Honors Thesis
Michael Parkin, Advisor
Paul Dawson, Second Reader
April 26, 2011
CHAPTER 1. Introduction

Education is a broad topic. It can be reasonably divided into many age categories, including but not limited to: early childhood education, elementary schools, secondary schools, two and four year universities, graduate and professional programs, and adult job training.

From there, even more nuances exist. Everyone agrees that the American education system needs to be improved, but how can improvement even be measured? Test scores, dropout rates, and other statistics are popular methods, but school systems and states can manipulate numbers—even unintentionally—to gain more money for their districts. Some critics say that standardized tests don’t truly measure learning. Test-taking in itself is a skill that must be taught for success, which brings the question: do tests measure learning or do they measure test-taking ability? And when teachers must take time out of their lesson plans to teach these test-taking strategies, are the students benefiting? Using grades or grade point averages to measure improvement and academic success brings the issue of grade inflation and non-standardized grading procedures. Perhaps acceptance to college could be a good measure, since most colleges examine academic records holistically and consider test scores, GPA, extracurricular activities, and writing ability as part of a larger picture. But college acceptance can be influenced by a high school’s guidance program, and many qualified students choose not to attend college. School attendance, Advanced Placement classes taken, attitudes toward school—all will most likely be associated with academic success, but are also heavily influenced by socioeconomic status and other family factors just as much as education policy.

In this paper, I have chosen to use the test scores available in my dataset, because test scores have become the “gold standard” used in most education policy research today. I will try to be conscious of the potential pitfalls involving these statistics and incorporate that into my
analysis. I will look specifically at how teacher quality affects student learning—whether it has any effect, and what particular characteristics or qualifications have a positive effect.
CHAPTER 2. Review of the literature

The need for reform is uniformly accepted, but theories on how reform can be achieved vary widely. In this review of the existing literature, I will examine the various hypotheses and research on education reform in the United States, as well as the gaps of knowledge. I am focusing only on K-12 education, and am largely ignoring the arguments for and against school vouchers for private schools. Though experimenting with school vouchers may well be a valid method for improving the education experience for some children, the fact remains that vouchers are impractical for all children, and so the public school system will still need to be fixed.

Why must education be reformed?

Statistics show that American children lag behind nearly every other developed nation: our high school seniors rank twenty-fifth in math and twenty-first in science out of thirty developed nations.¹

Our students are failing to meet the basic standards required to become productive members of society. Between 1983 and 1999, 10 million students reached their senior year of high school but could not read at a basic level—and nearly double that are unable to do basic math.² 40% of 17-year-olds do not have the skills in math and 60% do not have the skills in reading to successfully work at a manufacturing company.³ Dropping out of high school is correlated with much higher rates of criminal behavior: 68% of inmates in Pennsylvania state prisons are high school dropouts, and they cost the state an average of $132,000 over the course

---

¹ Karl Weber, Waiting for "SUPERMAN": How We Can Save America's Failing Public Schools (New York: PublicAffairs, 2010).


of their prison term. By contrast, thirteen years of private school tuition costs an average of $99,600. This study doesn’t prove that dropping out causes incarceration, but implies that if we fix what is causing the high dropout rates, we may see a similar benefit in reduced incarceration rates.

The achievement gap between minority students and white students is bigger than ever. The average black or Hispanic 17-year-old student has equivalent scores on nearly every measure of academic achievement to the average white 13-year-old student. Only 51% of African American students and 55% of Latino students graduate from high school, compared to a (still poor) 76% rate for white students.

Many students are poorly behaved and many teachers do not have the skills to manage their classrooms or teach appropriately. In the 1996-1997 school year, 57% of public schools reported moderate to severe discipline problems while 4 in 5 teachers reported feeling underprepared to teach to high academic standards. Only 60% of public high school science teachers majored or minored in their main teaching subject in their undergraduate years.

Clearly, there’s a problem. These grim numbers and the call for reform began with the

---

4 Karl Weber, Waiting for "SUPERMAN": How We Can Save America's Failing Public Schools (New York: PublicAffairs, 2010).


6 Karl Weber, Waiting for "SUPERMAN": How We Can Save America's Failing Public Schools (New York: PublicAffairs, 2010).


start of compulsory education in the United States in the mid-19th century and continue to today.
No program has yet managed to turn the tide, and in fact, reform efforts have shown little progress.

*Education reform’s change through history*

Classical education is the basis for modern forms of education (these techniques are particularly apparent in schools teaching philosophy and law), and is still taught in a fairly pure form in universities such as St. John’s College. Under the classical curriculum, primary schools teach grammar, logic, and rhetoric, known as the *trivium*. Secondary schools primarily taught astronomy, arithmetic, music, and geometry, but provided context through lessons in history, fine arts, and natural sciences. Tertiary education was usually an apprenticeship.

Grammar schools became popular as institutions in the 19th century under the leadership of two British teachers named Andrew Bell and Joseph Lancaster. They developed the Monitorial school, which depends on advanced students acting as tutors and teacher’s assistants to help others learn. The mark this left on our modern education system is obvious in the ubiquitous teaching assistants in universities and the idea of peer tutoring in secondary schools. This method of teaching kept costs low and accessible to poor children because it depended on a mutual exchange of information to ensure learning.

The late 1800s and early 1900s gave way to ideas of educational progressivism. This line of thinking was proposed as an alternative to classical learning, and emphasizes “learning by doing.” John Dewey was its greatest advocate and he refined the theory by developing hypotheses on the psychology of learning and then testing them. Progressive education is characterized by a flexible curriculum that is influenced by the students’ interests, a wide variety of learning activities and materials, education in civic and social responsibility, assessment
through projects instead of tests, an emphasis on problem solving and critical thinking, and community service projects integrated into the curriculum. Progressive education today is winning new converts who are promoting it in opposition to No Child Left Behind and the principles behind it.

Between classical, monitorial, and progressive theories, several new ideas were born. The modern liberal arts college was born out of all three. Experiential education is closely related to progressive learning and focuses on the relationship between teacher and student and places a premium on learning through direct, hands-on experiences. Humanistic education, the theory from which Waldorf Schools developed, focuses on the whole body as a mode of learning. Emotional and social health, intellect, and career skills are all developed. Notably, humanistic educators encourage self-evaluation instead of grades and believe that testing is not a good measure of learning.

More recent research establishing that there are a variety of personality types and learning styles have created doubts about many of these theories. Forms of education that teach merely theory work well for certain individuals but fail others. The progressive or experiential style of education that is based in hands-on experimenting is helpful for some learners and lost on others.

Major education reform approaches and theories of today

Constructivism is a popular theory underlying many modern education reform efforts today. Constructivism views the learner as an individual with a unique personality and background that affect his or her worldview. The learner is responsible for finding motivation and actively participating in the process, with the teacher taking a role as facilitator to help the learner reach his or her own understanding. The teacher is expected to make learning a social,
dynamic process that changes to suit the needs of the learner. Constructivism is associated with experiential learning.

Multiple intelligence theory is a widely known, though somewhat controversial, theory of learning. It states that a student who excels in one subject does not necessarily have a higher intelligence than a student who does not excel in that subject, because students have strengths in different areas. The accepted areas of intelligence are: spatial, linguistic, logical-mathematical, bodily-kinesthetic, musical, interpersonal, intrapersonal, naturalistic, and existential.¹⁰ This theory points to less emphasis on standardized testing, especially IQ tests, because they focus primarily on only one area (usually logical-mathematical) and provide an unfair assessment of a student’s intelligence. Psychologists have criticized this theory for its lack of empirical evidence, but it is commonly used in schools today for its practical lessons in teaching. No Child Left Behind has not incorporated lessons directly from this theory. The act encourages teaching based on “scientifically based research practices” but is vague as to the specific theories they support or what qualifies as acceptable research practices.

Site-Based Management (SBM) is an effort to lift the burden on state and local bureaucracies by delegating reform and implementation to the individual school. SBM operates under the assumption that local parents, teachers, and students will know what is best for their own school.

In recent years, SBM has escalated to the point of allowing schools to break ties entirely with the school district and form charter schools while retaining eligibility for public funding. Charter schools have garnered mixed reviews. In the presence of bureaucratic failure, the

creation of a charter school can completely turn around educational experiences for the local students, but their quality and oversight varies greatly. The National Education Association has voiced its concern over several aspects, including whether they will serve all students equally, whether standards of accountability and oversight are adequate, and whether the creation of charters will decrease funding for the rest of the school district. Nevertheless, anecdotal evidence has shown some students making significant progress, and charter schools certainly have their proponents.

One of the most controversial topics in education reform today is that of government-funded vouchers for private independent and parochial schools. Availability varies by state—and even region, as school vouchers are much more common in the West and Midwest than in the Northeast and South. Proponents say that converting to a free market competition between private and public schools improve outcomes in both. This claim has been supported in some research. On the other side, many public school teachers—and the official stance of the National Education Association, the largest labor union in the United States—have come out against vouchers. The NEA says vouchers allow for less accountability of public resources because private schools do not require uniform academic standards; private schools do not always adhere to the same civil rights protections guaranteed in public schools and may discriminate in their admissions processes; and of course, vouchers only provide school choice


for a minority of lucky students. Experiments with school vouchers in other countries have been mostly unsuccessful. “Free schools” in Sweden resulted in eroding academic standards in the country’s schools and increased social segregation. School vouchers are primarily supported by voters on the right. The issue certainly deserves further study, but it is out of the scope of this particular paper.

Most education reform efforts have incorporated lessons from many of these modern theories in some form. The primary goal of all of these efforts is the same—to improve student learning outcomes. But they also integrate other objectives as well: minimizing the achievement gap; compensating teachers fairly; eliminating school violence and bullying; making the system more efficient (educating students more effectively for less money); and preparing students to meet the new needs of the modern workforce. These secondary goals are what complicate the issue, and introduce politics into the mix.

The politics of education reform

The classic literature on modern education reform describes a diminishing consensus: 93% of respondents to a Gallup poll agreed with the idea that the quality of public education should be improved; 73% agreed with the idea that the quality of public education should be improved by government action; and 47% agreed with the idea that the quality of public education should be improved by government action and new taxes. The typical dilemma in policy reform emerges—nearly all citizens want public services to improve, but many fewer

---


want improvement at the expense of raised taxes.

Going even further into the economic realm, teachers’ unions are an extremely powerful yet not very well known coalition. The NEA has 2.7 million members, representing nearly every teacher in the country, and is the largest union in the United States.\(^\text{17}\) It is also one of the largest campaign contributors, spending more than ExxonMobil, Microsoft, Wal-Mart, and the AFL-CIO combined. The NEA and the American Federation of Teachers (AFT) together spent over $71 million on campaign contributions in the 2007-2008 election cycle, mostly to Democratic candidates.\(^\text{18}\) Teachers unions were originally started in the face of the extremely unfair treatment they received in the early 20\(^{th}\) century and before—pay for women was particularly inequitable—but it’s clear that the teachers have gained serious political clout.

With that in mind, it is necessary to examine some of the NEA’s policy positions. No Child Left Behind and Obama’s current “Blueprint” are among their top focuses. NEA says that the federal government should develop more effective and fair assessments, rather than current tests, and focus on providing equal expectations and support for all schools, rather than encouraging competition for funds.\(^\text{19}\) In particular, the Minority Community Outreach program advocates better and more equal resources for schools with large minority populations and works to establish stronger ties with the communities surrounding these schools, coordinating with an effort to close the achievement gap.\(^\text{20}\) They also advocate for a living wage for all education


professionals, and a $40,000 starting salary for all K-12 teachers nationwide, with appropriate raises. They cite pay issues as a major reason for high turnover rates among teachers who are just starting out. The NEA endorsed a bill, passed in August 2010, which secured funding for 161,000 teachers to keep their jobs. They have endorsed making high school graduation a national priority and outlined a 12-point plan to achieve this—including making graduation compulsory for all children under 21 (like compulsory attendance laws until the age of 16 become the norm in the early 20th century), decreasing class size, establishing more career education programs, and connecting with parents.

Apart from these policy recommendations, though, the NEA’s heaviest influence comes from its representation of teachers. This most recently came to the forefront when Michelle Rhee, Chancellor of Washington, D.C. schools, negotiated a revolutionary contract with the unions that offered pay raises and bonuses for high performing teachers in exchange for lessening their tenure protection. Taking advantage of the weakened tenure, she fired hundreds of teachers who had questionable disciplinary histories and who scored poorly on teaching evaluations. This sparked a huge controversy and led to her resignation (see case study, below). The NEA held the cards in this situation—the deal they granted Rhee was much more than had been expected, but its failure in public opinion has likely doomed the chances of a repeat in the near future.


How has the legislation of the 20th century led to the modern state of education?

Modern education reform efforts began in earnest during the civil rights movement in the 1950s and ‘60s. Schools were desegregated and affirmative action policies, including busing, began to be implemented.

The 1950s also marked the rough beginning of presidential attention to the nation’s education. Starting with the Eisenhower campaign, every major presidential candidate has made education a platform issue. In 1952, Eisenhower stated the problem and the solution very simply: students were not learning, and federal aid without interference was the answer.\(^{25}\) Federal aid had been discussed in the 1930s and 1940s, but no bill had been passed. The public was overwhelmingly in favor of Eisenhower’s approach. However, the following four years were filled with partisan bickering, unproductive committees, and failed bills.\(^{26}\) In 1995, a White House conference on education, focusing exclusively on elementary and secondary education, endorsed federal aid for school construction on the basis of demonstrated need. And in 1958, the National Defense Education Act passed, providing scholarships to college students and aid to institutions if both were in the field of national defense.

The NDEA broke the barrier for more legislation and sparked a national interest in education. President Kennedy won the NEA’s support and continued the program. He turned his sights toward building schools and raising teachers’ salaries. During Kennedy’s administration, the fight over aid in parochial schools began—Kennedy had proposed aid only to public schools, but Catholic groups objected—and stalled his progress.

The period from 1956 to 1962 was marked by a “crisis” in higher education as prices rose. This helped turn around attitudes in Washington and opened up the possibility of increasing


\(^{26}\) *Ibid.*
federal spending. In 1963, Kennedy introduced two dozen wide-reaching programs, mostly providing scholarships and loans. The easy ones, including a vocational education bill, passed first; the more controversial, such as “selective aid” for elementary and secondary schools (meaning a bigger piece of the pie for schools in poor areas), came later. President Johnson began championing his education legislation and broke the traditional resistance to expenditures. By 1965, more than 10% of public expenditures were going towards educational at all levels.\(^27\)

As part of his Great Society, Johnson pushed equal opportunities for poor students, improved school libraries, and special education and gifted programs, but maintained that “federal assistance does not mean federal control.”\(^28\) The 1965 Elementary and Secondary Education Act (ESEA, or Title I) is the largest K-12 program to date and it continues to govern the allocation of funds to schools in underprivileged areas. No Child Left Behind and Obama’s education plan are offshoots of ESEA. The literature is mixed on the success of increased funding. Since there are so many potential outcomes to measure and so many methodologies to use, the success of any given program is frequently difficult to evaluate.

The 1980s brought a shift towards the right. Reagan’s *A Nation at Risk: The Imperative For Education Reform* was published in 1983 and is considered a key event in reform efforts. It outlined the many ways in which American schools had failed our children and proposed a series of recommendations that included standardizing the curriculum and improving teacher quality. Though the recommendations were rather non-specific, they did advocate for improved teacher salaries. E.D. Hirsch, a prominent educator at the time, elevated the idea of “cultural literacy” in the minds of reformers. Cultural literacy was the idea that there are some key facts that every American once knew and that are essential for understanding in higher education and the job

\(^{27}\) Ibid.  
\(^{28}\) Ibid.
This idea remains prominent today and is commonly present in elementary and secondary curricula.

In the 1990s, philosophy shifted towards an outcome-based model, an approach that uses methods from constructivist teachings and focuses heavily on student achievement. Congress established National Education Goals and devised standardized tests to measure progress. States also established individual objectives. This theory of reform evolved into the standards-based method of today, of which No Child Left Behind is the most prominent example. The ultimate goal for every high school graduate is to have a minimum set of skills in reading, writing, and math with the eventual goal of being a useful employee. Standards-based reform recognizes individual learning styles, but holds all students to the same standards, regardless of poverty, race, or gender. This type uses criterion-referenced tests, which measure how well a student has mastered the material against a set standard, rather than norm-referenced assessments, which compare students’ knowledge and abilities with those of their peers.

*How have today’s reforms overlooked teacher quality?*

Today’s reform efforts have focused on a variety of factors: a longer school day or year, school choice (charter schools and vouchers), reducing class size, providing more computers and other resources, bilingual education, adjusting curricula and textbooks, improving teacher quality (through adjusting credential requirements, improving training, raising salaries, merit pay, or firing poor performers), reducing absenteeism, establishing programs that track children from birth to graduation in an effort to create a more supportive home environment, and funding (which is necessary to all the other factors, and also addresses issues of poor infrastructure).

No Child Left Behind (NCLB), proposed by President Bush immediately after taking office in 2001 and supported by Senator Ted Kennedy, became law in early 2002 with
overwhelming support from both parties. Its full title is: an act to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind. NCLB is an example of standards-based education reform, which places an emphasis on setting standards and devising a way to measure achievement. Specifically, NCLB requires states to set their own achievement standards and make their own assessments. These benchmark assessments are given to students in certain grades and are required for the state to receive federal funding. The bill led to increased funding for education: overall congressional appropriations for education increased from $42 billion in 2001 to $54.4 billion in 2007; funding for No Child Left Behind increased 40% between 2001 and 2007 to a total of $24.4 billion; and allotted $1.475 billion in remedial math and reading programs.  

A 2005 press release from the Department of Education happily declares, “No Child Left Behind Act is working.” They cite several subgroups’ test statistics that grew to the highest levels in history (for example, 9-year-olds reading and math and African American 13-year-olds math).  Though this may be true, education reformers have criticized the administration’s report for omitting other subgroups. The press release did not include a full report of the findings of the National Assessment of Education Progress, and it is possible that other subgroups did not perform at the same level. In addition, the report measured progress between 2000 and 2005, but the law took effect in early 2002 and funding did not trickle down until 2003. Therefore, not all the improved scores can reasonably be attributed to the single law.


NCLB does have several progressive and widely applauded aspects. In particular, it aims to narrow the achievement gap by creating common expectations for all students and requires states to develop a plan to improve the academic achievement of underserved populations, often through greater parent involvement. NCLB puts a strong emphasis on developing basic skills in reading, math, writing, and science as core areas, and encourages early literacy. By requiring the use of “scientifically-based” teaching methods, NCLB aims to standardize and improve overall teacher quality. Overall, the law demands greater accountability from schools and teachers and links inputs (teachers and content) with outcomes (assessment). By measuring schools’ success rates, the government can incentivize them to improve and also provide parents with valuable information to allow them to change schools if necessary.

Critics may applaud the lofty goals that NCLB seeks to accomplish, but believe that there are better approaches. There are three basic problems with the Act. First, standardized testing inherently does not measure merely learning and mastery of skills.\(^\text{31}\) Instead, scores are influenced by what students are learning in school, what students are learning outside of school, and students’ natural intelligence, whereas the goal is generally stated as only testing what students are learning in school.\(^\text{32}\) Factors such as parental involvement, poverty, gender, and even such varying conditions such as the amount of sleep the student got the night before can all play into standardized test scores—quite a few of these are outside of the control of the school, teachers, and students. Researchers have developed an alternative method called the value-added model, which statistically controls for these factors by measuring only the differences between the test scores of a teacher’s students at the end of the year and the test scores of those same


students at the end of the previous year, effectively isolating only the factors of that particular school year (primarily, all else being equal, the teacher).\textsuperscript{33} The value-added model is increasingly being adopted as a new method of evaluating teacher performance across the country. Of course, though this may be an insurmountable problem, any type of testing always invites the attitude of “teaching to the test,” which encourages teachers to emphasize and drill only the material known to be on these high-stakes tests, to the exclusion of other potentially useful and interesting material. Art, music, foreign language, and gifted programs can be seen as less important when the tests stress only achieving the bare minimum in the core academic subjects.

Second, incentivizing with funding may not be the best way to improve learning. Schools with lower initial scores are likely to be in impoverished neighbors with students who may not be fluent in English or have parental support. These areas are naturally more difficult and will probably see more modest improvements in student outcomes, even with the same amount of funding. By giving more funding to high-performing schools and allowing students with the most involved parents to move to one of these high-performing schools, are we not placing the underserved schools at an even greater disadvantage?

Third, NCLB does not put enough emphasis on the goal of improving teacher quality. NCLB specifies that educators must use “scientifically based research” practices (SBR) but does not specify what these are. SBRs have been identified through commonly used practices from previous pieces of legislation, but NCLB’s specific wording may limit the use of innovative new teaching practices.\textsuperscript{34} The law requires that schools inform parents if their child has a teacher who


has not met the requirements to be considered “highly qualified,” but does not move toward eliminating teachers who are not “highly qualified.” 25% of funding through a program to improve technology in schools under NCLB is pegged for professional development, but for the most part, responsibility for hiring and training good teachers still lies on the states and school districts. The federal government does not provide adequate support through specific guidelines.

Upon taking office, Obama proposed revising NCLB. In a 2007 campaign speech, he said, “…I’ll tell you what’s wrong with No Child Left Behind. Promising high-quality teachers in every classroom and then leaving the support and the pay for those teachers behind is wrong. Labeling a school and its students as failures one day and then throwing your hands up and walking away from them the next is wrong.”

He proposes funding a wider range of assessments to test skills other than merely the core academic subjects and develop special tests for students with disabilities and those who do not speak English as a first language. Obama also wants to improve the program’s accountability system by creating incentives for raising graduation rates, to discourage schools from kicking underperforming students out to improve their scores.

Obama has made other steps towards education reform in his presidency so far. The Health Care and Education Reconciliation Act of 2010—which included the Student Aid and Fiscal Responsibility Act as a rider—ended the practice of federally subsidized loans through private banks and gave that power directly to the Department of Education, increased the Pell Grant scholarship award, capped loans at a percentage of income, and increased funding for historically poor and minority schools and community colleges.

---


His American Recovery and Reinvestment Act of 2009 also increased funding to elementary and secondary education and early childhood programs. This law introduced the Race to the Top program, which requires states to apply for funding and scores these applications based on a 500-point system. The Act lists a number of criteria to be considered, in order from highest possible points to lowest: great teachers and leaders (improving quality and providing career improvement opportunities; equally distributing the highest quality teachers and administrators), state success factors (articulating the state’s plan for reform), standards and assessments (adopting common standards; developing other high-quality assessments), general selection criteria (encouraging charter schools), turning around the lowest-achieving schools, and establishing data systems to support instruction.\(^{37}\) The Act provides funds to the states that score the highest on their applications, and 48 states have taken action already.\(^{38}\) The program is so new that the reforms have yet to be proven, but Race to the Top has been criticized by some education leaders and states rights activists for pressuring states to adopt national standards rather than customizing their standards for the states’ own students.

*Current research on the effect of teacher quality*

The latest research on teacher quality often holds that teacher quality differs greatly and may certainly have a strong effect on student learning, but no single teacher characteristic determines quality: most of the variation in quality goes unseen. This school of thought (known as effect size) says that using teacher characteristics in policy-making decisions is unreliable and discretion should be given to individual principals to evaluate the quality of and hire their own

---


teachers.

A 2004 study by Nye, Konstantopulos, and Hedges investigates just this principle.\(^{39}\) They performed a four-year experiment in which teachers and students were randomly assigned, and students were compared only to other students in their own school, to classes in an effort to isolate the effect of the teacher. They did find a significant difference, backed up by older research, in teacher quality but it was only present among 2\(^{nd}\) grade reading and 3\(^{rd}\) grade math scores. The differences were more pronounced in schools where students had a lower socioeconomic status.

Some research on the effect of teacher quality on student learning exists. Andrew J. Wayne and Peter Youngs summarized the entire body of research on teacher quality at the time of publishing in 2003.\(^{40}\) They found that the research fell mostly into four major variables: teachers’ college ratings, test scores, course taking and degrees, certification status. Evidence for these first four is considered conclusive; the remaining studies were categorized as “all other characteristics” and they are not conclusive.

Research on the quality of a teacher’s own college is limited. A study from the 1970s by Summers and Wolfe tested a variety of teachers from different academic subjects and grade levels and compared students scores.\(^{41}\) They found significant positive relationships for 6\(^{th}\) grade teachers and 8\(^{th}\) grade Social Studies teachers, but no significant relationships for any 12\(^{th}\) grade teachers. A 1981 Murnane and Phillips study found no relationship between teachers’ colleges

---


and black elementary school students’ test scores. The latest research project that tested this relationship was completed in 1994 by Ehrenberg and Brewer. They found that teachers from highly rated undergraduate colleges were more effective for white and black students. Results were inconclusive for Hispanic students. Wayne and Youngs conclude that more research is needed on this variable, but they also note that results varied widely depending on the method of rating the teacher’s college. Overall, the results pointed to a positive effect for at least some students.

Studies examining teacher test scores fall into the categories of teacher licensure examination scores, achievement studies that involve tests of teachers’ verbal skills, and later studies that use other test scores. Two studies examined teacher licensure examination scores. One discovered that 6th graders actually learned less when their teachers scored higher on the National Teachers Examination and the results from older students were inconclusive. Ferguson found that the same cohorts of students had more improvement on reading tests over a four-year period in districts where teachers had higher scores on the Texas Examination of Current Administrators and Teachers.

Studies that examined student learning in relation to tests of their teachers’ verbal skills mostly used data from the 1960s or before. Ehrenberg and Brewer (1995) demonstrated that

---


teachers’ scores on a verbal facility test explained some of the variation in achievement gains from school to school. Murnane and Phillips (1981) found no relationship between teachers’ scores on word tests and student achievement gains when controlling for other factors such as race, sex, experience, degree, and rating of the teacher’s undergraduate institution. Hanushek (1992) found that teachers’ scores on word tests improved their elementary school students’ scores on a reading test but not a vocabulary test over the course of a single year.

Rowan, Chiang, and Miller (1997) found that high school math teachers who answered a single math question that appeared on a survey had students who achieved higher gains between 8th and 10th grades—but a single math question is not a very reliable source of data. Ferguson and Ladd (1996) obtained teachers’ composite ACT scores and found that 3rd-4th grade student reading scores were positive related to their teachers’ ACT scores, but found no similar gains in math scores. Overall, Wayne and Youngs found five positives and two negatives for this category. They attribute the inconsistency to the varying use of control variables—the only two studies that controlled for the teacher’s undergraduate college rating were the two with negative results. This implies that the lone variable of verbal skills may actually have a significant positive effect. Wayne and Youngs conclude that policymakers should consider tightening requirements for teacher licenses.


The literature on teacher degrees and coursework is confusing and complicated. A number of early studies were inconclusive and the remainder was split: 1 positive and 3 negative. Wayne and Youngs do not accept the negative findings by sheer numbers, though. The single positive study (Ferguson & Ladd, 1996) is the newest and makes use of the latest data collection, which improves the ability to determine whether an additional degree is related to the teaching subject area.

Goldhaber & Brewer (1997) showed that a degree in Mathematics has a positive impact on a Math teacher’s students’ scores, controlling for the highest level of education (i.e., a teacher with a higher degree showed no more positive results than teachers with bachelor’s degrees in the same subject). Goldhaber & Brewer (2000) and Rowan, Chiang, and Miller (1997) found similar results—mathematics and science teachers had a positive relationship, but the relationship for humanities (History and English) teachers were indeterminate.

Goldhaber & Brewer (1997 & 2000) show that teacher certification has a discernible effect only when the certification is in Mathematics and the teacher is teaching in Mathematics. Their 2000 study further showed that student learning was higher in Math when the teacher held a standard certification, as opposed to a private school certification or no certification at all. These findings are similar to those for degree and subject matter—these factors all affect student performance for certain subjects (mostly math and science) but not for others.

Wayne and Youngs consider the factors mentioned above (college ratings, teacher test...


scores, degrees and coursework, and certification status) as having enough conclusive research to establish policy based on it. They mention other factors that either have not been studied sufficiently or have only flawed studies that cannot be interpreted reliably. These factors include teacher experience (which inadvertently captures the effect of circumstances under which a teacher was hired—i.e., under a shortage or a surplus—and differences in motivation), race (studies on whether students learn more from a teacher of their own race have been mixed), and holding a degree in education.

There has not been a great deal of research on American teachers in recent years, but one notable study from 2005 does in fact examine teacher quality research within the frame of standards-based reform (NCLB, in particular).54 The text of NCLB says that all teachers should be “highly qualified” and defines that as having a bachelor’s degree, being fully licensed or certified, and demonstrating content knowledge in his or her subject area (new teachers must pass a state academic test, or have majored or completed coursework in that area; experienced teachers must also show proof of some degree or coursework, but the guidelines are much more lax). Smith, Desimone, and Ueno point out that these requirements are all “proxy measures”—they do not directly cause better teaching; they are merely indicators. They choose to focus on the quality and content of teaching rather than student test scores because they see test scores as just another proxy measure. The authors found that a teacher holding a certification in Math was not more likely to emphasize conceptual learning goals. These teachers are more likely to have a high level of preparedness to teach Math, but they are also more likely to hold a degree in Math

or Math Education, implying that certification status may not be an adequate indicator variable. They did find that this preparedness explains the discrepancy between degree and use of conceptual learning goals. New teachers are more likely to rely on teaching strategies centered on procedure and orderliness, focusing on classroom management and “learning the system.” Teacher training tends to emphasize these aspects and not conceptual learning. The authors, controlling for other commonly associated factors, found that participating in professional development in the teacher’s field increased use of reform teaching strategies, suggesting that professional development activities could become an important part of meeting NCLB requirements.

Palardy and Rumberger conducted a study in 2008, using the Early Childhood Longitudinal Study to look at the effects of teacher background qualifications, attitudes, and instructional practices on reading and math achievement among first graders. They found that instructional practices and attitudes had more robust associations with achievements than did background qualifications, and concluded that the NCLB “teacher quality” mandates should focus more on classroom practices and attitudes than on teacher education and experience.

A study on beliefs of second-grade students’ and teachers’ beliefs, by Murphy, Delli, and Edwards, found a set of subjective characteristics that make a good teacher. They defined a good teacher, though, by students’ evaluations, not through tests or other measures of learning, operating on the theory that good teaching results in happy and satisfied students. They found that good teachers are “caring, patient, not boring, and polite.” Student-centered instruction with

---


the teacher moving around the classroom frequently is seen as an engaging characteristic. I take these results with a grain of salt because they only measure student perceptions and not necessarily student learning, and I don’t believe they can be extrapolated to form education policy as a whole because the needs of second graders must vary dramatically from those of older students, when understanding of the more challenging subject material may be more important than classroom practices.

An evaluation of the Tennessee Value Added Assessment System (TVAAS) by Kupermintz shows that the value-added system may be the best we have in terms of evaluating teacher effectiveness right now, but it is not perfect and does not necessarily stay in line with most mainstream theories of education. The TVAAS calculates the average performance gains in each school district and the average performance of a certain teacher’s students in relation to the rest of the school district. Therefore, it contains three components: a system (district) model, a school model, and a teacher model. Because it is a norm-referenced measure, a teacher’s score in a value-added model depends on the performance of all the other teachers in the district—so the same weak teacher could have a much higher score in a weak school system than in a strong one. Of course, it is difficult to isolate the effect of the teacher in a system like this—differences between schools and districts will inevitably come into effect. Limited data for some teachers also skews the results. When data is scarce (like for a part-time teacher, or one with few students), the statistical analysis steers the result towards the average. Of course, all the problems that using standardized tests as a measure bring are still present with the value-added model.

Still, Kupermintz concludes that it is a model with promise, but results should not be considered

authoritative without other convincing evidence.

In my opinion, research on the effects of teacher quality is limited and scattered. Though a number of high quality studies do exist, the sheer number of variables to test means that there is potential for a lot more investigation. The topic is more important than ever with only marginal claims of success under No Child Left Behind and the recent controversy surrounding Michelle Rhee’s focus on improving teacher quality.
CHAPTER 3. Data analysis

The research questions

How much of an effect does teacher quality have on student learning? What teacher characteristics show the greatest effects? And how can we apply this knowledge to improve the hiring and compensation practices for teachers?

The study

To investigate the answers to these questions, I used data from the Progress in International Reading Literacy Study (PIRLS) 2006, gathered by the International Association for the Evaluation of Educational Achievement (IEA). The 2006 database includes the results of surveys given to over 210,000 fourth grade students, their parents, and their reading teachers in 40 countries. I looked only at the datasets for the United States.

The target population was 4\textsuperscript{th} graders in the United States, where students have a mean age of 9.5 years and are in the process of becoming independent readers. Researchers tried to include every primary school in the United States for their population but were forced to exclude some schools for feasibility issues. These exclusions included schools that were geographically remote, had few students, had a curriculum that differed greatly from the mainstream education system, or were intended for students with special needs. Within schools, they also excluded disabled students and non-native English speakers if they were unable to perform normally in the testing situation.

Because the United States sample size was so large, researchers performed a two-stage sample, first selecting regions, then schools. They employed a respected method of sampling known as systematic probability proportional-to-size to select the schools and the students.
The dependent variable

Though there are of course inherent flaws in using student test scores as a measure of student learning, it remains the standard used in nearly all education policy studies. Test scores are the only dependent variable found in most datasets (PIRLS included) that measures overall learning.

The IEA developed a series of 10 passages (five literary and five informational) and a total of 126 questions (12 per passage) that tested reading comprehension. The passages were divided into rotating booklets and each student was given 80 minutes to complete a booklet consisting of two passages and 24 questions.

Given the inevitable error that must occur when students are given different questions, PIRLS provides five imputed estimates (known as plausible values) for each student for the informational passage and five imputed estimates per student for the literary passage. The dataset also provides five plausible values for an overall estimated reading score. The average score is designed to be 500 with a standard deviation of 100. To obtain a single dependent variable, I created a composite variable that averaged the five plausible values in the overall estimated reading score. That variable is listed as Reading_avg in my dataset. More details on the procedure in developing the assessment and the reading achievement scale can be found at the PIRLS website at http://pirls.bc.edu/PDF/p06_technical_report.pdf

The independent variables

I conceptualized a broad list of teacher characteristics among the variables measured in the teacher questionnaire that I thought could potentially affect their students’ learning outcomes. They are detailed below. Of course, it would be impossible to measure variables such as career satisfaction with total accuracy due to the potential biases present in the survey process, but
PIRLS follows generally accepted industry practices in developing their surveys to minimize this bias.

In addition to the assessment, PIRLS worked to gather several other components. Researchers distributed a Student Questionnaire, a Learning to Read questionnaire (to parents and caregivers), a Teacher Questionnaire, a School Questionnaire, and a Curriculum Questionnaire. I merged the Student, Learning to Read, and Teacher questionnaires into a single dataset. Though the other questionnaires certainly contain valuable information to my research question, I narrowed my scope to focus on teacher characteristics as my independent variables and individual student and home characteristics as control variables.

I selected a wide variety of independent variables designed to identify teacher characteristics that are associated with higher scores on the reading achievement scale. Some are based on variables that have been tested in other studies (though I have not found any other studies that used the PIRLS 2006 data) and some are original variables that I have not seen tested in any other study. They include:

- **Teach_age**: Teacher’s age; an ordinal variable with values 1-6, youngest to oldest
- **Teach_sex**: Teacher’s sex; a nominal variable
- **Teach_satisfact**: The Index of Teacher Career Satisfaction (TCS) is based on teachers’ agreements with five statements: I am content with my profession as a teacher, I am satisfied with being a teacher in this school, I would describe the teachers at this school as a satisfied group, I had more enthusiasm when I began teaching than I have now, and I do important work as a teacher. Responses were on a four-point scale (1=Disagree a lot, 2=Disagree a little, 3=Agree a little, 4=Agree a lot; negative statements are reverse coded). For this index, the responses for the five statements were averaged. An average of 3-4 receives a “high”
value; an average of 2 to less than 3 receives a “medium” value; an average of 1 to less than 2 receives a “low” value. The indexing was done by the researchers and only the complete index (none of the individual components) values were included in the dataset.

- **Teach_edu**: Teacher’s highest level of formal education completed; an ordinal variable with values 1-6 (education levels are put in terms of International Standard Classification of Education so as to be comparable with other countries; 6 is the highest level)
- **Teach_yrs**: Number of years teaching (in general; not limited to teaching reading); interval
- **Teach_txtbks**: How often the teacher uses textbooks in reading class; ordinal
- **Teach_internet**: How often the teacher uses the internet in reading class; ordinal
- **Teach_cert**: Whether or not the teacher holds a teaching certificate; nominal
- **Teach_work**: Whether the teacher works full time or part time; nominal
- **Teach_semi**: How many hours the teacher has spent in seminars dealing with reading; ordinal
- **Teach_bks**: How often the teacher reads books related to teaching; ordinal
- **Teach_read**: How often the teacher reads for enjoyment; ordinal

**Control variables**

Though I could easily select hundreds of variables that could potentially interact with my independent variables, I chose to focus on the following three. I believe they reflect common thought regarding theories of how students learn best.

- **Stud_sex**: Sex of student; nominal
- **Teach_size**: Number of students in class; interval
• *Home_bks:* To obtain a general, representative measure of students’ access to resources and parental encouragement, I had intended to use either a measure of parents’ education, parents’ income, or an index of home education resources (HER), which is an index derived from the number of books in the home, educational aids in the home, and parents’ education. Unfortunately the education and HER variables had an overwhelming number of missing cases, and income was not provided. Instead, I used the number of books at home as reported by parents in the home questionnaire. This should not be seen as a stand-in for income but it provides a somewhat unusual and helpful measure of the home environment. Variable is ordinal.

• *School_loc:* The character of school location—rural, urban, or suburban. Nominal.

• *School_disad:* The percentage of students from economically disadvantaged homes in the school. Ordinal.

• *School_affl:* The percentage of students from economically affluent homes in the school. Ordinal.

• *School_enrl:* Total school enrollment
### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.547&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.299</td>
<td>.187</td>
<td>65.0159778</td>
</tr>
</tbody>
</table>

---

<sup>a</sup> Predictors: (Constant), TOTAL ENROLLMENT OF STUDENTS, PERC OF STUD/ECONOMICALLY AFFL HOME, YEARS TAUGHT/ALTOGETHER, TIME SPENT IN SEMINARS DEALT W RDG, Rdg for enjoyment, Teacher: holds a teaching certificate (dummy), Index teacher career satisfaction (TCS), Teacher: works full time (dummy), Using resources/material on internet, Student: female dummy, Teacher: female dummy, HIGHEST LEVEL OF FORMAL EDUCATION, School is located in urban area, Using resources/textbooks, Read/bks related to teaching in gen, STD IN CLS, HOW MANY BOOKS AT HOME, School is located in rural area, PERC OF STUD/ECONOMICALLY DISAD HOME, AGE OF TEACHER.
<table>
<thead>
<tr>
<th>Model</th>
<th>Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unstandardized Coefficients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>391.843</td>
<td>107.429</td>
<td>3.647</td>
</tr>
<tr>
<td></td>
<td>AGE OF TEACHER</td>
<td>-7.565</td>
<td>6.694</td>
<td>-.129</td>
</tr>
<tr>
<td></td>
<td>Index teacher career satisfaction (TCS)</td>
<td>-7.066</td>
<td>11.864</td>
<td>-.049</td>
</tr>
<tr>
<td></td>
<td>HIGHEST LEVEL OF FORMAL EDUCATION YEARS</td>
<td>-.273</td>
<td>9.890</td>
<td>-.002</td>
</tr>
<tr>
<td></td>
<td>TAUGHT/ALTOGETHER Using resources/textbooks</td>
<td>2.241</td>
<td>.894</td>
<td>.286</td>
</tr>
<tr>
<td></td>
<td>Using resources/material on internet Teacher: holds a teaching certificate (dummy)</td>
<td>1.751</td>
<td>6.359</td>
<td>-.023</td>
</tr>
<tr>
<td></td>
<td>Teacher: works full time (dummy) TIME SPENT IN SEMINARS DEALT W RDG</td>
<td>20.277</td>
<td>7.845</td>
<td>.206</td>
</tr>
<tr>
<td></td>
<td>Read/bks related to teaching in gen Rdg for enjoyment</td>
<td>64.333</td>
<td>42.199</td>
<td>.127</td>
</tr>
<tr>
<td></td>
<td>Student: female dummy</td>
<td>3.992</td>
<td>49.792</td>
<td>.006</td>
</tr>
<tr>
<td></td>
<td>STD IN CLS HOW MANY BOOKS AT HOME</td>
<td>-1.137</td>
<td>5.088</td>
<td>-.019</td>
</tr>
<tr>
<td></td>
<td>SCHOOL IS LOCATED IN RURAL AREA</td>
<td>-14.075</td>
<td>6.665</td>
<td>-.234</td>
</tr>
<tr>
<td></td>
<td>SCHOOL IS LOCATED IN URBAN AREA</td>
<td>16.138</td>
<td>5.096</td>
<td>.283</td>
</tr>
<tr>
<td></td>
<td>PERC OF STUD/ECONOMICALLY DISAD HOME</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Performing a linear regression with all independent and control variables results in only a few significant relationships—and many are negative, indicating that the characteristics may decrease students’ test scores. By running a linear regression with all independent and control variables, the significance for each variable represents the significance controlling for all other variables. The model for all variables is:

Predicted \( \text{Reading}_{\text{avg}} = 391.843 - 7.565\text{Teach}_{\text{age}} + 21.340\text{Teach}_{\text{female}} - 7.066\text{Teach}_{\text{satisfact}} - 0.273\text{Teach}_{\text{edu}} + 2.241\text{Teach}_{\text{yrs}} - 1.751\text{Teach}_{\text{txtbks}} + 20.277\text{Teach}_{\text{internet}} + 64.333\text{Teach}_{\text{cert}} + 3.922\text{Teach}_{\text{full}} - 1.137\text{Teach}_{\text{semi}} - 5.617\text{Teach}_{\text{bks}} + 16.215\text{Teach}_{\text{read}} + 4.725\text{Stud}_{\text{female}} + 0.919\text{Teach}_{\text{size}} + 16.138\text{Home}_{\text{bks}} - 5.072\text{School}_{\text{rural}} - 4.661\text{School}_{\text{urb}} - 14.075\text{School}_{\text{disad}} - 4.705\text{School}_{\text{affl}} - 0.018\text{School}_{\text{enrl}} \)

Out of my independent variables, the ones that look like they have merit include: number of years teaching (significant at the .05 level, and an increase in 1 year of teaching is associated with a score increase of 2.241), using internet resources in teaching (significant at the .05 level, showing an increase of 20.277 on the assessment score when internet use increases from never to monthly to weekly to daily), and teacher reading for enjoyment (significant at the .05 level, showing an increase of 16.215 when frequency of reading increases from never to monthly to weekly to daily).

Out of my control variables, several show significant relationships: how many books at home (significant at the .01 level, an average increase of 16.138 on score when number of books increases from 0-10 to 11-25 to 26-100 to 101-200 to 200+) and school percentage of students in
economically disadvantaged homes (significant at the .05 level, an average decrease of 14.075 on score when percentage increases from 0-10% to 11-25% to 26-50% to 50%+).

In addition, the adjusted r-square value of .187 indicates that all these variables together account for 18.7% of the variation in score.

Hypotheses

In light of my 11 independent and 8 control variables, I developed a few much more narrow hypotheses based on my theoretical framework.

Hypothesis 1: The students of teachers demonstrating higher career satisfaction have higher test scores than the students of teachers demonstrating low career satisfaction.

I find teacher career satisfaction to be the most intriguing variable—in my own public school experience, I found that many of my teachers did not put much effort into teaching and would frequently use movies or “busywork” to occupy us instead of preparing lesson plans. I am curious whether creating a more satisfied group of teachers would improve teaching quality. Of course, this variable will inevitably invite more questions than it answers, as the public policy implications of exactly how to create a “satisfied teacher” in a cost effective way are vastly complicated and not provided through this research. Instead, I seek to merely determine whether teachers that believe they are satisfied in their work improve students’ test scores.

Hypothesis 2: The students of teachers who have been teaching for longer have higher test scores than the students of newer teachers.

A question that has been discussed and debated is whether more experienced teachers are more effective. This has important policy implications because if it turns out to be true, lawmakers can look for ways to keep teachers in the career for longer and perhaps research further to find out exactly why experience creates effectiveness—perhaps a mentor program for
new teachers could help transfer some of these benefits. Programs such as Teach for America could be profoundly impacted by this finding.

*Hypothesis 3:* The students of teachers who have spent more time in seminars about teaching reading will have higher test scores than the students of teachers who have not attended seminars about teaching reading.

Another question with immediately useful policy results is whether time spent in seminars about teaching reading improves teachers’ students’ test scores. This is tricky because without a controlled experiment in which teachers are randomly assigned to attend a seminar, we may assume that teachers with poor performance may be attending these seminars through a desire to improve their performance or because of a mandate from their supervisors.

*Hypothesis tests*

For hypothesis 1, I first ran a linear regression on the variable alone.

<table>
<thead>
<tr>
<th>Model Summary</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>1</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), Index teacher career satisfaction (TCS)
| Model | Unstandardized Coefficients | Standardized Coefficients | | | |
|-------|----------------------------|---------------------------|-----------------|-------|
|       | B                           | Std. Error                | Beta            | t     | Sig. |
| 1     | (Constant)                  | 520.304                   | 5.312           | 97.941 | .000 |
|       | Index teacher career satisfaction (TCS) | 5.948 | 1.932 | .044 | 3.079 | .002 |

a. Dependent Variable: Average of Plausible Values 1-5 for Overall Reading

By itself, the variable is significant at the .01 level and indicates that for every increase in career satisfaction (from low to medium to high), student scores increase an average of 5.948 points. The adjusted r-square indicated that this variable account for only .2% of the variation in student test scores. When controlling for some logically related factors that I thought could influence teacher satisfaction, however—including number of years teaching, whether the teacher works full time, number of students in class, location of the school, and percent of students in economically disadvantaged or advantaged homes, the relationship entirely disappears, with the p-value jumping to .970 (indicating that variation of this scale could be observed due to random chance 97% of the time):
### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.337a</td>
<td>.114</td>
<td>.067</td>
<td>68.3412924</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), PERC OF STUD/ECONOMICALLY AFFL HOME, STD IN CLS, YEARS TAUGHT/ALTOGETHER, Index teacher career satisfaction (TCS), Teacher: works full time (dummy), School is located in urban area, School is located in rural area, PERC OF STUD/ECONOMICALLY DISAD HOME
### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Constant)</td>
<td>564.804</td>
</tr>
<tr>
<td></td>
<td>Index teacher career satisfaction (TCS)</td>
<td>.410</td>
</tr>
<tr>
<td></td>
<td>YEARS</td>
<td>1.245</td>
</tr>
<tr>
<td></td>
<td>TAUGHT/ALTOGETHER</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Teacher: works full time (dummy)</td>
<td>5.136</td>
</tr>
<tr>
<td></td>
<td>STD IN CLS</td>
<td>-1.015</td>
</tr>
<tr>
<td></td>
<td>School is located in rural area</td>
<td>7.357</td>
</tr>
<tr>
<td></td>
<td>School is located in urban area</td>
<td>7.192</td>
</tr>
<tr>
<td></td>
<td>PERC OF STUD/ECONOMICALLY DISAD home</td>
<td>-15.088</td>
</tr>
<tr>
<td></td>
<td>PERC OF STUD/ECONOMICALLY AFFL HOME</td>
<td>2.869</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Average of Plausible Values 1-5 for Overall Reading

Evidently, this relationship is spurious. I performed additional tests to see if I could identify the single variable that caused teacher satisfaction to become insignificant, indicating possible issues of multicollinearity:
I found that controlling for most individual variables meant that teacher satisfaction maintained its significance, but when I introduced the percentage of students from an economically disadvantaged home, the p-value for career satisfaction shot up.

For hypothesis 2, a test with the variable alone shows a significant relationship at the .05 level, but with an adjusted r-square that indicates it accounts for only .1% of the variation in test scores:
This time, controlling for the age of the teacher, career satisfaction, education, percent of students in economically disadvantaged homes, and school location still shows a statistically significant relationship:
## Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.351(^a)</td>
<td>.123</td>
<td>.077</td>
<td>68.0131418</td>
</tr>
</tbody>
</table>

\(^a\) Predictors: (Constant), PERC OF STUD/ECONOMICALLY AFFL HOME, YEARS TAUGHT/ALTOGETHER, Index teacher career satisfaction (TCS), School is located in urban area, HIGHEST LEVEL OF FORMAL EDUCATION, School is located in rural area, PERC OF STUD/ECONOMICALLY DISAD HOME, AGE OF TEACHER
### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Constant)</td>
<td>568.716</td>
<td>65.015</td>
<td>8.748</td>
</tr>
<tr>
<td></td>
<td>YEARS</td>
<td>2.267</td>
<td>.894</td>
<td>.291</td>
</tr>
<tr>
<td></td>
<td>TAUGHT/ALTOGETHER</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>AGE OF TEACHER</td>
<td>-9.152</td>
<td>6.481</td>
<td>-.158</td>
</tr>
<tr>
<td></td>
<td>Index teacher career satisfaction (TCS)</td>
<td>2.670</td>
<td>10.896</td>
<td>.019</td>
</tr>
<tr>
<td></td>
<td>HIGHEST LEVEL OF FORMAL EDUCATION</td>
<td>-2.060</td>
<td>9.621</td>
<td>-.017</td>
</tr>
<tr>
<td></td>
<td>School is located in rural area</td>
<td>8.234</td>
<td>13.826</td>
<td>.052</td>
</tr>
<tr>
<td></td>
<td>School is located in urban area</td>
<td>6.633</td>
<td>14.198</td>
<td>.043</td>
</tr>
<tr>
<td></td>
<td>PERC OF STUD/ECONOMICALLY DISAD HOME</td>
<td>-13.928</td>
<td>5.959</td>
<td>-.242</td>
</tr>
<tr>
<td></td>
<td>PERC OF STUD/ECONOMICALLY AFFL HOME</td>
<td>3.514</td>
<td>6.437</td>
<td>.053</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Average of Plausible Values 1-5 for Overall Reading

This hypothesis may have merit. It indicates that for every increase of 1 year in teacher experience, student test scores increase by an average of 2.277 points. This is not a large increase, but it is significant.

For hypothesis 3, the variable alone is statistically significant and represents 1% of the variation in test scores. It indicates that as time spent in seminars increases (from none to <6 hours to 6-15 hours to 35+ hours), student scores actually decrease an average of 6.021 points.
When controlling for plausible control variables—teacher sex, education, experience, teaching certificate, whether they teach full time, reading books about teaching, and school location and percent of students in economically disadvantaged homes—the relationship entirely disappears like in hypothesis 1:
## Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.398&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.159</td>
<td>0.106</td>
<td>67.0770938</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), PERC OF STUD/ECONOMICALLY DISAD HOME, Teacher: holds a teaching certificate (dummy), Read/bks related to teaching in gen, School is located in rural area, HIGHEST LEVEL OF FORMAL EDUCATION, Teacher: works full time (dummy), Teacher: female dummy, YEARS TAUGHT/ALTOGETHER, TIME SPENT IN SEMINARS DEALT W RDG, School is located in urban area
### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>528.588</td>
<td>85.052</td>
<td>6.215</td>
</tr>
<tr>
<td></td>
<td>TIME SPENT IN SEMINARS</td>
<td>-.014</td>
<td>4.535</td>
<td>.000</td>
</tr>
<tr>
<td></td>
<td>DEALT W RDG</td>
<td>19.933</td>
<td>14.540</td>
<td>.103</td>
</tr>
<tr>
<td></td>
<td>Teacher: female dummy</td>
<td>-8.156</td>
<td>9.289</td>
<td>-.066</td>
</tr>
<tr>
<td></td>
<td>HIGHEST LEVEL OF</td>
<td>YRS</td>
<td>1.187</td>
<td>.581</td>
</tr>
<tr>
<td></td>
<td>FORMAL EDUCATION</td>
<td>TAUGHT/ALTOGETHER</td>
<td>62.025</td>
<td>40.704</td>
</tr>
<tr>
<td></td>
<td>Teacher: holds a teaching</td>
<td>certificate (dummy)</td>
<td>3.927</td>
<td>49.546</td>
</tr>
<tr>
<td></td>
<td>Teacher: works full time</td>
<td></td>
<td>Read/bks related to teaching in gen</td>
<td>-0.071</td>
</tr>
<tr>
<td></td>
<td>(dummy)</td>
<td></td>
<td>School is located in rural area</td>
<td>6.607</td>
</tr>
<tr>
<td></td>
<td>School is located in urban area</td>
<td></td>
<td>12.339</td>
<td>13.849</td>
</tr>
<tr>
<td></td>
<td>PERC OF STUD/ECONOMICALLY DISAD HOME</td>
<td></td>
<td>-19.567</td>
<td>4.879</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Average of Plausible Values 1-5 for Overall Reading

The relationship appears to be spurious.

**Hierarchical Linear Modeling**

To further isolate the teacher characteristics from other potentially confounding factors, I used Hierarchical Linear Modeling (HLM), which separates each level of effects in a method developed specifically for education policy research. I then separated the variables into three
categories: individual (student) level, class (teacher) level, and school level. I standardized each variable on a scale from 0-1 and ran linear regressions.

Individual/student level:

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.219(^a)</td>
<td>.048</td>
<td>.048</td>
<td>66.9397976</td>
</tr>
</tbody>
</table>

\(^a\) Predictors: (Constant), how many books at home, Student: female dummy

### Coefficients\(^a\)

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1 (Constant)</td>
<td>501.511</td>
<td>2.603</td>
<td>192.677</td>
<td>.000</td>
</tr>
<tr>
<td>Student: female dummy</td>
<td>7.950</td>
<td>1.908</td>
<td>.058</td>
<td>4.166</td>
</tr>
<tr>
<td>how many books at home</td>
<td>56.974</td>
<td>3.813</td>
<td>.208</td>
<td>14.943</td>
</tr>
</tbody>
</table>

\(^a\) Dependent Variable: Average of Plausible Values 1-5 for Overall Reading
Class/teacher level:

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.179&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.032</td>
<td>.029</td>
<td>69.6393950</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), STD IN CLS, Read/bks related to teaching in gen, Rdg for enjoyment, Teacher: female dummy, HIGHEST LEVEL OF FORMAL EDUCATION, Teacher: works full time (dummy), Teacher: holds a teaching certificate (dummy), Using resources/textbooks, AGE OF TEACHER, Using resources/material on internet, Index teacher career satisfaction (TCS), TIME SPENT IN SEMINARS DEALT W RDG, YEARS TAUGHT/ALTOGETHER
<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
</tr>
<tr>
<td>1</td>
<td>(Constant)</td>
<td>581.087</td>
</tr>
<tr>
<td></td>
<td>AGE OF TEACHER</td>
<td>-1.114</td>
</tr>
<tr>
<td></td>
<td>Teacher: female dummy</td>
<td>-2.795</td>
</tr>
<tr>
<td></td>
<td>Index teacher career satisfaction (TCS)</td>
<td>8.284</td>
</tr>
<tr>
<td></td>
<td>HIGHEST LEVEL OF FORMAL EDUCATION YEARS</td>
<td>3.212</td>
</tr>
<tr>
<td></td>
<td>TAUGHT/ALTOGETHER</td>
<td>.191</td>
</tr>
<tr>
<td></td>
<td>Using resources/textbooks</td>
<td>-7.175</td>
</tr>
<tr>
<td></td>
<td>Using resources/material on internet</td>
<td>8.583</td>
</tr>
<tr>
<td></td>
<td>Teacher: holds a teaching certificate (dummy)</td>
<td>-10.601</td>
</tr>
<tr>
<td></td>
<td>Teacher: works full time (dummy)</td>
<td>-17.351</td>
</tr>
<tr>
<td></td>
<td>TIME SPENT IN SEMINARS DEALT W RDG</td>
<td>-4.997</td>
</tr>
<tr>
<td></td>
<td>Read/bks related to teaching in gen</td>
<td>-7.172</td>
</tr>
<tr>
<td></td>
<td>Rdg for enjoyment</td>
<td>-1.263</td>
</tr>
<tr>
<td></td>
<td>STD IN CLS</td>
<td>-.124</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Average of Plausible Values 1-5 for Overall Reading
School level:

### Model Summary

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
<th>Adjusted R Square</th>
<th>Std. Error of the Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.294</td>
<td>.086</td>
<td>.058</td>
<td>68.8828040</td>
</tr>
</tbody>
</table>

a. Predictors: (Constant), total enrollment of students, School is located in urban area, perc of stud/economically affl home, School is located in rural area, perc of stud/economically disad home

### Coefficients

<table>
<thead>
<tr>
<th>Model</th>
<th>Unstandardized Coefficients</th>
<th>Standardized Coefficients</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>Std. Error</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>1 (Constant)</td>
<td>581.333</td>
<td>28.193</td>
<td>.052</td>
<td>20.620</td>
</tr>
<tr>
<td></td>
<td>8.579</td>
<td>13.969</td>
<td>.054</td>
<td>.614</td>
</tr>
<tr>
<td></td>
<td>8.582</td>
<td>14.029</td>
<td>.055</td>
<td>.612</td>
</tr>
<tr>
<td>School is located in rural area</td>
<td>-62.527</td>
<td>23.836</td>
<td>-.267</td>
<td>-2.623</td>
</tr>
<tr>
<td>School is located in urban area</td>
<td>8.915</td>
<td>25.454</td>
<td>.034</td>
<td>.350</td>
</tr>
<tr>
<td>perc of stud/economically disad home</td>
<td>-42.528</td>
<td>35.007</td>
<td>-.094</td>
<td>-1.215</td>
</tr>
</tbody>
</table>

a. Dependent Variable: Average of Plausible Values 1-5 for Overall Reading

From the adjusted r square values, we can see that the school and individual variables explain much more of the variation in test scores than the teaching. This hierarchical approach is more appropriate than the earlier models because it eliminates a substantial amount of redundancy that existed in those first linear models.
In fact, the two variables for individual level are both significant at the .01 level, and together account for 4.8% of the variation. They show that female students average 7.95 points better than males, and that the number of books in the home (ostensibly a measure of parental encouragement of reading) has a large effect—going from 0-10 books to 11-25 books to 26-100 books to 101-200 books to 200+ books results in an average increase of 56.97 points.

At the school level, the only significant variable is the percentage of students in an economically disadvantaged home—apparently, the location and school enrollment are not factors. This level accounts for 5.8% of the variation, and indicates that an increase in percentage of students from economically disadvantaged homes from 0-10% to 11-25% to 26-50% to more than 50% accounts for a decrease of 62.53 points.

At the teacher level, a number of variables show up as significant, yet altogether they only account for 2.9% of the variation. Career satisfaction, significant at the .01 level, shows an average increase of 24.85 points when going from low to medium to high (as usual, the negative coefficient is misleading because the variable is coded high to low). Interestingly, textbook use and internet use are both significant at the .01 level but have opposite relationships. Textbook use shows an average decrease of 28.70 points when going from almost daily use to weekly use to monthly use to almost never. Internet use has an average increase of 34.33 points when following the same scale. Teacher efforts at improvement also have an effect, though not necessarily as predicted. As time spent in seminars about teaching reading increases from none to less than 6 hours to 6-15 hours to 16-35 hours to 35+ hours, student test scores decrease an average of 24.99 points. As time spent reading books about teaching goes from none to yearly to monthly to weekly, test scores decrease an average of 28.69 points. Whether the teacher works
full or part time is significant at the .05 level, showing a decrease of 17.35 points for a teacher that works full time.

Based on these results, my first hypothesis (The students of teachers demonstrating higher career satisfaction have higher test scores than the students of teachers demonstrating low career satisfaction) looks like it may have merit, but my second (The students of teachers who have been teaching for longer have higher test scores than the students of newer teachers) and third (The students of teachers who have spent more time in seminars about teaching reading will have higher test scores than the students of teachers who have not attended seminars about teaching reading) do not.
CHAPTER 4. Policy analysis and conclusion

My research focused on a very narrow aspect of education reform: whether teacher quality has an effect, and if so, what characteristics are associated with student achievement. In my own study, I found that teachers relatively have a quite small effect on student outcomes. When I tested all my teacher variables (which included teacher’s age, sex, career satisfaction opinions, education, years of teaching experience, use of textbooks, use of the internet, teaching certificate, employment status [full time or part time], time spent in seminars on teaching reading, time spent reading books about teaching, and how often the teacher reads for enjoyment), I found that all these variables combined accounted for 2.9% of the variation in test scores. This compares to the school level (5.8%) and individual student level (4.8%). These seem very low, but I restricted my variables quite a bit. There were a number of other variables I could have included but did not, for the sake of simplicity.

Other studies show higher percentages but comparable relative values: Hanushek et al show a minimum of 7.5% teacher effects (compared to 3-7% school effects)\(^\text{58}\); Rockoff shows 5-6.4% teacher effects (compared to 2.7-6.1% school-year effects and a whopping 59-68% student fixed effects)\(^\text{59}\); Rowan et al report 10-20% teacher effects\(^\text{60}\); and Nye et al find a range of 6.6-13.5% teacher effects (compared to 1.9-9.7% school effects and 31.2-43.9% student


This suggests that these authors were taking more variables into account to get these higher numbers, but otherwise confirms my own results: teacher quality does have a statistically significant effect and while it alone may not be enough to bring the drastic changes that are needed, it is one of the easiest components to affect through policy. Teacher quality is a rather intangible and difficult to operationalize variable. It does not especially lend itself to analysis with empirical data. Though this data suggests that teacher quality is less critical to student outcomes than I thought, this does not take into account the largely unmeasurable qualities that we’ve all seen in our own beloved teachers throughout our time in school: the teacher that inspired us to follow a particular career path or other passion, or the teacher that helped us through a difficult home life—even at the expense of helping us achieve higher scores on standardized tests. Analyzing test scores cannot be the be-all and end-all of developing a policy on teacher hiring and firing and compensation.

Obviously changing the fixed student characteristics (such as income and race) would require broad reforms in policy areas outside the sphere of education. Reducing poverty may well have an enormous effect on student achievement, but that task is surely easier said than done. Poverty reduction has long been a tenet of liberal politics and many have devoted their lives and research to solving the problem but saw little or no improvement. Unfortunately, putting all students on level footing before they even arrive at school is an ideal that remains in the distant future. Compared to the daunting prospect of reducing poverty, changing teacher hiring practices seems like an easy fix.

---

Based on my research and on my limited analysis of past research, I have some suggestions for policies and for further research. The results I present here do not suggest that we should completely redesign the hiring process, curricula, or teacher salaries at this point. I would suggest making small changes such as adjusting curricula to accommodate less textbook use and more internet use in the classroom when the curricula are up for renewal. Since my research shows that part time teachers tend to have students with somewhat better scores than full time teachers’ students, and part time teachers are probably financially beneficial to the school districts, it may be helpful to continue or even escalate hiring part time teachers.

My final finding on teacher quality shows that teachers who are satisfied in their careers have students with higher test scores. Though this finding is valuable and interesting, it is loaded for several reasons. The statements from the teacher questionnaire are rather vague. More research is certainly needed. A high score on this measure could indicate that some teachers are innately suited to be a teacher, and schools should seek these teachers that will be satisfied simply due to the joy of teaching. But a high score could also be interpreted to mean that the benefits and salaries distributed by a school have a large impact on how successful a teacher will be. But perhaps even more likely, a teacher’s career satisfaction is a reflection of his or her students: when I controlled for factors such as school location and percentage of students who come from an economically disadvantaged home, the relationship between career satisfaction and test scores was no longer significant. This suggests that teachers who work at schools with students from the suburbs and/or come from more advantaged backgrounds (and get higher test scores anyway) are more satisfied in their careers. This just brings us back to improving the

---

62 In my study, where teachers agree more strongly with the following four statements, the higher their career satisfaction is: I am content with my profession as a teacher, I am satisfied with being a teacher in this school, I would describe the teachers at this school as a satisfied group, and I do important work as a teacher; and the less strongly teachers agree with the following statement, the higher career satisfaction is: I had more enthusiasm when I began teaching than I have now.
poverty problem, and implies that improving teacher satisfaction should perhaps not be the end goal, but merely a side effect.

I also found teacher experience to be a significant variable, and this significance remained even after controlling for several other factors. This implies that programs intended to attract teacher prospects who will remain in the career track for a long time, or keep current teachers, may be worth the time and money. Though this would require more research, a system of mentoring so that experienced teachers can pass along whatever it is about their years that improves learning could be promising.

As controversial as merit pay has become, further study would be beneficial. Obama’s Race to the Top has embraced the issue, but the NEA remains staunchly against it. Of course, the major criticism of merit pay is that it only measures a small portion of a teacher’s success—his or her students’ test scores. Though it would require resources, crafting a more comprehensive method of measuring teacher success may be worth it to win over the critics. It could take into account criteria such as student evaluations, classroom visits (by the principal or independent evaluators), and even attendance and disciplinary records to identify and reward the teachers that are truly good (that is, imparting knowledge and impacting their students’ lives), and not merely “teaching to the test.” These factors could of course be problematic and difficult to measure as well, but will provide a more complete picture of the teacher’s performance. But of course, implementing merit pay is easier said than done, as we saw in Michelle Rhee’s case. Reforming the criteria and more careful negotiations (or at least, less drastic actions taken once negotiations are completed) could be the key, if further studies show that merit pay does in fact improve student performance.

Another issue that I did not study specifically, but that could be helpful in improving outcomes, is school funding. Some research has already dealt with this question of whether simply allocating more money will improve student performance, but this should of course be heavily influenced by how effectively the money is distributed. Further research on how spending on specific areas and programs within schools affects the school’s performance could be incredibly beneficial.

In assessing No Child Left Behind’s effectiveness and attention to teacher quality, I admit that teacher quality is not quite as large a factor in student performance as I had predicted. However, it should receive attention because it is one of the most easily influenced factors. Though NCLB does not put an extremely large emphasis on teacher quality, it does address it. The problem is that NCLB’s provisions in this area are vague. The reforms require that every teacher be “highly qualified,” but leaves it up to the states to determine what that means. NCLB specifies that teachers must use “scientifically based research” practices, but does not elaborate, once again leaving it up to states and, frequently, individual schools or school systems, to try to puzzle out what their own policies should be. More guidance, though perhaps not more requirements, would be helpful in this area.

I conclude that we should approach the question of assessing teacher quality and reforming the education system based on these assessments with caution. Teacher quality—and potentially the manner in which we treat and reward teachers—does have an impact on student success, even if it is not quite as strong as I had expected. It could quite possibly be more influential in ways that are difficult for policymakers and researchers to identify and measure, so it certainly shouldn’t be written off. Therefore, I recommend further research and sensible, cautious policy changes in this area based on my own findings and those of other researchers.