Advances in Evidence-Based Education

Volume 2

Education at the Crossroads: The State of Teacher Preparation

Advances in Evidence-Based Education

Volume 2

Education at the Crossroads: The State of Teacher Preparation

Edited by:
Ronnie Detrich
Randy Keyworth
Jack States

The Wing Institute
Oakland, CA

Published by The Wing Institute
2102 Dennison Street, Suite B
Oakland, CA 94606
510.534.9464
www.winginstitute.org

Volume 2 published Fall 2012.

ISSN 1941-1693
ISBN 978-0-9800177-1-7

Printed in the United States of America.
Printed on acid-free paper

Dedication

These proceedings from the 5th annual summit are dedicated to the memory of Ernie Wing, who is also the namesake of the Wing Institute. Ernie championed evidence-based education as an educator and child advocate. As an educator, he founded Spectrum Center, which has been a beacon for evidence-based practice and state-of-the art educational services since 1975. As an advocate, Ernie served hundreds of families with the most challenging special education needs, gaining the admiration and respect of both parents and school districts. Through his efforts, Ernie set the standards for professionalism, integrity, effectiveness, and caring as he helped thousands of children gain access to effective educational services. He was a good man and a good friend, and is missed.

Acknowledgments

We are indebted to the people who made this book possible. We wish especially to recognize the work of Jin An in managing the process of producing this publication. Without her patience and organization skills, we would be still in the conceptual phase of the book. We are grateful for the outstanding work of our copy editor, Susan Lang. Her ability to grasp the technical aspects of the topic while keeping the writing clear and concise was essential in completing the project. We would like particularly to recognize the work of the three speakers from the 2010 summit: Jim Kauffman, Larry Maheady, and Dan Reschley. It is on these three presentations that the contents of the papers are based. Jim Kauffman, Michael Jabot, and Larry Maheady warrant special recognition for writing the papers that are the core of this anthology. And, finally, we would like to acknowledge teachers for the outstanding but underappreciated job they do in educating our children.

Contents

Introduction

Proceedings from the Wing Institute's Fifth Annual Summit on Evidence-Based Education

Education at the Crossroads: The State of Teacher Preparation

RANDY KEYWORTH
RONNIE DETRICH
JACK STATES
The Wing Institute

In 1983, the U.S. Department of Education published a landmark report, *A Nation at Risk*, which identified a crisis in education performance so severe that it constituted a threat to the nation. Student achievement on standardized tests was well below proficiency standards, too few students were graduating from high school, there was a dramatic gap between the performance of White students and that of African-American and Hispanic students, and the performance of U.S. students compared to those of other industrialized nations was falling (Gardner et al., 1983). Enormous resources, energy, and focus were marshaled to take this challenge head on. As educational gains failed to materialize, this cycle of "call to action" and "education reform" has been replicated at regular intervals. Goals 2000, begun in 1994, was one of many programs launched with much the same fanfare, message, and intent. When that failed to produce the desired results, No Child Left Behind (NCLB) became the education reform law of the land in 2001. Now, once again, as the realization sinks in that we are failing to make progress in educating our children, new reforms are being contemplated.

ATTEMPTS AT SCHOOL REFORM THROUGH STRUCTURAL INTERVENTIONS

The past efforts to reshape education generated an enormous amount of action and change in the form of structural interventions: large-scale system changes that affect the organizational design of education systems without directly addressing the actual teaching that takes place in the classroom. The assumption has been that each of these structural interventions would improve teacher and student outcomes. Several of the most recent structural interventions include increased education funding, class size reduction, school choice, and, most recently, charter schools. As the following data show, (a) the effort and resources expended to carry out these interventions have been significant; (b) the structural interventions have, by and large, been implemented on a large scale; and (c) they have had little or no impact on student outcomes at the macro level.

Increased Education Funding

At the national level, education spending has increased dramatically over the past 40 years (Figure 1). Annual K–12 per pupil funding has increased by 140% from the 1969–70 school year to the 2007–08 school year (from $4,637 per pupil to $11,134), when adjusted for inflation and benchmarked at 2007–08 dollars. Funding increased 22% over the 10-year period ending in 2007–08.

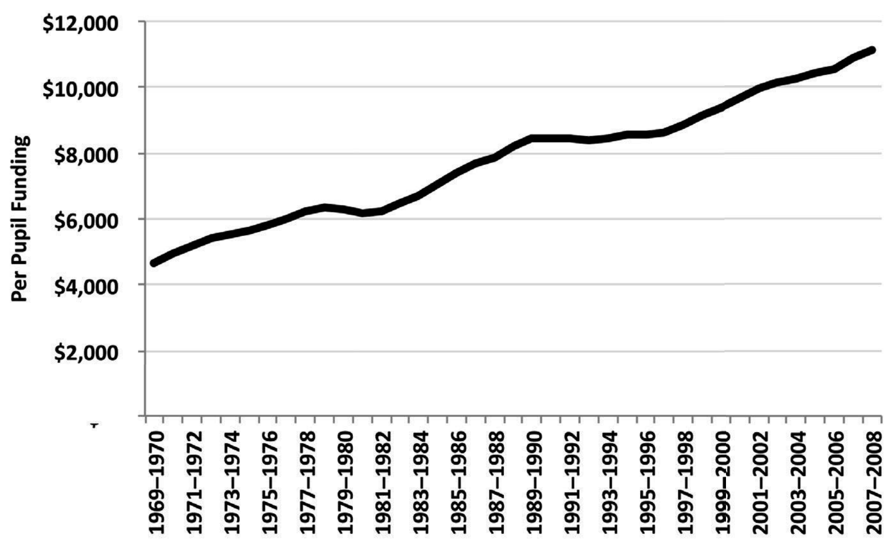


Figure 1. Annual K–12 per pupil funding in U.S. 1970–2008 (adjusted for inflation). Data are drawn from Snyder and Dillow (2011, p. 274).

Internationally, the United States spends more per student than any other nation in the world except Luxembourg. The metric used by the Organisation for Economic Co-operation and Development (OECD) is the total dollars spent over the K–12 life of a child. In 2009, the United States spent an astounding $105,752 (Organisation for Economic Cooperation and Development [OECD], 2010c) The top five nations with the highest reading scores averaged only $66,792 in spending over the K–12 life of a child. The top five nations with the highest mathematics scores averaged only $78,995 (Organisation for Economic Cooperation and Development [OECD], 2010a).

As a structural intervention, increased funding for education fits the pattern identified above. The intervention has been extremely costly, it has been implemented by successfully, and, as will be demonstrated later, there has not been a corresponding impact on student outcomes.

Class Size Reduction

Few structural interventions have garnered more public support than class size reduction, and the resources committed to this intervention over recent years have been significant. A 2007 survey showed that 77% of Americans favored spending educational dollars on decreasing class size rather than increasing teacher salaries (Howell, Peterson, and West, 2007). As of 2010, 36 states have laws restricting the number of students in a general education classroom, in some or all grades (Zinth, 2010). In 1996, California launched an ambitious initiative to reduce K–3 class sizes to 20 students per class. It spent over $20 billion from 1996–97 through 2009–10 on reduced class sizes, averaging $1.75 billion per year for last the 5 years (Luckie, 2009). In 2003, Florida adopted a class size reduction constitutional amendment. It is projected to have spent $21.6 billion from 2003–04 through 2011–12, averaging $2.94 billion per year for the last 5 years (Florida Department of Education, n.d.). As a result of these and many other initiatives, pupil-teacher ratios in public schools have fallen by about 30% since 1970 (Figure 2).

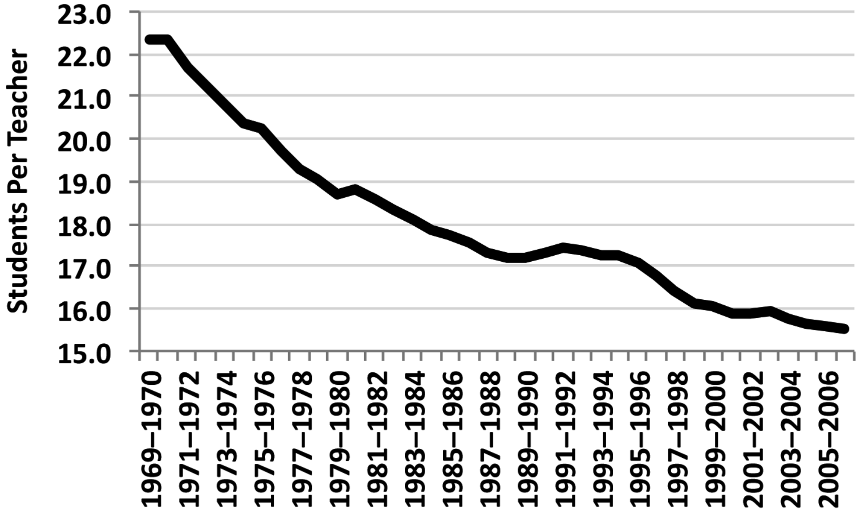


Figure 2. Pupil-teacher ratios in public schools. Data are drawn from Snyder and Dillow (2010, p. 100).

Class size reduction represents another structural intervention that has been implemented at a significant scale and cost without corresponding changes in student outcomes.

School Choice

Another structural intervention has been to increase the amount of choice that parents have in selecting their children's school placement. The theory is that increased competition (driven by choice) will improve school performance. These choices typically include charter schools, private schools, public magnet schools, and other public school programs that provide options. As of 2010, 33 states had passed legislation mandating school districts to implement intra-district or inter-district school choice programs, which allow parents to send their children to traditional public schools outside of the neighborhoods in which they reside (Nichols & Özek, 2010). The growth in charter schools is discussed later. The percentage of students enrolled in schools offering choice increased from 20% to 27% during the 15-year period between 1993 and 2007 (Figure 3).

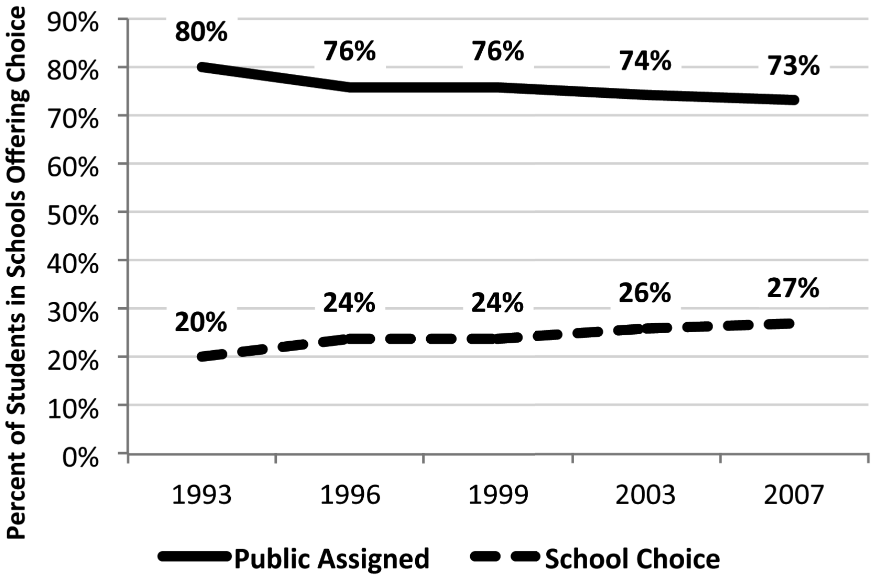


Figure 3. Percent of students enrolled in assigned public schools and schools offering choice. Data are drawn from Grady and Bielick (2010, p. 7).

While public assigned schools still make up the majority of student placements, the school choice structural intervention has continued to increase.

Charter Schools

The most recent and popular structural intervention is the charter school. Usually publicly funded and governed by organizations or groups under contract with the state, charter schools have greater autonomy than public schools and are often exempted from selected state or local rules and regulations. As of November 2010, charter schools operated in 40 states and the District of Columbia (National Center for Education Statistics [NCES], 2011a).

From 1999–2000 to 2008–09, the number of students enrolled in charter schools more than tripled from 340,000 to more than 1.4 million (Figure 4).

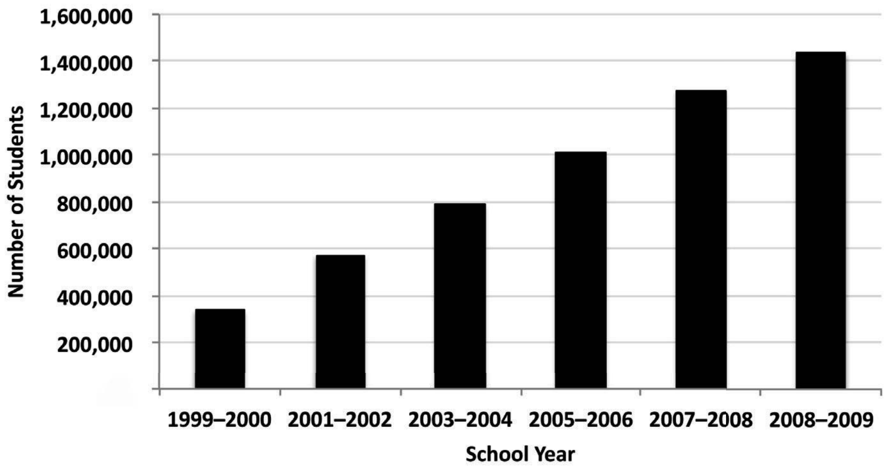


Figure 4. Number of students enrolled in public charter schools (1999–2000 to 2008–09). Adapted from *The Condition of Education 2011,* (p. 25), by S. Aud, W. Hussar, G. Kena, K. Bianco, L. Frohlich, J. Kemp, and K. Tahan, 2011, Washington, DC: U.S. Department of Education. In the public domain.

During this period, the percentage of all public schools classified as charter schools increased from 2% to 5%. In 2008–09, there were 4,700 public charter schools in the United States (NCES, 2011a).

THE IMPACT OF STRUCTURAL INTERVENTIONS ON STUDENT PERFORMANCE

While these structural interventions — greater funding, smaller classes, more choice, and more charter schools — have been successful in terms of changing the public education landscapethere has been virtually no corresponding improvement in student performance at the national level. This conclusion comes from three well-established sources of student performance data: the National Assessment of Educational Progress (NAEP), the Program for International Student Assessment (PISA), and the U.S. 4-year adjusted cohort graduation rate:

> NAEP has often been called the "gold standard" for standardized academic testing because of its constant rigorous scrutiny (Gorman, 2010). It was established in 1964, with the first tests administered

in 1969. It provides a continuing assessment of what America's students know and can do in math, reading, science, writing, the arts, civics, economics, geography, and U.S. history. NAEP is administered by the National Center for Education Statistics (NCES), a division of the Institute of Education Sciences in the U.S. Department of Education. Panels of technical experts within NCES and other organizations continually scrutinize tests for reliability and validity, keeping them similar from year to year and documenting changes. It is one of the only common metrics for all states, providing a picture of student academic progress over time.

PISA is a carefully constructed and well-documented test instrument for measuring student academic performance across nations (Organisation for Economic Cooperation and Development [OECD], 2006). Coordinated by the Organisation for Economic Co-operation and Development (OECD), this international study is conducted every 3 years. It measures the performance of 15-year-old students in 64 countries (34 member nations and 30 participating nations) in reading, mathematics, and science. Tests have been given since 2000. In addition to reporting on test scores, PISA collects data on a large number of education system characteristics and identifies statistical correlations between results and selected variables.

The 4-year adjusted cohort graduation rate is the number of students who graduate in 4 years with a regular high school diploma divided by the number of students who entered high school 4 years earlier. It was adopted in 2008, when the U.S. Department of Education enacted regulations establishing a uniform and more accurate measure for calculating the rate at which students graduated from high school. Prior to this mandate, many states failed to account for students who left school prior to the 12th grade, often dramatically skewing the data (Hall & Gutierrez, 1998). The 4-year adjusted cohort graduation rate captures all students, including those who drop out in earlier grades. Above all, it is a metric that is now uniform across all 50 states and can be used over time.

There is much debate in our education system about what constitutes a quality education and how best to measure many of the non-academic outcomes such as creativity, social intelligence, and problem solving. There is also much cynicism about such macromeasures as standardized tests. However, while standardized tests may not measure every education outcome, they do assess one of the most important outcomes: what students have learned in selected content areas such as reading and math. And while some of the standardized tests used in different states and localities may merit cynicism, the NAEP and PISA tests are consistently analyzed to meet the highest standards of reliability, validity, and social relevance. Data from these tests provide a clear and unambiguous picture of how well the U.S. education system is educating students on selected measures. The cohort graduation rate data provides an additional critical indicator of overall performance of the system.

Student Performance Data (NAEP)

The richest set of student achievement data comes from the NAEP, which makes available test data in mathematics and reading going back to 1970 (Long-Term Trend Assessment) and up to 12 different subject areas going back to 1992 (main NAEP Assessment). The Long-Term Trend Assessment data provides test scores at age 9, 13, and 17. The main NAEP Assessment tests by grades 4, 8, and 12.

NAEP provides data on subject matter achievement in two ways: scale scores and achievement levels. Scale scores provide a numeric summary of what students know and can do in a particular subject and are presented for groups of students. NAEP subject area scales for reading and math range from 0 to 500. Achievement levels are used to report results in terms of what students should know and are able to do. The Long-Term Assessment data only report scale scores, but show a remarkable lack of student achievement progress over the last 40 years in both subjects (Figures 5 and 6). This occurred despite numerous and significant school reform initiatives (A Nation at Risk, Goals 2000, NCLB) and the aforementioned structural interventions.

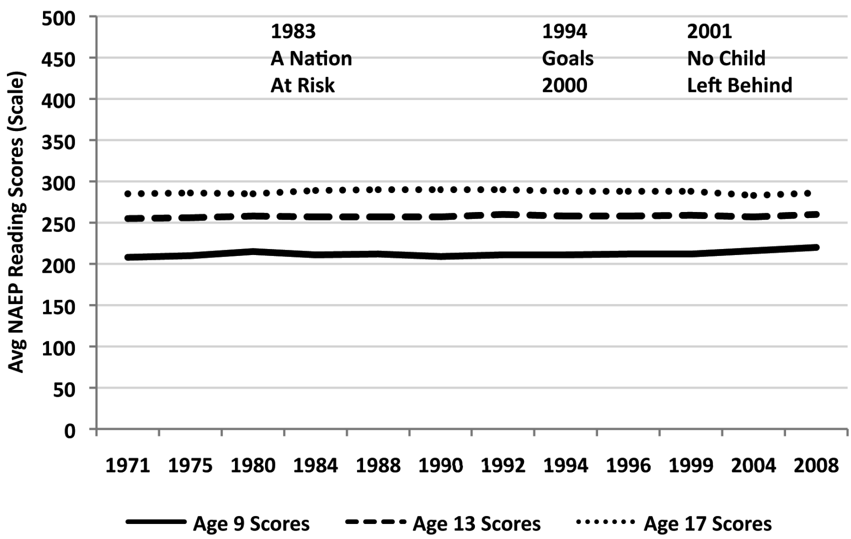


Figure 5. Reading Scores: National Assessment of Education Progress (NAEP) Long-Term Trend Assessment. Data are drawn from *National Center for Education Statistics Data Explorer for Long-Term Trend* [Data file].

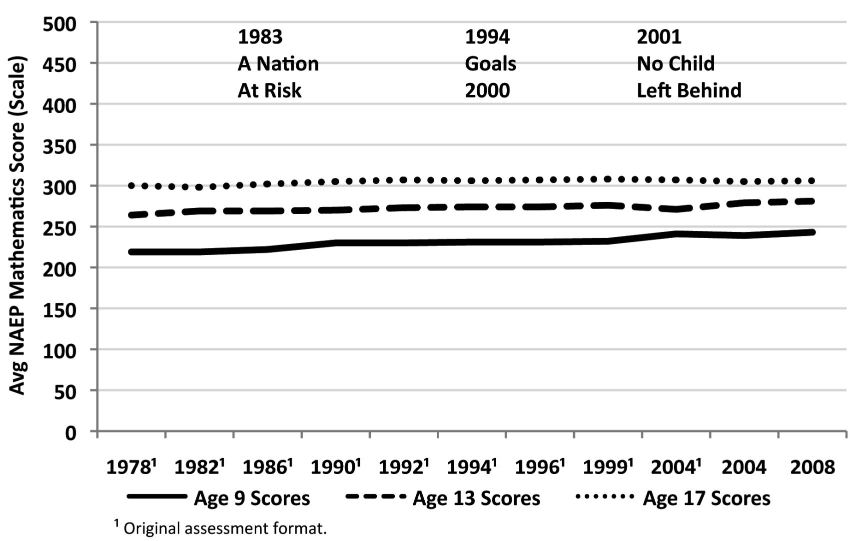


Figure 6. Mathematics Scores: National Assessment of Education Progress (NAEP) Long-Term Trend Assessment. Data are drawn from *National Center for Education Statistics Data Explorer for Long-Term Trend* [Data file].

The data become even more alarming when analyzed in the context of achievement levels. The main NAEP Assessment standards identify three achievement levels, or benchmarks, for student performance at each grade: "advanced" represents superior performance, "proficient" signifies solid academic performance, and "basic" denotes partial mastery of prerequisite knowledge and skills fundamental for proficient work. "Proficiency" becomes a critical benchmark because it is the level at which students have met the standards for a subject area. It is also the benchmark by which the No Child Left Behind law holds school districts accountable. While the law allows for states to use their own tests and proficiency cut scores (a flaw in the system), one of NCLB's fundamental goals is that *all* children are to be proficient in reading and math by 2014. Proficiency standards are critical in evaluating education effectiveness.

NAEP data can also be analyzed to identify the percentage of students at a given grade level who are at or above proficiency. Again, "proficiency" means that students at this level have demonstrated competency over challenging subject matter for their grade level. Below proficiency means students have only partial mastery. Figure 7 shows the percentage of fourth-grade children who can read at or above proficiency level.

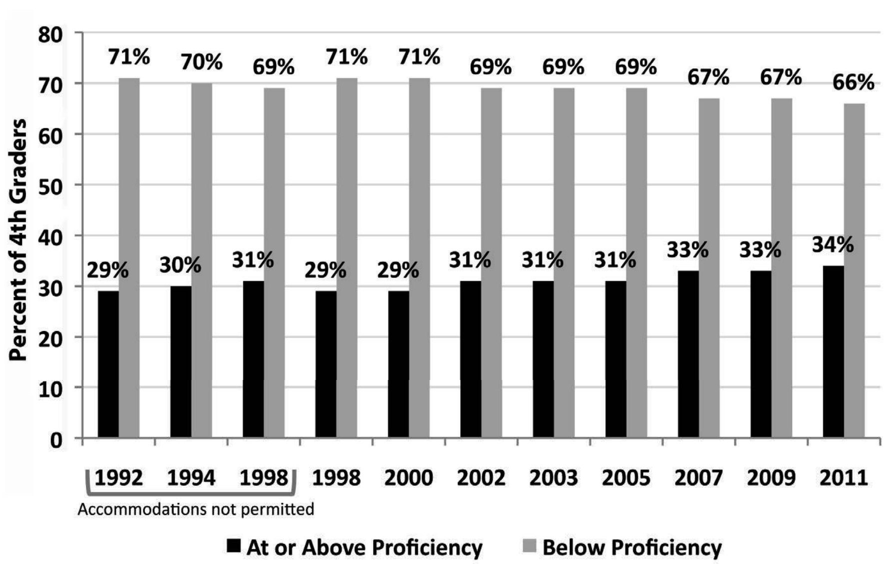


Figure 7. Percent of 4th graders reading at or above proficiency. Adapted from *The Nation's Report Card: Reading 2011,* (p. 10), by the National Center for Education Statistics, 2011, Washington, DC: U.S. Department of Education. In the public domain.

In 2011, only one third of fourth-grade students read at or above proficiency level, which represents only a 5% point improvement since 1992. Reading proficiency data varied significantly across states, with New Mexico and Mississippi having the lowest percentage of proficient readers at 20% and 22%, respectively. The state with the greatest percentage of proficient readers was Massachusetts, with 51% (National Center for Education Statistics [NCES], 2011c).

The data did not improve significantly when it came to the percentage of 12th-grade students who read at or above proficiency (Figure 8).

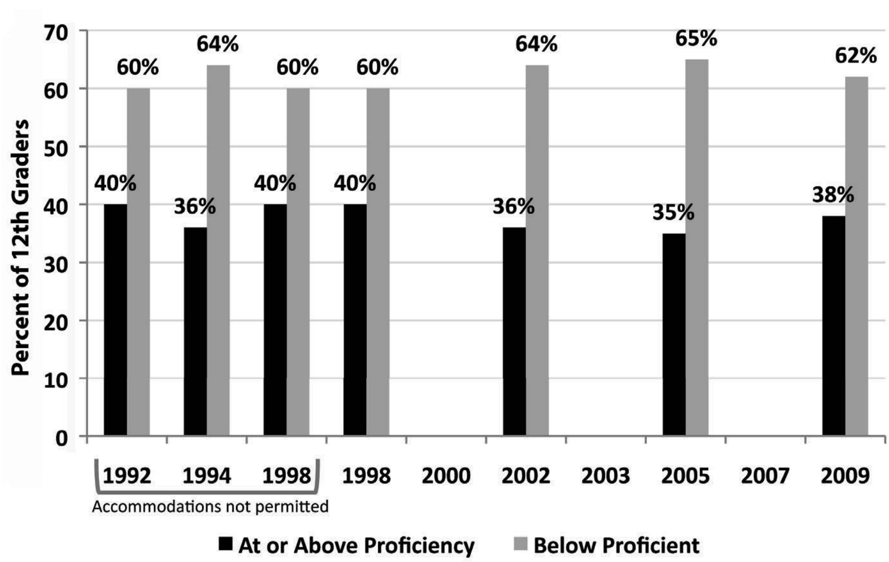


Figure 8. Percent of 12th graders reading at or above proficiency. Adapted from *The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results,* (p. 9), by the National Center for Education Statistics, 2010, Washington, DC: U.S. Department of Education. In the public domain.

Only 38% of 12th-grade students were reading at or above proficiency in 2009, which is actually a decrease in performance from 40% in 1992. While 12th grade achievement data have not historically been collected at the state level, 11 states volunteered to participate in a pilot program (National Center for Education Statistics [NCES], 2010). Once again, individual states had widely differing performances. West Virginia (29%), Arkansas (32%), and Florida (32%) had the lowest percentages of proficient readers among 12th graders. New Hampshire (44%) and Massachusetts (46%) had the highest (NCES, 2010).

Achievement levels in mathematics painted a very similar picture. While there was a significant improvement in test scores between 2000 and 2007, there has been little subsequent change, leveling out at 39% to 40% proficiency (Figure 9).

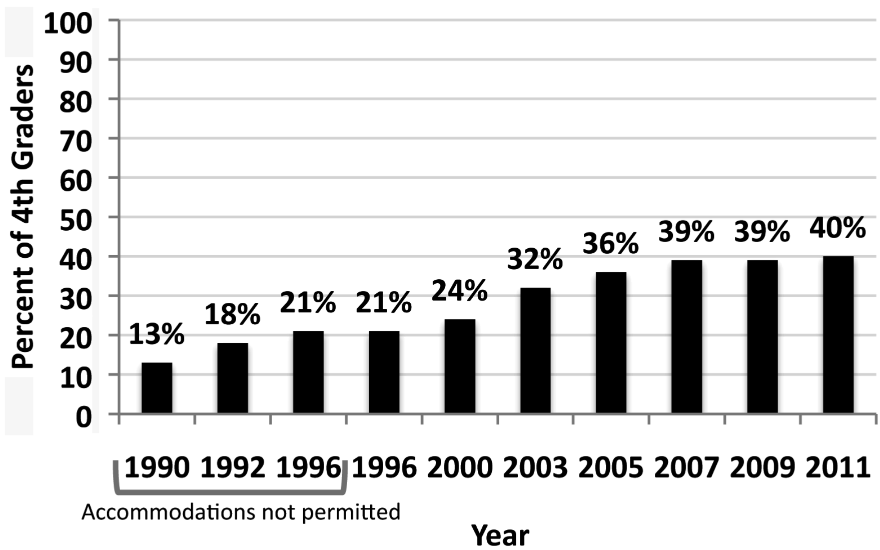


Figure 9. Percent of 4th graders at or above proficiency in mathematics. Adapted from *The Nation's Report Card: Mathematics 2011,* (p. 11), by the National Center for Education Statistics, 2011, Washington, DC: U.S. Department of Education. In the public domain.

Mathematics achievement data for 12th-grade students is only available for 2005 and 2009, as a change in the mathematics framework for the assessment necessitated a new trend line for that subject at grade 12. A total of 23% of 12th-graders performed at or above the proficient level in mathematics in 2005, 26 % in 2009 (NCES, 2010). As with reading achievement data, the only individual state data came from the 11 state pilot programs in 2009. West Virginia (13%) and Arkansas (15%) had the lowest percentage of 12th-grade students at or above proficiency in mathematics. New Hampshire (32%) and Massachusetts (36%) had the highest (NCES, 2010).

NAEP data also show a significant gap in the performance of children of color. Figure 10 illustrates the stark contrast between 12th-grade White students and students of color in reading proficiency. In 2009, 46% of White students were at or above proficiency, while only 22% of Hispanic students and 17% of Black students were at or above proficiency.

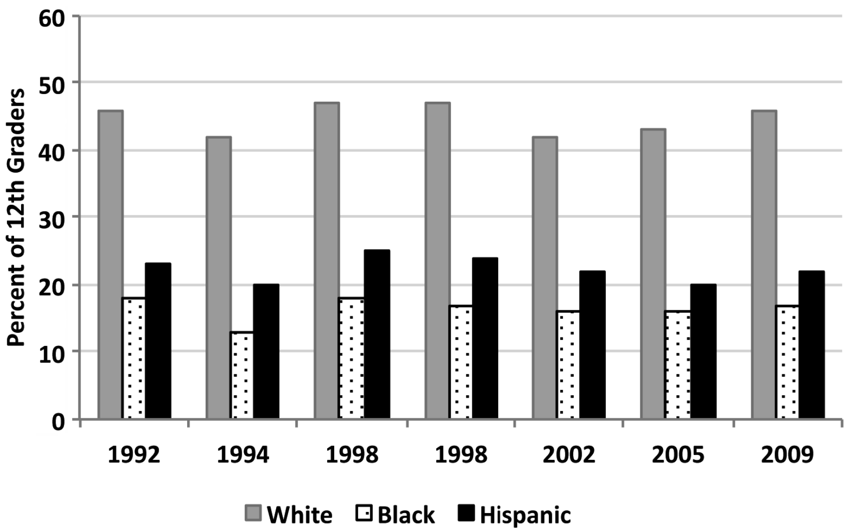


Figure 10. Percent of 12th graders by race at or above proficiency in reading. Data are drawn from the National Assessment of Educational Progress Reading Assessments of 1992, 1994, 1998, 2002, 2005, and 2009 (National Center for Education Statistics, 2012b).

The data also show that the reading proficiency gap continued without improvement. In 1992 the proficiency gap between White and Black students was 28%; in 2009 it was 29%. In 1992 the gap between White and Hispanic students was 23%; in 2009 it was 24%.

NAEP achievement data in mathematics show the same level of discrepancy in proficiency between races. In 2009 only 33% of White students were at or above proficiency in mathematics, while Black and Hispanic students had staggeringly low proficiency levels of 6% and 8%, respectively (Figure 11).

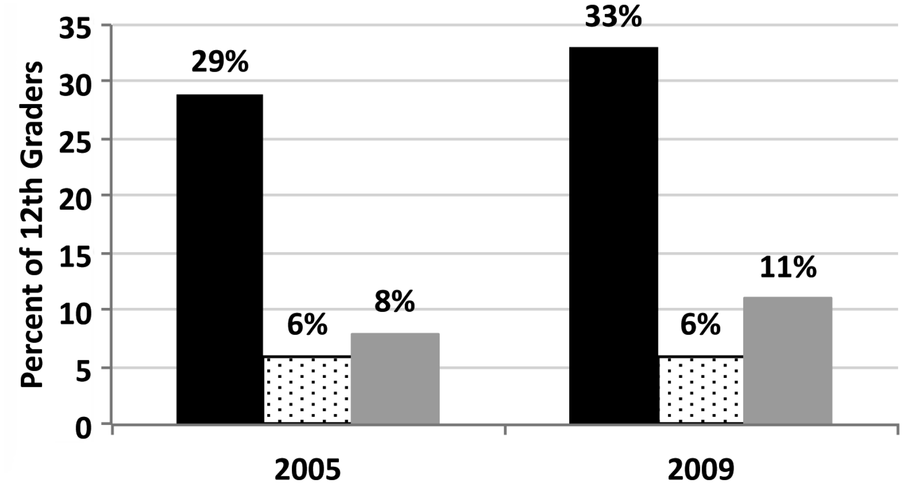


Figure 11. Percent of 12th graders by race at or above proficiency in mathematics. Data are drawn from the National Assessment of Educational Progress Mathematics Assessments of 2005 and 2009 (National Center for Education Statistics, 2012a).

As with reading, there was no improvement in this gap since the previous test in 2005. In fact, it got worse, with the gap between White and Black students increasing from 23% points in 2005 to 27% in 2009, and that between White and Hispanic students increasing from 21% points to 22% points.

Student Performance Data (PISA)

The other student performance outcome test data come from PISA results, which show the United States trailing 13 nations in reading, 16 in science, and 24 in mathematics (Table 1). The United States test scores actually dropped by 5 points between 2000 and 2009 PISA tests (Organisation for Economic Cooperation and Development [OECD], 2010b).

Table 1
2009 PISA reading, science, and mathematics scores.

Rank	Reading	Score	Science	Score	Mathematics	Score
1	South Korea	539	Finland	554	South Korea	546
2	Finland	536	Japan	539	Finlnd	541
3	Canada	524	South Korea	538	Switzerland	534
4	New Zealand	521	New Zealand	532	Japan	529
5	Japan	520	Canada	529	Canada	527
6	Australia	515	Estonia	528	Netherlands	526
7	Netherlands	508	Australia	527	New Zealnd	519
8	Belgium	506	Netherlands	522	Belgium	515
9	Norway	503	Germany	520	Australia	514
10	Estonia	501	Switzerland	517	Germany	513
11	Switzerland	501	United Kingdom	514	Estonia	512
12	Poland	500	Slolvenia	512	Iceland	507
13	Iceland	500	Poland	508	Denmark	503
14	**United States**	**500**	Ireland	508	Slovenia	501
15			Belgium	507	Norway	498
16			Hungary	503	France	497
17			**United States**	**502**	Slovak Republic	497
18					Austria	496
19					Poland	495
20					Sweden	494
21					Czech Republic	493
22					United Kingdom	492
23					Hungary	490
24					Luxembourg	489
25					**United States**	**487**

Adapted from *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I)* (p. 15), by the Organisation for Economic Cooperation and Development, 2010, Paris: OECD. Copyright 2010 by OECD.

Student Performance Data (Graduation Rates)

The 4-year adjusted cohort graduation rate data paint a grim picture. In the 2007–08 school year, approximately 25% of all students nationwide (one in four) who entered high school 4 years earlier as freshmen failed to complete high school graduation requirements. This translated to 1.3 million students failing to earn diplomas. In addition to documenting extremely poor graduation rates, the data show very slight improvement over the previous 14 years (Figure 12)

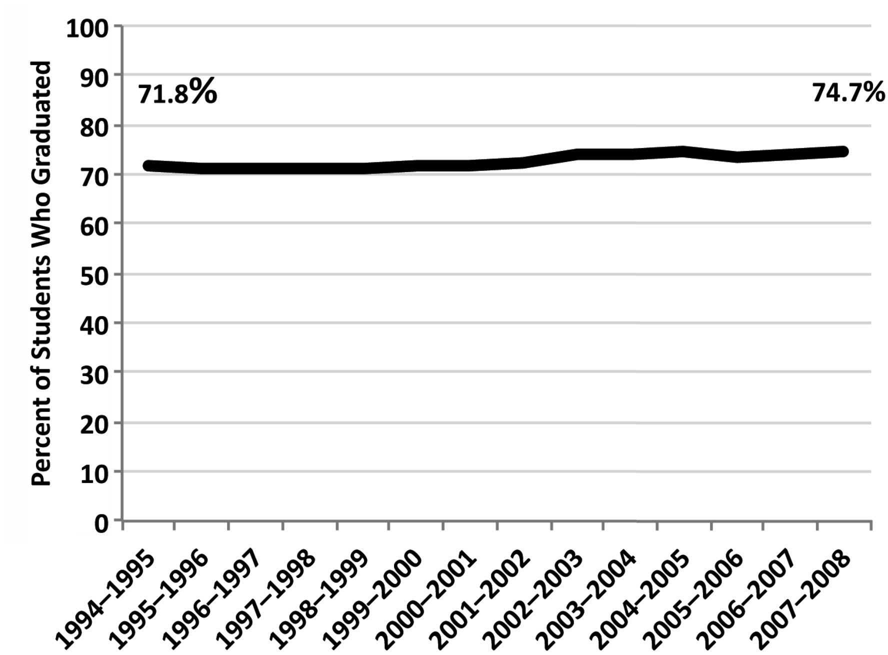


Figure 12. Average high school freshman graduation rate. Data are drawn from Snyder and Dillow (2011, p. 177).

As with test scores, graduation rates varied dramatically from state to state. They ranged from the graduation percentages in the low 50s (Nevada 51.3%, District of Columbia 56%) to the high 80s (Vermont 89.3%, Wisconsin 89.6%) (Snyder & Dillow, 2011).

As with test scores, student performance based on graduation rates shows significant inequality when analyzed by race. Asian/Pacific students and

White students had the highest percentage of graduate rates (91% and 81%, respectively). Other ethnic groups had much lower percentage of graduation rates: American Indian/Alaska Native/Asian Pacific Islander, 64%; Hispanic, 64%; and Black,62%. (Figure 13).

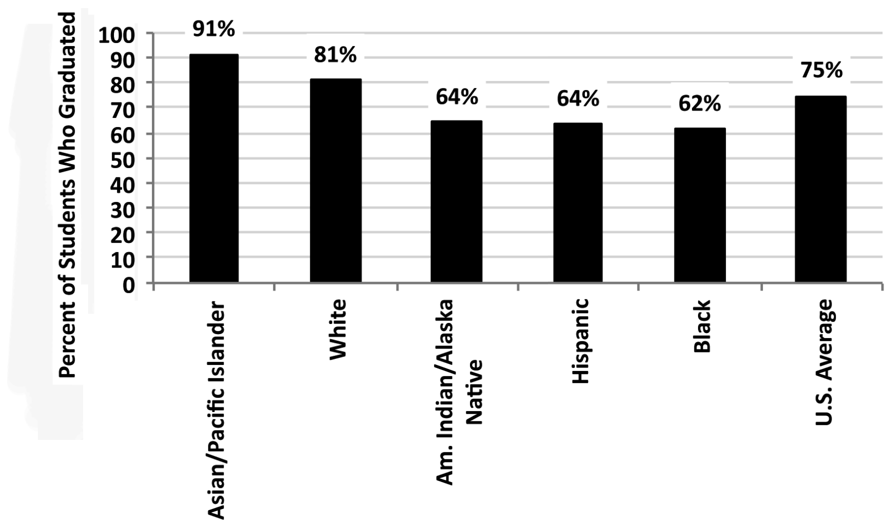


Figure 13. High school graduation rate by ethnicity (2007–08). Data are drawn from Stillwell (2010, p. 7).

Student Performance Data (Summary)

The poor performance of 12th graders nationwide in achieving reading and math proficiency is a clear indicator of the deficiencies of the U.S. education system. However, the percentage of 18-year-olds who are proficient in reading drops even more dramatically when graduation rate data are factored in. In other words, the 2009 NAEP proficiency reading rate of 38% for 12th-grade students (Figure 8) leaves out the 25% of students who failed to graduate and most likely fell below proficiency in reading. Factoring in those students produces the statistics shown in Table 2.

Table 2
Reading proficiency of all 18-year-olds

	% of 12-Grade Students At or Above NAEP Reading Proficiency	Graduation Rate	% of all 18-Year-Olds At or Above NAEP Reading Proficiency
All	38	75	28
White	46	81	37
Hispanic	22	64	14
Black	17	62	11

The data in column 1 are drawn from the National Assessment of Educational Progress Reading Assessments of 1992, 1994, 1998, 2002, 2005, and 2009 (National Center for Education Statistics, 2012b). The data from column 2 are drawn from Stillwell (2010, p. 7).

This analysis suggests that as few as 28% of all 18-year-olds in 2009 were reading at or above proficiency levels. When the data are broken down further by ethnicity, the results are staggering. Only 14% of Hispanic children, and 11% of Black children were reading at proficiency by age 18! While this is a rough calculation and doesn't count any 18-year-old dropouts who may have been proficient in reading or any 18-year-olds who were tested and didn't graduate, the essence of the outcome is clear. The United States is failing to educate the vast majority of its 18-year-olds in reading. The proficiency scores were even worse for math; just 29% of 12th-grade students were at or above proficiency. With graduation rates factored in, only 22% of 18-year-olds were proficient in math.

EDUCATION AT THE CROSSROADS: THE STATE OF TEACHER PREPARATION

It is clear that education is at yet another crossroads. Despite the investment of an enormous amount of time, money, and energy, we face the exact same problems identified almost 30 years ago in *A Nation at Risk*. Student achievement on standardized tests is well below proficiency standards, too few students are graduating from high school, there is a dramatic gap between the performance of White students and that of African-American and Hispanic students, and student performance lags far behind that of other industrialized nations. The data suggest we have made no progress whatsoever.

This stunning lack of improvement in student performance in the face of such an enormous effort leaves us with the question: what have we missed? The answer takes us back to the most essential component of education, teaching. While focusing on structural interventions, we failed to examine and improve what actually takes place in the classroom between teachers and students. Structural interventions by themselves do not necessarily impact the quality of teaching. Increased funding, smaller class sizes, school choice, and charter schools have no impact if teachers are not given the skills to be effective. This was the focus of the Wing Institute's Fifth Annual Summit on Evidence-Based Education, *Education at the Crossroads: The State of Teacher Preparation.*

The Wing Institute's annual evidence-based education summits were created to help answer the question of what is missing in education reform. They bring together education stakeholders from a wide range of professions, disciplines, organizations (academic, service, education, research, and advocacy), and consumers in a 2-day working session built around a specific topic. The goal is to share the very latest data and research on the topic, facilitate discussion and problem solving among a diverse group of participants, and establish action steps for dissemination of the resulting information into real-world settings. Past summit topics have included:

- **Building an Evidence-Based Education Roadmap**
- **Response to Intervention (RtI): An Evidence-Based Education Review**
- **Sustainability: Implementing Programs That Survive 100 Years**
- **Data-Based Decision Making: The Achilles' Heel of Evidence-Based Education**

The following chapters are the proceedings from the Wing Institute's 2010 summit, *Education at the Crossroads: The State of Teacher Preparation.* The summit focused on the critical role of teacher preparation in any reform effort, including the importance of linking student outcomes to teacher performance, and linking teacher quality to teacher preparation, induction, and support. A review of the state of the art on teacher preparation was provided by three speakers whose professional accomplishments have significantly advanced our knowledge: Dr. James Kauffman (Professor Emeritus of Education, University of Virginia), Dr. Dan Reschly (Professor of Education and Psychology, Vanderbilt University), and Dr. Larry Maheady (Professor, Department of Curriculum and Instruction, SUNY Fredonia).

In an attempt to answer the question of what is missing in education reform, the Wing Institute has been conducting an extensive and ongoing search of existing databases, research studies, policy analyses, and other sources of scientific and performance data for clues. Historically, the biggest obstacle to answering this question has been a lack of data on how we are doing (student and school performance outcome data) and what works in education (efficacy and effectiveness research on education interventions). When performance outcome data were present, they seldom measured relevant outcomes consistently and empirically over time. Where research has existed, it has often been qualitative (subjective), not quantitative (objective). As a result, most reform efforts have been flying blind, with little empirical feedback to evaluate their impact and effectiveness.

This situation has been changing recently, as an abundance of useful performance outcome and research data are becoming available. The bad news is that these data question the value of many of our education reform efforts. The good news is that they are starting to paint a picture of where we are and what went wrong. The best news is that they provide guidance for where we need to go to make effective school reform a reality. That guidance points toward the importance of teachers, and to new and more effective strategies for teacher preparation.

In the first chapter, ***Effective Teachers Make a Difference****,* Jack States of the Wing Institute reviews the most recent research and data on teacher preparation, including the impact of teachers on student achievement, the critical skills that make teachers effective, the evidence-based strategies for producing effective teachers through teacher preparation programs, and strategies for transitioning teachers from preservice to classroom.

In the second chapter, ***Science and the Education of Teachers*****,** James Kauffman discusses the importance of making teacher preparation as scientific as possible and urges not just adopting but embracing a scientific and mathematical approach to improving education. He emphasizes that professions based on scientific evidence and field tests develop manuals and checklists to guide their practices, and argues that education must do the same.

In the third chapter, ***Comprehensive Teacher Induction: What We Know, Don't Know, and Must Learn Soon!***, Larry Maheady and Michael Jabot review how teacher induction programs have failed to support new teachers, improve their teaching skills, or positively impact student learning. They discuss what we know and don't know about teacher induction, and describe the promising efforts of one regional state college to improve teacher induction.

Taken together, these papers begin to build a roadmap for actually linking school reform initiatives to student performance outcomes.

REFERENCES

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). *The Condition of Education 2011* (NCES 2011-033). Retrieved from http://nces.ed.gov/pubs2011/2011033.pdf

Florida Department of Education. (n.d.). *Class Size Reduction Amendmen*t. Retrieved from http://www.fldoe.org/classsize/

Gardner, D. P., Larsen, Y. W., Baker, W. O., Campbell, A., Crosby, E. A., Foster, C. A., Jr., ...Wallace, R. (1983). *A Nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education*. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED226006

Gorman, S. (2010). *An Introduction to NAEP.* (NCES 2010-468). Retrieved from http://nces.ed.gov/nationsreportcard/pdf/parents/2010468.pdf

Grady, S., & Bielick, S. (2010). *Trends in the Use of School Choice: 1993 to 2007* (NCES 2010-004). Retrieved from http://nces.ed.gov/pubs2010/2010004.pdf

Hall, D., & Gutierrez, A. S. (1998). *Getting Honest about High School Graduation*. [PowerPoint slides]. Retrieved from http://www.edtrust.org/sites/edtrust.org/files/publications/files/Session14GettingHonestAboutHSGraduation.ppt

Howell, W., Peterson, P. E., & West, M. (2007).What Americans think about their schools: The 2007 Education Next—PEPG Survey. *Education Next, 7*(4), 12-26. Retrieved from http://educationnext.org/what-americans-think-about-their-schools/

Luckie, M. S. (2009). *California's class-size-reduction program: Total spending since 1996*. [Interactive Graph]. Retrieved from http://californiawatch.org/k-12/californias-class-size-reduction-program-total-spending-1996

National Center for Education Statistics (NCES). (2010). *The Nation's Report Card: Grade 12 Reading and Mathematics 2009 National and Pilot State Results.* (NCES 2011–455). Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2009/2011455.pdf

National Center for Education Statistics (NCES). (2011a). *Fast Facts*. Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=30

National Center for Education Statistics (NCES). (2011b). *The Nation's Report Card: Mathematics 2011*. (NCES 2012–458). Institute of Education Sciences, U.S. Department of Education, Washington, DC.

National Center for Education Statistics (NCES). (2011c). *The Nation's Report Card: Reading 2011*. (NCES 2012–457). Institute of Education Sciences, U.S. Department of Education, Washington, DC.

National Center for Education Statistics (NCES). (2011d). *Data Explorer for Long-Term Trend.* [Data file]. Retrieved from http://nces.ed.gov/nationsreportcard/lttdata/

National Center for Education Statistics (NCES). (2012a). *Mathematics: Grade 12 National Results*. Retrieved from http://nationsreportcard.gov/math_2009/gr12_national.asp?subtab_id=Tab_3&tab_id=tab1#tabsContainer%20information

National Center for Education Statistics (NCES). (2012b). *Reading: Grade 12 National Results*. Retrieved from http://nationsreportcard.gov/reading_2009/gr12_national.asp?subtab_id=Tab_3&tab_id=tab1#tabsContainer information

Nichols, A., & Özek, U. (2010). *Public School Choice and Student Achievement in the District of Columbia*. (Working Paper 53). Retrieved from National Center for Analysis of Longitudinl Data in Education Research - Urban Institute website: http://www.urban.org/url.cfm?ID=1001499

Organisation for Economic Cooperation and Development (OECD). (2006). *PISA 2006 Technical Report.* Paris: Author.

Organisation for Economic Cooperation and Development (OECD). (2010a). *PISA 2009 Results: Learning to Learn – Student Engagement, Strategies and Practices (Volume III).* http://dx.doi.org/10.1787/9789264083943-en

Organisation for Economic Cooperation and Development (OECD). (2010b). *PISA 2009 Results: What Makes a School Successful? – Resources, Policies and Practices (Volume IV).* Retrieved from http://dx.doi.org/10.1787/9789264091559-en

Organisation for Economic Cooperation and Development (OECD). (2010c). *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (Volume I).* Retrieved from http://dx.doi.org/10.1787/9789264091450-en

Snyder, T. D., & Dillow, S. A. (2010). *Digest of Education Statistics, 2009*. (NCES 2010-013). Retrieved from http://nces.ed.gov/pubs2010/2010013.pdf

Snyder, T. D., & Dillow, S.A. (2011). *Digest of Education Statistics 2010* (NCES 2011-015). Retrieved from http://nces.ed.gov/pubs2011/2011015.pdf

Stillwell, R. (2010). *Public School Graduates and Dropouts From the Common Core of Data: School Year 2007–08*. (NCES 2010-341). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010341

Zinth, K. (2010). *Class-Size Policies*. Retrieved from Education Commission of the States website: http://www.ecs.org/clearinghouse/85/21/8521.pdf

Chapter 1

Effective Teachers Make a Difference

JACK STATES
RONNIE DETRICH
RANDY KEYWORTH
The Wing Institute

Abstract: The failure of the American education system to meet expectations, as well as the failure of school reform efforts to alter this picture, has increasingly turned the focus of school improvement to teachers. Research supports the important role that teachers play in student achievement. Given the pivotal position of teachers in student success, the question becomes, are teacher preparation programs doing their part to produce quality teachers? This chapter examines the available research on effective teaching, how to impart these skills, and how to best transition teachers from preservice to classroom with an emphasis on improving student achievement. We review current preparation practices and examine the research evidence on how well they are preparing teachers. We are fortunate that sufficient research is available that suggests how teacher training can be improved and successful classroom teachers produced.

There is a commonsense belief that good teachers make a difference in a child's life. This notion is not surprising since most of us have benefited from a teacher who inspired and challenged us. Critical questions need to be asked: Is this impression supported by rigorous research evidence? How much influence does one teacher have in improving student achievement?

In the 1960s, the prevailing wisdom emphasized the importance of home and socio-economic status on student achievement. The impact of school and, in particular, teachers was downplayed (Cochran-Smith & Zeichner, 2005). Since that time, the importance placed on teachers has gained traction. Public policy has directed greater resources to teachers. Improvements in the quality of research are increasingly providing decision makers with a convincing body of evidence on the topic of how to effectively train teachers. This research corroborates what was once only an intuitive notion: A quality teacher can significantly affect a child's education and improve student achievement (Sanders & Rivers, 1996). The remainder of this chapter considers this evidence in considerable detail.

Much of school improvement over the past 40 years has been disappointing. Despite clear evidence of the impact that well-prepared teachers have on student achievement, most major reform has not addressed how teachers teach. Figure 1 illustrates that reforms in the guise of structural interventions have had, at best, a minimal impact on achievement as measured by high-stakes tests and graduation rates (Yeh, 2007).

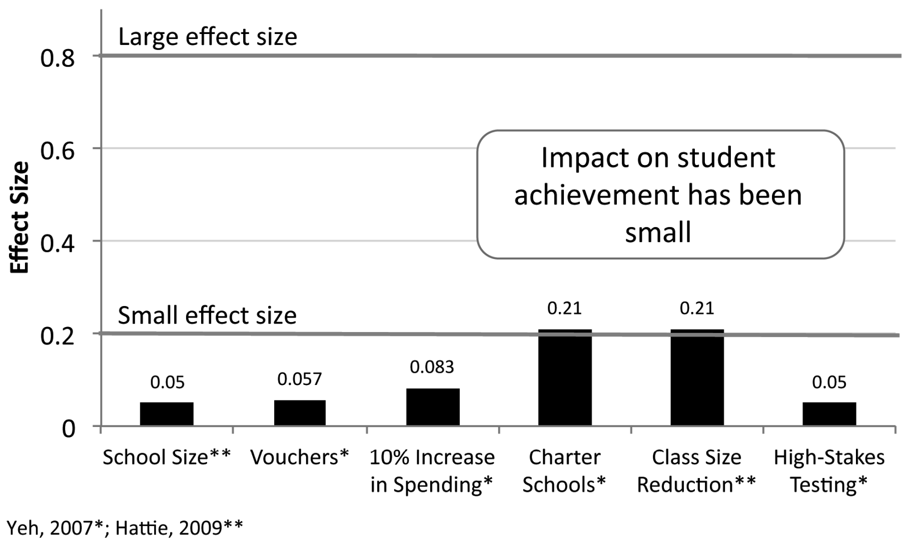


Figure 1. Impact of strucural reform interventions. Data are drawn from Hattie (2009, Appendix B) and Yeh (2007, p. 431).

As stakeholders in education, we are fortunate to have reliable evidence — to be addressed in the remainder of this chapter — that supports the strategy of improving teacher performance as a cornerstone of future reform efforts. The goal of this strategy is to address deficits in education noted in policy reviews such as *A Nation At Risk* (Gardner et al., 1983), while being consistent with the reform goals delineated in the No Child Left Behind Act of 2001, which calls for improved standards for teacher training and credentialing.

WHAT RESEARCH TELLS US ABOUT THE IMPORTANCE OF TEACHERS

Figuring out what research tells us about the significance of teachers has not been without serious challenges. Prior to the 1980s, qualitative research predominated the field of education, and quantitative research methods were not often applied to examining this issue (Cochran-Smith & Zeichner, 2005). Not until the 1990s was quantitative research commonly seen in the literature or methods such as value-added modeling employed in studies on the importance of teachers. An advantage of this trend toward quantitative measures is that these measures can be used to establish causal relations between interventions and outcomes. The results of these studies can be analyzed for effect size, allowing for reliable comparison of results across studies.

Although qualitative research can be effective in describing phenomena, the results cannot be separated from the individual or case studied, making the data inherently subjective. In contrast, quantitative research relies on measurements of events that can be expressed as a specific quantity or unit and whose results can be generalized to populations, settings, treatment variables, and measurement variables used to predict future events. Quantitative and qualitative methods are valuable tools when used to answer questions for which they were designed.

Table 1
Effect Size

Cohen's d*	Effect Size
Small	d=0.2
Medium	d=0.5
Large	d=0.8

Note: Effect sizes range from minus to positive. A small effect is commonly defined as d = 0.2, medium as d = 0.5, and large as d = 0.8, but it is not uncommon to see effect sizes that exceed 1.0. The terms "small," "medium," and "large" are relative. Researchers accept the risk of using relative terms in the belief that they have more to gain than lose by offering a common conventional frame of reference when no better way to estimate the impact of a practice or intervention is available. Effect sizes in the 0.4 range or smaller are often considered minimal levels for educational purposes (Gersten et al., 2005).

* The accepted benchmark for effect size comes from Jacob Cohen (1988), a U.S. statistician and psychologist.

Among the first to use effect size to address the importance of teachers in improving student achievement were Johnson and Zwick (1990). Using data compiled by the National Assessment of Educational Progress (NAEP), they calculated that teachers had an average effect of 0.24 per year on students ages 9, 13, and 17, in the subject areas of reading, writing, civics, U.S. history, mathematics, and science.

Hattie (2009) worked for 15 years to research and synthesize over 800 meta-analyses on the influences on achievement in school-aged students. He offered an effect size for each of the educational practices and interventions. He also reported that research conducted in New Zealand identified an effect size of 0.35 for teacher effectiveness across three subject areas: reading, mathematics, and writing.

The importance of a teacher's contribution to student performance was demonstrated in a randomized controlled trial conducted by Nye, Konstantopoulos, and Hedges (2004). The results of this study showed substantial differences among teachers in their capacity to produce achievement gains in students. Simply stated, they found that 7% to 21% of student gains could be attributed to teacher effectiveness.

Together, teacher effect size and percentage of student gains build the case for the importance of teachers in student success.

Another attempt to ascertain a teacher's impact used value-added modeling. Sanders and Rivers (1996) wanted to understand the effect on students of prolonged exposure to effective teachers compared with prolonged exposure to ineffective teachers (Figure 2).

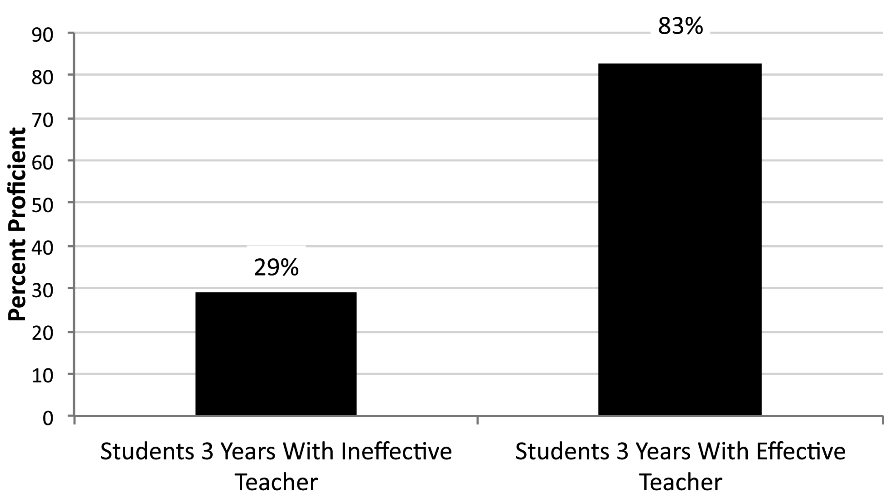


Figure 2. Teacher effectiveness: Gains in 8th-grade math. Data are drawn from Sanders and Rivers (1996, p. 3).

They employed the Tennessee Value-Added Assessment System (TVAAS), designed to determine an individual teacher's influence on the rate of academic growth. The study found that students with similar aptitude and initial achievement scores performed significantly differently depending on the quality of the teachers to whom the students were assigned. The effects of being taught by effective and ineffective teachers were still measurable 2 years after the initial study. The results suggested that the teacher effects on students were additive and cumulative, and offered little evidence that more effective teachers in later grades would make up for years of ineffective instruction.

A study in Texas elementary schools estimated that teachers accounted for 3% of the variance in student achievement (Mendro, Jordan, Gomez, Anderson, & Bembry, 1998). A large-scale U.S. government study reported teacher impact on student test scores between 4% and 18% (Rowan, Correnti, & Miller, 2002). An American Education Research Association (AERA) policy paper on the topic of value-added research on teacher effectiveness, *Teachers Matter: Evidence from Value-Added Assessments* (2004), concluded that "value-added measurement has proven that very good teaching can enhance student learning; that family background does not determine a student's destiny; and that decisions made about teacher hiring, placement, and training make a difference for academic achievement."

In summary, the available research supports the notion that teachers make an important contribution to student success in school. The importance of teachers to student achievement gains offers educators a powerful leverage point in reform efforts. The research further supports vigorously pursuing interventions targeted at what happens in the classroom through improving how teachers teach.

A BRIEF HISTORY OF TEACHER PREPARATION

The need to provide students with qualified teachers has been an issue of concern for well over 150 years. Teacher preparation programs, commonly called "normal schools," provided undergraduate training during the 19th and early 20th centuries. Teaching preparation following this model remained basically unchanged for 100 years. No single model of pedagogy or skills to be taught teachers emerged; each state set its own credential requirements, and preparation programs tended to design their own models of training.

This situation began to change in the 1980s, when disappointment with student test scores coincided with a shortage of trained teachers, reinforcing the belief that the shortage of qualified teachers contributed to the poor performance of schools. The result of the undersupply of fully trained teachers was an increasing dependence on the use of alternatively credentialed teachers

(Constantine et al., 2009).

In 2001, concerns regarding the quality of teachers in classrooms culminated in the landmark intervention of the federal government with legislation titled PL 107–110, No Child Left Behind (NCLB). Among the many issues addressed by NCLB was the insertion of incentives for reducing the use of underqualified teachers. The law required states to provide highly qualified teachers to all students by 2014. The legislation and subsequent regulations were the first national attempts to control the quality of teachers and teacher training. NCLB regulations hold school districts accountable by requiring that their teachers meet the following standards: (a) have a bachelor's degree, (b) be fully certified and/or licensed by the state in which they teach, and (c) be competent in the subject matter they teach.

The establishment of these higher standards led to predictions of serious shortages of qualified teachers. In spite of the challenges posed in filling positions with fully credentialed personnel, schools have been successful in staffing classrooms with appropriately credentialed teachers. According to Department of Education data, by 2008 more than 95% of public school teachers had acquired the necessary teacher certification (Figure 3). It should be noted that each state has been given the flexibility to establish its own standards for "highly qualified," so the term does not have a consistent meaning. A teacher who is highly qualified in one state may not meet the standards of another state.

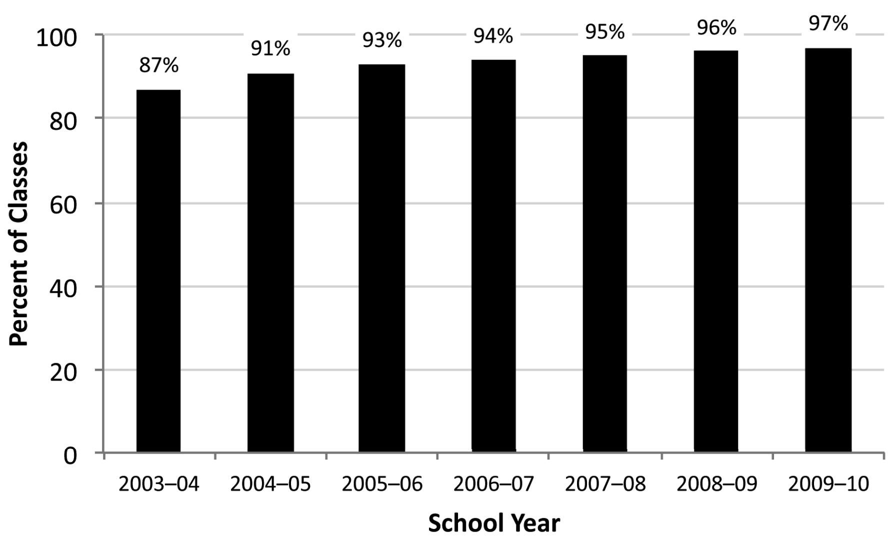


Figure 3. Core academic classes taught by highly qualified teachers. Data are drawn from ED Data Express (U.S. Department of Education, 2011).

Stiffer credential standards were not the only change in teacher education over the past 40 years, a period that witnessed a major increase in teacher education levels. In 1971, the majority of teachers, 70%, possessed a bachelor's degree, and fewer than 30% held a master's or higher degree. Today the trend has reversed itself, and now a majority of teachers, 56%, hold a master's degree (National Education Association [NEA], 2003) (Figure 4). This reversal represents a significant increase in the formal education of teachers.

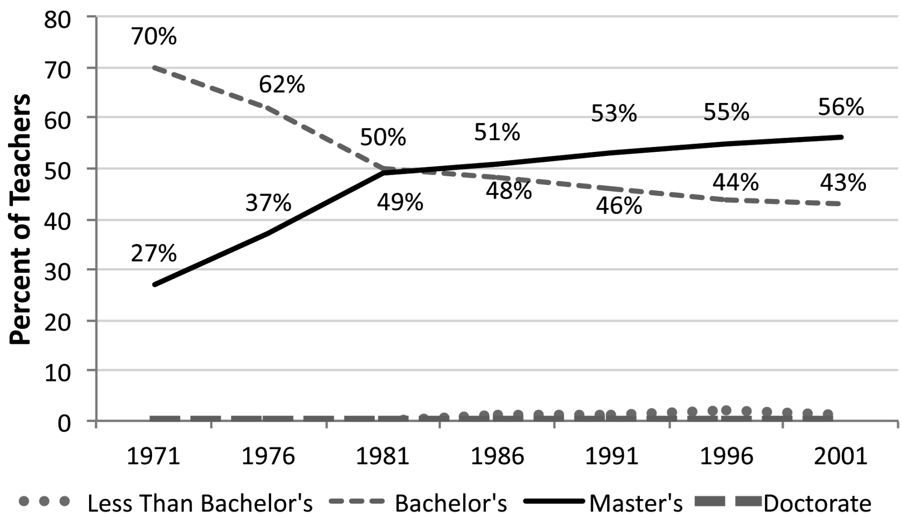


Figure 4. Public school teachers: Highest degree held. Data are drawn from the National Education Association (2003, p. 5).

Most research conducted before 2000 on the importance of education level of teachers on student achievement is correlational or qualitative. Unfortunately, much of the impetus for the shift toward post-bachelor's degree teacher education was driven by a desire to make preparation programs appear more professional, because of a lack of respect often accorded the programs rather than a desire to improve the pedagogy or teacher training models (Zeichner & Conklin, 2005).

In spite of the significant increase in the number of teachers with master's degrees, little improvement in critical student outcomes, such as test scores or graduation rates, was evident in the data from 1971 through 2001 (National Assessement of Educational Progress [NAEP], 2009).

During this time, the NAEP test scores have remained essentially unchanged across all grade levels. It is clear we have missed something critical in our attempts to improve teacher training. Mandating that teachers have credentials and increasing their time spent in higher education have not improved student performance.

The problem with reform efforts such as mandating credentials and a shift toward higher education is fundamental. The interventions were designed as simple structural modifications that did not address how teachers teach. They offered a change in the facade of teacher preparation, but not the substance of the interactions between teachers and students. To make a difference, as discussed earlier, teacher preparation reform must make changes to practices and pedagogy: what we teach teachers and how we teach them. Until practices with a strong evidence base for effectiveness are adopted and student teachers are given the opportunity to master them by working with real students, we should not be surprised when reform efforts fail. An emerging body of knowledge about what works will help to build how to teach teachers will help to build a new model of teacher education (Brophy, 2004; Joyce & Showers, 2002).

WHAT WE SHOULD TEACH: TEACHER SKILLS

If we want to provide teachers with the skills that offer the best chance for success in the classroom, we must start with the premise that the skills we teach should derive from the best available evidence on what works. Education literature is full of recommendations for what teachers should be taught. Unfortunately, much of what we have been teaching in preparation programs is based on fad, folk wisdom, and shoddy research methodology (Kauffman, 2010). Snider (2006) described the typical experience of many teacher preparation students: "I learned very little in my undergraduate teacher education program about how to teach… I knew very little about curriculum, effective teaching, or principals of classroom management..." Anecdotal evidence indicates that many teachers feel their training experience was similar. Some say they felt lost when they began teaching. They were poorly prepared to handle student conduct, assess student performance, or effectively implement teaching strategies (Cochran-Smith & Zeichner, 2005).

An excellent place to start a discussion of what works for students is the research of Wang, Haertel, and Walberg (1997), which identifies 28 categories of variables that influence student learning. By combining the effect size of different practices derived from research along with a content analysis and a survey of educational experts, Wang and her colleagues established a weighted score for each category. All three data sources produced sufficient agreement that the variables could be ranked according to impact on student learning. In Figure 5, the domains of effective instruction and the relative impact of each are summarized.

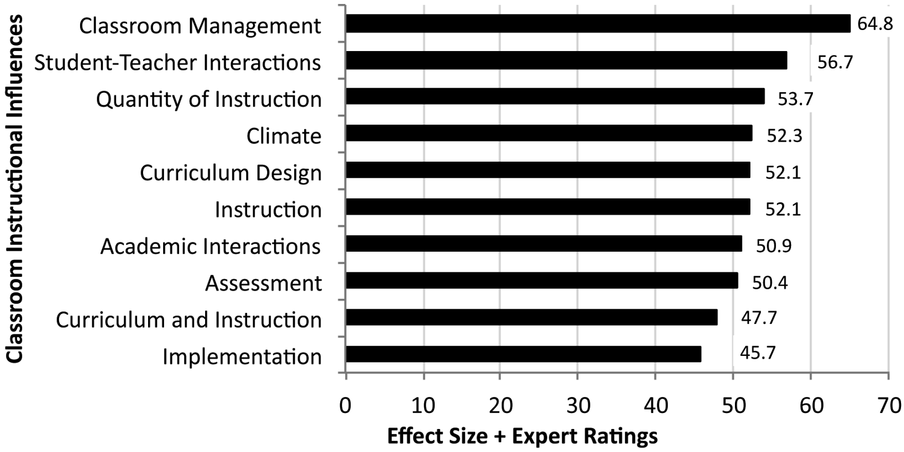


Figure 5. Impact of instructional influences on learning. Data are drawn from Wang, Haertel, and Walberg (1997, p. 201).

These findings are supported in two subsequent meta-analyses, (Hattie, 2009; Kavale, 2005), each of which corroborates the Wang et al. 1997 study. These meta-analyses build a case for the importance of assessment, classroom management, teaching strategies, and well-designed curriculum. In Figure 6, the effect sizes for different instructional practices are presented. All of the effect sizes are in the range to be considered clinically or socially significant (Gersten et al., 2005).

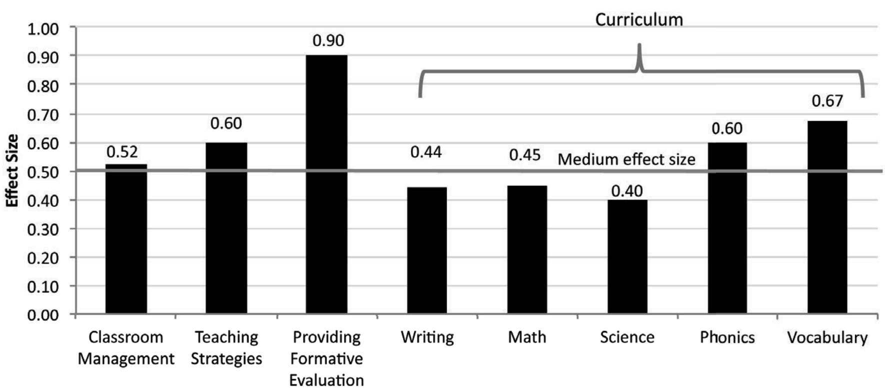


Figure 6. Impact of effective classroom interventions. Data are drawn from Hattie (2009, Appendix B).

Formative Assessment

When it comes to critical skills for teachers, few are as important or powerful as formative assessment. Also known as progress monitoring, formative assessment is frequent ongoing assessment of student performance. Research consistently ranks formative assessment in the top tier of variables that make a difference in improving student achievement (Hattie, 2009; Marzano, 1998). It is not surprising that approaches such as Response to Intervention (RtI), Data-Based Decision Making (DBDM), and Positive Behavior Interventions and Supports (PBIS) depend heavily on frequent progress monitoring.

A meta-analysis by Fuchs and Fuchs (1986) demonstrated the impact of formative assessment on student performance (Figure 7). The study provided evidence for monitoring student progress through the systematic collection of performance data. The effects of progress monitoring were found to be significantly enhanced when the data were collected weekly and when teachers interacted with this information by graphing the data and analyzing the information using set decision rules.

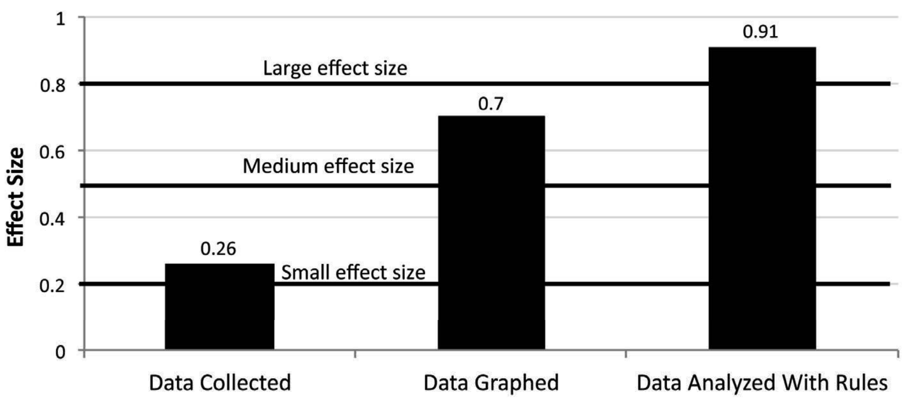


Figure 7. Impact of formative assessment (progress monitoring) on student achievement. Data are drawn from Fuchs and Fuchs (1986, p. 204).

Subsequent research (Table 2) has built a persuasive body of knowledge supporting the early work by Fuchs and Fuchs (1986) in this area. Formative assessment provides indicators to verify and maintain student progress and can act as an important diagnostic tool pointing to when and how to adjust instruction. The take-home message is that formative assessment coupled with graphing and following rules for analyzing and responding to data can be a powerful educational intervention.

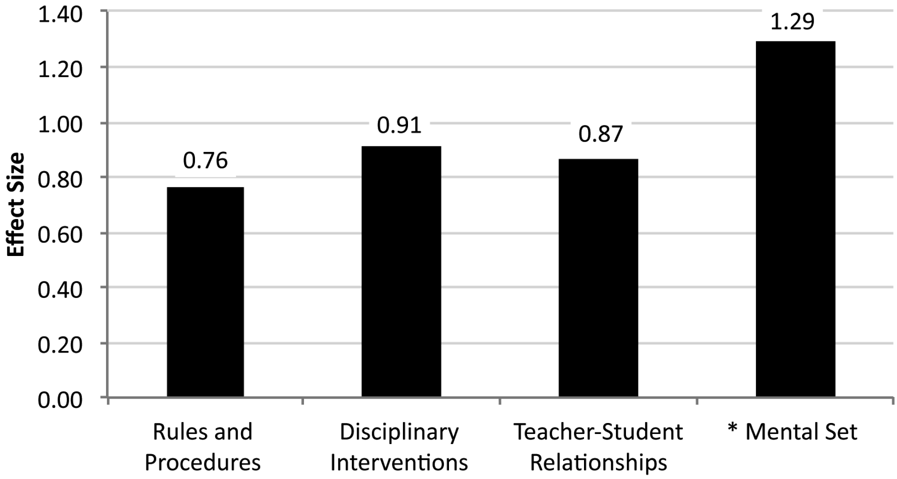
Table 2
Effect size for formative assessment

Study	Average Effect Size
Black and Wiliam, 1998	0.4–0.7
Bloom, 1976	0.54
Haller, Child, and Walberg, 1998	0.71
Hattie, 2009	0.90
Fuchs and Fuchs, 1986	0.90
Kavale, 2005	0.70
Kumar, 1991	1.31
Scheerens and Bosker, 1997	1.09
Walberg, 1999	0.94

Classroom Management

When surveyed, principals and teachers cited classroom management and student conduct near the top of the list of issues impeding the effective running of a classroom. Hattie (2009) ranked classroom management fifth among school issues affecting student performance. Classroom conduct problems have a debilitating effect on schools, impacting staff morale as well as contributing to lower student achievement (Marzano, Marzano, & Pickering, 2003). Major educational interventions such as PBIS and the Good Behavior Game (GBG) were designed specifically to mitigate the impact of misconduct by reducing behavior problems and indirectly student academic performance.

Marzano et al. (2003) conducted a meta-analysis that included 134 effect sizes derived from 100 studies on the topic of behavior management. The results from this meta-analysis are presented in Figure 8. The overall impact on student achievement in this study was an effect size of 0.521. The study reported a 20% increase in achievement when systematic rules and procedures were implemented. In the original report, the effect sizes were reported as negative numbers because the measures were a reduction of behavior problems relative to comparison conditions. For ease of understanding, the effect sizes here are reported as positives to more clearly communicate the benefits of effective classroom practices. The values remain the same.



* Teacher's keen awareness of disruptive or potentially disruptive behavior and immediate attention to it

Figure 8. Impact of behavior management factors on student achievement. Data are drawn from Marzano, Marzano, and Pickering (2003, p. 8).

Teaching Strategies

What do some teachers do that makes them better teachers? This section examines the importance of teaching strategies that make a significant difference in student learning. Hattie (2009) reviewed 14 meta-analyses of 5,667 studies to derive an effect size of 0.60 for teaching strategies. In his meta-analysis, Marzano (1998) arrived at a similar effect size of 0.52 for teaching strategies.

Unfortunately, it is not enough to know that teaching strategies make a difference. As educators, we need to know what strategies work and under what conditions they are effective. To do this, we must create a knowledge base that identifies specific interventions as well as the core strategies from which these interventions have been constructed.

For example, Swanson and Hoskyn (1998) emphasized sequencing, drill repetition, and strategy cues as effective teaching strategies. In particular, they found that reading skills (i.e., comprehension, vocabulary, and creativity) were responsive to this approach and produced large effect sizes above 0.8. If we want to increase the success of student reading, we must build reading programs based on proven core strategies described in the report of the National Reading Panel (National Institute of Child Health and Human Development [NICHD], 2000).

Unfortunately, existing research has not made the task of constructing a list of key strategies simple. We are hampered in this effort for a number of reasons. Different meta-analyses define "strategies" differently. Also, strategies are often combined in ways that make direct comparisons difficult. The teaching strategy chart below (Figure 9) offers a side-by-side look at some of the important strategies with medium to large effect sizes that teachers should master.

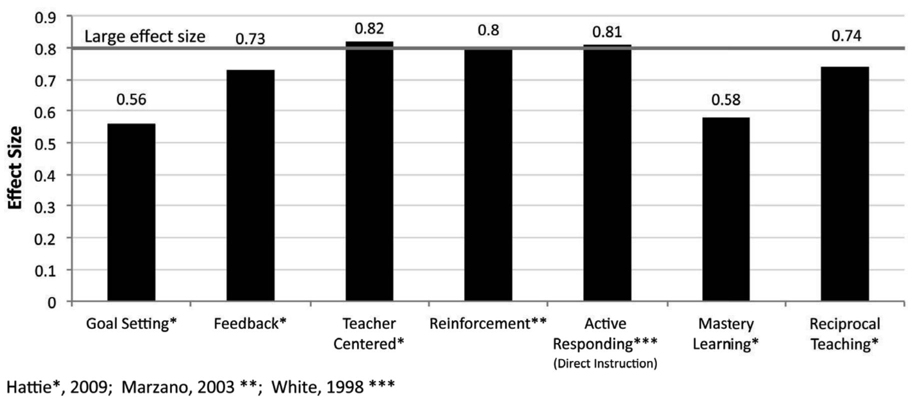


Figure 9. Impact of teaching strategies on student achievement. Data are drawn from Hattie (2009, Appendix B), Marzano, Marzano, and Pickering (2003, p. 8), and White (1988, p. 368).

Teaching Strategy Definitions (Hattie, 2009; Marzano et al., 2003; White, 1988)

Goal setting: The process of establishing a direction for learning.

Feedback: Information provided to teachers on student performance as well as information provided to students on their own performance that functions to correct or maintain performance.

Teacher centered: Having teachers establish the learning plan and criteria for successful completion, making expectations clear, demonstrating skills, checking students for skills acquisition, and having students demonstrate skill fluency over time.

Reinforcement: Rewarding student effort and providing recognition for desirable performance.

Active responding: Requiring students to talk, write, solve problems, or otherwise respond rather than sit and listen. Active responding allows students to receive more frequent and immediate feedback.

Mastery learning: Ensuring that each student masters prerequisite materials before moving on to more complex or advanced materials. One way to do this is by breaking down material into manageable units.

Reciprocal teaching: Requiring students to summarize, answer questions, clarify points of confusion, and predict what to anticipate in future lessons. The teacher and students take turns assuming the role of teacher in leading this dialogue.

Curriculum

What role does the curriculum play in fostering student achievement? Gauging the influence is often difficult. A curriculum is generally more than one teaching strategy, and studies look at the impact of the curriculum as a whole and not at each strategy and practice on its own. An examination shows that many learning strategies are shared across curricula, whether the subject matter is reading, math, science, or history.

There is a growing body of research available to educators through resources such as What Works Clearinghouse on what curricula are and are not effective. Training teachers in the use of effective curricula is challenging since different districts use different curricula. It is impossible for a teacher preparation program to train new teachers to effectively implement all of the possible curricula they may be required to use. To facilitate the process of training teachers to be effective, it may be wise to train teachers in the common, shared strategies. Below is an extended discussion of effective teaching strategies across a number of different content areas or skills.

Skills: Reading

A substantial body of research exists on how to teach reading, a fortunate circumstance because reading is pivotal to success in most subjects taught in school. Research shows that students who are poor readers in the early years are likely to continue to fall behind in future years (Juel & Leavell, 1988; Chard & Kameenui, 2000).

In 1997, Congress asked the director of the National Institute of Child Health and Human Development to convene a panel to assess the status of research-based knowledge on reading. The report of the National Reading Panel (NICHD, 2000) identified five areas with a sufficient evidence base for inclusion in reading programs. Figure 10 describes effect sizes associated with each component and compares them with effect sizes from Hattie (2009). The data from these two sources strongly suggest the importance of these components of reading: phonemic awareness, phonics instruction, fluency, vocabulary, and exposure to reading comprehension strategies. An effective reading curriculum should contain these elements, although the elements alone are not sufficient to ensure that the curriculum will be effective. It is all a matter of how the elements are combined and how the instruction is conducted.

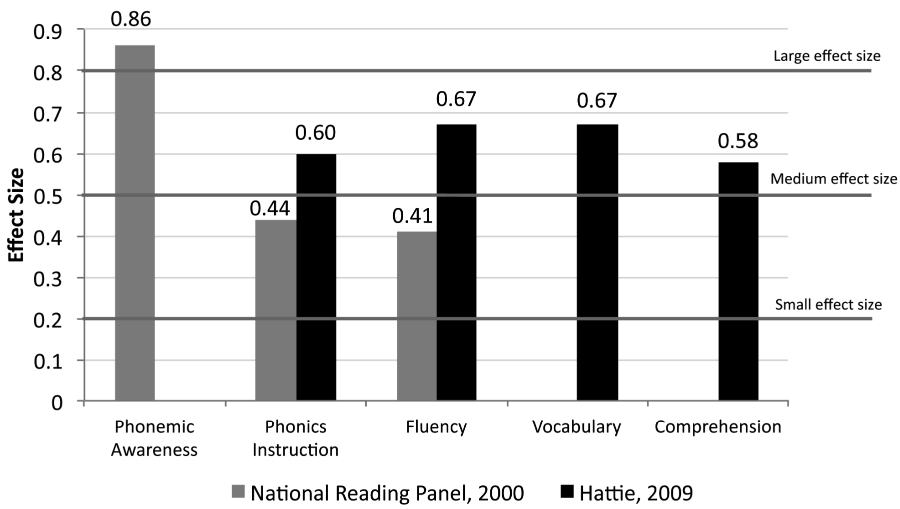


Figure 10. Effect size for components of reading. Data are drawn from Hattie (2009. Appendix B) and National Institute of Child Health and Human Development (2000, pp. 2–3, 2–112, 3–16).

Skills: Other

The evidence is not as clear for guiding curriculum selection in subject areas other than reading. Research by subject area reveals effect sizes that are generally in the medium range (Hattie, 2009; Marzano, Pickering, & Pollack, 2001). On the other hand, the available research does provide compelling evidence that certain approaches are unlikely to be effective. Perceptual motor training and whole language are examples of practices with a small effect size (Figure 11).

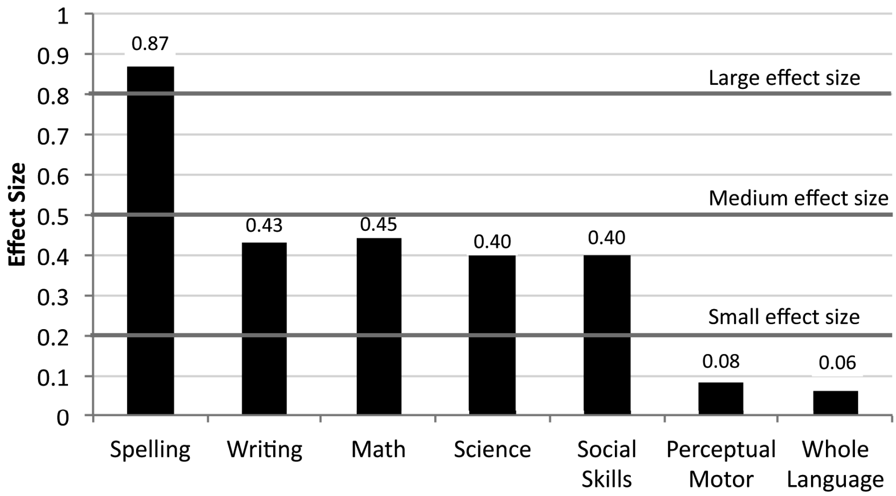


Figure 11. Effect size for curriculum other than reading. Data are drawn from Hattie (2009, Appendix B).

Subject Matter Expertise

This section looks at the evidence supporting subject matter training as a requirement of teacher preparation. Subject matter expertise is frequently identified as essential training for teachers, and a great deal of emphasis has been placed on ensuring that teachers have adequate training in the subject areas they teach. NCLB lists "knowledge of subject matter area" as one of only three critical features of a highly qualified teacher. Given the limited training time available in teacher preparation programs, is subject matter important?

The Education Commission of the States (Allen, 2000) found little evidence to support subject matter training as critical to effective teacher preparation. Wilson and Floden (2003) and Floden and Meniketti (2005) found little data supporting subject matter training as significant in producing successful teachers. Two comprehensive studies, by Ahn and Choi (2004) and Hattie (2009), looking at the effect size of teacher subject matter training on student achievement, found the impact to be no greater than 0.09 for all subjects (Figure 12). The greatest effect size was in math, and even then the impact was only 0.12, still below what is considered a small effect size of 0.2.

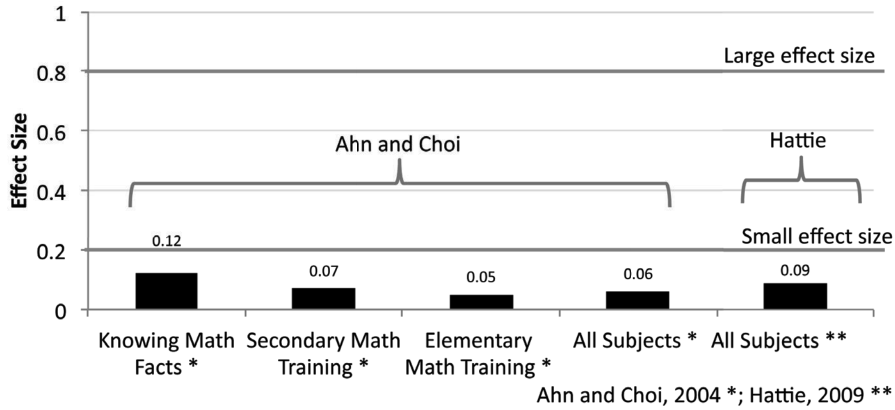


Figure 12. The impact of subject matter training on student achievement. Data are drawn from Hattie (2009, p. 297) and Ahn and Choi (2004, p. 30).

WHAT SKILLS TEACHER PREPARATION PROGRAMS TEACH

Knowing to what extent teacher preparation programs are teaching formative assessment, classroom management, teaching strategies, and curriculum is important to determine if these programs are equipping teachers with the training they most need. Each state certifies teachers within that state and operates training independently of the other states. No national standards exist for teacher preparation. There are two national organizations whose mission is to improve programs through accreditation: the National Council for the Accreditation of Teacher Education (NCATE) and Teacher Education Accreditation Council (TEAC). NCATE and TEAC established standards for programs, which include requiring schools to complete an audit consisting of paper compliance and site visits. Unfortunately, neither has looked at the effectiveness of graduate teachers from universities that NCATE or TEAC approved and the achievement of the students they instruct. Furthermore, accreditation is not mandatory for preparation programs. NCATE accredits fewer than half of the programs in the nation (650 of the over 1,500 programs). TEAC has a little over 200 accredited members.

Another way to discover what preparation programs are teaching is to survey teachers about their programs. It is important to note, survey data of this type have their limitations. In this instance, it is what teachers said about their programs, not what the programs did. When asked to describe their satisfaction with the preparation program they had completed, teachers often gave contradictory responses. General questions regarding satisfaction elicited positive responses, but queries about specific areas of training drew answers that were not always as affirmative (Figure 13).

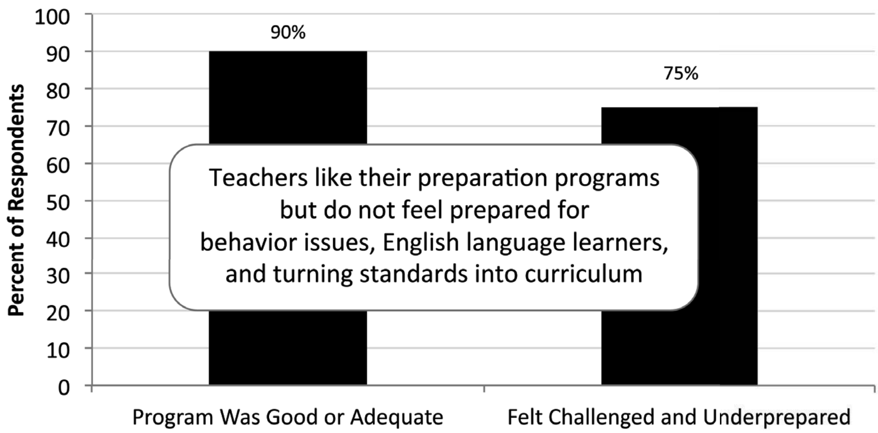


Figure 13. Teacher survey of preparedness. Data are drawn from Hart Research Associates (2010, p. 5).

Reading

Given the importance of reading for students, knowing what preparation programs are doing to prepare teachers to teach reading is crucial. One study that examined preparation programs surveyed course syllabi from a representative sample of 72 U.S. teacher preparation programs about what they offered prospective teachers in reading training (Walsh, Glaser, & Wilcox, 2006). Despite 60 years of rigorous research into what works in teaching reading, many teacher preparation schools fail to teach the fundamental components of reading.

As noted earlier in this chapter, the National Reading Panel report (NICHD, 2000) substantiated the need for phonemic awareness, phonics, fluency, vocabulary building, and exposure to reading comprehension strategies. The report found that only 15% of the sampled schools provided training in all the components. Figure 14 describes the number of components these schools taught. The fact that NCATE accredited a program did not increase the likelihood the school would teach scientifically based reading. The teaching of phonics was the most frequently taught component of reading, but much of reading instruction did not make use of the other critical components. The study found that teacher preparation faculty often portrayed scientifically based reading instruction as one of many approaches no more valid than others.

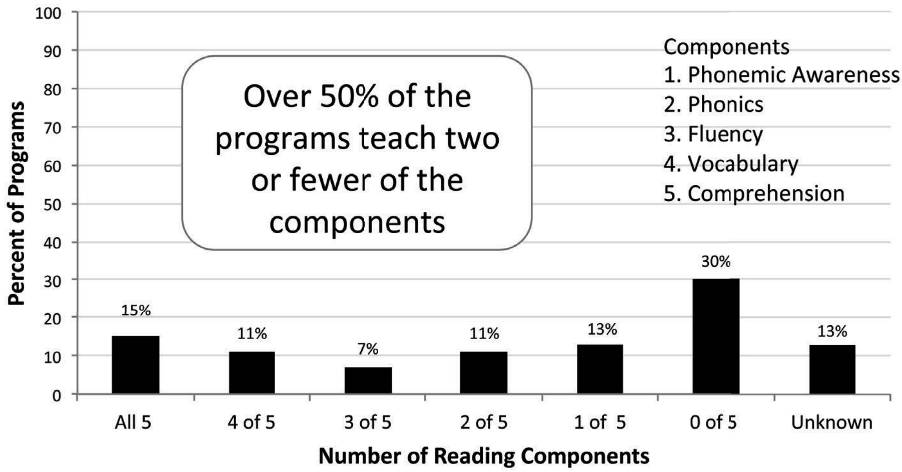


Figure 14. Percent of teacher preparation programs teaching evidence-based components of reading. Data are drawn from Walsh, Glaser, and Wilcox (2006, p. 24).

Formative Assessment

Earlier in this chapter we described the importance of formative assessment in improving outcomes for students. It is therefore vital to know how well teacher preparation programs are doing in training future teachers about formative assessment. Spear-Swerling (2008) surveyed 13 teacher preparation schools in Connecticut to find out whether they were teaching formative assessment. She identified the frequency of the term "formative assessment" (or comparable terms "progress monitoring," "rapid assessment," or "ongoing assessment") in course descriptions.

Despite formative assessment's great potential for improving student achievement scores, teacher preparation programs surveyed in Connecticut did not emphasize this powerful tool. The Spear-Swerling study found only 14.3% of the preparation programs surveyed included formative assessment and none incorporated Dynamic Indicators of Basic Early Literacy Skills (DIBELS), a program that relies heavily on formative assessment, into their syllabi (Figure 15).

If what occurred in Connecticut can be generalized to other states and other preparation programs, inadequate training in formative assessment has the potential to undermine major reform efforts such as Response to Intervention (RtI) built around ongoing assessment of students. Limited training in formative assessment risks the inadequate education of a generation of teachers who are increasingly held accountable for the failure of students.

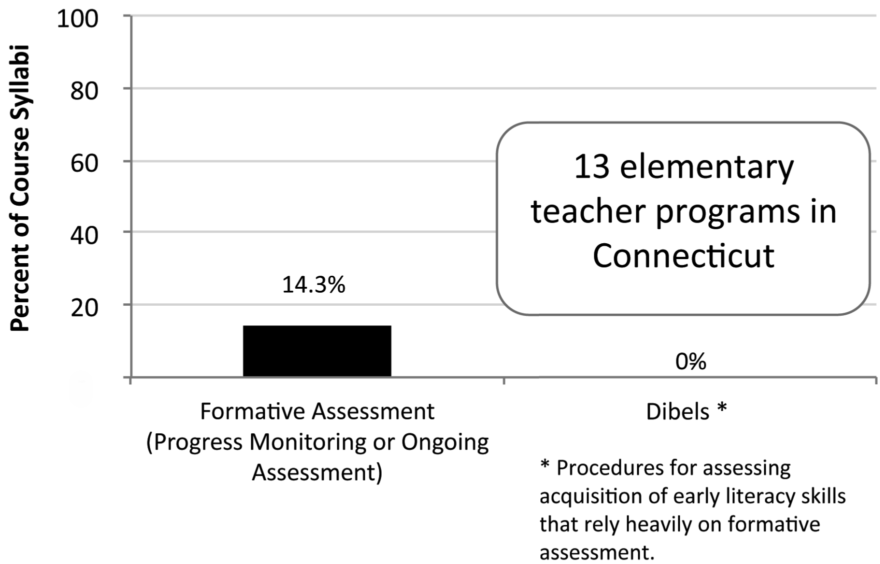


Figure 15. Survey of courses teaching formative assessment. Data are drawn from Spear-Swerling (2008, p. 285).

Behavior Management

Instruction in behavior management has been shown to have a significant impact on student achievement (Marzano et al., 2003), although qualitative reports from teachers suggest they are unprepared to handle conduct problems when they begin teaching (Hart Research Associates, 2010).

More rigorous methods designed to get at this issue have been difficult to find. One study (Begeny & Martens, 2006) does provide insight into this area. It looked at teacher course work and applied training in behavior management practices for elementary, secondary, and special education students in six teacher preparation programs in the Northeast. A major finding was that students received "little training in behavioral instruction concepts, strategies, programs, and assessment practices." Participants reported "no training" for 43% of the behavioral items surveyed in the study.

HOW TEACHER PREPARATION PROGRAMS PREPARE TEACHERS

Discussions of teacher preparation generally focus on the content teachers should be taught rather than on how best to instruct teachers. The way we prepare teachers has varied little over the past 100 years. We rely on lectures provided by professors in university settings along with a traditional 8 weeks of student teaching, which generally happens at the end of the preparation process. In recent years, there have been calls to change this model. A report commissioned by NCATE (2011) proposed a radical departure from the university-based model to a clinically based approach that emphasizes field experience over didactic training. In this section, we will examine research on approaches to teaching teachers that increase the likelihood that skills learned in the preparation setting will be mastered and used when the new teacher enters the classroom.

Didactic Presentation (Lecturing)

Lecturing prospective teachers is the most common form of instruction found in teacher preparation. This method remains in use for a number of important reasons: It is efficient and flexible, it gives instructors greater control of the material to be presented to students, and it offers easy methods — tests and papers — to assess mastery of the material (Friesen, 2011). Unfortunately, there are also disadvantages associated with lectures (Heward, 2004). Among these is the fact that listening to a lecture is a passive experience. Research suggests that requiring frequent responses during instruction is the most effective way to improve student performance (Heward, 2008). If this is true for schoolchildren, it may also hold true for prospective teachers. More importantly, answering questions during a lecture is a far cry from being able to demonstrate effective use of a skill in the field.

Coaching

Joyce and Showers (2002) looked at the question of how best to train teachers so that new knowledge is transferred to classrooms. Their research examined four methods of training teachers.

1. <u>Discussion</u>: Theories, facts, and information presented through discussion, readings, or lectures.
2. <u>Demonstration</u>: Modeling a skill for the persons being trained.

3. Practice and feedback: Using a skill under simulated conditions.
4. Coaching: Collaborative work between trainer and trainee to solve problems or answer questions that arise in the classroom.

The traditional lecture method did not result in teachers applying newly acquired skills in the classroom. The introduction of skill demonstrations by the trainer was insufficient to ensure the transfer of the skill from the demonstrator to the trainee. Even the introduction of practice was not enough to see the skill put into use in the classroom. Only when coaching was added did a significant transfer of skills to the classroom occur (Figure 16).

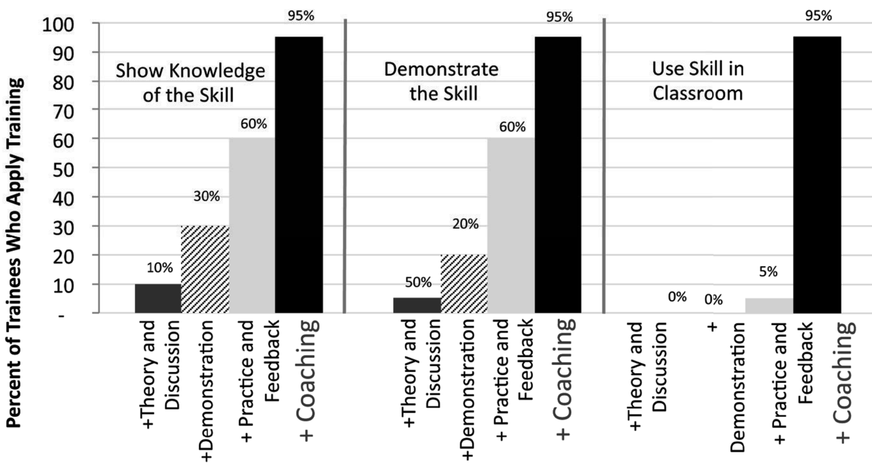


Figure 16. Coaching: Teacher training method producing the best results. Data are drawn from Joyce and Showers (2002, p. 78).

This study shows how critical it is that teacher preparation programs balance the traditional university-based training with effective field experience to give new teachers the necessary skills to be successful.

Field Experience

The question is, what types of field experience result in the best skill acquisition by teachers in training? Field experience (student teaching) is a set of training experiences occurring in actual school settings or in a clinical or laboratory environment. It is designed to bridge the gap between the university setting and actual classroom teaching and to integrate educational theory, knowledge, and skills in practice under the direction of a qualified supervisor. While the Joyce and Showers data (2002) suggest that coaching used in teacher training is critical for ensuring that new skills are actually used in the classroom, it is not enough to argue that all field experience techniques are effective.

Student teachers directly observe teaching, participate in teaching, and independently teach students. They are meant to work with a mentor teacher from an active K–12 classroom and/or preparation program faculty in order to receive feedback designed to hone skills previously taught in the university setting. Effective field experience requires a high level of coordination between the K–12 placement site and the preparation program. Unfortunately, such coordination requires substantial time and effort. In practice, insufficient time and resources are allocated to field experience, and teachers often receive inadequate coaching and are left to fend for themselves.

The importance that educators place on field experience is evidenced by the ubiquitous presence of the practice throughout teacher preparation. In spite of the acceptance of the need for field experience, there is little agreement on methodology, frequency, duration, and supervision of field experience placements (Clift & Brady, 2005). The lack of agreement on this practice is borne out by the different state standards for the amount of field experience states require of new teachers (American Association of Colleges for Teacher Education [AACTE], 2010) (Figure 17).

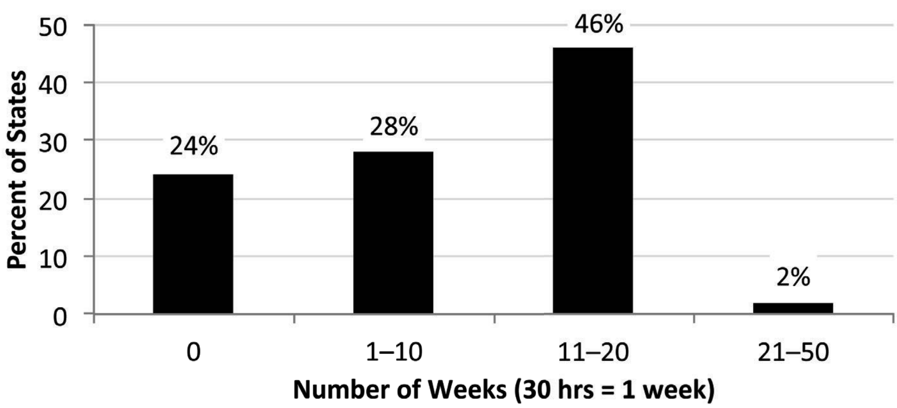


Figure 17. State field experience requirements (student teaching and clinical experience). Data are from American Association of Colleges for Teacher Education (2010, p. 10).

Microteaching

Microteaching is a technique used in field experience training in which the student teacher along with a supervising faculty or mentor teacher reviews video of lessons the student taught. The coach or instructor provides the student teacher with specific feedback on the implementation of the lessons taught, what worked, and corrective feedback on how to improve performance. This method, used in laboratory settings or in real classrooms, can be an effective technique to enhance field experiences. Microteaching is helpful both in improving the teacher's performance and increasing student achievement. Hattie (2009) found an effect size of 0.88 for microteaching (Figure 18).

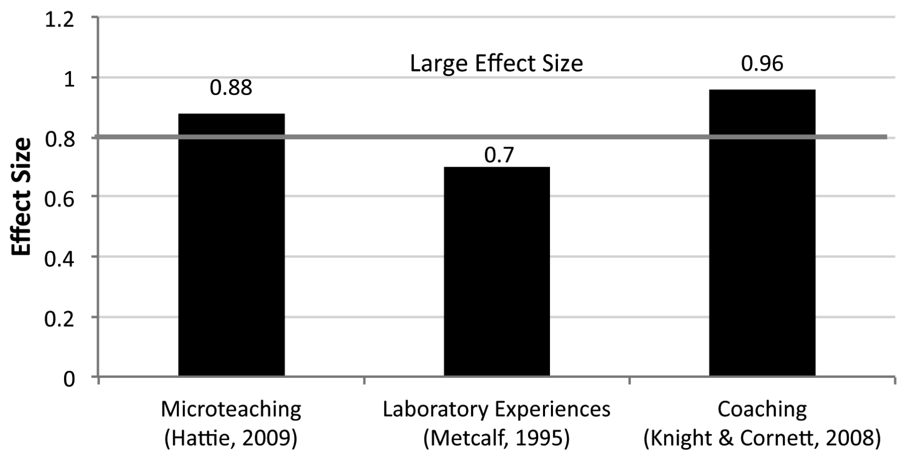


Figure 18. Impact of teacher training methods. Data are drawn from Hattie (2009, Appendix B), Metcalf (1995, p. 12), and Knight and Cornett (2008, p. 13).

Overall, the evidence in support of the current approach to field experience in teacher preparation is inadequate. There are not enough studies, and few of those were experimental or used rigorous methodologies. A summary of studies by Floden and Meniketti (2005) found them to be overwhelmingly qualitative, with the vast majority focused on the teacher's attitude change or perception of the field experience rather than on critical outcomes such as the effect of field experience on student achievement (Figure 19).

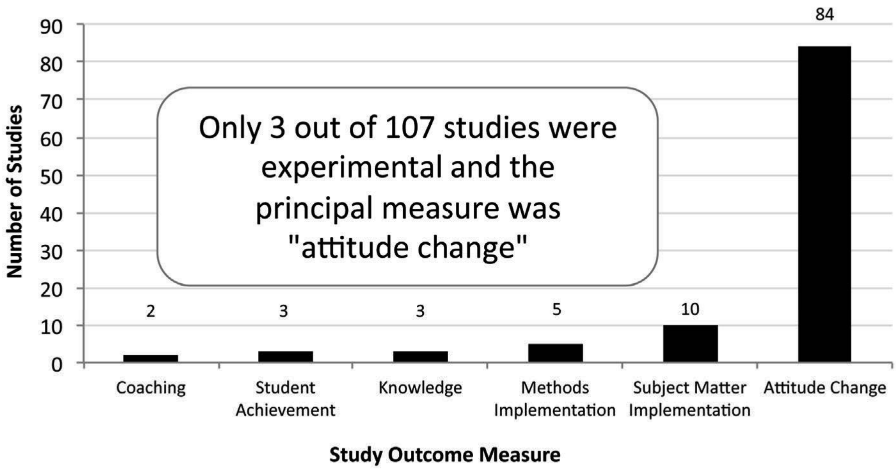


Figure 19. Survey of field experience research topics. Data are drawn from Floden and Meniketti (2005, p. 288).

WHAT WE KNOW ABOUT TEACHER PREPARATION MODELS

Substantial time and resources have been allocated to producing state and national standards for teacher preparation in order to create a model of teacher preparation that can reliably produce teachers who make a difference. What does research tell us about the evidence supporting these different approaches to training teachers?

Teacher preparation models generally fall into three categories: (a) 4-year bachelor's degree credential, (b) 5-year post-bachelor's degree credential, and (c) alternative credential.

Four-Year Bachelor's Degree Credential

This undergraduate model requires the student to spend 4 years obtaining a bachelor's degree built around a prescribed course of education study. The requirements for a 4-year credential model vary by state. This credential was founded on the "normal school" model with a focus on teaching subject matter and methodology of education.

Five-Year Degree Credential

A relatively modern concept that gained momentum in the 1960s, the 5-year credential model requires teacher candidates to obtain a bachelor's degree before beginning a course of education study. The driving force behind the adoption of the model was a belief prevalent in the education community that teachers were not respected (Zeichner & Conklin, 2005). It was assumed that making teacher certification a post-bachelor and/or graduate degree model would confer greater esteem on the profession of teaching. The trend was widely embraced by teacher preparation programs across the nation, as well as being adopted by a number of states including California.

By the mid-1980s, organizations such as the American Association of Colleges for Teacher Education and the Carnegie Forum on Education and the Economy were actively advocating the 5-year program as a solution to unsatisfactory student achievement. It has been estimated that upward of 25% of American teachers receive credentials from post-baccalaureate programs (Zeichner & Conklin, 2005). Evidence supporting the efficacy remains weak. The most extensive research comparing the effectiveness of 4-year and 5-year credential teacher programs was conducted by Andrew (1990) and Andrew and Schwab (1995). These two studies reached similar findings. Unfortunately, this research did not directly examine student achievement or teacher performance but instead relied heavily on surveys (Figure 20).

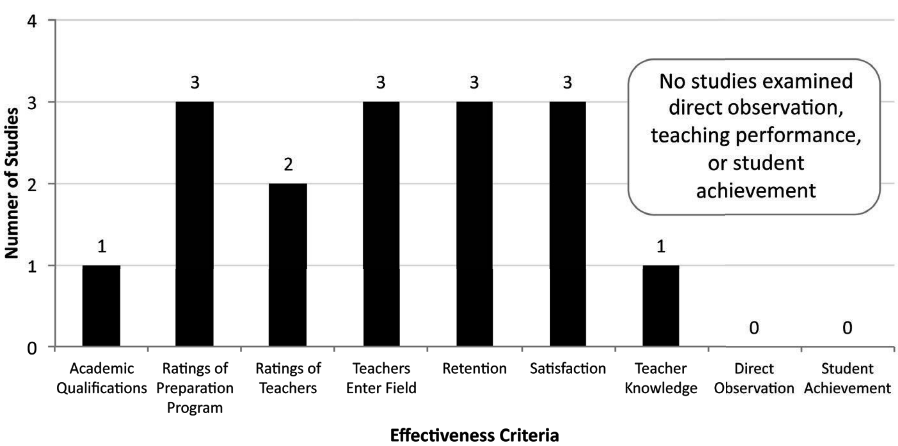


Figure 20. Outcome measures used in studies evaluating 4-year and 5-year teacher preparation programs. Data are drawn from Zeichner and Conklin (2005, p. 705).

The take-home message is that without examining student achievement, current research on the effectiveness of 5-year programs cannot answer the important question of whether the benefits of the additional year outweigh the costs.

Comparing Traditional Credentials (TC) And Alternative Credentials (AC)

To fill a critical teacher shortage in the 1980s, schools began to hire teachers enrolled in alternative credential programs. The alternative credential proved very popular, and the number of teachers with this type of credential increased substantially during the 1980s (Constantine et al., 2009). The principal distinction between the TC and AC models is that TC teachers complete the credential program before being hired to teach students, whereas AC teachers are enrolled in programs that provide formal teacher preparation coursework while those teachers are already employed in the classroom. A prime example of this approach is Teach For America. The program places over 8,000 recent college graduates or professionals in classrooms in low-income communities for 2 or more years. The goal is to provide underperforming schools with teachers who are motivated to make a difference and willing to be trained while on the job.

The issue has been a lightning rod for those concerned with the stagnant performance of schools as measured by NAEP scores and high dropout rates. Those resistant to the AC route are generally opposed on the basis that putting untrained personnel in classrooms will result in lower student performance (Constantine et al., 2009). These concerns prompted changes in regulations across the country, culminating in the NCLB mandate requiring teachers to hold full state certification. Unfortunately, when the regulation was ordered, an important question was left unanswered: Does full credentialing actually increase student achievement?

Constantine et al. (2009) shed light on the issue with the results of their 2-year randomized controlled study funded by the Institute of Education Sciences (IES). The study concluded that there was no statistically significant difference in performance between students of TC and AC teachers (Figure 21). Variation in student achievement was not strongly linked to the teacher's chosen preparation route or to other measured teacher characteristics. The study found no meaningful difference in the performance of teachers when it came to student achievement in mathematics and reading. Neither route to certification was found to be superior.

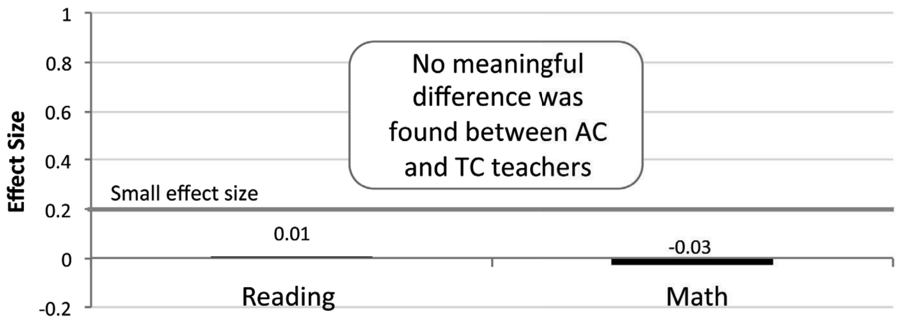


Figure 21. Difference in achievement of alternative credentialed (AC) teachers compared with traditional credentialed (TC) teachers. Data are drawn from Constantine et al. (2009, Appendix A: exhibit A.7).

WHAT WE KNOW ABOUT TEACHER PREPARATION STANDARDS

National Teacher Certification

National Board for Professional Teaching Standards was established in 1987 to foster "high and rigorous standards for what accomplished teachers should know and be able to do" (NBPTS mission statement). As a voluntary national system, NBPTS certifies that a teacher has taught for at least 3 years, has submitted a teaching portfolio that includes video recordings of classroom teaching, and has successfully responded to essay questions assessing pedagogical knowledge. The process requires teachers to pay a substantial fee and can take from 3 months to several years to complete. With the advent of NCLB and greater accountability, school districts have come to view the process as a way to improve student achievement, allocating scarce resources in the form of performance compensation to encourage teachers who acquire certification.

Cantrell, Fullerton, Kane, and Staiger (2008) examined whether the certification by NBPTS correlated with teacher impact on student achievement. The study reviewed the available literature on the topic, including the performance of NBPTS-certified teachers and the role certification played in improving student achievement. The analysis provided a summary of effect sizes from six studies between 2004 and 2006. There were no statistically significant differences between the math and reading test scores of students assigned to NBPTS-certified teachers and those of students assigned to teachers who did not apply for NBPTS certification. It also provided results from recent research that looked for a correlation between NBPTS certification and teachers with the largest estimated impact on student achievement. The research generally found very small effect size differences of 0.05 to 0.1 between the impact on student achievement of certified teachers and applicants who failed to obtain certification. Cantrell et al. found no studies with an effect size above 0.1 (Figure 22). This research offers little to recommend NBPTS certification as an effective strategy for improving teacher effects on student performance.

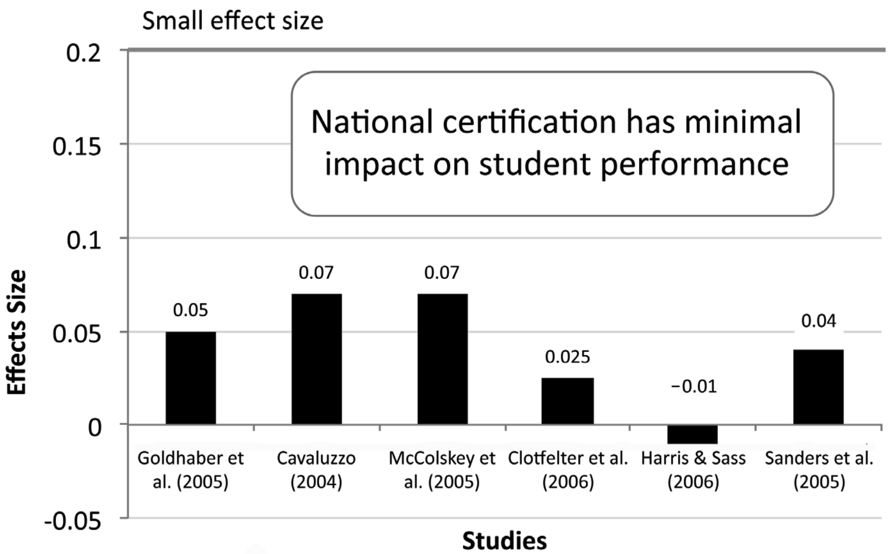


Figure 22. Comparison of NBPTS-certified teachers with non-certified teachers on student math achievement Data are drawn from Cantrell, Fullerton, Kane and Staiger (2008, Table 3).

Program Accreditation

Program accreditation is a common quality control practice used in higher education as a means of holding colleges and universities accountable to standards of excellence. The National Council for the Accreditation of Teacher Education (NCATE), founded in 1954, and the Teacher Education Accreditation Council (TEAC), founded in 1997, have a mission to improve teacher education through accrediting preparation programs. NCATE accredits fewer than half of the programs in the nation, just 650 of the over 1,500 programs. TEAC has a little over 200 accredited members.

Both work to improve quality by requiring preparation programs to meet best practices standards through compliance procedures and periodic site visits. Neither organization includes in its mission statement that the goal of accreditation is to improve schoolchildren's performance. Despite the best attempts of both bodies to improve the quality of teachers entering the workforce, there is little research to support that the programs are having a significant impact on the quality of teachers. The research found on the organizations' websites offers little evidence that teachers graduating from accredited programs are any more effective than teachers coming from unaccredited institutions (ncate.org; teac.org).

A primary study by Gitomer, Latham, and Ziomek (1999) on the NCATE website promoting accreditation effectiveness showed that graduates of NCATE-accredited colleges of education passed Education Testing Service (ETS) content examinations for teacher licensing at a higher rate than did graduates of unaccredited colleges. The results of this study are described in Figure 23. There are two issues of concern regarding the study. First, it does not offer evidence that passing the Praxis II, a teacher certification exam, makes for better teachers in the classroom as measured by student academic outcomes. Second, we do not know if the 8% difference in the scores between NCATE-trained teachers and non-NCATE teachers is statistically or socially significant.

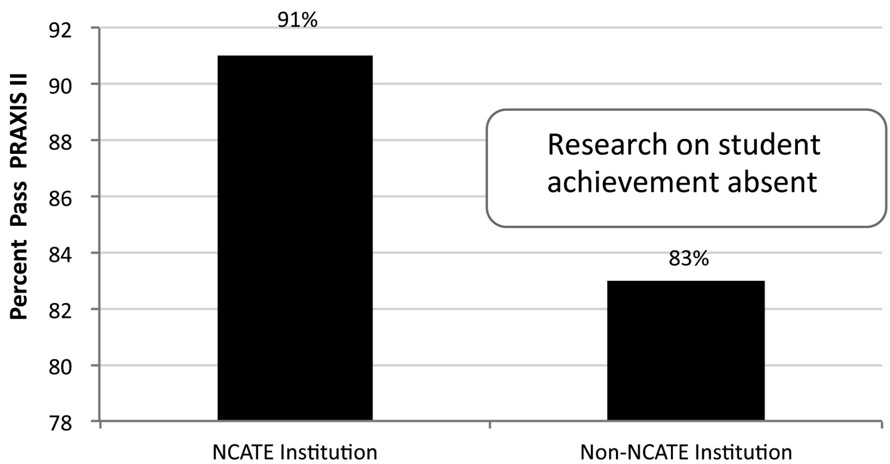


Figure 23. Comparison of NCATE-trained teachers and non-NCATE teachers passing the Praxis II. Data are drawn from Gitomer, Latham, and Ziomek. (1999, p. 25).

Unfortunately, too few studies have been done on the subject, and the research that has been conducted offers insufficient evidence to know whether being accredited by either NCATE or TEAC will result in preparation programs producing teachers who can make a difference in the lives of students.

WHAT WE KNOW ABOUT PREPARATION PROGRAM EFFECTIVENESS RESEARCH

We previously examined the importance of assessment of students as an essential strategy to improve student performance. Assessing graduates of teacher preparation programs and their impact on student achievement is another important strategy for improving the performance of the education system. To date, few studies have been conducted linking preparation programs and the quality of preparation program graduates. Until recently, few incentives or contingencies were placed on preparation program administrators by regulators, funding sources, or consumers to conduct this type of research.

School administrators and those involved in the hiring of teachers would benefit greatly by knowing which preparation programs produce the best teachers. It would help all involved to know which preparation programs incorporate evidence-based practices in their required course work. Correlating teachers' course of study to outcome performance data would be crucial in expanding our knowledge base and assisting other preparation programs to improve performance. This would prove invaluable in determining which course of study produced teachers whose students had the best outcomes and to use that program's curriculum as a template for other preparation programs. Information about what works and what practices to avoid is sorely needed as pressure has increased for greater accountability for preparation programs.

The studies evaluating preparation programs over the past 30 years have typically been qualitative and provide little information that can inform stakeholders which preparation programs produce the best results. In the past, we did not have the tools to conduct the research we require. Now, tools such as value-added modeling are being more widely used to answer these questions. Researchers are attempting to answer questions such as which preparation programs are the best at producing teachers who raise student achievement scores, as Noell and Burns (2006) did in their study of preparation programs in Louisiana. Their analysis suggest that it may be possible to use achievement and educational personnel databases to assess the effectiveness of teacher preparation programs.

WHAT WE KNOW ABOUT TEACHER INDUCTION

Before induction was introduced in the 1980s, after teachers completed preservice training they would be dropped into an education system that emphasized independence in deciding how to run their own classrooms. The traditional preparation model offered new teachers approximately 8 weeks of field experience in a real classroom to get them ready for this day (American

Association of State Colleges and Universities [AASCU], 2010). New teachers were provided a few hours of orientation and then given control of the classroom.

By the 1980s, this picture began to change. Stagnant student graduation rates and declining test scores elicited concerns from educators and the public (Maheady & Jabot, in press, this volume). In the education literature, terms such as "preservice training," "in-service teacher training," "induction," and "mentoring" increasingly appeared as a way to improve teacher performance. By 2000, over 80% of public school teachers received some form of formal induction training (Smith & Ingersoll, 2004). Induction consists of practices that help new and beginning teachers become competent and effective classroom professionals who also understand school and community cultures (AACTE, 2010).

Maheady and Jabot (in press, this volume) provide a thorough look at teacher induction and the remarkable growth of the practice. They make a compelling case for induction services. They also analyze the available research on current models and offer solutions to remediate problems evident in today's comprehensive teacher induction services.

To begin a discussion of induction, it is reasonable to ask one key question: Are new teachers less effective than veteran teachers? Research on the preparedness of new teachers, not surprisingly, supports the common wisdom of the existence of a "rookie" phenomenon (Kane, Rockoff, & Staiger, 2006;) (Figure 24). The rookie is an inexperienced teacher who requires on-the-job training before mastering the skills needed to be effective. Even though the data in Figure 24 capture the effect of additional years of experience, it is important to note that the difference between a beginning teacher and a second-year teacher is only .06 of a standard deviation, which is not a very large effect.

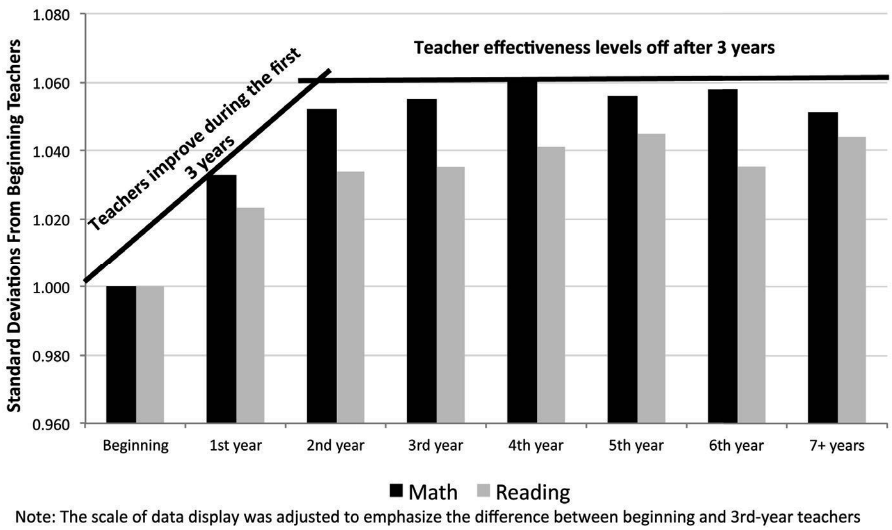


Figure 24. Value-added returns with years of experience. Data are drawn from Kane, Rockoff, and Staiger (2006, Table 10).

Research offers insight into how long it takes to get new teachers up to speed. An analysis of math and reading scores, when correlated with teacher experience, provides credence to the notion that time in the classroom makes for better teachers. The data support the following facts: Students of first-year teachers, on the whole, produce lower test scores; improvement in teacher performance happens over the first 3 years; and additional experience does not result in continuing improvement after the third year. It is also important to note that the impact on student achievement is small, as measured by an effect size of 0.2 or less.

When first proposed, teacher induction was offered as an answer to stagnant student achievement scores and as a way to stem the loss of teachers through turnover. Chronic shortages of qualified teachers in the 1980s made the issue of turnover even more important, as it contributed to the number of new and undertrained teachers entering the workforce. Astonishingly, by the fifth year of teaching, almost 50% of new teachers leave the profession (Ingersoll, 2003) (Figure 25). If the trend is to be reversed, understanding why so many teachers leave in the first few years might lead to a solution. If induction could make teachers feel better prepared and increase job satisfaction, the practice might possibly reduce turnover (Hart Research Associates, 2010).

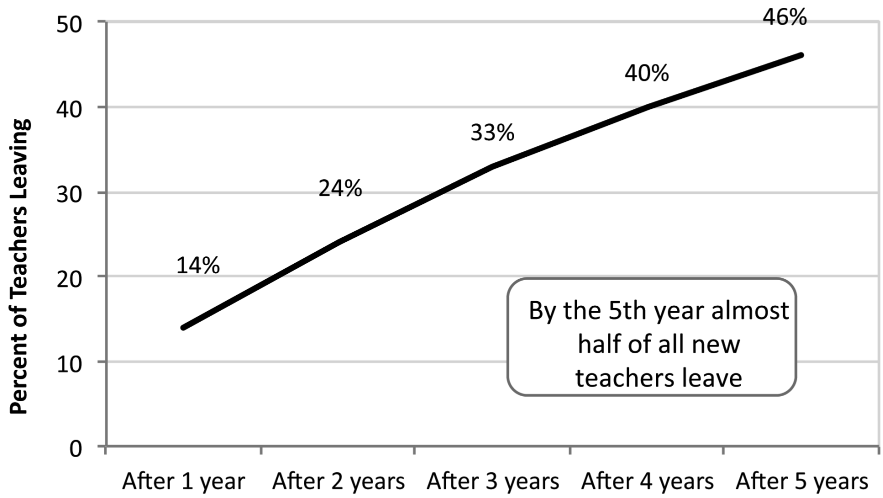


Figure 25. Teachers leaving the profession. Adapted from "Is there really a teacher shortage?," by R.M. Ingersoll, 2003, *A research report co-sponsored by the Center for the Study of Teaching and Policy and the Consortium for Policy Research in Education,* p. 14. Copyright 2003 by the Center for the Study of Teaching and Policy. Adapted with permission.

The first large-scale induction program in the United States was established in Florida in 1980. Induction offered a way for schools to systematically deliver instruction to orient new personnel, train staff to the standards unique to a specific school, and use veteran teachers to mentor and coach beginning teachers. Proponents of the practice suggested other valuable benefits from induction, notably, improved morale and enhanced communication between administrators and teachers, problems that affected most school systems (Ingersoll & Kralik, 2004).

Representing a possible solution to multiple problems, induction had great appeal. The attraction was so compelling that over the next 10 years most states followed the example of Florida and established induction models. Induction in one form or another rapidly grew from a relatively unknown niche service to involving over 80% of public school teachers by 2000, and it is expected to soon engage 90% of all beginning teachers (Smith & Ingersoll, 2004; Maheady & Jabot, in press, this volume). Figure 26 describes the growth of teacher participation in induction programs between 1990 and 2000.

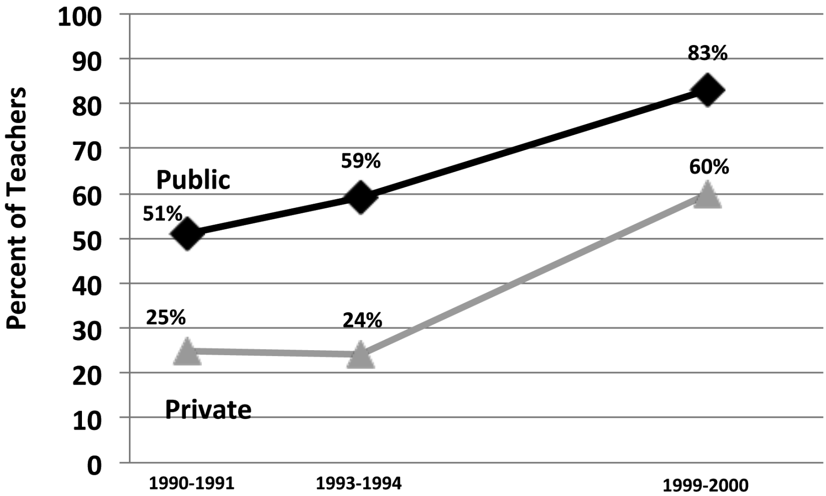


Figure 26: Percent of teachers participating in induction programs in public and private schools. Adapted from "What Are the Effects of Induction and Mentoring on Beginning Teacher Turnover?", by T. M. Smith and R. Ingersoll, 2004, *American Educational Research Journal,, 41,* p. 691. Copyright 2004 by the American Educational Research Association. Adapted with permission.

Despite its promise, induction poses many challenges. It is one of the more complex interventions attempted over the past 30 years. As a systemic approach to school reform, induction necessitates added investments in resources, time, and money to ensure the intervention sustains over time. Successful induction involves significant changes in the school practices, including hiring, preparation of a curriculum specific to the school, an orientation procedure, protected initial assignments, mentor and other support, frequent coaching, and ongoing evaluation (Cherian & Daniel, 2008; New Teacher Center, 2006).

The question should be asked, is the practice a smart use of scarce resources? It has been estimated that the annual cost of induction in California and Connecticut is, on average, $4,000 per trained teacher (Alliance for Excellent Education, 2004). A cost effectiveness ratio analysis (Yeh, 2007) suggests that induction, as implemented in a recently released randomized controlled study (Glazerman et al., 2008), is not a cost-effective intervention (Figure 27). When induction, structural interventions (e.g. class size reduction, charter schools, increased spending, and high-stakes testing), and an instruction intervention (i.e., formative assessment) were compared, induction along with the structural interventions compared poorly with formative assessment.

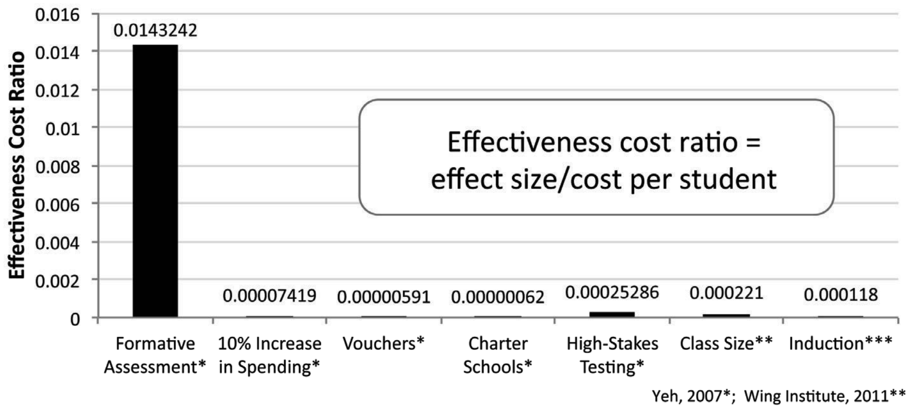


Figure 27. Cost benefit of educational interventions. Data are drawn from Yeh (2007, p. 431). Data from columns with ** were calculated by the Wing Institute based on Yeh's formula for effectiveness cost ratio.

Unfortunately, this massive paradigm shift, along with a commitment of substantial time and resources, happened despite the lack of rigorous research to guide school systems in what works and what to avoid in induction. Given the considerable costs, it would seem wise for education stakeholders to design induction programs based on the best available research.

What Research Reveals About Induction

Most of the available research on the impact of induction either lacks rigor or is inconclusive (Maheady & Jabot, in press, this volume). The exception is the Institute of Education Sciences (IES) report summarizing the results of a randomized controlled study on induction (Glazerman et al., 2008). The study examined the effects of induction programs on student achievement for second through sixth graders in mathematics and reading. The study also looked at the effects of induction on teacher practices and teacher retention. The comprehensive induction model studied included (a) yearlong intensive and structured support for beginning teachers, (b) weekly meetings for new teachers with trained mentors, (c) ongoing classroom observations and constructive feedback, and (d) monthly professional development sessions. The outcomes of this large-scale randomized clinical trial suggest that comprehensive induction has very little, if any, statistically positive impact on beginning teacher effectiveness, satisfaction, or teacher turnover. Unfortunately, issues with the implementation of the study such as a failure to measure treatment integrity, limited time allocated to classroom observation, and the control and experimental groups' similar exposure to mentoring provide sufficient reasons to question the results of the study.

A major weakness in the Glazerman study is the lack of a control group that received no form of induction training. The preponderance of induction training in American schools resulted in both the control group and the experimental group receiving a form of induction. It is always a more stringent test of an intervention to compare it with an alternative intervention rather than treatment as usual. If we are to really identify what works, then the proper comparison will need to be with alternative forms of induction, but only after determining the efficacy of the practice against no treatment. As yet, the data do not support the assertion that induction is more effective than providing no induction services.

Another problem with the study is the imprecise definition for mentoring and the drift in the mentoring experience resulting from ineffective safeguards for treatment integrity. The actual mentoring of the induction group varied significantly among teachers within the group. The number of minutes on average a group member received was ultimately not significantly greater than what was provided the control group. Given the quality control issues with the induction services provided the control group and the experimental group along with fundamentally little difference in the induction experience, it is not surprising that the effect sizes for reading and math for both groups differed little and were found to be small (Figure 28).

Manuals can be valuable tools for defining practices and setting expectations for alternate forms of induction (Kauffman, in press, this volume). Only when a model is implemented consistently and with integrity can the real effectiveness of the intervention be assessed. Otherwise, it is impossible to know if the intervention is effective or not and whether poor implementation accounts for a poor outcome. Even efficacious practices are likely to produce poor results when treatment integrity is neglected.

Student Outcomes

Glazerman et al., (2010) found no significant effects of induction on student reading or math achievement scores. The overall average effect size across all grades after 3 years was 0.11 for reading (Figure 28). Students who were taught by teachers receiving comprehensive induction demonstrated no improvement in either the first or second years of reading or math scores. It was not until the third year that gains occurred, but at no time did these effect sizes climb above 0.2, a small effect.

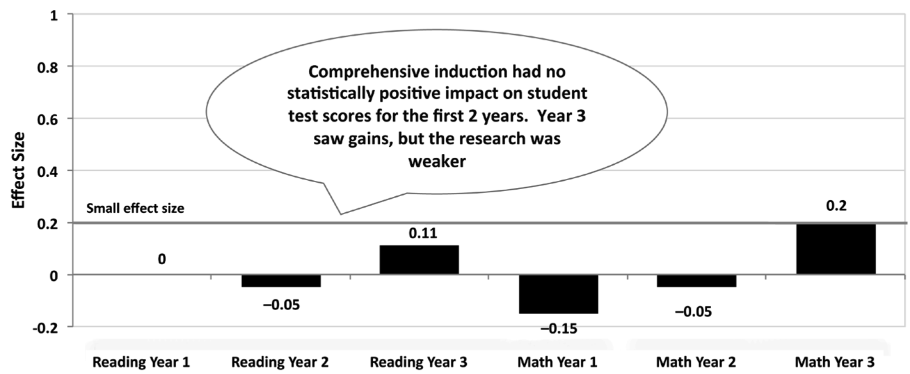


Figure 28. Impact of comprehensive induction on student reading and math test scores. Data are drawn from Isenberg et al. (2009, p.99) and Glazerman et al. (2010, p. 93).

Teacher Outcomes

Glazerman et al. (2010) found no statistically significant effect on teacher retention rates after the first year. On average, only 70% of the induction and control group teachers returned to teach a second year. There was no effect on the proportion of teachers still employed in the profession after 1 year, and no practical difference in the amount of time the control and the induction groups remained (Figure 29).

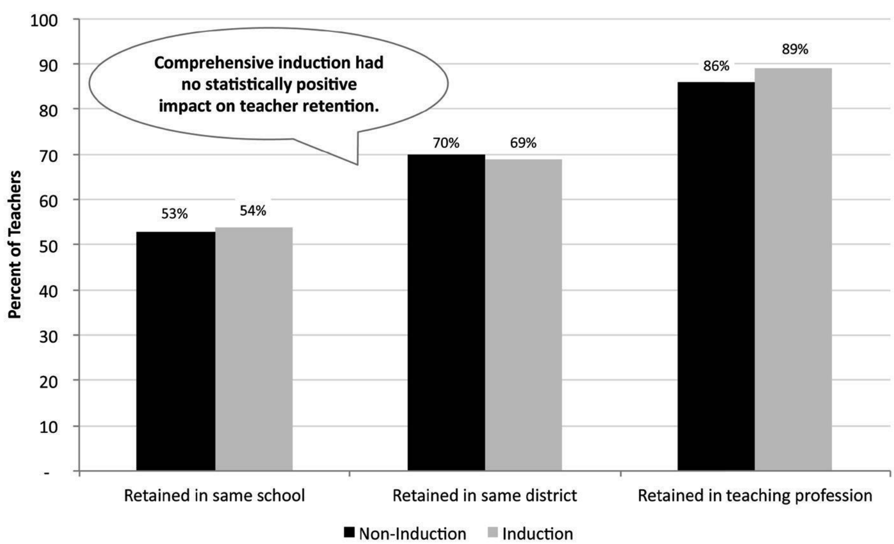


Figure 29. Impact of comprehensive induction on teacher retention. Data are drawn from Glazerman et al. (2010, p. D-11, Table D.9).

The results of the Glazerman study are disappointing when considering the substantial costs in time and money invested nationally in induction services. There are few quantitative studies on induction, and the results of this randomized controlled trial had been eagerly awaited. The conclusions from the study offer little to suggest that induction is having a significant impact on the two key outcome measures: student achievement and teacher turnover. At first glance, the results suggest induction as just another education fad that didn't work, to be pushed aside and then forgotten.

However, there may be reasons for the study's failure to find evidence to support induction as an effective intervention. Closer examination of the study highlights the difficulty encountered in implementing a complex intervention in real-world settings. Specific problems with implementation included trouble in objectively defining the intervention and the control group practices, overreliance on surveys to obtain data, infrequent data collection, and poorly designed measurement tools.

Induction Summary

The variability in results from induction at this time is not surprising. Our limited knowledge of what works and what does not impedes our ability to state what components of induction are effective. It may be too soon in the study of induction to expect a standard model.

The models in use vary significantly in purpose, duration, intensity, activities, assessment, content, and use of mentoring (Maheady & Jabot, in press, this volume). These impediments make it very difficult to draw conclusions about comprehensive induction and its impact.

The Maheady and Jabot chapter (in press, this volume) examines induction thoroughly and insightfully. It provides a history of induction and a summary of the available research on the topic, and discusses the issues impacting implementation of induction programs. Finally, it presents guidelines for overcoming obstacles, emphasizing the importance of building critical partnerships between teacher preparation programs and receiving K–12 schools. Induction has the potential to be a significant force in improving student achievement, strengthening teacher morale, and reducing unwanted teacher defections. However, additional research must be conducted to better understand how to overcome serious problems in the design of current models and practices.

SUMMARY: EFFECTIVE TEACHERS MAKE A DIFFERENCE

The recently released Programme for International Student Assessment (PISA) international 2009 test scores found that the United States ranked 14th in reading, 17th in science, and 23rd in mathematics. The continued disappointing performance, which shows scores changing little over 30 years, has the potential to place the United States at a serious competitive disadvantage in the coming decades. To reverse the trend, American educators must look to new solutions. We must look beyond the structural interventions of the past such as class size reduction, charter schools, or smaller schools that have failed to produce results and begin to read the available evidence on what works.

Over the past 30 years, the process of building a science of education for teachers has been underway. We have accumulated a significant knowledge base of strategies and practices that do make a difference. We also have over 100 years of data on various teacher preparation models (Cochran-Smith & Zeichner, 2005). We are gathering valuable information on what works as well as what to avoid, to help schools avoid repeating past failures. When asked what essential skills teachers need to be successful, we can provide a list of skills, based on increasingly rigorous research, that improve student achievement.

We know with increasing confidence that formative assessment, effective classroom management, and effective teaching strategies improve student achievement. We can answer questions regarding the role subject matter plays in making a good teacher. Unfortunately, subject matter exposure makes very little difference. We continue to acquire knowledge of the most effective pedagogical strategies for teaching teachers. We need to balance past overreliance on didactic instruction with effectively coached in-class training and field experiences for new teachers. We also need to assess the performance of teacher preparation programs to identify which schools can act as models for other programs. Finally, we need to provide teachers with support that offers feedback on a timely and regular basis. Ultimately, it all comes down to effective implementation. The best teachers can lose their edge if not provided effective supervision and feedback.

All too often stakeholders punish teachers for not meeting expectations. Until teachers are trained in the correct skills and supported in using those skills, it is difficult to hold them accountable for underperforming. Teaching is a very challenging job requiring the mastery of many skills. It is important to remember teachers are only one important piece of the puzzle. The remaining pieces include effective management practices, parental support, the selection of and implementation of evidence-based practices, and sound leadership.

REFERENCES

Ahn, S., & Choi, J. (2004, April). *Teachers' subject matter knowledge as a teacher qualification: A synthesis of the quantitative literature on students' mathematics achievement.* Presented at the annual meeting of the American Educational Research Association, San Diego.

Allen, M. (2000). *The progress of education reform: Teacher preparation and induction*. Denver, CO: Education Commission of the States. Retrieved from http://www.ecs.org/clearinghouse/22/38/2238.pdf

Alliance for Excellent Education. (2004). *Tapping the potential: Retaining and developing high-quality new teachers*. Washington, DC: Author.

American Association of Colleges for Teacher Education (AACTE). (2010). *The clinical preparation of teachers: A policy brief.* Washington, DC: Author.

American Association of State Colleges and Universities (AASCU). (2006). Teacher induction programs: Trends and opportunities. *Policy Matters*, *3*(10), 1–4.

American Education Research Association (AERA). (2004). Teachers matter: Evidence from value-added assessments. *Research Points*, *2*(2). Retrieved from http://www.aera.net/Portals/38/docs/Publications/Teachers%20Matter.pdf

Andrew, M. D. (1990). Differences between graduates of 4-year and 5-year teacher preparation programs. *Journal of Teacher Education, 41,* 45–51

Andrew, M., & Schwab, R. L. (1995). Has reform in teacher education influenced teacher performance? An outcome assessment of graduates of eleven teacher education programs. *Action in Teacher Education, 17*(3), 43–53.

Begeny, J., & Martens, M. (2006). Assessing pre-service teachers' training in empirically-validated behavioral instruction practices. *School Psychology Quarterly, 21*(3), 262–285.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139–148.

Bloom, B. S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Brophy, Jere, E. (2004). *Motivating students to learn*. Mahwah, NJ: Lawrence Erlbaum Associates.

Cantrell, S., Fullerton, J., Kane, T. J., & Staiger, D. (2008). *National board certification and teacher effectiveness: Evidence from a random assignment experiment* (NBER Working Paper No. 14608). Retrieved from the National Bureau of Economic Research website: http://www.nber.org/papers/w14608

Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the 21st century.* Report of the Task Force on Teaching As a Profession. Washington, DC: Author.

Chard, D., & Kameenui, E. (2000). Struggling first-grade readers: The frequency and progress of their reading. *Journal of Special Education, 34*(1), 28–38.

Cherian, F., & Daniel, Y. (2008). Principal leadership in new teacher induction: Becoming agents of change. *International Journal of Education Policy & Leadership, 3*(2).

Clift, T., & Brady, P. (2005). Research on methods and field experiences. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 309–336). Mahwah, NJ: Lawrence Erlbaum Associates.

Cochran-Smith, M., & Zeichner, K. M. (Eds). (2005). *Studying teacher education: The report of the AERA panel on research and teacher education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Constantine, J., Player D., Silva, T., Hallgren, K., Grider, M., and Deke, J. (2009). *An Evaluation of Teachers Trained Through Different Routes to Certification, Final Report* (NCEE 2009-4043). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Floden, R., & Meniketti, M. (2005). Research on the effects of coursework in the arts and sciences and the foundation of education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 261–308). Mahwah, NJ: Lawrence Erlbaum Associates.

Friesen, N. (2011). The lecture as a transmedial pedagogical form: A historical analysis. *Educational Researcher, 40*(3), 95–102.

Fuchs, L. S., & Fuchs, D., (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, *53*(3), 199–208.

Gardner, D. P., Larsen, Y. W., Baker, W. O., Campbell, A., Crosby, E. A., Foster, C. A., Jr., ...Wallace, R. (1983). *A nation at risk: The imperative for educational reform. An open letter to the American people. A report to the nation and the secretary of education*. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED226006

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149–164.

Gitomer, D. H., Latham, A. S., & Ziomek, R. (1999). *The academic quality of prospective teachers: The impact of admissions and licensure testing*. Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RR-03-35.pdf

Glazerman, S., Dolfin, S., Bleeker, M., Johnson, A., Isenberg, E., Lugo-Gil, J.,…Britton, E. (2008). *Impacts of comprehensive teacher induction: Results from the first year of a randomized controlled study* (NCEE 2009-4034). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Glazerman, S., Isenberg, E., Dolfin, S., Bleeker, M., Johnson, A., Grider, M., & Jacobus, M. (2010). *Impacts of comprehensive teacher induction: Final results from a randomized controlled study* (NCEE 2010-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Haller, E. P., Child, D. A., & Walberg, H. J. (1998). Can comprehension be taught? A quantitative synthesis of "metacognitive studies." *Educational Researcher, 17*(9), 5–8.

Hart Research Associates (2010). *Career changers in the classroom: A national portrait.* Woodrow Wilson National Fellowship Foundation. Retrieved from http://www.woodrow.org/images/pdf/policy/CareerChangersClassroom_0210.pdf

Hattie, J., (2009). *Visible learning: A synthesis of over 800 meta-analyses related to achievement*. New York: Routledge.

Heward, W. (2004, January). *Want to improve the effectiveness of your lectures? Try guided notes.* Retrieved from The Ohio State University Academy of Teaching website http://hdl.handle.net/1811/34578

Heward, W. (2008). *Exceptional children: An introduction to special education* (9th ed.). Upper Saddle River, NJ: Prentice Hall.

Ingersoll, R. M. (2003). *Is there really a teacher shortage?* A research report co-sponsored by the Center for the Study of Teaching and Policy and the Consortium for Policy Research in Education. Seattle, WA: Center for the Study of Teaching and Policy.

Ingersoll, R. M., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says* (pp. 1–23). Denver, CO: Education Commission of the States.

Isenberg, E., Glazerman, S., Bleeker, M., Johnson, A., Lugo-Gil, J., Grider, M.,…Britton, E. (2009). *Impacts of comprehensive teacher induction: Results from the second year of a randomized controlled study* (NCEE 2009-4072). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Johnson, E. G., & Zwick, R. (1990). Focusing the new design: The NAEP 1988 technical report. *Journal Of Educational and Behavioral Statistics, 17*(2), 95–109.

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Juel, C., & Leavell, J. (1988). Retention and nonretention of at-risk readers in first grade and their subsequent reading achievement. *Journal of Learning Disabilities*, *21*(9), 571–580.

Kane, T. J., Rockoff, J. E., & Staiger, D. O. (2006). *What does certification tell us about teacher effectiveness? Evidence from New York City.* (Working Paper #12155.) Cambridge, MA: National Bureau of Economic Research.

Kauffman, J. M. (2010). *The tragicomedy of public education: Laughing, crying, thinking, fixing*. Verona, WI: Attainment.

Kauffman, J. M. (in press). Science and the education of teachers. In R. Detrich, R. Keyworth, & J. States (Eds.), *Advances in evidence-based education. Volume 2: Education at the crossroads: The state of teacher preparation.* Oakland, CA: The Wing Institute.

Kavale, K. A. (2005). Effective intervention for students with specific learning disability: The nature of special education. *Learning Disabilities: A Multidisciplinary Journal, 13*(4), 127–138.

Knight, J., & Cornett, J. (2009). *Studying the impact of instructional coaching.* Retrieved from the Instructional Coaching Kansas Coaching Project website http://instructionalcoach.org/research/tools/paper-studying-the-impact-of-instructional-coaching

Kumar D. D. (1991). A meta-analysis of the relationship between science instruction and student engagement. *Review of Educational Research, 66*(2), 99–136.

Maheady, L., & Jabot, M. (in press). Comprehensive teacher induction: What we know, don't know, and must learn soon! In R. Detrich, R. Keyworth, & J. States (Eds.), *Advances in evidence-based education. Volume 2: Education at the crossroads: The state of teacher preparation.* Oakland, CA: The Wing Institute.

Marzano, R. J., (1998). *A theory-based meta-analysis of research on instruction.* Aurora, CO: Mid-Continental Regional Educational Laboratory.

Marzano, R. J., Marzano, J. S., & Pickering, D. J. (2003). *Classroom management that works: Research-based strategies for every teacher.* Alexandria, VA: Association for Supervision and Curriculum. Development.

Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works: Research-based strategies for increasing student achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mendro, R., Jordan, H., Gomez, E., Anderson, M., & Bembry, K. (1998). *An application of multiple linear regression in determining longitudinal teacher effectiveness*. Dallas, TX: Dallas Public Schools.

Metcalf, K. K. (1995, April). *Laboratory experiences in teacher education: A meta-analytic review of research.* Presented at the annual meeting of the American Educational Research Association, San Francisco. Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED388645

National Assessment of Educational Progress (NAEP). (2009). *The nation's report card.* Retrieved from http://nces.ed.gov/nationsreportcard/

National Council for Accreditation of Teacher Education (NCATE). (2010). *Transforming teacher education through clinical practice: A national strategy to prepare effective teachers.* Retrieved from http://www.ncate.org/LinkClick.aspx?fileticket=zzeiB1OoqPk%3D&tabid=715

National Education Association (NEA). (2003). *Status of the American public school teacher, 2000–01.* Retrieved from http://www.nea.org/assets/docs/HE/status.pdf

National Institute of Child Health and Human Development (NICHD). (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Report of the National Reading Panel. U.S. Department of Health and Human Services, National Institutes of Health. Retrieved January 20, 2011, from http://www.nichd.nih.gov/publications/nrp/smallbook.cfm

New Teacher Center (2006). *Understanding New York City's groundbreaking induction initiative: Policy implications for local, state, and national educators.* Santa Cruz, CA: New Teacher Center. Retrieved from http://newteachercenter.org/pdfs/NYCPolicyPaper.pdf

Noell, G. H., & Burns, J. L. (2006). Value-added assessment of teacher preparation: An illustration of emerging technology. *Journal of Teacher Education, 57*(1), 37–50.

Nye, N., Konstantopoulos, S., & Hedges, L. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*(3), 237–257. Retrieved from http://steinhardt.nyu.edu/scmsAdmin/uploads/002/834/127%20-%20Nye%20B%20%20Hedges%20L%20%20V%20%20%20Konstantopoulos%20S%20%20(2004).pdf

Rowan, B., Correnti, R., & Miller, R. (2002). *What large-scale research tells us about teacher effects on student achievement: Insights from the prospects study of elementary schools.* Consortium for Policy Research in Education (Research Report Series RR-051). Retrieved from http://cw.marianuniversity.edu/mreardon/755/document%20repository/Teacher%20Effects%20on%20Student%20Achievement.pdf

Sanders, W. L., & Rivers, J. C. (1996). *Cumulative and residual effects of teachers on future student academic achievement.* University of Tennessee Value-Added Research and Assessment Center. Retrieved from http://heartland.org/policy-documents/cumulative-and-residual-effects-teachers-future-student-academic-achievement

Scheerens, J., & Bosker, B. (1997). *The foundations of education effectiveness.* New York: Pergamon.

Smith, T., & Ingersoll, R. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal*, *41*(3), 681–714.

Snider, V. E. (2006). *Myths and misconceptions about teaching: What really happens in the classroom.* Lanham, MD: Rowman & Littlefield.

Spear-Swerling, L. (2008). Response to intervention and teacher preparation. In E. L. Grigorenko (Ed.), *Educating individuals with disabilities* (pp. 273–294). New York: Springer Publishing.

Swanson, H. L., & Hoskyn, M. (1998). Experimental intervention research on students with learning disability: A meta-analysis of treatment outcomes. *Review of Educational Research*, *68*(3), 277–321.

U.S. Department of Education. (2011). *National — All schools — Percent of core academic classes taught by highly qualified teachers*. [Data file]. Retrieved from http://eddataexpress.ed.gov/data-element-explorer.cfm/tab/trend/deid/1262/state/US/

Walberg, H. (1999). Productive teaching. In H. C. Waxman & H. J. Walberg (Eds.), *New directions for teaching practice and research* (pp. 75–104). Berkeley, CA: McCutchen Publishing.

Walsh, K., Glaser, D., & Wilcox, D. D. (2006). *What education schools aren't teaching about reading and what elementary teachers aren't learning.* Washington, DC: National Council on Teacher Quality (NCTQ).

Wang, M., Haertel, G., & Walberg H. (1997) Learning Influences. In H. Walberg & G. Haertel (Eds.), *Psychology and educational practice* (pp. 199–211). Berkeley. CA: McCutchen Publishing.

White, W. A. T. (1988). A meta-analysis of the effects of direct instruction in special education. *Education and Treatment of Children, 11*(4), 364–374.

Wilson, S. M., & Floden, R. E. (2003). *Creating effective teachers: Concise answers for hard questions.* An addendum to the report, Teacher preparation research: Current knowledge, gaps, and recommendations. American Association of Colleges for Teacher Education. Retrieved from http://eric.ed.gov/PDFS/ED476366.pdf

Yeh, S. S. (2007). The cost-effectiveness of five policies for improving student achievement. *American Journal of Evaluation, 28*(4), 416–436.

Zeichner, K. M., & Conklin, H. (2005). Teacher education programs. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA panel on research and teacher education* (pp. 645–735). Mahwah, NJ: Lawrence Erlbaum Associates.

Chapter 2

Science and the Education of Teachers

JAMES M. KAUFFMAN
University of Virginia

ABSTRACT: *The preparation of teachers should be as scientific as we can make it. Preparing teachers in the scientific tradition requires embracing scientific and mathematical views of things that are impossible and those that are inevitable. It also requires knowing what science is and is not. Making teaching an applied science will require developing manuals and checklists for instruction and adopting the notion that teaching should become a manualized profession. All professions basing their work on scientific evidence and field tests develop manuals and checklists to guide their practices. Education must do the same.*

We must soon decide whether education — particularly teacher education — is going to be made significantly more scientific or continue very much as it has been practiced. In many ways, teacher education is at a crossroads. Teacher education's inadequacy was summed up well by Snider's (2006) description of her undergraduate training in education:

> I learned very little in my undergraduate teacher education program about how to teach; and for those first 8 years I relied on luck, trial and error, and the competence of colleagues for my professional development. I regret that I didn't know more from the beginning because despite my earnest efforts, my students didn't achieve as much as they could have. I knew very little about curriculum, effective teaching, or principles of classroom management beyond what I learned on the job. (p. 2)

This description is not very different from what many of us have experienced. Lack of training in direct instruction or Direct Instruction (see Kauffman, 2010, for a description of differences between di and DI) is understandable for those of us trained before these effective ways of teaching were described clearly. But why are most prospective teachers *not* taught to use DI today? DI is a readily available, scientifically and logically derived, field-tested program for teaching that, to my dismay, is still widely ignored. To me, this is inexcusable. We in higher education must do better. We must make education and the preparation of teachers an applied science (Kauffman, 2011) and a logically derived endeavor (Engelmann & Carnine, 2011). And we must do better than alternatives such as Teach For America and Troops to Teachers. If teacher preparation in higher education remains what it has been, it will very likely be completely discredited, as it probably should be.

This paper is divided into three major sections. The first section is about some impossibilities that people seem to hanker for but that just aren't possible — and then some inevitabilities that people tend to ignore. We ought to get over the fantasy that we can achieve the impossible or avoid the inevitable. I highlight impossibilities and inevitabilities because if we cannot get comfortable with them, then we have no real hope of making education a science. The second section describes a few of the many common misconceptions about science, with special attention on how they apply to education. Common misconceptions about science can lead us to false conclusions. The third section is about why we need manuals and checklists in education, especially in teacher education. It makes a case for teaching teachers to teach by the book and for using devices to help keep us from making common errors that are especially costly to learners.

IMPOSSIBILITIES AND INEVITABILITIES

Impossibilities

Some things just aren't possible. We know this because of some very basic realities of math and science. True, some things once considered impossible are possible today, and some of the things we consider impossible today might be possible some day. However, some things will always remain impossible, such as adding two positive integers and obtaining a sum less than either of them or talking about something without using a word or words for it. Unfortunately, some people either implicitly or explicitly assume that we do not really have to deal with unchangeable realities, that these realities can either be ignored or treated as inconvenient truths. For example, some people seem to think that universal proficiency, something impossible by definition, is achievable and raise questions only about the year in which we might reasonably expect it to be achieved. One newspaper editorial — without irony, obviously not noticing the mathematical impossibility, practical nonsense, or self-contradiction of its statement — praised the goal of 100% proficiency in 2014 set by the No Child Left Behind Act (NCLB), saying that the goal of universal proficiency "while laudatory, may be unrealistic" (*The Washington Post*, 2007).

To understand what I've said, you first have to think about what "proficiency" means. Proficiency in any skill (say, swimming or driving or math) is defined by what most people can do after specific training; it isn't a level of performance just pulled out of the air without reference to what people can do. Proficient/not proficient isn't a distinction based on what only a few of the highest performers can do, nor is it based on the performance of the most inept. So, to say that *all* people will do what *most* people can is simply a self-contradiction, a logical impossibility. It's as comical as saying that we're going to have all the children above average. Don't misunderstand. We can often help *more* people become proficient at something, but *all* people? Well, *universal* proficiency — *all* students becoming proficient in an academic skill, for example — just isn't in the cards. Getting more people proficient at something could be very hard but possible for some skills. Truly universal proficiency? No. Won't happen. Those of us who work with students who have severe disabilities understand that universal proficiency in reading, for example, just isn't possible. We don't approve of terms like "all" or "universal" being used as if our kids aren't considered. In education, it's important to think about the meaning of what we say and to say what we mean as precisely as possible. We want our students to do that. We should do that too.

Actually, the impossible is not a laudatory goal. That is, it's not good to set our sights on something that's logically, mathematically impossible. And it's not good to say things we don't mean. It's tempting to make sarcastic remarks about ill-considered comments like those in the *Post* or the contention that *all* children, regardless of their level of ability, should go to college or be prepared for a career — verbal equivalents of waving to Ray Charles (Kauffman, 2005). Real-world talk about education is more likely to better the lives of children than fantasy talk is.

Another impossibility is measuring something reasonably precisely without getting a statistical distribution. In education, this means a distribution of scores ranging from lowest to highest and having an average. Measuring educational performance accurately without getting a distribution with what statisticians call "moments" — mean, standard deviation, skew, and kurtosis, for example — is impossible. There are no exceptions. And this means that it is impossible to find that all of the individuals measured are at or above any location on that distribution except the lowest point. So, finding that all of the students are at or above the 20th percentile, for example, is impossible. Regardless of what a secretary of education or the United States Congress or someone with a Ph.D. hopes for, sets as a goal, or decides *should* happen, it is just not possible with the kind of mathematics we have on planet Earth. Consequently, NCLB was dead on arrival because it assumes that *all* students — or very nearly all, even excluding 2% or so of students who have disabilities — can be judged proficient by their state's test scores.

Now, probably I should explain a little about that 20th percentile statement I made. We can almost always improve students' performance. And, depending on the comparison we make, we could have more than 80% of students scoring above the 20th percentile — *of a distribution of test scores other than the one in which they were included*. That is, it's possible to have all of a particular group of students who took a test be above the 20th percentile of a different group of students who took the same test. For example, we could find that in a given school all of the students who took the SAT in a particular year scored above the 20th percentile on the SAT norm (i.e., above the 20th percentile of the group that took the test for norming purposes and established the 20th percentile for the norm). So, there's always the question of what comparison we want to make. Do we want to compare the students in the group that just took the test to each other on that test, or do we want to compare them to another group? Sure, we might get all students above a percentile greater than zero if the percentile refers to the percent in a different group, but not if it's the group we have. Let's think a little more about this.

It's possible to "play games" with statistics, even to play a game that makes something look good. Sometimes the game is played fairly. Comparison to an existing norm, perhaps even an old one, *can* make sense. But, suppose we want to make a group look good in such a comparison, even if it's sensible, by showing that everyone in the group is above the Xth percentile (i.e., any percentile greater than zero) for another group. We could do three things, and here's where the game gets really tricky and can be played to mislead people. First, we could choose a lower percentile; the lower the percentile, the greater our chance of getting everybody above it. Second, we could compare a smaller group to a larger group; the smaller the group we compare to a larger group, especially the normative group, the better our chances of getting everybody above a given percentile of the larger group. Third, we could make a comparison to a group that includes a lot of low performers; the greater the percentage of low performers in the comparison group, the better our chances of looking good by comparison.

Which reminds me of another thing we might consider: If we get all of the students above, let's say, the 20th percentile of some older test or normative group, then should we consider the older test outdated because the old norms aren't valid? The point is that we could make a comparison that isn't really sensible. And sometimes it's illogical, not sensible, to make a comparison to another group of test takers. But let's get back to the meaning of 20th percentile. For any given group that takes a test, we can't have more than 80% of the students who take that test above the 20th percentile of their group simply because the meaning of 20th percentile is that 20% of those who took the test got that score or a lower one. It's impossible to have more than 80% above the

20th percentile of that group for the same reason that we can't have more than 100% of a group.

Other examples of impossibilities that people call for can be found, like *an elite education for everyone*, which by definition is impossible (Kauffman, 2010). Garrison Keillor's description of Lake Wobegon is funny because we realize that it is impossible to have all of the children above average. We should not allow silly statements about children or schools, such as a goal of universal proficiency, to carry any legitimacy in serious talk about education. Lake Wobegon talk and the goal of universal proficiency are the stuff of comedy, not of serious thinking about educational outcomes.

Inevitabilities

Some things happen whether we want them to or not; they are inevitabilities, just facts of life that we should acknowledge and not think we can ignore. They are the flip side of impossibilities, realities that won't go away even if we wish they would. One example of the inevitable is the reality that some children are not going to learn to read. Ever. Even basic sight words. And some are not going to learn to read with what we consider fifth-grade comprehension. No matter what program we use or who teaches them.

What makes me especially angry is that many people do not include these children when *talking* about the education of *all* children — which, presumably, reflects their *thinking* about what is involved in teaching *all* children. These children, who don't "measure up" to the expectation that *all* children will reach a certain level of educational performance, are just written off, apparently. So when NCLB or some other misbegotten policy calls for *all* children to learn… whatever… these children are not even considered. They're assumed not to count, to be insignificant, and they and their teachers are assumed to be irrelevant or incompetent. As a special educator, I realize that there are children of school age who cannot walk or talk or communicate, cannot feed or toilet themselves, and need care and supervision around the clock. Yet these children are to be loved and respected and taught all the skills they can learn. But supposing that they can be made ready for college or a career is just preposterous. I also realize that there are children at every point on the distribution of ability.

The "bell curve" or normal distribution is often condemned, but the realities of statistical distributions of whatever we care enough about to measure will not go away (Kauffman & Lloyd, 2011). True, people sometimes make appalling assumptions about individuals related to distributions, but equally appalling is the assumption that the bell curve (i.e., a statistical distribution) can be ignored, called irrelevant in making policy decisions or training teachers, or simply wished out of existence. Although people may well be mismeasured (Gould, 1996b), regardless of the way they're measured we must consider what Gould

(1996a) called the "full house" — in education, all of the children to be taught including those far below average in whatever skill is measured.

Then there is the problem of prevention. I think the following bears special attention: Most people love the idea of prevention in the abstract, but they do not want to face its inevitabilities in practice. They seem to forget that it is impossible to prevent what has already happened. Prevention requires *anticipation* of whatever is supposed to be prevented (Kauffman, 2003). People often argue that we now misidentify many students as having disabilities but that we should practice prevention. Perhaps they do not understand that if many students are now misidentified as having disabilities, then prevention inevitably means that we are going to misidentify even more.

Imagine what would be inevitable if we actually practiced prevention. First, think about measuring whatever it is you want to prevent — maybe reading failure, maybe behavior problems, maybe something else. If you think about measuring it, then you are going to visualize the result — a distribution of scores that might approximate a normal curve. But even if the distribution you imagine is skewed (lopsided) or leptokurtotic (bunched up) or platykurtotic (spread out), think about what prevention requires. Prevention requires keeping as many individuals as possible from getting far from the central tendency of that distribution.

The basic idea of prevention is catching students earlier, before their problems get so bad. The idea is that if we catch a problem reader in first grade, for example, and we offer really effective instruction, then that student won't be so far behind come fourth grade. So, prevention requires moving the criterion or trip point for doing something about the problem (usually, we call this an intervention) closer to the central tendency. The distribution means, mathematically, that prevention requires including more individuals, not fewer, in the intervention. It requires increasing the risk of a false positive — identifying a child for help who doesn't really need it. Thus, complaints that we already serve too many students in special education and expressions of the unacceptability of misidentification are really arguments *against* prevention. Now, admittedly, if we move the criterion for receiving special education to more severe cases and leave prevention to the province of general education, then complaints that too many children are receiving special education can make some logical sense. However, then the argument for prevention becomes one of suggesting that more children should be identified earlier in general education, and also that the criterion for getting special education should require children to fail as much as they do now.

But for now, consider another problem that often upsets people when it comes to identifying children for special education or any other intervention — the mistakes in identification called false positives and false negatives. People generally don't like drawing a line, because it's arbitrary and some children are always close to it. The children who are in the close-to-cut-point (CTC) areas of a distribution are the ones who don't quite make it into the special program for some reason or who are in a special program when they perhaps shouldn't be. It is always possible to argue that some children were selected for a special program when they should not have been (false positives) and some should have been selected but were not (false negatives).

A line or criterion for a special program is an inevitable part of having a special program. Lines, labels, sorting — they are all necessary, inevitable aspects of having a special program. Who should get help with their mortgage? Having a program of financial assistance requires a line for qualification, a label describing those who participate in the program, and a means of sorting those who need help from those who do not. The same problems go with the designation of Tier 2 or Tier 3 in response to instruction or levels as does any other program in education that does not include every single child (Kauffman, 2010; Kauffman & Lloyd, 2011).

Another inevitability is that every line has margins (we might call them standard errors). No exceptions. Those who complain of misidentification may suggest multiple lines. But the problem is that every additional line creates more margins and more mistakes. Always. This is just basic math and science. Some cases are always CTC. Having three tiers instead of two in response to instruction increases the chance of making mistakes by about 50%.

Examples

We might consider an example of measuring reading ability and designating a level of performance that signifies reading failure. Remember, measuring without getting a distribution is impossible, and drawing a line or cut point for qualifying for a special program is inevitable. Figure 1 depicts a distribution of reading scores, with lower reading scores on the left of the curve. Possible cut points representing reading failure, A and B, are shown, along with their margins. If the cut point is moved from A to B (i.e., from more severe to less severe reading problems), then more students are included in the definition of reading failure. Moreover, by having two cut points, A *and* B, each a different tier, we double our chances of making an error — a false positive or a false negative. Two points about inevitability are noteworthy. First, moving the cut point toward less severe problems inevitably involves more individuals (i.e., a greater area under the curve). Second, every cut point has margins, its CTCs — areas of uncertainty on both sides of the cut point, simply because no test or other means of judgment is faultless, containing no error.

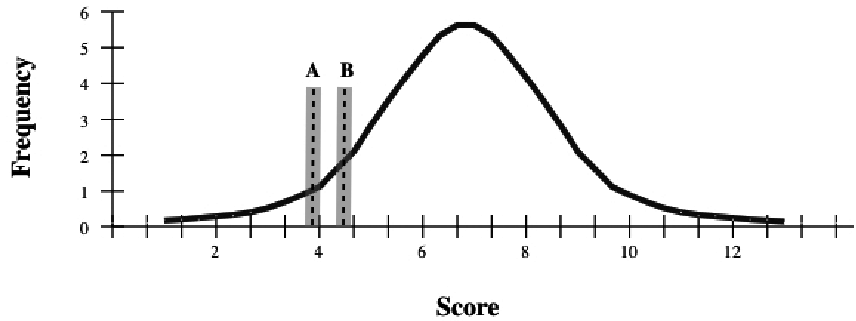


Figure 1. A hypothetical distribution of reading scores with alternative cut points A and B (and their associated margins or borders of uncertainty) indicating reading failure.

Figure 2 is a depiction of a curve of behavioral problems or disorders. It is another way of showing that as we move a cut point for defining disorder toward less severe cases (in this case, from A to B or C or from B to C), we inevitably include more children in our definition.

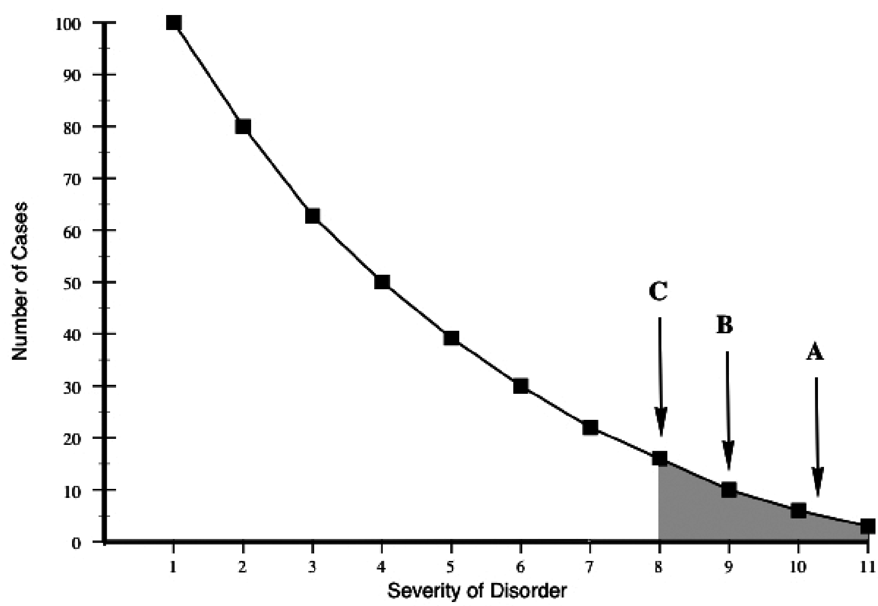


Figure 2. Number by serverity. Hypothetical curve of emotional or behavioral disorders showing alternative cut points A, B, and C defining disorder.

In no way does acknowledgment of the inevitabilities depicted by Figures 1 and 2 deny the agony of making difficult decisions that affect children's lives. However, we must acknowledge realities, like cut points and CTCs, in talking and writing about education. We have more than enough evasion of realities in rhetoric about education already. We do need precision in teaching, but we also need precision in our language and *thinking* about teaching. Perhaps we should start with careful thinking about what is possible and what is inevitable. The call for evidence-based practices must be consistent with reality-based thinking. "Never-never land" thinking does not help us or children.

The impossibilities and inevitabilities I have described are very fundamental ideas about realities, about how the world works, about mathematical functions that we cannot wish away. If we find it inconvenient to deal with them, then I think we have no real hope of adopting a scientific approach to education.

COMMON MISCONCEPTIONS ABOUT SCIENCE

It is very hard to get people who have not been trained as scientists to think scientifically about *most* things, including education (Kauffman, 2011; Landrum & Tankersley, 2004; Sagan, 1996; Specter, 2009). A very common belief is that teaching and learning cannot be researched in a scientific way. Some will argue that even if certain aspects of teaching can be made a science, the scientific research on teaching is trivial or meaningless — that the truly important things that happen between teachers and students, especially in students' minds, just can't be a matter of scientific study with important implications for teaching.

True, a science of education is very difficult, but it is not impossible. A science of education is in many respects more difficult than the "hard" or "bench" sciences because the kinds of *control* that can be achieved in most physics experiments and the kind of *stability* that characterizes earth science are impossible in education. Berliner (2002) was correct in saying that a science of education is the hardest science of all. Some scientists say that education is not a science at all. And to some extent, they're correct because education is not reliably scientific now and never has been. But I do not agree with the scientists who say that education is beyond the *reach* of science, for reasons that I hope will become clearer as you read this section of my paper.

A science of education is extremely unlikely without an understanding of what science is and how it works in the general case. Therefore, in this section I review a few of the basic principles that apply to *any* science and give some examples of how these principles might apply to education. (For a discussion of more principles than those covered here and a fuller treatment of science in education, see Kauffman, 2011.)

A serious science of education could disenthrall us from magical thinking about impossibilities and inevitabilities. It could help us find out *what* works and help us understand *why* something works or the *conditions* under which it works and doesn't work. Following the path of science does not guarantee that we won't make any missteps. In fact, if we take a scientific view of education, we'll make some mistakes. However — and this is important — our being scientific means that we will eventually, if not immediately, recognize our errors. As my friend, biologist Dan Burke, commented, "Science is not a steady parade of 'truth' but more a tortured path of six steps forward and five steps back, but generally moving in the right direction" (personal communication, December 27, 2009).

Science is commonly misunderstood. Not just by teachers, but by the general public. People would often rather be illiterate or disbelieving when it comes to science, even in areas like medicine (Sagan, 1996; Specter, 2009). We educators face a tremendous challenge in trying to help people understand how science might be applied to teaching and the advantages of a scientific approach to instructional problems.

Judgment Versus Certainty

One of the first things to come up in talk of science and education is *judgment versus certainty*. Contrary to popular opinion, data do not speak for themselves. Scientists must speak for data to make sense of their findings. Educators must use their judgment to urge action based on what they see as the *preponderance* of evidence rather than unarguable results. For example, DI ought to be supported because the preponderance of evidence suggests it is more effective than whole language.

Disproof Versus Proof

Scientists understand the idea of *disproof versus proof*. Science is not really the pursuit of direct proof but of things that can't be disproved — indirect proof. Scientists try to find something wrong with findings or explanations, and if they can't find anything wrong, then accept what they have found as their best guess — a tentative truth (Baldwin, 2008). In education, we very seldom can assume that something has been proved. More often, it's possible to draw the conclusion that something has been *disproved* beyond any reasonable doubt. A given procedure may be shown *not* to produce the desired result. So, then, we conclude it's no good or doesn't work. If something is shown *not* to work, then scientists accept the evidence that it doesn't work; only findings that scientists try their best to disprove but can't disprove pass scientific muster.

"Facilitated communication" (FC, in which a "facilitator" helps someone who is communicatively impaired type messages on a keyboard but does not influence the messages) illustrates this principle. Someone who says that the real task in research on FC is to show that it *does* work, not that it doesn't, is simply wrong. Science doesn't work that way. Someone who actually understands the scientific way of looking at problems knows that the way to show that FC apparently works is to try very hard to show that it doesn't work and to fail. Only if researchers can't disprove FC are they allowed, using the methods and assumptions of science, to assume that FC works. If researchers are successful in showing that FC *doesn't* work, then scientists will conclude that FC is hokum. Those trying to show that FC *does* work are wasting their time and the time of anyone who listens to them. They're wasting time because FC has already been shown resoundingly *not* to work. The suggestion that the real scientific task is to prove that FC works in the face of overwhelming evidence that it doesn't is much like saying that the real task of science is to prove that there is a raccoon at the dinner table when overwhelming evidence indicates that there is not.

Another example of this principle is the claim that cold fusion had been achieved. The real task of scientists was never to show that cold fusion worked or had been achieved. The task of science was to show that it didn't work and wasn't achieved. Only if people failed in every attempt to show that cold fusion did *not* work would we be led by science to conclude that it must have been achieved.

Contingent Versus Noncontingent Statements

The principle of *contingent versus noncontingent statements* is very important to scientists. Scientists usually qualify their statements by specifying contingencies. They might say something will probably happen only *if* or *when* the conditions are right. The idea is that they describe the *conditions* under which something is likely to happen and those under which it isn't. Almost always in education, results have to be called contingent. For example, the claim that rewards *always* work is baloney; the claim that rewards work has to be qualified. Most teachers know and all scientific investigations have found that rewards have their desired effect only under certain circumstances. It's true that children might be either punished or reinforced by presumed rewards, depending on the circumstances and just how the attention or other rewards are given (Kauffman, Pullen, Mostert, & Trent, 2011). Of course, claiming a contingency that can't be disproved, such as "only if you *really* believe," as is sometimes done with FC, isn't enough. Disproof is still the key.

Replication Versus Idiosyncratic Data

Another misconception about science with special relevance for education is *replication versus idiosyncratic data*. The finding of a single scientist or lab isn't at all convincing to actual scientists. Real scientists are not satisfied unless other people working in other labs can replicate a finding. This was one of the big problems with the cold fusion claim: Other people couldn't make it happen in their labs; only the researchers who claimed they produced it, only those who didn't doubt it, could do it.

Education is especially susceptible to claims that can't be replicated. Finding an effect that can be replicated by other teachers in other locations is particularly important. A single study means relatively little unless it was extraordinarily large and well designed. Even then, definitive evidence can be had only by replication. One of the reasons the programs known as DI are scientifically sound is that the curricula and instructional methods are replicable, and replication has confirmed DI's superior effectiveness (Carnine, Silbert, Kame'enui, & Tarver, 2010).

Observation, Measurement, Reason, and Experiment Versus Philosophy or Ideology

Many people seem not to understand that scientists are concerned about the use of reason or rationality as well as observation and measurement. Actually, scientists are interested in this contrast or competition: *observation, measurement, reason, and experiment versus philosophy or ideology*. Too many educators pride themselves most in their philosophy or ideology and take too little pride in the four contrasting demands of science. Consider at this point just the matter of reason or logic. A science of education requires logic. It requires *more* than logic, but it can't *ignore* logic. Experiment is critical, but so is logical analysis of problems (Engelmann & Carnine, 2011; Engelmann, Bateman, & Lloyd, 2007). Remember that data do not speak for themselves, but when scientists speak they must make sense. That is, they have to be rational.

I recently found an excellent example of *illogic* in reading about standards-based Individual Education Programs (IEPs) in a publication of the National Association of State Directors of Special Education, in which Ahearn (2006) quotes a professor of education: "We must understand that 'ready means never.' If we wait until students are ready to work on challenging standards by virtue of having mastered basic skills, they will never work on challenging standards" (p. 12).

Is this true *only* for teaching children with disabilities, or is it a generalizable principle that we could apply to other problems of education? Think about the training of athletes, musicians, scientists, and, in fact, training in anything in which there are prerequisites for working on more challenging tasks. Ask yourself some reasonable questions. Would you say that basketball players don't really need basic skills in passing, dribbling, shooting, and so on before they play in competition and that judging them not ready for competition because they haven't mastered the basic skills means they will never play in competition? Would you suggest that beginning piano players ought to tackle difficult pieces of music first so that they don't waste time on basic music skills, because if they must first master basic skills then they'll never work on challenging pieces?

Clearly, there is a serious disconnect between ordinary logic applied to other problems of teaching or learning and Ahearn's quotation. Perhaps instructing children with disabilities is a unique case, in that the acquisition of fundamental skills is not necessary for acquiring more advanced skills. But I doubt it. Or maybe the professor of education Ahearn was quoting meant to say that some students are often not appropriately challenged. That assertion may be true. For some students, fundamental skills are a challenge, and some students aren't challenged by fundamentals. But saying that some students are not appropriately challenged is quite different from stating that if students are required to master basic skills before attempting more challenging tasks, then they'll never be asked to take on challenging (i.e., advanced) tasks or standards.

You might also wonder whether the professor of education Ahearn quoted is in la-la land and actually believes that all students can learn whatever is challenging for the majority of students, regardless of what they've mastered previously. In any case, the statement reflects outrageously poor, illogical thinking or careless language or both. I repeat: Science requires more than logical thinking, but it does require logical thinking. There is no illogical science of anything. And when it comes to education, prior learning is the single most important factor to consider in what a student should be expected to learn next. The statement that Ahearn quotes suggests that students with disabilities who are working on tasks that are challenging for *them* don't need to master more basic skills before working on their challenges. I can only hope that someone intended to say that students are always ready to learn their next challenging task only after they've learned more basic skills. But that is not what the statement says.

Gradual Change Versus Paradigm Shift

A lot of education reformers seem to misapprehend the issue of *gradual change versus paradigm shift*. Paradigms do not change often in science, and they are not changed simply by demand, assertion, or act of will. In science, paradigms are changed by data that can't be explained by an older paradigm. And a new paradigm does not necessarily invalidate an old one but might just add to it. For example, quantum mechanics adds to certain aspects of subatomic physics, but it does not overturn or invalidate or replace Newtonian mechanics for macroscopic objects. Many education reformers are particularly fond of the "break the mold" or "breakthrough" idea of educational reform. However, in education, just as in other scientific endeavors, actual paradigm shifts are *extremely* rare. Gradually accumulated evidence is more likely to be a reliable guide to good teaching than is something paradigmatically different from anything we already know.

Theory Versus Fact

Lots of people misunderstand the scientific meaning of *theory versus fact*. In science, a theory is a way of making sense of facts. To a scientist, a theory is not just a guess. A theory is something that for a scientist organizes facts and helps the scientist predict phenomena. In a science of education, theories should help us make sense of research data. "Theory" in education must come to mean what it means in better established sciences. It can't be a euphemism for ideology or mere guesswork.

PREPARING TEACHERS TO USE A SCIENCE OF EDUCATION

The preparation of teachers has been a highly controversial issue for a very long time. And preparing teachers to put a science of education into practice is just one more controversial aspect of it. One obvious fact about preparing teachers to use a science of education is that we have to have a science of education for them to be prepared to use! Everyone wants better teachers, so that is not the issue. The issue is how to prepare teachers better. If you ask people who don't know much about teaching what we should do, they are likely to say something like, "Well, get smarter teachers" or "We need teachers who know their subjects, and that's more important than the kind of teacher training they get."

How to Train Teachers

How should prospective teachers be selected and trained? That is not an easy question to answer unless you are going to just repeat the same tired old nonsense we've been hearing from education reformers for more than 50 years. Trying to answer that question requires some actual knowledge of teaching and schools. If the answer were really simple, either those trying to answer it are too dimwitted to figure it out or there is some sort of conspiracy to keep education from being what it should be.

What are the essential personal characteristics of good teachers? Just how smart does a teacher have to be? Are smarter people better teachers, or at what point does intelligence become irrelevant because just being smarter doesn't make a person a better teacher? What are the essential skills teachers need to be successful? What role does knowledge of each of the following play in making a good teacher: (a) subject matter to be taught, (b) child development, (c) pedagogy or instruction, and (d) behavior management? What other areas of knowledge or expertise are required? To what extent can teachers be prepared before they enter a classroom, and how much (and what) do they simply have to learn on the job? How can we distinguish better teachers from those not as good; that is, how should we rank teachers for reward or recognition and identify those who should be fired for their incompetence or, at least, be told they'd better improve dramatically if they want to keep their jobs? These are not trivial questions. They go to the heart of what teacher preparation is and to the root of controversies about teachers and teaching.

Lots of assumptions and ideologies are related to these questions, but not lots of good thinking and not lots of scientific evidence. Aside from a few obvious characteristics such as not being abusive to students, being fairly intelligent, being reasonably sensitive to the needs of others, and having a relatively high level of energy, we just don't know much about what kind of person makes a better teacher. Aside from the logical assumption that a person can't teach something he or she doesn't know, we are in the dark about how important subject knowledge is. Clearly, people can be failures at teaching what they *do* know. So just knowing something isn't all that's important; knowing how to *teach* it is important if someone is going to be a successful teacher. Teachers who know their subject could be taught to use DI, but that has long been neglected by the education world (Engelmann, 2007).

One thing we can do if we want to make teacher preparation more scientific is look at other types of work that are essentially applied sciences to see how they have made use of the scientific method and put science into practice (Carnine, 2000). It might be impossible to find another profession in which this has been done completely or flawlessly, but that is not essential. We do know that some other professions are way ahead of the teaching profession in making use of science and getting their practitioners to be more consistent in using the practices that science tells them are more effective than just going with their intuition or preferences or some other seat-of-the-pants way of deciding what to do and how to do it.

For example, piloting airplanes and performing surgery are manualized in many ways. Many professions give their trainees manuals because complicated work is involved. The basic idea of a manual is that other people have done this complicated task before and found out how to do it without making a mess of things or creating a disaster. In fact, the manual usually tells trainees as well as experienced professionals how to do something safely, if not best. The reason for following a manual is that responsible professionals do not want to make a fatal error or do something that creates a crisis or unnecessary risk. The manual explains how to avoid a crisis, how to avoid risking disaster, how to do something so that success is more likely than failure. This is why we want the people we entrust with our lives or our health to follow the manual — we want them to do it, as we say, "by the book."

A good manual gives step-by-step instructions based on scientific knowledge and field tests. It is based primarily not on a philosophy or guesswork but rather on what science and logic and experience recommend. A good manual tells us not only how to do something but how to solve problems — how to troubleshoot if something doesn't go right. Why is it taking us so long to manualize the profession of education?

Another way of avoiding disasters that we ought to adapt for education is the checklist. In *The Checklist Manifesto: How to Get Things Right*, Gawande (2009) describes the value of checklists for things like flying airplanes and building skyscrapers in addition to performing surgery. In fact, he suggests, anything that is very complicated can be done far more safely with a really good checklist. The checklist has to be short, focused on the most important things that science and experience tell us, and useful for practitioners. Some people in every profession resist using checklists, but virtually no one receiving professional services thinks that professionals can do without them. Passengers want *their* pilot to use a checklist. Physicians having surgery want *their* surgeon to use a checklist. Using a good checklist is just a way of avoiding an unnecessary calamity. Why haven't checklists become an important part of teaching and preparing teachers?

Perhaps the reason is that education often is not based on scientific information and field tests, as Engelmann (2007) points out so painfully. Educators can't seem to develop a consensus about lots of things, like what they believe children should be able to do and how best to get them to do it. What educators seem to want to do is argue philosophy, not solve problems in a scientific manner. We ought to be aware of what other professions do; for the most part, they apply science, they prepare step-by-step manuals based on practice, they use checklists to help practitioners remember important things, and the more complex the task they undertake the more they see the need for manuals and checklists. In education, we simply don't need to mislead teachers into thinking they can just "wing it" in the classroom.

Suppose we are going to get serious about using manuals and checklists in training teachers. What do we need to know about how a checklist works? Here are some things to remember: (a) A checklist isn't any help if you don't have a specific outcome in mind; (b) you have to know whether what you check off has been done; (c) a checklist does not mean you can be competent without being artful in practicing your profession; (d) a checklist has to be short and designed to avoid common, serious mistakes; and (e) a good checklist does not concentrate power in a particular person, but increases communication and helps people function better as a team.

Now, with all the advantages of checklists, why do so many people despise them, especially for teaching? Well, checklists require close attention to what we're doing, they may make us feel regimented, they point out human frailty — and we like to see ourselves as creative people who are able to improvise and don't need checklists.

Someone could ask, very reasonably, whether we have any manuals and checklists in education. I think we do. For example, much of DI is pretty well manualized, and it has been demonstrated to be a highly effective way of teaching arithmetic and reading, especially to students who are low performing or at risk of academic failure. But we need to develop manuals and checklists for teachers in many more areas of their work. Figure 3 is an example of a *possible* checklist derived from a behavior management text (Kauffman et al., 2011). It may have serious flaws. Before it could be judged sound — reliable and useful — it would need to be field tested like all other checklists that pass muster.

When giving instructions, have I:

- ☐ Made the instruction as simple and clear as possible?
- ☐ Given the instruction in a clear, firm, but polite way?
- ☐ Obtained students' attention before giving the instruction?
- ☐ Given only essential instructions?
- ☐ Given one instruction at a time?
- ☐ Waited a reasonable time for compliance?
- ☐ Monitored compliance?
- ☐ Provided appropriate positive consequences for compliance?

Figure 3. Possible checklist for giving instructions (Kauffman et al., 2011).

CONCLUSION

A science of teacher education is difficult but possible. It first requires a science of education. Such a science requires recognizing impossibilities and inevitabilities, understanding what science is and isn't, and devising manuals and checklists. We must get on with the task of creating useful manuals and checklists for our work. These must be based on reason, field tests, and scientific evidence of effectiveness.

REFERENCES

Ahearn, E. (2006, May). *Standards-based IEPs: Implementation in selected states*. Alexandria, VA: Project Forum, National Association of State Directors of Special Education. Retrieved January 16, 2009, from http://www.projectforum.org/docs/Standards-BasedIEPs-ImplementationinSelectedStates.pdf

Baldwin, J. D. (2008). *Ending the science wars*. Boulder, CO: Paradigm Publishers.

Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher, 31*(8), 18–20.

Carnine, D. (2000). *Why education experts resist effective practices (and what it would take to make education more like medicine)*. Washington, DC: Fordham Foundation.

Carnine, D. W., Silbert, J., Kame'enui, E. J., & Tarver, S. G. (2010). *Direct instruction reading* (5th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Engelmann, S. (2007). *Teaching needy kids in our backward system: 42 years of trying*. Eugene, OR: ADI Press.

Engelmann, S., Bateman, B. D., & Lloyd, J. W. (2007). *Educational logic and illogic*. Eugene, OR: Association for Direct Instruction.

Engelmann, S., & Carnine, D. (2011). *Could John Stuart Mill have saved our schools?* Verona, WI: Attainment.

Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York: Metropolitan Books.

Gould, S. J. (1996a). *Full house: The spread of excellence from Plato to Darwin*. New York: Three Rivers Press.

Gould, S. J. (1996b). *The mismeasure of man* (revised & expanded ed.). New York: Norton.

Kauffman, J. M. (2003). Appearances, stigma, and prevention. *Remedial and Special Education, 24,* 195–198.

Kauffman, J. M. (2005). Waving to Ray Charles: Missing the meaning of disability. *Phi Delta Kappan, 86,* 520–521, 524.

Kauffman, J. M. (2010). *The tragicomedy of public education: Laughing, crying, thinking, fixing*. Verona, WI: Attainment.

Kauffman, J. M. (2011). *Toward a science of education: The battle between rogue and real science*. Verona, WI: Attainment.

Kauffman, J. M., & Lloyd, J. W. (2011). Statistics, data, and special education decisions: Basic links to realities. In J. M. Kauffman & D. P. Hallahan (Eds.), *Handbook of special education* (pp. 27–36). New York: Routledge.

Kauffman, J. M., Pullen, P. L., Mostert, M. P., & Trent, S. C. (2011). *Managing classroom behavior: A reflective case-based approach* (5th ed.) Upper Saddle River, NJ: Pearson.

Landrum, T. J., & Tankersley, M. T. (2004). Science in the schoolhouse: An uninvited guest. *Journal of Learning Disabilities, 37*, 207–212.

Sagan, C. (1996). *The demon-haunted world: Science as a candle in the dark.* New York: Random House.

Snider, V. E. (2006). *Myths and misconceptions about teaching: What really happens in the classroom*. Lanham, MD: Rowman & Littlefield.

Specter, M. (2009). *Denialism: How irrational thinking hinders scientific progress, harms the planet, and threatens our lives*. New York: Penguin.

The Washington Post. (2007, September 10). Save school standards: Congress should resist attempts to water down the No Child Left Behind law, p. A14.

Chapter 3

Comprehensive Teacher Induction: What We Know, Don't Know, and Must Learn Soon!

LARRY MAHEADY
MICHAEL JABOT
SUNY Fredonia

ABSTRACT: *Comprehensive teacher induction refers to those practices that help new and beginning teachers become competent and effective classroom professionals who also understand school and community cultures. Induction programs were designed to support new teachers and facilitate their socialization into the profession. Additionally, these programs were seen as productive ways to improve retention in the profession, refine instructional practice, and ultimately improve pupil learning. To date, induction programs have failed to accomplish such lofty educational goals. This article describes what educators know and don't know about comprehensive teacher induction and offers some general guidelines for improving both research and practice. The paper also describes the efforts at one regional state college to improve new and beginning teacher practice and to provide empirical evidence to support such efforts. The overall message is that comprehensive teacher induction can positively impact teaching practice and pupil learning; to do so, however, will require careful reconsideration of its conceptual, procedural, and empirical underpinnings.*

Too many American children are plagued by unacceptable educational outcomes, declared Secretary of Education Arne Duncan (2009). Almost one third of students drop out or fail to complete school on time, and only 60% of African American and Latino pupils graduate when expected. In many large cities, half or more of low-income teens drop out of school. Children who attend our neediest schools are likely to have the least qualified teachers, and over the next 4 years one third of our veteran teachers may retire. Duncan noted that teaching has never been more difficult or more important.

Yet these adverse outcomes and educational inequities are not new. Indeed, countless others have lamented America's academic decline, persistent achievement gaps, and increases in disruptive and destructive student behavior (e.g., Abell Foundation, 2001; Ballou & Podgursky, 2000; Carnine, 2000; Coalition for Evidence-Based Educational Policy, 2002; U.S. Department of Education, 2002; Walker, Ramsey, & Gresham, 2003–2004). Educators were warned, as well, that persistent educational failure may lead ultimately to societal questioning of teacher education's efficacy and its sole right to prepare teachers (Greenwood & Maheady, 1997). Secretary Duncan (2009) commented soberly on these possibilities by noting that over 60% of teacher education graduates reported that their training programs did *not* prepare them adequately for work in contemporary classrooms. He went on to cite specific shortcomings in classroom and behavior management, working with high-needs students, and using data to improve instruction and student learning. A clear gap exists between the educational realities of P–12 schools and preparation efforts in many teacher education programs (Cibulka, 2009).

This is not to suggest that policy makers, teacher educators, and education leaders and researchers sat by idly while educational outcomes deteriorated. On the contrary, in the past 40 years numerous educational reforms were undertaken to improve pupil outcomes. Keyworth (2010) and States (2010) highlighted many of these structural reforms (e.g., increased funding for students; higher pay for teachers; more teachers with advanced degrees, credentials, and professional certifications; smaller class and school sizes; charter schools; vouchers; high-stakes testing; and school reform initiatives like Goals 2000 and No Child Left Behind) and noted that their overall impact on pupil learning has been disappointing at best. Despite massive increases in funding, smaller class sizes, more qualified and better credentialed teachers, and extensive state and federal legislative reforms, student achievement in reading and math has remained relatively stable over the past three decades, and the achievement gaps and differential graduation and drop-out rates among high- and low-income students have persisted or escalated.

Perhaps one of the most promising structural reform efforts to emerge in the past few decades is teacher induction. Educational leaders, researchers, and policy makers heralded induction and mentoring programs as indispensable vehicles for supporting new teachers, increasing retention in the profession, refining instructional practice and quality, and ultimately improving pupil learning (e.g., Alliance for Excellent Education, 2004; Arends & Ragazio-DiGilio, 2000; Fletcher, Strong, & Villar, 2008; Guarino, Santibanez, & Daley, 2006; Howe, 2006; Strong, 2005). For many, teacher induction and mentoring programs were seen as ways to bridge the gap between preservice education and the classroom and to help new teachers make a successful transition into the profession. Evidence suggested further that these programs were received favorably in schools and that they had a positive impact on teacher retention (e.g., Ehrich, Hansford, & Tennent, 2004; Ingersoll & Kralik, 2004). Yet, significant questions remain regarding the effects of induction programs on teaching practice and student learning, the most salient variables in applied educational research.

This article describes what is known and isn't known about teacher induction and mentoring programs. These impressions were derived from an illustrative rather than comprehensive literature review and as such must be interpreted cautiously. The analysis examines conceptual, procedural, and empirical issues and discusses potential implications for policy makers, teacher educators, practitioners, and applied researchers. The remainder of the paper offers guidelines for improving induction research and practice and describes the modest efforts of a regional state university to do so. The overall message is clear: Comprehensive teacher induction *can* positively impact teaching practice and pupil learning; to do so, however, will require careful reconsideration of its conceptual, procedural, and empirical underpinnings.

PREVIOUS RESEARCH

Although teacher induction and mentoring emerged only during the past few decades, they have already generated extensive literature. A simple Google search, for example, yielded over 253,000 hits on the topics. In addition, at least 12 comprehensive reviews appeared in the general (Arends & Ragazio-DiGilio, 2000; Ehrich et al., 2004; Feiman-Nemser, Schwille, Carver, & Yusko, 1999; Gold,1996; Howe, 2006; Huling-Austin, 1992; Humphrey et al, 2000; Lopez, Lash, Schaffner, Shields, & Wagner, 2004; Wang, Odell, & Schwille, 2008; Whisnant, Elliott, & Pynchon, 2005) and special education (Billingsley, Griffin, Smith, Kamman, & Israel, 2009; Griffin, Winn, Otis-Wilborn, & Kilgore, 2003) literature. Here, we provide a working definition for comprehensive teacher induction, highlight its primary purposes and components, discuss the adequacy of the existing evidence base, and summarize conclusions and implications.

Comprehensive Teacher Induction: Defined

When examining the literature, the words "preservice and in-service teacher training," "induction," and "mentoring" appear frequently. Quite often the latter two terms are used interchangeably. There are, however, important distinctions among the terms that must be articulated initially. Induction programs, for example, were viewed as distinct *theoretically* from preservice and in-service preparation in that they did not provide additional training but rather offered support to new employees who had already been trained (Ingersoll & Smith, 2004). However, given that over 60% of new teachers reported being inadequately prepared when exiting their preparation programs, a fundamental problem may exist between theory and practice. School districts, for example, may have spent significantly more time "backfilling" for what was not taught in teacher education programs and as a result failed to attain their extensive goals. Induction was also conceived as a *broader developmental process* than teacher training, a process that served as a bridge from "student of teaching to teacher of students" (Ingersoll & Smith, 2004; p. 29).

Mentoring, on the other hand, was defined more narrowly as one-on-one assistance and support given by experienced professionals to novice educators (American Association of State Colleges and Universities [AASCU], 2006). The emphasis here was on the *personal guidance* that veteran teachers could provide for their novice colleagues (Ingersoll & Smith, 2004). Most often, mentoring was seen as one component of a more comprehensive approach to beginning teacher support and development (i.e., induction). Darling-Hammond & Sykes (2003) noted that although induction was often associated with mentoring, it also encompassed careful hiring procedures, protected initial assignments, steady provision of mentor and other support, and improved evaluation to help novices.

The term "comprehensive teacher induction" emerged from a national report by the Alliance for Excellent Education (2004). This document emphasized the broader and more comprehensive nature of induction and identified the following critical components: (a) high-quality mentoring, (b) shared planning time and collaboration, (c) ongoing professional development, (d) participation in an external network of teachers, and (e) standards-based evaluation. High-quality mentoring typically meant carefully screened and trained mentors selected from common disciplines who expressed interest in helping novice colleagues. Ample time to meet and plan instruction, observing one another's teaching, and ongoing administrative support were also seen as essential to successful induction programs (Arends & Rigazio-DiGilio, 2000; Barlin, 2010).

Additionally, AASCU recommended that professional development curricula and activities be geared to individual novice needs and that ongoing feedback be provided throughout the induction period (AASCU, 2006). In this paper, comprehensive teacher induction is defined as *those practices that help new and beginning teachers become competent and effective classroom professionals who also understand school and community cultures*. A distinction is made as well between *beginning* teachers who are in their first year of teaching and *new* teachers who have previous teaching experience but are in their first year in that school system. A tabular display of what does and does not constitute comprehensive teacher induction is provided in Table 1.

Table 1
Comprehensive Teacher Induction: What It Is and What It Is Not

Comprehensive teacher induction is...	Comprehensive teacher induction is not…
High-quality mentoring from trained mentor in common discipline	Part-time, informal guidance from untrained and often overextended colleague
Shared planning time and collaboration	No shared release time for collaboration and planning
Ongoing and targeted professional development	General professional development activities not linked to new teacher needs
Participation in an external network of teachers (i.e., professional learning community	Relative isolation of new teachers from their more experienced peers
Explicit administrative support	Unknown or unspecified support; giving difficult assignments (out-of-discipline subject matter, extracurricular duties, or multiple preparations)
Standards-based evaluation	Informal and/or summative, or no evaluation
Reduced course preparation and limited extracurricular activities	

Comprehensive Teacher Induction: Purposes, Components, and Benefits

Comprehensive teacher induction is a multifaceted and multipurpose process that can potentially benefit students, teachers (novice and experienced), administrators, policy makers, and the community. Table 2 highlights purposes, components, and potential benefits associated with comprehensive teacher induction programs. Administratively, induction is seen as a constructive policy response to problems of teacher turnover and the inadequate preparation of preservice teachers (Glazerman et al., 2008). Providing new and beginning teachers with the ongoing support and guidance of more experienced colleagues makes a lot of sense to practitioners and administrators alike. Experienced teachers have an opportunity to share their professional wisdom and expertise, and novices can learn the ropes from their more successful colleagues. Induction programs are also seen as ways to socialize new teachers into the profession, improve their teaching practice, reduce teaching-related stress and frustration, navigate unwritten district policies, and ultimately improve pupil learning. This represents a tall order for one reform effort, even when implemented under ideal conditions.

The literature suggests, however, that teacher induction programs are often implemented under less-than-ideal circumstances. AASCU (2006) reported that there is little consistency in induction programs across schools, districts,

Table 2
Primary Purposes, Components, and Potential Benefits Associated With Comprehensive Teacher Induction

Primary purposes of comprehensive teacher induction	Major program components	Potential benefits to mentors, mentees, school, and community
Improve teacher performance	Preorientation and orientation sessions that describe teaching assignments; curriculum and resources; historical and cultural information; community and regional culture; salary and benefits	**Benefits to students:** Improved teacher performance; higher academic achievement; improved continuity of instruction; enhanced class and school climate
Retain competent teachers in the profession	Systematic and sustained supports including formal mentor program; new and beginning teacher communication network; team planning/teaching; resource files; master teacher observations; study groups; resource personnel	**Benefits to new and beginning teachers:** Accelerated instructional success and effectiveness; greater self-confidence; enhanced job satisfaction; improved personal and professional well-being (e.g., reduced stress and frustration); increased opportunities for making connections with faculty, staff, and community; improved level of comfort and support
Promote the professional and personal well-being of new and beginning teachers	Targeted professional development with content most needed by new and beginning teachers. Activities might include workshops; formal course work with or without university involvement; online learning; committee work; staff meetings; research; curriculum development projects	**Benefits to mentors:** Development of leadership skills; increased professional growth and job satisfaction; enhanced collaboration skills; enhanced self-image; more opportunities to share instructional expertise
Build a foundation for continued professional growth	Explicit administrative support that might include protected initial teaching assignments (e.g., minimum preparation, teaching in areas of strength, no extracurricular assignments); formative and standards-based evaluation procedures	**Benefits to administrators:** Improved principal and teacher interactions and relationships; retention of presumably competent teachers
Transmit school and community culture	Orientation to district and school policies and procedures; participation in school-community events; membership on school climate committees	**Benefits to school and community:** Collegial communication network designed to facilitate interactions among experienced and new teachers; retention of competent teachers; increased student success; enhanced understanding of local community and culture
Socialize new teachers into the teaching profession		
Improve pupil learning	Staff development activities related to curriculum- and instruction-related practices; ongoing pupil progress monitoring; and structured, decision-making policies and practices	**Benefits to students:** Better academic and behavioral performance; greater access to advanced coursework; access to higher education and scholarship; improved life circumstances **Benefits to teachers, administrators, and community:** Direct evidence of teacher impact on pupil learning; better instructional decision making; well-educated community members

and states. Some programs are limited to one-on-one informal mentoring designed simply to help new teachers "survive" their first year, whereas others include carefully selected and trained mentors, systematic professional development, explicit instructional feedback, and formative evaluation procedures. Implementation efforts have also been hampered by a lack of ongoing administrative support, undertrained and overextended mentors, and inadequate and unstable funding patterns. Glazerman et al. (2008) noted that the most common arrangement was the pairing of new and experienced teachers without training, supplemental materials, or release time for induction. Potential benefits of any intervention are hampered in the presence of such implementation barriers.

One reason that school districts may not offer more support to new teachers is that comprehensive teacher induction is expensive (Alliance for Excellent Education, 2004; Villar & Strong, 2007). Induction programs were estimated to range from $1,660 to $6,605 per teacher per year. Moreover, there is not compelling evidence that investing more resources in comprehensive teacher induction will attract and retain more competent teachers than less expensive, informal mentoring alternatives. Finally, there are immeasurable costs associated with removing experienced teachers from their own classrooms to help others that may serve as disincentives. Villar and Strong (2007) conducted a systematic benefit-cost analysis and concluded that increases in teacher effectiveness, which presumably would result from comprehensive induction, could yield greater savings for school districts over the costs normally associated with teacher attrition. Collectively, the literature suggests that educators still lack commonly agreed-upon definitions for induction and mentoring. Those programs that do exist vary greatly in intensity and potential utility and as such cannot be viewed as common interventions or practices. Challenging economic times may further hamper efforts to move the field forward.

Despite definitional confusion, multiple and potentially competing purposes, and extreme program variability, induction and mentoring programs have increased dramatically. The number of new teachers who received some form of formal induction and mentoring expanded considerably over the past two decades (Smith & Ingersoll, 2004). During the 1990–1991 school year, 40% of beginning teachers said that they had participated in a formal teacher induction program. By 1993–1994, this participation rate increased to 51% of all new teachers in public schools, and by the 1999–2000 school year, the percentage of new and beginning teachers engaged in induction and mentoring programs reached 79%. Given such growth rates, one would predict that at least 90% of all new and beginning teachers in our public schools are currently involved in some form of induction program. This is important because induction and mentoring programs have been institutionalized to some extent in schools, and they may provide a necessary infrastructure for addressing the formidable educational challenges delineated by Secretary Duncan (2009). If these programs can be implemented effectively and efficiently, then they may provide a viable mechanism for improving teacher practice and student learning.

Evidence of Effectiveness

The obvious question confronting education professionals is, do comprehensive teacher induction programs work? That is, do they increase teacher retention, facilitate socialization into the profession, improve new teachers' practice, and ultimately enhance pupil learning? The literature (e.g., Fletcher et al., 2008; Ingersoll & Kralik, 2004) suggests generally that comprehensive teacher induction does improve teacher retention under certain conditions (e.g., adequate administrative support, use of well-selected and well-trained mentors from common disciplines, and sufficient opportunities for novice educators to participate in instructional decision making), and that most participants are satisfied with their induction-related experiences. Unfortunately, the literature is much less clear about induction's impact on teaching practice and pupil learning. These topics have received considerably less attention, and outcomes have been modest at best. Unambiguous interpretations of the literature are hampered, as well, by an overall lack of methodological rigor (Ehrich et al., 2004; Humphrey et al., 2000; Johnson, Berg, & Donaldson, 2005). Table 3 highlights some primary limitations associated with induction research.

Table 3
Primary Limitations in Existing Research on Comprehensive Teacher Induction

1. Use of research designs that cannot establish functional relationships (i.e., qualitative versus quantitative methodologies).
2. Overemphasis on retention and satisfaction outcomes and underemphasis on teaching practice and pupil learning.
3. Almost exclusive use of indirect (e.g., opinion and attitude surveys) rather than direct outcome measures (e.g., observation and achievement measures).
4. Failure to quantify independent variables (i.e., nature of mentoring content and pedagogy) and to measure fidelity of implementation.
5. Selection bias and lack of internal controls in evaluative and quantitative studies.
6. Typical lack of direct observational measures of novices' teaching practice and/or formative measures of pupil performance.

According to research reviews (Ingersoll & Kralik, 2004; Lopez et al., 2004; Totterdell, Woodroffe, Bubb, & Hanrahan, 2004), most teacher induction studies were inconclusive and/or lacked appropriate rigor. To begin, most studies were qualitative rather than quantitative and, therefore, could not answer the types of cause-and-effect questions raised regarding induction efficacy. Moreover, induction researchers have relied heavily on indirect (e.g., attitude and other self-report methods) rather than direct measures of teacher and pupil performance (e.g., direct observation and student achievement). Ehrich et al. (2004), for example, noted that the literature was dominated by professional testimonials and personal opinions. Positive study outcomes included receiving empathy, getting good ideas for teaching, discussing strategies with peers, and getting feedback on one's teaching. Veteran and novice teachers generally liked their experiences and felt that they were helpful for professional growth.

Johnson et al. (2005) reported further that most quantitative induction studies were also limited by *selection bias* and a *lack of control groups*. Since schools that had induction programs were more likely to support teachers in other ways, they would also be more likely to retain them even without induction programs (i.e., selection bias). The failure to include control groups precluded researchers from ruling out typical professional growth as a contributor to subsequent induction outcomes. Conventional wisdom suggests that teachers normally improve their practice during the first 3 to 5 years of teaching (Lopez et al., 2004). Whether or not comprehensive induction programs can accelerate this growth cannot be answered without more rigorous research methods.

The limitations in the induction literature are most troubling with regard to impact on teaching practice and pupil learning. Humphrey et al. (2000) noted that student achievement was the *least* studied outcome in induction research. What may arguably be the most important outcome for professionals and parents (i.e., improvement in student learning) appeared to be the least studied induction outcome. Equally distressing was researchers' failure to view teaching practice as a necessary mediating variable in pupil learning. It was rare, indeed, for researchers to measure directly novice teachers' instructional practice before, during, and after induction training. In fact, teaching practice was an underrepresented dependent variable in most induction research. This generalized failure to measure teaching practice makes it virtually impossible to determine induction's impact on pupil learning (Goe & Coggshall, 2007).

To address inadequacies in the induction literature, the Institute of Education Sciences (IES) funded a 3-year, randomized control trial (Glazerman et al., 2008; Isenberg et al., 2009) to examine the impact of traditional versus comprehensive teacher induction programs on five dependent measures: (a) teaching practice, (b) student achievement, (c) teacher retention, (d) process-related variables (e.g., amount of time working collaboratively), and (e) composition of district work forces. Research-related questions, procedures, and outcomes are summarized in Table 4. The study involved 17 school districts,

Table 4

Research questions, designs, and outcomes associated with three-year, IES-funded, randomized control trial study

Research Questions	Participants and Settings	Independent Variable	Dependent Variables	Outcomes
What is the effect of comprehensive teacher induction on the types and intensity of induction services teachers receive compared to the services they receive from the districts' current induction programs?	17 school districts across 13 states; All districts had at least 50% of pupils who qualified for free and/or reduced meals;418 schools: 100 treatment and 103 control school in 9 ETC districts;110 treatment and 105 control schools in NTC districts; Classroom practices data collected on 698, K–6 teachers in self-contained classes; Formal test scores collected for 261 language arts and 281 math teachers	Comprehensive teacher induction included: Carefully selected and trained mentors; On average, 1 mentor worked with 12 novice teachers; curriculum of intensive and structured teacher supports; Direct observations in mentor classes; Formative assessment tools for novice and experienced teachers' Outreach for district leaders; explicit focus on instruction	Participant demographic variables; Direct observations of novice practice using the Vermont Classroom Observation Tool (VCOT); Student achievement on district-adopted measures; Teacher mobility surveys	Noticeable impact on induction-related process variables (e.g., treatment teachers reported receiving more mentoring than control teachers; treatment teachers participated in more specific induction activities than control teachers; treatment teachers spent more time in certain professional activities (e.g., keeping written journals, working with study groups, and observed others' teaching more often) during year 1; fewer significant effects reported during year 2;
What are the impacts on teachers' classroom practices?				No impacts on teaching practices during 1st and 2nd year
What are the impacts on student achievement?				No positive impacts on student achievement during 1st or 2nd year
What are the impacts on teacher retention?				No impacts on teacher retention after 1st and 2nd years
What is the impact on the composition of the district's teaching workforce?				No positive impacts on districts teacher composition after 1st and 2nd year

serving primarily low-income students, across 13 states. All the districts had at least 50% of pupils who qualified for free and/or reduced-cost meals. Roughly half of all teachers in each district received "typical" or existing induction services (i.e., control group) while the other half received comprehensive induction services (i.e., experimental group) that were developed by the Educational Testing Service (ETS) or New Teacher Center (NTC).

Methodologically noteworthy in the IES study were the explicit descriptions of induction services and the direct measurement of fidelity of implementation. Findings from the first 2 years were disappointing. Although some noticeable impact was reported on process-related variables (e.g., frequency of mentor-mentee contacts), no effects were found on teaching practice, student achievement, teacher retention, and/or composition of district work forces. A more recent *Education Week* article (Sawchuk, 2010) reported, however, that a third-year IES evaluation showed modest improvements in pupil achievement as a result of comprehensive teacher induction.

Summary and Conclusions

Collectively, what do educators know about comprehensive teacher induction? Induction and mentoring programs can be found in most public schools across the country, yet there is little consistency in what they look like from place to place. These programs are quite costly but may actually save money for school districts in the long run (Villar & Strong, 2008). Some evidence suggests that comprehensive induction increases teacher retention and that participants are typically satisfied with the training and support they receive. On the other hand, there is insufficient evidence to suggest that induction programs improve teaching practice and even less to show that student learning is improved. Conventional wisdom suggests that comprehensive induction programs will persist in our schools in one form or another. The fact that they have not impacted teaching practice and pupil learning should stimulate educators' collective efforts to demonstrate how they can do so.

Unfortunately, more remains unknown than known about comprehensive teacher induction. How do educators answer the basic question, do induction programs work? If they qualify their answers to include only retention and participant satisfaction, then they might respond affirmatively. However, if they examine issues of practice and learning, then their response is less clear or confident. We must ask, as well, is increased retention in itself a sufficient outcome or should retention be linked to improved pupil learning? In the absence of efficacy data, will schools end up retaining ineffective teachers? Should experienced teachers continue to "mentor" novice colleagues even if their efforts do not impact practice and pupil learning? Are the costs associated with removing highly effective veteran teachers from the classroom worth the benefits of improved retention and satisfaction? Would better induction programs improve practice and student learning? If so, what should be included in these better induction programs? If not, should new teachers still be supported?

If comprehensive teacher induction provides a viable infrastructure for improving instructional practice and pupil learning, then what knowledge and skill bases are most relevant and how should they be imparted to novice teachers? Do different types of new teachers (e.g., primary, intermediate, and secondary; general and special education; and traditional and alternative certification) need different knowledge and skills? Or do all new teachers need a common knowledge and skill base to improve pupil learning? With regard to mentoring, do structures exist in districts to identify mentors who are unusually effective in improving pupil learning? How can highly effective teachers be convinced to leave their classrooms, and will there be any instructional costs to pupils? Can replacing highly effective teachers with unproven instructors be justified?

Empirically, would better research methods improve induction outcomes? Can more direct measures of practice and pupil learning be used effectively and efficiently in induction research and practice? What roles, if any, can single-case research designs play in documenting induction outcomes? Finally, for those in teacher education, what roles should preparation programs play in comprehensive teacher induction? Can teacher educators and P–12 personnel share induction roles and responsibilities, and, if so, in what ways? Can formal induction-related experiences begin earlier in preservice preparation, and what would that look like? Can preservice teachers learn to assess and adjust their own practices in response to ongoing measures of pupil performance?

THE SOLUTION

As noted, there are many unanswered questions in the induction literature. Here, we argue that comprehensive induction programs can have a more visible impact on practice and learning if major changes occur in how these programs are conceptualized, implemented, and evaluated. Greenwood and Maheady (1997) offered three plausible explanations for educators' inability to noticeably improve practice and learning: (a) failure to use existing technologies to measure *changes* in pupil learning, (b) inability to use research methods that were linked *directly* to student learning, and (c) an overreliance on advocacy rather than research to guide educational reform efforts. We reiterate the importance of these ideas and suggest that they undergird future efforts to refocus induction programs in schools. Here, we offer six basic guidelines for improving induction research and practice (Table 5) and provide three examples of

Table 5
Primary Limitations in Existing Research on Comprehensive Teacher Induction

1. Make teaching practice and pupil learning the overarching goals of comprehensive teacher induction programs, and measure them directly.
2. Reconceptualize teacher induction as an ongoing performance feedback system for all education professionals.
3. Align content and processes in induction around evidence-based knowledge and skills.
4. Use more rigorous research methods.
5. Use comprehensive induction programs as vehicles for bridging the research-to-practice gap in education.
6. Link teacher education programs and P–12 schools in the collaborative design, implementation, and evaluation of comprehensive induction.

how teacher educators, researchers, and school personnel have worked collaboratively to improve teacher practice and pupil learning.

Guidelines for Improving Research and Practice

Perhaps the most fundamental change is to make improved pupil outcomes — academic and behavioral — the overarching goal of induction efforts. Induction works only if and when pupil performance improves as a function of comprehensive induction services. To date, participant satisfaction and retention have taken precedence over pupil learning and improved teaching practice. These priorities must be reversed in future research and practice. While retention and satisfaction are important outcomes, their utility is linked directly to whether or not children benefit from their teachers' instruction. Retaining ineffective and satisfied teachers is not an acceptable outcome; nor is retaining effective teachers who are dissatisfied with existing working conditions. The highest priority, therefore, must be to retain well-satisfied teachers who are unusually effective in promoting pupil learning. Elevating better pupil outcomes to the forefront of induction research and practice will also require the development and identification of more and better progress-monitoring systems and empirically supported teaching practices.

It would also be useful to reconceptualize comprehensive teacher induction as one component of a larger professional development system in which *all* educators receive ongoing performance feedback and support for improving pupil outcomes. While novice teachers may need additional monitoring and support, such assistance can be provided in a system that recognizes and rewards success and addresses instructional challenges in a proactive and constructive manner. While the exact nature and structure of such systems are not complete, noteworthy exemplars can be found in mental health (Chorpita, 2008; Fixsen, Blasé, Duda, Naoom, & Van Dyke, 2008), positive behavioral support (Sugai & Horner, 2008), and school psychology (Tilly, 2009) literature. In effect, a roadmap to evidence-based education in schools must be created (Detrich, Keyworth, & States, 2008). The fundamental purpose of induction (i.e., improving pupil learning), therefore, will be aligned with broader, systemwide policies and practices to support all personnel for improving student learning. If pupil learning drives instructional decision making, then comprehensive induction services must contribute positively to the schoolwide agenda to make better educational decisions about children, particularly the most fragile learners.

A third important guideline is to align induction content and pedagogy with empirically supported practices. There are glaring omissions throughout the induction literature regarding what was taught to novice teachers and how such instruction was provided. When content was described it was typically in generalities such as "classroom management," "assessment and instruction," "inquiry-based approaches to learning," and "school-related policies and procedures." Pedagogy was described similarly as "informal information sharing sessions," "weekly meetings," "mentor observations," and/or "written teaching summaries." There was no mention that mentors modeled and/or provided systematic feedback on novice teachers' use of empirically supported practices. One major problem for researchers and practitioners is that the literature provides very little guidance about what content to include in induction programs and how to transform this knowledge into teaching practice. The good news is that some educational practices are more effective than others and that, whenever possible, these practices should be used over those without comparable evidence. Indeed, scientifically based practices are mandated by federal legislation (e.g., No Child Left Behind; Individuals with Disabilities Educational Improvement Act) and serve a consumer protection function for educators (Detrich, 2008).

One important criterion for induction content might be the following: *Curricular programs and instructional practices that are used in induction programs should have empirical support*. Several analyses (Holdheide & Reschly, 2008; Oliver & Reschly, 2007; Smartt & Reschly, 2007) identified a number of empirically supported practices in reading, mathematics, and classroom organization and management and examined their relative use in teacher education programs. These initial assessments were quite sobering and suggested that many, perhaps most, teacher education programs were *not* promoting the use of empirically supported practices among new teachers. Teacher educators' failure to promote practices that benefit children are clearly reflected in Secretary

Duncan's comments about new teachers' ill-preparedness for contemporary classrooms.

Equally important as *what* is taught in induction programs is *how* that content and skill base are delivered. The good news once again is that more is known about how to change teaching practice than is employed in school-based professional development. Educators know, for example, that lecture-based, inservice training does little, if anything, to change practice. In contrast, in-class assistance in the form of modeling, coaching, and performance-based feedback does help teachers to improve their instruction (e.g., Buysse & Wesley, 2006: Joyce & Showers, 2002). Odom (2008) also described a variety of hot topics at the forefront of contemporary professional development. These topics included practice-based reviews of evidence, implementation science, and the use of enlightened professional development activities (e.g., peer coaching, web-based video and visual access, and communities of practice) to improve teaching practice. Combining empirically supported content with scientifically validated professional development strategies provides a potentially constructive framework for changing practice at the classroom, school, and system levels.

It is also obvious that induction practice will not improve much until the quality of research that undergirds its use improves as well. Currently, the literature is dominated by qualitative and quantitative studies that lack rigor and do not address directly or adequately the issues of practice and pupil learning. Even the most rigorous, experimental effort to date (IES-funded, randomized control trials) has not produced meaningful outcomes for policy makers, researchers, or practitioners.

A fourth guideline, therefore, is to use more rigorous research methodologies, preferably those that can be adapted to local, consumer-driven interests and needs and can provide meaningful opportunities for replication and wide-scale dissemination. Single-case research designs provide one powerful way for practitioners to demonstrate the effects of explicit teaching practices on educationally important and reliably measured instructional outcomes (Kennedy, 2005). These designs require that induction strategies and outcomes be defined operationally and measured for fidelity of implementation and reliability of outcomes. Pupil performance is assessed across adjacent phases where interventions are present or absent, and determinations are made about the success or failure of different teaching practices. The value in single-case research lies in its sensitivity to behavioral change, the rigor of its measurement systems, and its flexibility for application at the student, classroom, school, or system levels (Kennedy, 2005).

The use of single-case research designs may also help to bridge the gap between research and practice in education. Single-case designs allow teachers to study issues of practice at the child and classroom levels and permits administrators to examine similar issues at the school and district levels. Some induction programs also require novice teachers to engage in formal professional development activities that document their abilities to improve pupil learning. Single-case research designs would be particularly useful for meeting such professional requirements and producing a useful database on effective and ineffective practices in local schools. Indeed, some states are already mandating that teacher reappointment and tenure be linked to improved pupil performance.

The final guideline is that teacher preparation programs should work collaboratively with P–12 schools in the creation of new, data-based decision-making cultures in the schools. To do so, teacher educators must become more involved and responsive to the needs of public schools. A first step in that direction may have been the searing indictment of existing practice by top administrative officials (Duncan, 2009). A second step was reflected in comments from the president of the National Council for the Accreditation of Teacher Education. Cibulka (2009) said that teacher educators must create seamless transitions between preservice and in-service education, wrap their university coursework around P–12 educational needs, and substantially increase future teachers' clinical experiences. Those additional teaching opportunities were to be (a) intensive, (b) provided in our neediest schools, and (c) accompanied by data collection efforts that showed their impact on pupil learning.

Evidence of Effectiveness

We have been engaged to varying degrees in induction-related activities for almost 20 years now. Both of us have taught methods courses at the undergraduate and graduate levels, provided professional development to elementary and secondary teachers, and conducted research on the impact of empirically supported practices on pupil learning and behavior. Our audience has been primarily general education teachers, many of whom were in their first years of teaching. Their challenges, similar to those of most teachers, included accommodating the wide range of skill levels in their classes, increasing student productivity and accuracy, and solving a myriad of behavioral and interpersonal conflicts every day. They all had mentors, some of whom were more helpful than others. They received generic professional development, often delivered in workshops, and were required to earn a master's degree within 5 years of initial program completion.

Here, we describe three partnership projects that examined novice teachers' abilities to use empirically supported practices in real-life settings and to collect data on the effects of their instruction on pupil performance. These projects are offered as exemplars of the kind of collaboration needed to improve induction research and practice.

Preservice teachers' use of empirically supported practices

This project involved large groups of freshmen and sophomores who were completing their first formal field experiences in an inclusive general education program (Maheady, Jabot, Rey, & Michielli-Pendl, 2007). As part of course requirements, preservice teachers taught two formal lessons, collected pre- and postteaching data, graphed those data to reflect entire class, small group, and individual pupil performance, and then made written data-based instructional decisions. As part of the project, preservice teachers were also required to use one of six empirically supported practices (response cards, Numbered Heads Together, guided notes, graphic organizers, 3-step interview, and think-pair-share) and to collect data on the fidelity with which the selected practices were implemented. All student teachers were assigned to 10-week placements in pairs. They shared instructional planning, implementation, and evaluation responsibilities and were required to formally collect fidelity and outcome data. A total of 422 preservice teachers, 78% of whom were placed in high-needs schools, provided almost 17,000 hours of in-class assistance over four semesters. They taught more than 800 lessons and used empirically supported practices with a high degree of accuracy (M = 92%; range = 88% to 96%).

Pupil outcome data indicated that students made noticeable or marginal improvements in over 85% of preservice teachers' sampled lessons. Outcomes were determined on the basis of pupil improvements on pre- and postteaching assessments. Social validity data indicated that preservice teachers found their early teaching opportunities to be very important and useful. They also rated all project requirements as acceptable and reported high levels of satisfaction with program outcomes. This project was noteworthy because it was a collaborative arrangement between teacher educators and P–12 schools that produced mutual benefits for preservice and classroom teachers as well as the students they served. For teacher educators, it also provided an opportunity to measure *directly* novice teacher practice and its impact on pupil learning.

Preparing Student Teachers To Use Classwide Peer Tutoring

The second project (Maheady, Harper, Mallette, & Karnes, 2004) involved 10 preservice general educators who volunteered to use classwide peer tutoring (CWPT) (Delquadri, Greenwood, Whorton, Carta, & Hall, 1986) during their final student teaching experience. All student teachers were trained to implement CWPT with a high degree of accuracy using both on-campus and in-class assistance. Overall, it took about 2 hours of initial training, including 1 hour of in-class assistance (i.e., modeling, performance feedback, and coaching) to help preservice teachers reach a preestablished fidelity criterion. While they used CWPT, pupils' weekly spelling scores averaged 94%

(pretest M = 69%; range 52% to 89%) and only 8 out of 1,028 spelling tests administered resulted in failing grades. Further analyses revealed that preservice teachers implemented CWPT with a high degree of accuracy (M = 88%; range = 82% to 96%), but that they made some procedural adaptations at their classroom teachers' request. These procedural adaptations resulted in smaller achievement gains and less pupil satisfaction, an outcome that suggests caution when adapting empirically supported practices. This project provided a second example of how teacher educators might impact preservice teachers' practice and simultaneously examine the effects on pupil learning. The study also highlighted the importance of procedural adaptations and their potential impact on pupil performance and satisfaction.

Graduate Research-To-Practice Studies

The third partnership project was a graduate-level requirement for all teachers completing their master's degree in curriculum and instruction. The graduate program had a required 9-hour research sequence designed to help practitioners understand, design, and implement applied educational research. During the second course, novice teachers designed a single-case research study using guidelines articulated by Horner and colleagues (2005). They then carried out the project in either their own or other teachers' classrooms during the third course in the sequence. All research-to-practice studies included (a) identification of educationally and/or socially important problems; (b) brief and illustrative literature reviews; (c) operational definitions of target behaviors; (d) direct, frequent, and reliable measurement of target behaviors; (e) selection of empirically supported practices and direct measurement of intervention fidelity; (f) use of rigorous research designs (e.g., A-B-A-B, multiple baseline, and alternating treatments); and (g) assessment of social acceptability of intervention goals, procedures, and outcomes.

Here, we provide one example of a recently completed research-to-practice study. This particular study was completed by a first-year teacher working in a large urban setting in northeastern Ohio (Hiller, Maheady, & Jabot, 2010). The investigator taught a combined fifth- and sixthth-grade class with 18 students with a wide range of reading (second- to ninth-grade levels) and math (second- to seventh-grade levels) skills. In addition, many pupils had documented behavior problems and poor homework completion rates. When they did complete homework, their performance was below average. The investigator developed an intervention called the "mystery motivator game," which consisted of three primary components: (a) interdependent and dependent group contingencies, (b) spinners, and (c) unknown rewards in the form of mystery motivators (Rhode, Jenson, & Reavis, 1996).

First, the students were told that they were going to play a game designed to improve their homework performance. To win, they had to meet two criteria: (a) 100% homework completion and (b) 85% homework accuracy. If all students turned in completed assignments on time (100% completion), then the teacher would randomly pick a number from 1 to 18 from an opaque jar to determine whose paper would be checked *privately* to see if the second criterion (85% accuracy) was met. If the privately scored paper was 85% correct or better, then the entire class would be allowed to twirl a spinner to determine the type of reward. Spinners contained five pie-shaped wedges of differing widths, with higher preference rewards corresponding to *narrower* pie slices. The narrowest pie slice bore a question mark. On days when the spinner landed on the question mark, students were allowed to pick one of 15 mystery motivator envelopes hanging from the ceiling. Each decorated envelope contained slips of paper specifying the rewards (e.g., free time, dress-down days, and lunch in the room). Possible rewards were generated earlier by students through a suggestion box placed in the classroom. If the class or randomly selected pupils failed to meet either criterion, then they were encouraged to try harder the next day. The names of students whose papers were reviewed were never revealed.

The investigators used an A-B-A-B design and showed that the mystery motivator game produced immediate and noticeable improvements in pupils' homework completion and accuracy rates (Figure 1).

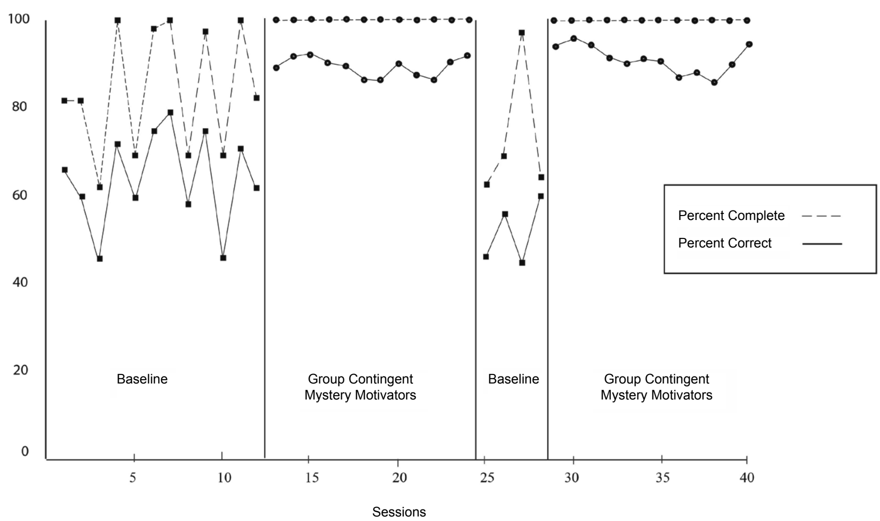


Figure 1. Mean completion and accuracy rates for fifth- and sixth-grade inclusion class across experimental phases.

During initial baseline, about 85% of the class turned in daily homework assignments (range = 42% to 100%). Homework accuracy, however, averaged only 64% (range = 46% to 79%). When the mystery motivator game was put into effect, both completion and accuracy rates improved immediately and noticeably. In fact, all students (100%) turned in *every* homework assignment during both intervention phases, and the class averages were 89% and 91% for each intervention phase. When the intervention was removed briefly, students reverted to inconsistent homework completion (M =75%), and their accuracy rates fell to an average of 52% (range = 45% to 65%). It is significant to note, as well, that there were no overlapping data points across any experimental phase.

During the past 3 years, over 40 research-to-practice studies were completed by general education teachers in their own and other classrooms. The number and types of studies are depicted in Table 6. These studies, using a variety of empirically supported practices in general education classes, produced consistent improvements in pupils' academic and/or behavioral performance. Most studies replicated findings from other researchers and showed that selected practices were also effective under typical teaching conditions. Novice teachers usually selected practices that could be used on a classwide basis and focused on increased student productivity and accuracy, active participation in class, and/or reducing common disruptive behaviors (e.g., talk-outs, out-of-seat, and noncompliance). Obviously, many studies were limited by short duration, lack of generalization and maintenance data, and occasionally fewer reliability and fidelity assessments. Their impact on pupil performance, however, was consistently positive, and the procedures and outcomes were well accepted by novice teachers and their public school colleagues.

SUMMARY AND IMPLICATIONS

The state of the art in comprehensive teacher induction is not pretty, at least not in terms of its documented impact on teacher practice and, more important, on student learning. This does not mean that comprehensive teacher induction cannot impact practice and learning. Rather, it suggests that a more concerted effort must be made to do so. Teacher practice and pupil learning can no longer remain a secondary variable of interest for practitioners and researchers. Indeed, impact on pupil learning should be viewed as the "gold standard" for determining if induction or any professional development programs are working (Greenwood & Maheady, 1997). Similarly, comprehensive induction programs must be reconceptualized as one component of a larger data-based culture dedicated to the improvement of all teaching practice. Much more thought and effort must also go into what is taught in professional development and how teachers, novice and veteran, can put newly acquired knowledge and skills into practice.

Table 6

Number of Research-to-Practice Studies Completed During Academic Years 2007–2010

	Intervention Strategies					
	CWPT	**Response cards**	**Self-monitor-ing**	**Group contingent mystery motivators**	**Numbered Heads Together**	**Other**
Early childhood		1				1
Childhood (Grades)						
K–2				2		1
3–6	6	2		6	1	1
Adolescence (Grades)						
7–12 Math	1	1	2			2
7–12 Science	3		2	1		
7-12 Social studies	1	1	2	2		2
Totals	**11**	**5**	**6**	**11**	**1**	**7**

In effect, a roadmap for building an evidence-based culture is needed (Detrich, et al., 2008). Finally, teacher educators and applied researchers must work collaboratively with P–12 schools to identify common educational problems and to develop effective, efficient, and socially acceptable strategies for preventing and/or ameliorating these instructional challenges. Given the increasing role of science in education, the rise of evidence-based federal policies, and the urgent need to improve educational outcomes in our country, there is no better time for such revolutionary changes to occur. Comprehensive teacher induction provides one vehicle for making such sweeping changes a reality.

REFERENCES

Abell Foundation. (2001, October). *Teacher certification reconsidered: Stumbling for quality.* Baltimore: Author. Retrieved from http://www.abell.org

Alliance for Excellent Education. (2004). *Tapping the potential: Retaining and developing high-quality new teachers.* Washington, DC: Author.

American Association of State Colleges and Universities. (2006). Teacher induction programs: Trends and opportunities. *Policy Matters, 3*(10), 1–4.

Arends, R. I., & Rigazio-DiGilio, A. J. (2000, July). *Beginning teacher induction: Research and examples of contemporary practice*. Paper presented at the annual meeting of the Japan–United States Teacher Education Consortium.

Ballou, D., & Podgursky, M. (2000). Reforming teacher preparation and licensing: What is the evidence? *Teachers College Record, 102*(1), 5–27.

Barlin, D. (2010, March). Better mentors, better teachers: Three factors that help to ensure successful programs. *Education Week, 29*(27). Retrieved from http://www.edweek.org/archive/ew/articles/2010/03/23/27barlin.html

Billingsley, B. S., Griffin, C. C., Smith, S. J., Kamman, M., & Israel, M. (2009). *A review of teacher induction in special education: Research, practice, and technology solutions.* (NCIPP Doc. No. RS-1). University of Florida, National Center to Inform Policy and Practice in Special Education. Retrieved from http://ncipp.org/reports/rs_1.pdf

Buysse, V., & Wesley, P. W. (2006). Evidence-based practice: How did it emerge and what does it mean for the early childhood field? In V. Buysse & P. W. Wesley (Eds.), *Evidence-based practice in the early childhood field* (pp. 1–34). Washington, DC: Zero to Three.

Carnine, D. (2000, April). *Why education experts resist effective practice: Report of the Thomas B. Fordham Foundation*. Washington, DC: The Thomas B. Fordham Foundation.

Chorpita, B. F. (2008, April). *Evolution of the revolution: How can evidence-based practice work in the real world?* PowerPoint presentation at the Third Annual Summit on Evidence-Based Education. Berkeley, CA: The Wing Institute.

Cibulka, J. G. (2009, June). *Meeting urgent national needs in P–12 education: Improving relevance, evidence, and performance in teacher preparation.* Washington, DC: National Council for Accreditation of Teacher Education

Coalition for Evidence-Based Educational Policy. (2002). Rigorous evidence: The key to progress in education. Retrieved from http//www.excelgov.org/displayContent.asp?Keyword=prppcEvidence

Darling-Hammond, L.. & Sykes, G. (2003). Wanted: A national teacher supply policy for education: The right way to meet the “highly qualified teacher” challenge. *Education Policy Analysis Archives, 11*(33). Retrieved from http://epaa.asu.edu/epaa/v11n33/

Delquadri, J. C., Greenwood, C. R., Whorton, D., Carta, J. J., & Hall, R. V. (1986). Classwide peer tutoring. *Exceptional Children, 52,* 535–542.

Detrich, R. (2008, November). *The ethical and legal basis for evidence-based education: Implications for the profession.* PowerPoint presentation at the annual conference of the Teacher Education Division of the Council for Exceptional Children, Dallas, TX.

Detrich, R., Keyworth, R., & States, J. (2008). A roadmap to evidence-based education: Building an evidence-based culture. In R. Detrich, R. Keyworth, & J. States (Eds.), *Advances in evidence-based education* (pp. 3–18). Oakland, CA: The Wing Institute.

Duncan, A. (2009, October). *A call to teaching!* An address by the Secretary of Education at the Rotunda at the University of Virginia, Charlottesville.

Ehrich, L. C., Hansford, B., & Tennent, L. (2004). Formal mentoring programs in education and other professions: A review of the literature. *Educational Administration Quarterly, 40*(4), 518–540.

Feiman-Nemser, S., Schwille, S., Carver, C., & Yusko, B. (1999, July). *A conceptual review of literature on new teacher induction*. Paper presented at the meeting of the National Partnership for Excellence and Accountability in Teaching, Washington, DC.

Fixsen, D. L., Blasé, K. A., Duda, M. A., Naoom, S. F., & Van Dyke, M. (2008, April). *Sustainability: The first thing. The only thing?* PowerPoint presentation at the Third Annual Summit on Evidence-Based Education. Berkeley, CA: The Wing Institute.

Fletcher, S., Strong, M., & Villar, A. (2008). An investigation of the effects of variations in mentor-based induction on the performance of students in California. *Teachers College Record, 110*(10), 2271–2289.

Glazerman, S., Dolfin, S., Bleeker, M., Johnson, A., Isenberg, E., Lugo-Gil, J.,… Ali, M. (2008). *Impacts of comprehensive teacher induction: Results from the first year of a randomized controlled study* (NCEE 2009-4034). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Goe, L., & Coggshall, J. (2007, May). *The teacher preparation > teacher practices > student outcomes relationship in special education: Missing links and necessary connections.* Washington, DC: National Comprehensive Center for Teacher Quality.

Gold, Y. (1996). Beginning teacher support: Attrition, mentoring, and induction. In J. Sikula, T. Buttery, & E. Guyton (Eds.), *Handbook for research on teacher education* (2nd ed., pp. 548–594). New York: Macmillan.

Greenwood, C. R., & Maheady, L. (1997). Measurable change in student performance: Forgotten standard in teacher preparation? *Teacher Education and Special Education, 20,* 265–275.

Griffin, C. C., Winn, J. A., Otis-Wilborn, A., & Kilgore, K. L. (2003). *New teacher induction in special education: Review of the literature*. Retrieved from http://www.coe.ufl.edu/copsse/docs/RS-5/1/RS-5.pdf

Guarino, C. M., Santibanez, L., & Daley, G. A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Educational Research*, *76*(2), 173–208.

Hiller, L., Maheady, L., & Jabot, M. (2010). *The effects of group contingent mystery motivators and spinners on the homework completion and accuracy of a 5th- and 6th-grade inclusion class*. Unpublished master's project, Department of Curriculum and Instruction, SUNY Fredonia, Fredonia, New York.

Holdheide, L. R., & Reschly, D. J. (2008, June). *Teacher preparation to deliver inclusive services to students with disabilities*. Washington, DC: National Comprehensive Center for Teacher Quality.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S. L., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practices in special education. *Exceptional Children, 71*(2)*,* 165–179.

Howe, E. R. (2006). Exemplary teacher induction: An international review. *Educational Philosophy and Theory, 38*(3), 287–297.

Huling-Austin, L. (1992). Research on learning to teach: Implications for teacher induction and mentoring programs. *Journal of Teacher Education, 24* (3), 173–180.

Humphrey, D. C., Adelman, N., Esch, C., Riehl, L. M., Shields, P. M., & Tiffany, J. (2000). *Preparing and supporting new teachers: A literature review*. Menlo Park, CA: SRI International. Washington, DC: Institute of Education Sciences.

Ingersoll, R. M., & Kralik, J. M. (2004). *The impact of mentoring on teacher retention: What the research says* (pp. 1-23). Denver, CO: Education Commission of the States.

Ingersoll, R. M., & Smith, T. M. (2004). Do teacher induction and mentoring matter? *NASSP Bulletin*, *88*(638), 28–40.

Isenberg, E., Glazerman, S., Bleeker, M., Johnson, A., Lugo-Gil, J., Grider, M., …Ali, M. (2009). *Impacts of comprehensive teacher induction: Results from the second year of a randomized control trial* (NCEE 2009-4072).Washington, DC: U.S. Department of Education.

Johnson, S. M., Berg, J. H., & Donaldson, M. L. (2005, February). *Who stays in teaching and why: A review of the literature on teacher retention*. Harvard Graduate School of Education: The Project on the Next Generation of Teachers.

Joyce, B. R.., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Kennedy, C. H. (2005). *Single-case designs for educational research.* Boston: Pearson/Allyn & Bacon.

Keyworth, R. (2010, April). *Education at the crossroads: The state of teacher preparation*. PowerPoint presentation at the Fifth Annual Summit on Evidence-Based Education. Berkeley, CA: The Wing Institute.

Lopez, A., Lash, A., Schaffner, M., Shields, P., & Wagner, M. (2004). *Review of research on the impact of beginning teacher induction on teacher quality and retention*. Menlo Park, CA: SRI International.

Maheady, L., Harper, G. F., Mallette, B., & Karnes, M. (2004). Preparing preservice teachers to implement class wide peer tutoring. *Teacher Education and Special Education*, *27*(4), 408–418.

Maheady, L., Jabot, M., Rey, J., & Michelli-Pendl, J. (2007). An early field-based experience and its impact on pre-service candidates' teaching practice and their pupils' outcomes. *Teacher Education and Special Education 30*(1)*,* 24–33.

Odom, S. L. (2008). The tie that binds: Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education, 29*(1)*,* 53–61.

Oliver, R. M., & Reschly, D. J. (2007, December). *Effective classroom management: Teacher preparation and professional development.* Washington, DC: National Comprehensive Center for Teacher Quality.

Rhode, G., Jenson, W. R., & Reavis, H. K. (1996). *The tough kid book: Practical classroom management strategies.* Longmont, CO: Sopris West.

Sawchuk, S. (2010) (2010, June 28). Teacher induction found to raise student scores. *Education Week.* Retrieved from www.edweek.org/ew/articles/2010/06/28/36induction.h29.html

Smartt, S. M., & Reschly, D. J. (2007, June). *Barriers to the preparation of highly qualified teachers in reading.* Washington, DC: National Comprehensive Center for Teacher Quality.

Smith, T., & Ingersoll, R. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal, 41*(3), 681–714.

States, J. (2010, April). *Education at the crossroads: Data mining and teacher preparation.* PowerPoint presentation at the Fifth Annual Summit on Evidence-Based Education. Berkeley, CA: The Wing Institute.

Strong, M. (2005). Teacher induction, mentoring, and retention: A summary of the research. *New Educator, 1*(3), 181–198.

Sugai, G., & Horner, R. (2008, April). *Sustainability & scaling & failure of Friday in-service day*. Powerpoint presentation at the 3rd Annual Summit on Evidence-Based Education. Berkeley, CA: The Wing Institute.

Tilly, W. D., III. (2009). *Big ideas in data-driven decision making at a system level.* PowerPoint presentation at the Third Annual Summit on Evidence-Based Education. Berkeley, CA: The Wing Institute.

Totterdell, M., Woodroffe, L., Bubb, S., & Hanrahan, K. (2004). *The impact of newly qualified teachers (NQT) induction programmes on the enhancement of teacher expertise, professional development, job satisfaction or retention rates: A systematic review of research literature on induction*. London: University of London Institute of Education, EPPI Centre.

U.S. Department of Education. (2002). *Meeting the highly qualified teachers challenge: The secretary's annual report on teacher quality*. Washington, DC: Author.

Villar, A., & Strong, M. (2007). Is mentoring worth the money? A benefit-cost analysis and five-year rate of return of a comprehensive mentoring program for beginning teachers. *ERS Spectrum, 25*(3), 1–17.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2003–2004, Winter). Heading off disruptive behavior: How early intervention can reduce defiant behavior — and win back teaching time. *American Educator,* 6-15, 18-21, 45.

Wang, J., Odell, S. J., & Schwille, S. A. (2008). Effects of teacher induction on beginning teachers' teaching: A critical review of the literature. *Journal of Teacher Education, 59*(2), 132–152.

Whisnant, E., Elliott, K., & Pynchon, S. (2005). *A review of literature on beginning teacher induction*. Retrieved from http://www.cstp-wa.org/Navigational/Policies_practices/Teacher_induction/ A_Review_of Literature.Pdf

About the Wing Institute

The Wing Institute is a 501(c)(3) nonprofit operating foundation dedicated to the promotion of evidence-based education policies and practices. It was founded in 2004 and named after Ernie Wing, an outstanding special education advocate who was an early champion of evidence-based education and quality services for children.

The Wing Institute is a "catalyst" organization, designed to bridge the research-to-practice gap by supporting and accelerating collaboration among researchers, educators, policy makers and consumers across disciplines. Its goal is to accumulate and disseminate existing knowledge, help create new knowledge, and facilitate the effective application of that knowledge to real-world settings.

It achieves this through the following strategies:

Evidence-Based Education Knowledge Network. The Institute operates an interactive, internet-based Knowledge Network designed to link education stakeholders interested in promoting evidence-based education. It facilitates the open exchange of ideas, resources, expertise, and support.

Information Clearinghouse. The Institute's Web site (www.winginstitute.org) houses an online information clearinghouse that reviews research, analyzes current issues and policies, provides links to related organizations and activities, and includes resources for education stakeholders.

Professional Forums. The Institute organizes, sponsors, and facilitates professional summits, seminars, and conferences to encourage dialogue, collaboration, and communication among researchers, decision makers, and educators across disciplines and organizations.

Publications. The Institute researches and disseminates journal articles, policy analyses, research reviews, books, newsletters, position papers, and tools related to application of evidence-based education to assist education decision makers.

Research. The Institute supports, promotes, reviews, and completes research in education policies and practices, including funding for graduate research.

Public Policy. The Institute engages in general public policy activities such as analysis, research, and development.