



120 SE 10th Ave.
Topeka, KS 66612-1182
785-296-6177
www.ksde.org

**A Proposal
to Select, Place, and Keep,
the Best Teachers
in the Neediest Schools**

A Kansas State Dept. of Education White Paper

Tony Moss
Data and Special Projects Workgroup
Career, Standards, and Assessment Services Team

16 May 2011

Contents

Why a Consortium?	2
The Consortium's Initial Plan	3
A New Look at the Problem	4
A History of Oversupply	4
Other Countries' Solutions	5
Where's the Theory?	9
Causes from the Literature	12
Better Than Value-Added	16
We Need a Second Experiment	17
Design Guided by Theory	18
Costs vs. Benefits	18
Conclusions	19
References	20

Why a Consortium?

Teacher selection, retention, and evaluation form a contentious policy knot. The direct annual costs of teacher turnover in the U.S. are estimated to be \$5 to \$7 billion. Six states—Arkansas, Kansas, Missouri, South Dakota, Utah, and West Virginia—have formed a consortium to improve teacher recruitment and retention. Annually, the consortium states lose more than \$272 million due to teacher turnover (Alliance for Excellent Education, 2005 and 2008). The indirect costs are difficult to estimate, but teacher turnover rates are most acute in high-poverty schools (Ingersoll, 2001).

One prominent strand in the knot is the lack of objective measures of teacher effectiveness. How can districts know which teachers to select or retain without some proven measures of their effectiveness?

Using value-added techniques, economists have shown that some teachers can significantly retard academic achievement, while others can greatly accelerate it (Rivkin, Hanushek, and Kain, 2005). For those who believe one of the central problems of American schooling is poor teaching protected by tenure and a due-process barricade, an objective measure of teaching effectiveness is very appealing. But the science of measuring teaching effectiveness is imprecise, controversial, and under development (Rothstein, 2007; Shulman, 2009; Baker, et al, 2010; Harris, 2009; Boyd, Goldhaber, Lankford, and Wyckoff, 2007; Ravitch, 2010; Sanders and Wright, 2008). Value-added models are unable to show what they are actually measuring. Improvements in students' scores could be due to factors other than the teacher—a good after-school tutor, another teacher or adult working closely with the student—just as declines in students' scores

could be due to changing schools, family problems, or a host of other causes (Baker, et al, 2010).

To overcome these shortcomings, advocates for new teacher evaluation instruments have proposed hybrid measures that emphasize student test gains, but integrate them with structured classroom observations, other student outcome criteria, and student survey data (Bill and Melinda Gates Foundation, 2010).

But even if scholars build a hybrid, at the very best, it will only marginally improve the selection of teachers and student test performance. And it won't reduce teacher turnover. Why?

- Because it will ignore the underlying causes of high turnover: teacher labor market and working conditions, and the environments shaping children;
- It doesn't explicitly include some teacher characteristics, like certain aspects of social intelligence; and
- It doesn't offer a way to place and retain our most capable teachers with our most disadvantaged children.

The Consortium's Initial Plan

While banding together increases their likelihood of success, the six states face a daunting task. They must design reforms based on limited and contested information. They must also coordinate the reforms among the institutions preparing teachers, the schools and districts hiring them, the teachers' unions defending them, and the federal officials demanding accountability. They will have to meet the funding conditions of supporting foundations, and respond to

advocates of free-market remedies. Under such diverse political pressures, designing sound research and institutional reforms—through committees—is so difficult, its attempt seems daring.

The consortium's initial work has:

- defined stages of skill advancement as professional standards that can be used in evaluations; and
- endorsed teacher-to-teacher mentoring and apprenticeship models.

It has also put stakes in the sand, accepting as goals:

- teacher pay-for-student performance; and
- creating some objective measures of teaching effectiveness.

The consortium's initial plan to delineate teaching standards into a four-stage professional model faces the same obstacles as all other teaching evaluation instruments: they haven't been empirically linked to improved student outcomes. If it isn't shown to measure teaching effectiveness, the professional continuum model is unlikely to win support as the basis of a new teacher salary ladder. While the consortium's evaluation model could provide principals with a new means of promoting or dismissing teachers, adding a new evaluation instrument to those now on the shelf is unlikely to reduce turnover, unless it is used as a guide in mentorship programs. The standards can be used to help select the most experienced teachers, and help monitor the professional progress of apprentices through the four stages.

Correlational and survey evidence strongly suggests that well-designed and well-implemented mentoring can reduce new-

teacher attrition by half (Smith and Ingersoll, 2004; Markow and Cooper, 2008). Regular planning and collaboration with same-subject teachers, and having an external network of teachers, were also associated with improved retention. Policies that commit the districts and unions to long-term mentoring—at least two years—and provide the time resources and integrated data to evaluate those policies, also appear more likely to improve new-teacher retention and skills (Parson, Lupe, and Bosserman, 2002; Markow and Cooper, 2008).

Pay-for-student performance is unlikely to reduce turnover. Motivation theory (Pink, 2009), and the results from some countries where pay-for-student performance has been tried (Martins, 2009), suggest it can have negative teacher and student effects.

A New Look at the Problem

This paper:

1. reviews the recent history of teacher labor markets;
2. describes the teacher selection, preparation and retention policies of three countries that successfully reformed their systems;
3. proposes a different theoretical guide; and
4. reviews key teacher survey data and a large, current experiment in teacher evaluation.

Based on this evidence, it then proposes a more promising experiment in teacher selection, preparation, and assignment.

A History of Oversupply

Prior to the 1970s, gender discrimination and racial discrimination actually

contributed to the quality of the teaching workforce (Bacolod, 2007; Murnane and Steele, 2007). Often the best professional choice for many talented women and minorities was teaching. But World War II, and the successes of the civil rights and women's movements, opened up new opportunities. With the rising inequality that began in the 1980s, the relative wages in some professions began to rise. As a percentage of GDP, the average teacher salary has decreased at about 2 percent per year since 1970, while salaries in professions requiring similar levels of education have increased. In particular, science and math teachers have more lucrative opportunities than teaching (Hampden-Thompson, Herring, and Kienzl, 2008). Today in Kansas, teachers can expect to earn about 70 percent of the earnings of people with comparable levels of education. The earnings of teachers in the other consortium states have also eroded relative to comparable professions:

Arkansas	74%
Missouri	75%
South Dakota	82%
Utah	81%
West Virginia	82%

(2006 figures and dollars; Mishel, Allegretto, and Corcoran, 2008).

Teacher demand has been a function of changes in student population and student-teacher ratios. In the past 15 years, Kansas enrollments increased modestly, by about 20,000. The composition of the student body also changed, with the proportions of low-income, students with disabilities, and English Learners (EL) going up by more than 14, 5, and 8 percentage points, respectively. These changes in the characteristics of students have increased the difficulty and complexity of teaching. Relatively more teachers were hired, so student/teacher ratios fell from 14.4 in the

fall of 2000 to 13.2 in the fall of 2007 (National Center of Education Statistics, 2009).

Some of these trends are also reflected in national data: from 1987-88 to 2007-08, K-12 enrollments increased 19 percent, but the number of teachers increased by 48 percent, mostly due to reduced class sizes, the growth in special education, and increasing specialization by subject. Some of the growth in teacher numbers is due to policy: the Individuals with Disabilities Act, and expanding high-school graduation and college entrance requirements, appear to have fueled the need for subject specialists (Ingersoll and Merrill, 2010).

At the same time, some universities and colleges have set or maintained their teacher education admission standards low to keep enrollments high and generate revenue. Nationally, there is a chronic oversupply of teachers: each year, more than 50,000 newly certified teachers do not go into teaching (Ingersoll, 2003; Auguste, Kihn, and Miller, 2010). Yet in some subjects, like math, at current salary levels, demand comes close to eclipsing supply (Ingersoll and Perda, 2010). In some schools with difficult working conditions, there can be a dearth of quality applicants. At current pay levels, and under current conditions, many teachers will choose other employment, or no employment, over working in schools with poor conditions. There can be scarcity of qualified applicants in one district and abundance in the next (Ingersoll and Perda, 2010).

The oversupply at relatively lower salaries across a range of working conditions becomes a revolving door, as teachers, especially new teachers, try to sort themselves into positions that best match their requirements and skills, and districts

and schools try to select the best available candidates. Many districts appear to have committed to a revolving door strategy to contain labor costs—starting teachers cost less, they help fund pensions for retiring teachers, and their reduced salaries can help districts meet local demands to restrict taxes (Ingersoll, 2003). The hidden costs of revolving door strategies are a decline in the experience, and quality, of teachers. In 1987-88, the modal teacher had 15 years of experience, but by 2007-08, she had one (Ingersoll and Merrill, 2010).

As a result of the long-term trends of declining relative pay, oversupply, reduced admission standards, and, in some schools, increasingly difficult working conditions, the academic quality of the workforce entering education has declined (Bacolod, 2007). Currently, about “23% of new teachers overall—and about 14% of those in high-poverty schools—come from the top third of graduates” (Auguste, Kihn, and Miller, 2010). Higher-ability teachers, as measured by their test scores and other characteristics, have been leaving the profession at higher rates than those who had lower assessment scores (Bacolod, 2007; Guarino, Santibañez, and Daley, 2006). While higher-paying professions are becoming increasingly gender balanced, teaching is become increasingly female: 66 percent of teachers were female in 1980; 76 percent were in 2007-08. In the past, feminization of a profession has coincided with its decline in pay and status (Ingersoll and Merrill, 2010).

Other Countries' Solutions

International comparative studies have provided examples of countries that have successfully reformed their systems. Two studies from McKinsey education consultants, *How the World's Best-Performing School Systems Come Out on*

Top, (Barber and Mourshed, 2007) and *Closing the Talent Gap: Attracting and Retaining Top-Third Graduates to Careers in Teaching* (Auguste, Kihn and Miller, 2010), provided the data and examples here.

The latter makes the case that:

- increasing teacher pay to the market rate for students in the top-third of their academic cohorts;
- restricting the selection of candidates based on market demand, academic and particular social skills; and
- improving the quality of teacher preparation,

can solve the teacher attrition problem and greatly improve student learning. Though we've noted some of the problems with using student assessments to judge teacher quality, and emphasized the importance of national labor market conditions, let's use

the shorthand of student assessments to look at the components of other countries' successful reforms.

Based on international student assessments, concentrations of the top-performing teachers are in Singapore, Finland, and South Korea. All are turnaround stories: fifty years ago, none of the three were educationally remarkable. Today, the lowest performing tenth of Finnish schools outperform all the Organization for Economic Co-operation and Development (OECD) countries.

The three countries select prospective teachers based on a telling mix of characteristics:

- all candidates are in the top third or higher of their academic cohorts;
- they persevere in the face of difficulty;
- they know how to motivate others;

- they are passionate about children; and
- they are skilled communicators and organizers.

These selection criteria fit with some American findings: a teacher's advanced degree is not predictive of better student performance (Rivkin, Hanusek, and Kain, 2005) but a teacher's GPA is (Murnane and Steele, 2007).

Once teacher candidates have been selected, all three countries train them something like the United States trains doctors.

Notably, Singapore uses a data-intensive process to monitor student achievement, but Finland does not. While Singapore awards merit, performance, and outstanding contributions with bonuses, some as high as 30 percent of base pay, Finland has none (see table below). As a percentage of per capita GDP, the starting salaries of teachers in Korea and Singapore are 62 and 24 points higher than their counterparts in the U.S. With 15 years of experience, Korean and Singaporean teachers are earning more than twice as much as similarly experienced teachers in the U.S. Finland's teachers' pay is only moderately better than that of U.S.

Country Comparisons of Policies Aimed at Attracting and Retaining Teachers

✓ significant priority in the country, best-in-class practice

polices to attract/retain top teachers	Singapore	Finland	S. Korea	U.S.
1 selective admissions to teacher education programs	✓	✓	✓	most programs not selective
2 government paid teacher preparation	✓	✓		students finance own education
3 government regulates supply of teachers to match demand	✓	✓	✓	oversupply of teachers
4 professional working environment	✓	✓	✓	variable working conditions
5 competitive compensation	✓		✓	compensation not attractive to many students
6 cultural respect accorded to teaching	✓	✓	✓	comparatively low
7 teaching considered as a career	✓	✓	✓	relatively high attrition in early years
8 robust opportunities for career advancement	✓			limited opportunities for advancement
9 performance pay for teachers	✓		✓	limited performance pay

Source: Interviews, McKinsey research (Auguste, Kihn and Miller, 2010)

teachers, but is comparable to that of other Finnish professionals with masters' degrees, plus the government pays the cost of their education and living expenses while they study.

The three countries have annual teacher turnover rates of 1 to 3 percent compared to U.S. rates of about 15 percent (see table below). In the U.S., almost half of new teachers leave within their first 5 years; those with 4 to 9 years of experience leave at a rate of 4.5 percent per year (Ingersoll, 2003; Murnane and Steele, 2007).

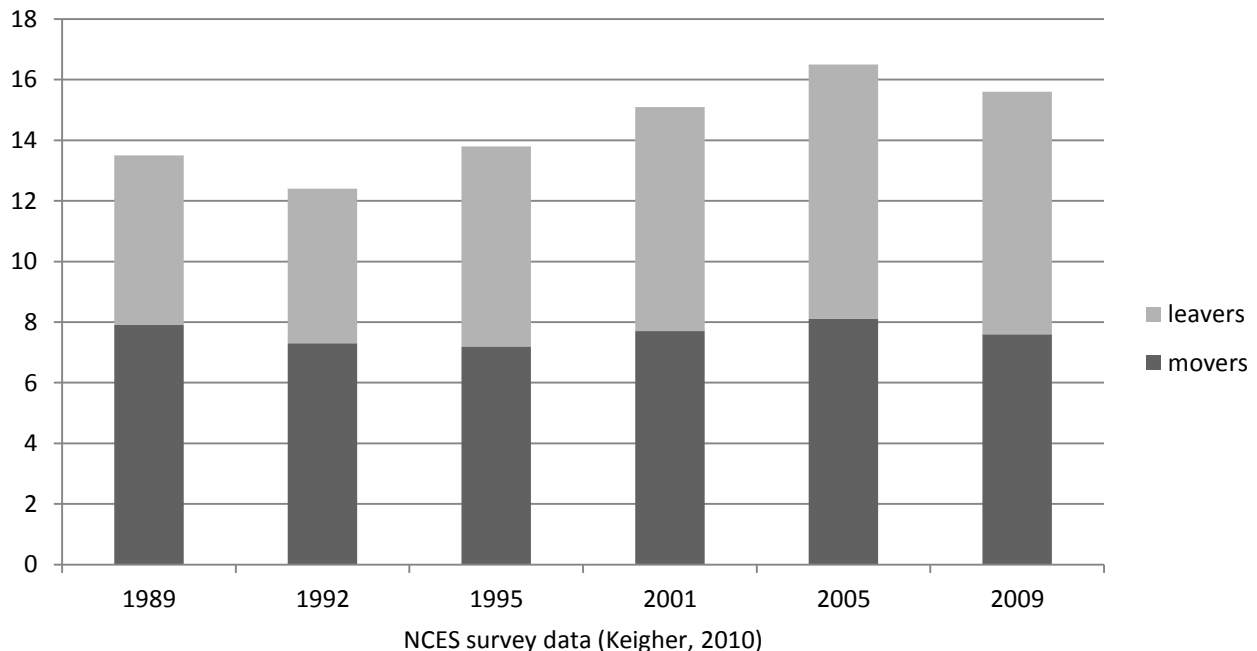
Because teachers are so highly selected and thoroughly trained, they are trusted with a high degree professional autonomy. Many administrative functions that, in the United States, are the responsibilities of non-teaching administrators, in Finland or Korea, are divided and managed by the teachers themselves. Currently, non-teacher salaries in the U.S. are 23 percent of educational expenses, while in Finland

and Korea, only 10 and 7 percent are. (Auguste, Kihn and Miller, 2010).

The dedication of high-performing countries to professional collaboration is also markedly more intense. American teachers average only about 2.7 hours per week for collaborative planning (Wei, Darling-Hammond, Adamson, 2010). At twenty hours per week, Singaporean teachers have seven times as much time for collaboration. Finnish teachers teach about half as many hours as American teachers do, which gives them time to plan and integrate lessons (Rothman and Darling-Hammond, 2011).

By international comparisons, American trends in professional development are also going the wrong way. Over the last decade, our predilection for the one-day workshop has increased (Wei, Darling-Hammond, and Adamson, 2010). But for professional development to measurably improve student learning, the minimum duration of

U.S. Teacher Turnover Trends



training in a *single subject* should be something greater than 49 hours (Yoon, Duncan, Lee, Scarloss and Shapley, 2007). The average number of hours American teachers spend in professional development *in all areas*—is about 44 hours per year. In Singapore, the government pays for a hundred hours of professional development per year (Rothman and Darling-Hammond, 2011).

These facts suggest:

- selecting teacher candidates *before* they enter school and matching supply to demand is more cost-effective than permitting oversupply and deterioration in the quality of applicants and retained teachers;
- if a country wants the best, it must pay the market price for the best—successful systems pay wages that at least match the salaries that comparable talent would find in other professions;
- certain non-academic social abilities—to empathize, to accurately read a child, and then to constructively respond, to motivate, to communicate, the ability to organize groups, and commitment—are as important to teacher effectiveness as high academic skills and content expertise;
- the U.S. is under-investing in teacher collaboration and professional development; and
- professional autonomy, good working conditions, government-paid higher education, stipends while studying, and prestige—are also important in attracting and retaining the best.

Where's the Theory?

When one compares the learning theories of almost any social or medical science—evolutionary psychology, sociology, the many branches of neurology—to the theories invoked by educational accountability and the teacher evaluation literature—one is struck by the latter's reductionism. Advocates of value-added models often invoke some free-market ideas—higher pay for higher student test scores—but ignore others—oversupply and pay scales below market rates have resulted in declines in applicant quality (Weisberg, Sexton, Mulhern, and Keeling, 2009). At the theoretical level, those advocating the use of student test scores to evaluate teachers make the following assumptions:

1. Poor teachers cause poor student outcomes.
2. State assessments, together with other measures, will provide principals with instruments that accurately measure teacher effectiveness.
3. With these new instruments, principals will remove poorly performing teachers with greater accuracy and frequency than they currently do informally.
4. Poorly performing teachers will be replaced by more effective ones.
5. The new instruments will provide teachers with insight that will guide their professional development.
6. Districts and states will pay higher-performing teachers for their improved performance.
7. Recognition and pay-for-performance will boost the retention and performance of good teachers.
8. The gaps in student academic achievement—by class and ethnicity—will gradually disappear.

Others have reviewed the strong as well as the broken links in this theoretical chain (Baker, et al, 2010; Ravitch, 2010; Braun, 2005). Here we'll include the students in a better working theory.

Theory directs attention to some causal factors to the exclusion of others. We need some explanation of just what happens between teachers and students, something to explain *how* and *why* learning happens. A fundamental understanding of what happens between teachers and students, and between students and others, is essential if we expect reforms to improve teaching and student learning. Working theories that ignore fundamental causes aren't likely to work.

For comparison's sake, let's consider the learning theories of human evolutionary psychology. It describes the brain as an instrument predisposed to learning in domains that have improved human survival. In the social domain, learning has sensitive or privileged periods that are formative. There is a hierarchical scaffolding of social interactions:

- the face-reading and voice responsiveness at birth;
- the bonding and attachment that shape habits of relation;
- the proto-conversations and the emotional exchanges of language acquisition;
- the ability to read faces, interpret voices and gestures, and anticipate what others are thinking at three and four (theory of mind);
- the joy and cognitive preparation of fantasy, play, and mastery of new skills;
- the ability to play, cooperate, and work in groups, the competition

for high status, and the stress of low status;

and so forth. Our brains are prominently designed for and shaped by social interaction (Geary, 2007). The essential model for effective schools isn't corporate but familial and tribal; its essential technology isn't testing but complex social interactions. The child development literature describes the building blocks of human competence in terms of *social environments and interactions* that can:

- suppress or limit gene expression (epigenetics);
- shape long-term habits of relating, motivation, and mental health risks (bonding and attachment and mental health research);
- shape concentration and self-control (neurology and behavioral genetics; psychology); and
- suppress or enhance IQ, academic achievement and learning readiness (psychology; see Turkheimer, et al, 2003).

In contrast, value-added, gain, and growth models mostly come from the field of economics. They ignore the complexities of social environments, human development, and the social complexities of teaching (Shulman, 2009). All of a child's developmental history is reduced to the child's past test scores, and big proxy control variables, like poverty and race (Kane, Taylor, Tyler, and Wooten, 2010). Learning on the part of students is a matter of skilled exposure to materials and practice. All children are assumed to have equal and constant plasticity in their responsiveness, and all teachers to have a constant, fixed ability to teach all students. A good teacher can be reduced to someone who consistently raises students' assessment scores (Hanushek, 2002).

One effect of this reductionism is conceptual exclusion of the larger environments shaping children and their abilities. Nowhere is it noted that children accurately reflect the environmental forces that shape them. We don't need to be concerned about the world-wide trend toward increasing proportions of working, single parents, often with weak social and economic supports, or the developmental, moral, and cognitive effects of replacing face-to-face interactions with intensified electronic and commercial substitutes. The national, state, and community obligations of creating smart environments for child development are shifted to a very narrow focus on teachers and a utilitarian goal of raising test scores. The fundamental building blocks of child competence—the causes behind our larger goals—are pushed aside for more limited, politically-defined goals.

At the same time, companies have learned to exploit evolutionary, hard-wired predilections and are influencing child development and capacities—think of Facebook, cell phones, or Halo—while outside of sports and clubs, schools are often asking students to learn increasingly complex materials divorced from what evolutionary theory would call primary learning domains (Geary, 2007). We might expect that the long-term or latent effects of a system that alternately pays and threatens teachers for students' learning or not learning of academic skills, divorced from social context and relationships, would result in less curiosity, creativity and joy in learning, more alienation, and probably more bullying and cheating. Portugal's emphasis on individual teacher effectiveness, as measured by student performance, has been linked to grade inflation and declines in student achievement (Martins, 2009).

The importance of intensive social and emotional interactions from highly-trained and responsive teachers and improved parenting has been experimentally confirmed by early childhood interventions like the Carolina Abecedarian and the HighScope Perry Preschool projects (Heckman, 2006; FPG Child Development Institute; HighScope Perry Preschool Study).

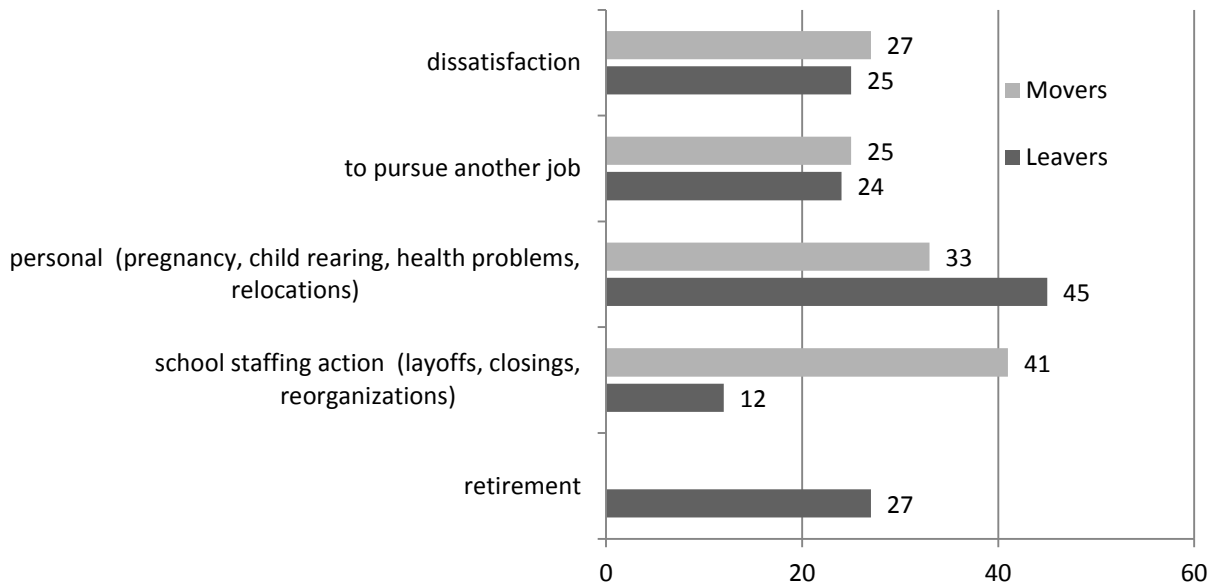
In sum, human development theory suggests that:

- certain measures of teacher social attributes—for now, let's say empathetic accuracy, constructive responsiveness, and commitment—should be included in teacher selection and evaluation;
- improvements in early child environments, particularly bonding, attachment and relational habits, language enrichment and positive responsiveness in child-rearing, are complementary to improved teacher effectiveness;
- we should expect that smarter environments and stronger developmental foundations will amplify teacher effectiveness, improve school environments, and improve student outcomes across the life span;
- environmental interactions and social relationships, e.g., the social status of students, trust between teachers and students, encouragement from peers, and collaboration between teachers, will influence student motivation and achievement and improve teacher effectiveness (Allensworth and Easton, 2007; Geary, 2007); and
- child developmental histories and teacher dispositions may be

necessary control variables when attempting to measure teacher effectiveness.

working close to home (Boyd, Lankford, Loeb, and Wyckoff, 2003), and exits to care for a family, high turnover rates should be

**Why Did Teachers Who Have Transferred (Moved), or Quit (Leavers), Say They Left?
Answers from a Nationally Representative Sample
(Ingersoll, 2001)**



Causes from the Literature

Ingersoll showed that the push of job dissatisfaction, and the pull of labor-market opportunities, were collectively much more important causes of teacher turnover than retirement (see chart above). Except in the case of layoffs and other school staff changes, movers' reasons for transfer were similar to career leavers' reasons for leaving the profession. In a labor market characterized by oversupply, declining relative pay, declining quality of entrants, strong personal preferences for

expected.

The specific reasons the dissatisfied gave for leaving (see chart on following page) suggest that turnover can be reduced by improving:

- salaries;
- administrative support for teachers;
- student behavior and motivation; and
- by using collaborative management styles.

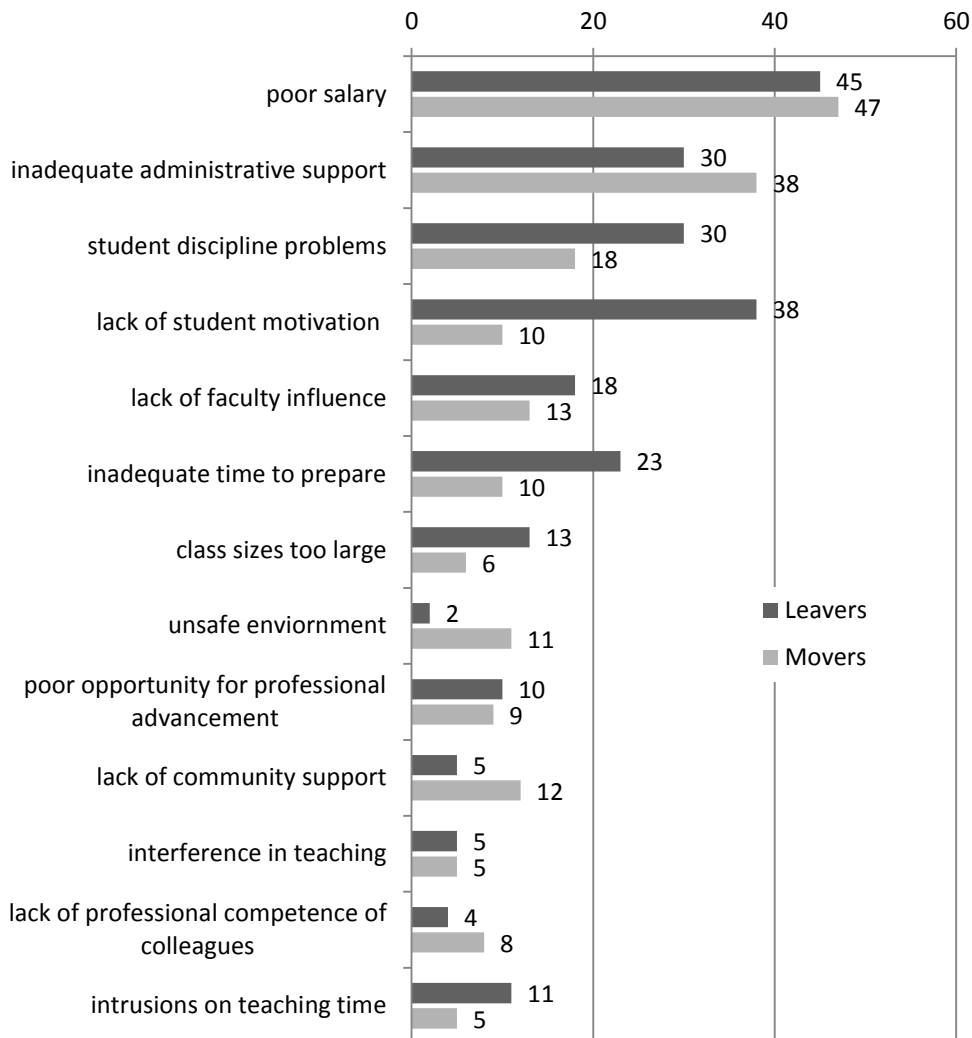
The reasons leaving and moving teachers gave for transfer only weakly support the consortium’s focus on identifying stages of professional development. Lack of opportunity for professional advancement was ninth on the list.

Some researchers say retirement is a more important source of attrition than Ingersoll does. When compared to other professions, teachers’ relatively high ratio

of pensions to salaries, they say, pulls teachers into retirement earlier (Harris and Adams, 2007). This suggests changes in retirement policies could slow retirement losses.

Overall, Ingersoll’s findings have been widely supported (Guarino, Santibañez, and Daley, 2006; Marvel, Lyter, Peltola, Strizek, and Morton, 2007; Borman and Dowling, 2008), but circumstances have

Self-Reported Reasons for Leaving Among Teachers Who Left Because of Dissatisfaction, from a Nationally Representative Sample (Ingersoll, 2001)



changed since Ingersoll published his results ten years ago—No Child Left Behind (NCLB) has been in effect for a decade, and a near financial collapse and Great Recession have wrecked state budgets.

Is there political will to experiment with new teacher selection criteria and higher salaries? Relatively low U.S. salaries to GDP per capita suggest that the U.S. could pay higher salaries, and that current salaries may not be high enough to attract quality applicants (Ladd, 2007). But the U.S. also spent 7.6 percent of GDP on education in 2007, well above the OECD average of 5.7 percent (OECD, 2011). These facts suggest that the U.S. may be caught in a cultural trap: we are already spending more than countries whose students are getting better test results, but we neither want to spend more, nor seem to know how to rearrange resources to get the results we want.

We can get a more current picture of teacher work conditions from the 2008-09 Teacher Follow-up Survey (Keigher, 2010). In the table below, former public school teachers who left teaching and were working in non-K-12 jobs were asked which was better—teaching or their current job—across a number of work conditions.

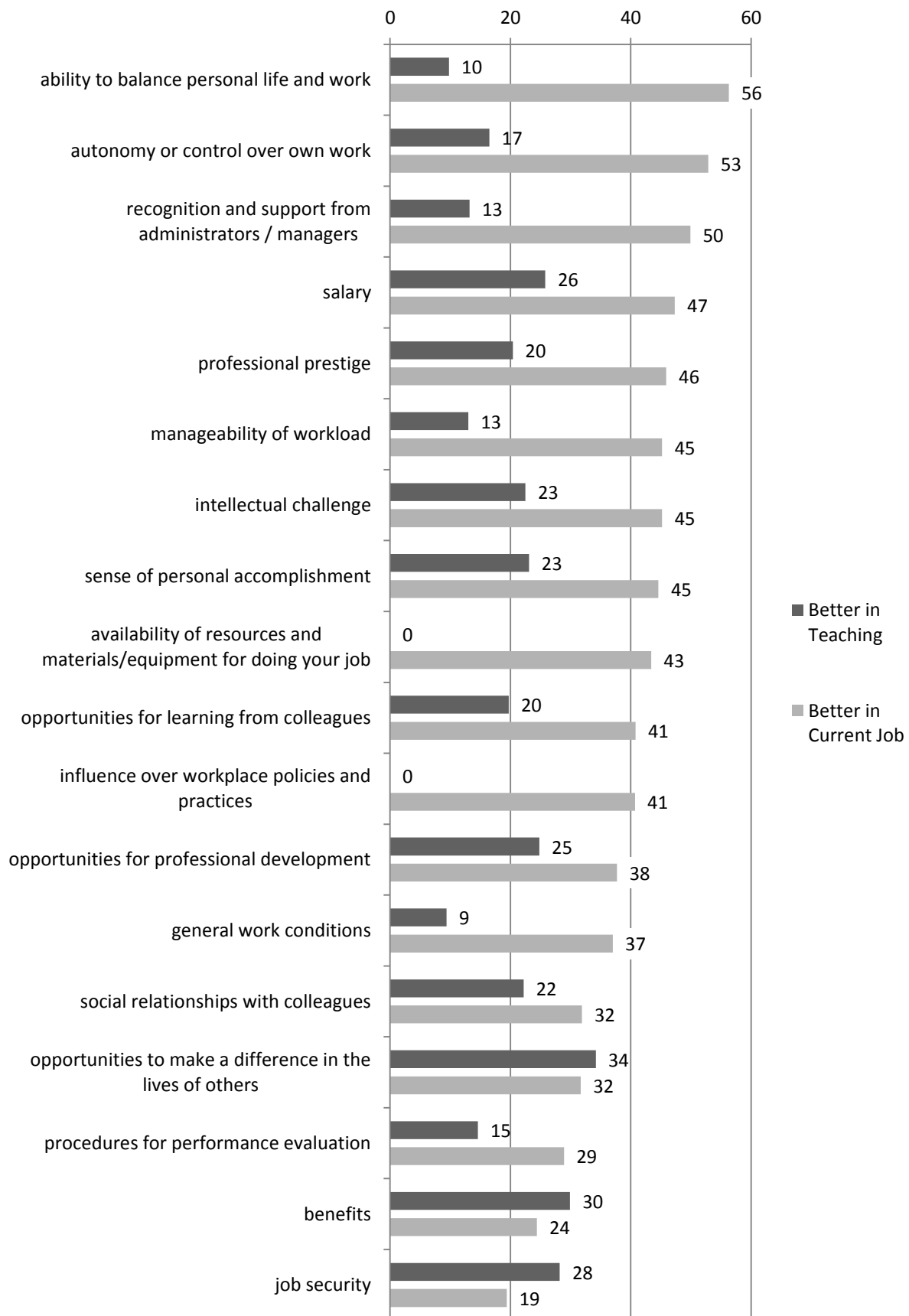
Interpretive caution is warranted. We don't know which factors were most prominent in respondents' decision to leave teaching. Though salary seems less prominent than autonomy, it could be that salary had more influence on the decision to leave. We also don't know about the climate and administration of the schools

they've left, these teachers' characteristics, nor how the two interacted. The error terms for some of the teaching-is-better responses were pretty large, too.

Nevertheless, the responses describe some of the most prominent deficits in working conditions that are associated with driving teachers into other professions:

- teaching workloads were heavier and intruded more into personal lives;
- in the post-NCLB world, non-teaching jobs offered much more autonomy, support from management, intellectual challenges, prestige, and a greater sense of accomplishment;
- salary differences continue to pull people away from teaching;
- underscoring the need for authentic mentorship, these former teachers had more opportunities to learn from and socialize with their new colleagues than they had from their former teacher-colleagues;
- they had more influence over workplace policies and practices in their new employment too, which supports the Ingersoll survey in suggesting that collaborative workplaces will reduce turnover;
- complaints about student behaviors and motivation are absent where they were prominent in Ingersoll's 2001 analysis; and
- teaching was superior in only three categories: job security, benefits, and the opportunity to make a difference in the lives of others.

Former Teachers Compare Their Current Positions to Teaching (Percentages, 2008-09 Teacher Follow-up Survey, Keigher, 2010)



Better Than Value-Added

The Gates Foundation has funded an experiment within six mostly urban districts. The experiment will identify what combination of teacher evaluation instruments are the best predictors of teachers' ability to raise student assessment scores (Bill and Melinda Gates Foundation, 2010). The goals include providing teachers with feedback that can help them improve their teaching, and the creation of a validated instrument that will guide principals in hiring, granting tenure, and firing teachers.

Here's what will be thrown into the analysis blender:

- state assessments, the Stanford 9 Open-Ended Reading, the Balanced Assessment in Mathematics, and ACT's QualityCore;
- 20,000 panoramic, digital, videotaped classroom observations using Teachscape's technology and scored by professionals at the Educational Testing Service (ETS), with subsets scored by experts at the National Board for Professional Teaching Standards (NBPTS) and the National Math and Science Initiative (NMSI) using the UTeach Observation Protocol;
- an ETS assessment to measure teachers' pedagogical knowledge and ability to identify errors in student reasoning;
- the Tripod student survey by Ron Ferguson to measure student perceptions of the classroom environment;
- a teacher survey from the New Teacher Center about working conditions, school environments,

and the instructional support teachers have received; and

five classroom observation instruments:

1. the Classroom Assessment Scoring System (CLASS) by Bob Pianta and Bridget Hamre;
2. Charlotte Danielson's Framework for Teaching;
3. Mathematical Quality of Instruction (MQI), by Heather Hill and Deborah Loewenberg Ball;
4. Protocol for Language Arts Teaching Observations (PLATO), by Pam Grossman; and
5. Quality Science Teaching (QST) Instrument by Raymond Pecheone.

With this mountain of data, and an expanded number of statistical analysis techniques, the contracted RAND experts will surely identify stronger predictors of teacher effectiveness.

The composites they create:

- are not likely to be cheap—classroom observations, even by automated video, will be expensive to code (Goe and Croft, 2009) and the creators of the proprietary instruments will want to be paid for their use;
- are likely to have some unintended consequences, for example, student survey and test fatigue, or, in some cases, video monitoring of classrooms;
- will be test, data and analysis intensive and require some unknown amount of contracted services and increases in the number of staff analysts and researchers; and
- will be focused on assessment results, at the expense of some

difficult to measure predictors that have been excluded, like salaries relative to other professionals, early childhood environments and relationships, child behaviors and motivation, teacher autonomy, teacher-student trust, teacher-to-teacher collaboration, and peer influences on academic achievement.

If efforts to reduce turnover and improve teacher effectiveness are focused only on developing new teacher evaluation instruments, we are likely to get more widely used teacher evaluations, but miss the greater causes of teacher attrition and poor student academic achievement. It is as yet unknown whether formal teacher evaluations will be more accurate in their sorting of teachers than the informal self-selection and school and district selection that help drive our current, high, teacher turnover rates.

We Need a Second Experiment

The amount of data that can be collected in an atheoretical experiment is vast. Guided mostly by political impulses, continuing to extend a data-intensive business model into schools could create a giant reporting machine with meager benefits and some large, unintended consequences. The Gates Foundation experiment is being swept forward by the momentum of new value-added measures, the political need for teacher evaluation instruments, the business-model orientation of the foundations themselves, and the accountability movement.

Child developmental theory and international comparisons suggest a second experiment.

In much greater detail, and with survey data and various scenarios, in their paper *Closing the Talent Gap: Attracting and Retaining Top-Third Graduates to Careers in Teaching*, the McKinsey analysts laid out policy options for improving the quality of teachers entering the profession (Auguste, Kihn and Miller, 2010). Below, I outline one possible scenario based on their work:

1. Create an elite school of education that adapts and improves upon the Finn, South Korean, and Singaporean examples.
2. Use ACT, SAT, and, in the case of non-academic specialized skills, parallel tests, to identify potential teacher candidates before they enter higher education, and invite them to apply to become charter members of a new teacher core;
3. Select those in the top 30 percent in specific disciplines and in the social attributes that matter most, like empathetic accuracy, responsiveness, perseverance, and commitment to children.
4. Match the number of applicants to positions opening in the highest need schools in the consortium states.
5. For those applicants accepted, pay their college expenses through the master's level, including a stipend that covers living expenses while they study.
6. Give them deep training and residencies, like doctors, and mentored practice in working with high-need populations, including early childhood, special education, low-income urban and rural students, and English Learners.
7. Graduates working in high-need schools must be paid a salary competitive with other professionals with their level of preparation and talents—no chiseling allowed. In

their international comparative study, McKinsey researchers estimated that a starting salary of \$65,000 and a maximum of \$150,000, together with improved teaching conditions, would more than quadruple the percentage of applicants from the top third of performers;

8. Time for collaborative planning and lesson integration should be comparable to the investments made by the most effective international exemplars;
9. Professional development must be substantial—more than 50 hours per year in each subject area; and
10. Like pilots trained and retained by the military, obligate graduates to a minimum of ten years of service in designated high-need schools. Other expectations—for example, a longer school year and longer school days—could also be part of the contracts. The highest-need schools could have the rights of first choice among graduates. Transfers would have to be restricted to other designated high-need schools and communities.

Design Guided by Theory

The McKinsey Top-Third model could be improved by first concentrating the new teaching core graduates in early childhood services, when children and families are most sensitive and responsive to improved environments, social influences, and learning. The influence of the teachers with students could also be increased, and the transitional stress on students reduced, if the same teachers work with the same families from early childhood through primary school. Transitions between social environments—from day care to kindergarten, primary school to middle

school, etc., could be deliberately smoothed.

With the objective of improving early childhood environments and childrearing norms, we would want to select and train a parallel group of social and health workers with the same selectivity, quality, and improved working conditions that were provided to the new teaching core. To cultivate community support and collaboration between teachers and social and health services, we would want the services placed in the schools. We would also want to guarantee their continuity in the community and not lose any of these highly-trained workers to attrition or transfer.

In exchange for receiving these specialists with subsidized training and salaries, participating states, schools and districts would have to guarantee high levels of autonomy, professional working conditions and resources.

Costs vs. Benefits

What about the costs of higher salaries and elite schools of education and social services?

The McKinsey consultants have suggested two ways to offset the costs:

1. Administrative positions can be cut. Non-teacher, administrative salaries in the U.S. are 13 percentage points higher than in Finland, and 16 points higher than in South Korea. Because of the greater autonomy of the teachers, some administrative functions are not needed or can be assumed by the teachers (Auguste, Kihn and Miller, 2010). But we want these highly trained professionals to continue to teach.

Any career ladders, additional responsibilities, and compensation for experience, should be designed to reduce turnover and facilitate teaching, and not offer a premature exit from teaching.

2. By focusing superior social, health, and early childhood services on improving early child development, and the building blocks of later competence, cost savings are realized as children age. As students age, if needed to reduce costs, classes can shift to higher student-teacher ratios and not increase student risks.

There are many savings that will not be immediately apparent. Consider improvements in students' social skills. Labor market economists remind us that the most consistent predictors of workplace success are non-academic—conscientiousness, cooperation or teamwork, and emotional stability (Sackett, 2010). This experiment requires long-term commitments and sustained resources to measure its real benefits. But it does begin to build a smarter, preventive system that promises higher returns throughout students' lives—improvements in social and academic skills, greater social engagement, and lower dropout rates, crime, and teenage pregnancy rates. The early childhood experiments cited above and the economist James Heckman have shown that effective early interventions are high-return investments (2006).

This second experiment also gives us the opportunity to test a model that can slow, and then stop, the revolving door in the neediest schools. Students as well as districts will not face the social and instructional disruptions of high turnover.

Instead of building a system that will depend on expensive and expansive data collection and analysis, testing can be minimized and money currently flowing to testing contractors can be redirected to teaching and direct child services. The experiment also would allow us to counter the erosion of professional autonomy set in motion by No Child Left Behind and its over-emphasis on testing. Much higher levels of trust and autonomy can be extended to teachers who are well-selected and well-trained from the beginning of their careers. We should expect that the increased professional autonomy will enhance their motivation and give them the latitude for creative problem solving based on local conditions and individual child needs. Based on the survey data cited above, improved autonomy should also reduce teacher turnover.

Conclusions

As with No Child Left Behind, some of the causal theories behind the current momentum to develop teacher evaluation instruments have strong political support, but many unanswered evidential questions. Labor market trends, international comparisons of successful reform efforts, child development theory, and surveys of teachers who have left the profession, strongly suggest that successful solutions should include:

1. pay levels comparable to those with similar preparation and talent *outside* the teaching profession;
2. careful and early selection of teacher candidates based on academic talent, and selected social attributes like empathetic accuracy, responsiveness, and commitment;
3. providing professional working conditions, including professional levels of autonomy;

4. building well-designed mentorship and collaborative working environments and making them the norm;
5. matching teacher demand to supply; and
6. the use of child development theory to focus the right services and environments where they are most cost-effective, especially early in children's lives.

This proposed experiment offers an opportunity to overcome some cultural and institutional character flaws. It would systematically select, train, and place our most capable teachers with our highest-need children and families. It deliberately limits turnover. It limits costs in several ways:

- by setting in place preventive approaches based on child development theory;
- by reducing administrative overhead;
- by redirecting resources from testing companies to better teacher selection, preparation, and retention; and
- by limiting the experiment to the highest-need schools.

Do we have what it takes to look at these problems clearly, and take the most rational steps to improve our institutions?

References

Alliance for Excellent Education (2005). Teacher attrition: A costly loss to the nation and to the states. Issue Brief. <http://www.all4ed.org/files/archive/publications/TeacherAttrition.pdf>

Alliance for Excellent Education (2008). What keeps good teachers in the classroom? Understanding and reducing teacher

turnover. Issue Brief. <http://www.all4ed.org/files/TeachTurn.pdf>

Allensworth, E.M., and Easton, J.Q. (2007). What matters for staying on-track and graduating in Chicago public high schools. Univ. of Chicago: Consortium on Chicago School Research.

Auguste, B., Kihn, P. and Miller, M. (2010). Closing the talent gap: Attracting and retaining top-third graduates to careers in teaching. Washington, DC: McKinsey. http://www.mckinsey.com/App_Media/Reports/SSO/closing_the_talent_gap_september_2010.pdf

Bacolod, M.P. (2007). Do alternative opportunities matter? The role of female labor markets in the decline of teacher quality. *The Review of Economics and Statistics*, 89, 4, 737-751. http://www.socsci.uci.edu/~mbacolod/Bacolod_teachq.pdf

Baker, E.L., Barton, P.E, Darling-Hammond, L., Haertel, E., Ladd, H.F., Linn, R.L., Ravitch, D., Rothstein, R., Shavelson, R.J., and Shepard, L.A. (2010). Problems with the use of student test scores to evaluate teachers. Washington, DC: Economic Policy Institute. <http://www.epi.org/publications/entry/6276/>

Barber, M. and Mourshed, M. (2007). How the world's best-performing school systems come out on top. McKinsey. http://www.mckinsey.com/client-service/Social_Sector/our_practices/Education/Knowledge_Highlights/Best_performing_school.aspx

Bill & Melinda Gates Foundation (2010). Working with teachers to develop fair and reliable measures of effective teaching. Also see: Learning about teaching: Initial findings from the measures of effective teaching project. Downloaded on 19 Jan 2011 from: <http://www.gatesfoundation.org/learning/Pages/2010-reliable-measures-effective-teaching.aspx>

Borman, G.D., and Dowling, N.M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research*. 78, 3, 367-409.

Boyd, D., Goldhaber, D., Lankford, H., and Wyckoff, J. (2007). The effect of certification and preparation on teacher

- quality. *The Future of Children*, vol. 17, no. 1, Spring, pp. 45-68.
http://www.princeton.edu/futureofchildren/publications/docs/17_01_03.pdf
- Braun, H.I. (2005). Using student progress to evaluate teachers: A primer on value-added models. Princeton, NJ: Educational Testing Service. Downloaded from www.ets.org/research/pic
- FPG Child Development Institute, see the Carolina Abecedarian Project website at <http://www.fpg.unc.edu/~abc/#home>
- Geary, D.C. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. J. Carlson and J.R. Levin (editors) *Psychological Perspectives on Contemporary Educational Issues*. Greenwich, CT: Information Age Publishing.
- Goe, L. and Croft, A. (2009). Methods of evaluating teacher effectiveness. National Comprehensive Center for Teacher Quality, Research-to-Practice Brief. <http://www.tqsource.org/tqtopractice.php>
- Guarino, C.M., Santibañez, L., and Daley, G.A. (2006). Teacher recruitment and retention: A review of the recent empirical literature. *Review of Education Research*, 76, 2, 173-208.
- Hampden-Thompson, G., Herring, W.L., and Kienzl, G. (2008). Issues brief: Attrition of public school mathematics and science teachers. U.S. Dept. of Education: National Center for Education Statistics.
- Hanushek, E.A. (2002). The long-run importance of school quality. National Bureau of Economic Research: Working Paper No. W9071.
- Harris, D.N. (2009). The policy uses and “policy validity” of value-added and other teacher quality measures. In D.H. Gitomer (editor), *Measurement Issues and the Assessment of Teacher Quality*. Thousand Oaks, CA: Sage.
- Harris, D.N. and Adams, S. (2007). Understanding the level and causes of teacher turnover: A comparison with other professions. *Economics of Education Review*, 26, 325-337.
- Heckman, J.J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, June 30, vol. 312, 1900-1902.
- Hewlett Foundation, see grants to promote deeper learning at <http://www.hewlett.org/grants/grantseekers/deeper-learning-grantseekers>
- HighScope Perry Preschool Study, see research summaries at <http://www.highscope.org/Content.asp?ContentId=219>
- Ingersoll, R.M. (2001). Teacher turnover, teacher shortages, and the organization of schools. University of Washington: Center for the Study of Teaching and Policy. <http://depts.washington.edu/ctpmail/PDFs/Turnover-Ing-01-2001.pdf>
- Ingersoll, R.M. (2003). Is there really a teacher shortage? University of Washington: Center for the Study of Teaching and Policy. <http://depts.washington.edu/ctpmail/PDFs/Shortage-RI-09-2003.pdf>
- Ingersoll, R.M. and Merrill, L. (2010). The changing face of the teaching force. *GSE@Penn: A Review of Research*, 7, 3, 1-10.
- Ingersoll, R.M., and Perda, D. (2010). Is the supply of mathematics and science teachers sufficient? *American Educational Research Journal*, 20, 3: 1-32.
- Kane, T.J., Taylor, E.S., Tyler, J.H., and Wooten, A.L. (2010). Identifying effective classroom practices using student achievement data. National Bureau of Economic Research: Working Paper No. 15803. <http://www.nber.org/papers/w15803>
- Keigher, A. (2010). *Teacher Attrition and Mobility: Results From the 2008-09 Teacher Follow-up Survey* (NCES 2010-353). Washington, DC: National Center for Education Statistics. <http://nces.ed.gov/pubsearch>.
- Markow, D., and Cooper, M. (2008). The MetLife survey of the American teacher: Past, present and future. pp. 144-145.

- <http://www.metlife.com/assets/cao/contributions/citizenship/teacher-survey-25th-anniv-2008.pdf>
- Martins, P.S. (2009). Individual teacher incentives, student achievement and grade inflation. IZA Discussion Paper No. 4051. <http://ftp.iza.org/dp4051.pdf>
- Marvel, J., Lyter, D.M., Peltola, P., Strizek, G.A., Morton, B.A. (2007). Teacher attrition and mobility: Results from the 2004-05 teacher follow-up survey. Washington, DC: National Center for Education Statistics.
- Mishel, L., Allegretto, S., and Corcoran, S. (2008). *The Teaching Penalty: Teacher Pay Losing Ground*. Washington, DC: Economic Policy Institute. www.epi.org.
- Murnane, R.J., and Steele, J.L. (2007). What is the problem? The challenge of providing effective teachers for all children. *The Future of Children*, vol. 17, no. 1, Spring, pp. 15-43. http://www.princeton.edu/futureofchildren/publications/docs/17_01_02.pdf
- National Center for Education Statistics (2009). Table 66. Teachers, enrollment, and pupil/teacher ratios in public elementary and secondary schools by state or jurisdiction: Selected years, fall 2000 through fall 2007. *Digest of education statistics*. http://nces.ed.gov/programs/digest/d09/tables/dt09_066.asp
- OECD (2011). *Education at a Glance: OECD Indicators*. 2007 data from Table B2.1. http://www.oecd.org/document/52/0,3746,en_2649_39263238_45897844_1_1_1_1,00.html
- Parsons, B., Lupe, C., and Bosserman, C. (2002). Using data to improve teacher induction programs. Washington, DC: The NEA Foundation for the Improvement of Education. www.nfie.org
- Pink, D.H. (2009). *Drive: The Surprising Truth About What Motivates Us*. New York, NY: Riverhead.
- Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*. New York, NY: Basic Books.
- Rivkin, S.G, Hanushek, E.A., and Kain, J.F. (2005). Teachers, schools, and academic achievement. *Econometrica*, 73, 2, 417-458.
- Rothman, R. and Darling-Hammond, L. (2011). Issue Brief: Teacher and school leader effectiveness: Lessons learned from high-performing systems. Washington, D.C.: Alliance for Excellent Education.
- Rothstein, J. (2007). Do value-added models add value? Tracking, fixed effects, and causal inference. Downloaded on 19 Jan. 2011 from <http://www.uh.edu/~adkugler/Rothstein.pdf>
- Sackett, P. (2010). What we know about non-cognitive constructs and work outcomes. Lecture slides available at Educational Testing Service, www.ets.org/c/15481/ppt/sackett_sessionII.ppt
- Sanders, W.L., and Wright, S.P. (2008). A response to Amrein-Beardsley (2008) "Methodological concerns about the education value-added assessment system." Document downloaded 16 May 2011 from http://www.oxydiane.net/IMG/pdf/Sanders_Wright_response_to_Amrein-Beardsley_4_14_2008.pdf
- Shulman, L.S. (2009). Assessment of teaching for teaching? Reflections on the invitational conference. In D.H. Gitomer (ed.) *Measurement Issues and Assessment for Teaching Quality*. London: Sage, pp. 234-244.
- Smith, T. M., and Ingersoll, R.M. (2004). What are the effects of induction and mentoring on beginning teacher turnover? *American Educational Research Journal*, 41, no. 3, 681-714.
- Teach for America (2011) selection criteria are available at <http://www.teachforamerica.org/admissions/who-were-looking-for/>
- Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., and Gottesman, I.I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*,

14:623-628.

[http://people.virginia.edu/~ent3c/papers2/Articles%20for%20Online%20CV/\(38\)%20Turkheimer%20et%20al%20\(2003\).pdf](http://people.virginia.edu/~ent3c/papers2/Articles%20for%20Online%20CV/(38)%20Turkheimer%20et%20al%20(2003).pdf)

Wei, R.C., Darling-Hammond, L., and Adamson, F. (2010). *Professional development in the United States: Trends and challenges*. Dallas, TX: National Staff Development Council.

Weisberg, D., Sexton, S., Mulhern, J., and Keeling, D. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Second Edition. Brooklyn, NY: The New Teacher Project.
<http://tntp.org/publications/reports/the-widget-effect/>

Yoon, K.S., Duncan, T., Lee, S., W.-Y., Scarloss, B., and Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Washington, D.C.: U.S. Dept. of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Southwest Laboratory.