The Accidental Revolution:
Teacher Accountability, Value-Added, and the Restructuring of the American School System

February 7, 2012

Chris Thorn
Douglas N. Harris

There is little question that test-based school accountability, driven by *No Child Left Behind*, has been one of the most influential policies in the nation's education history. It has changed what teachers teach, how they teach it, how teachers think and feel about their work, and how school leaders allocate resources. While still early in its development, NCLB's offspring—*Race to the Top*—may be having an even greater impact. With more advanced and interconnected data systems spawned by NCLB, it is increasingly possible to evaluate not just whole schools, but individual teachers and leaders on their "value-added" to student achievement. This shift represents a sea change in how schools operate, ending the half-century long autonomy that teachers have come to expect. The balance of power in school decision-making has shifted away from teachers, unions, and schools of education—what some call "the establishment"— toward testing companies, data managers, district department heads, school principals, and state and federal policymakers.

These changes have been substantially intentional. The goals of test-based accountability advocates have been to increase accountability for school leaders and individual educators, and to provide the analytic tools to help them respond to those pressures and improve learning. Perhaps without even realizing it, however, educator accountability is re-structuring roles and relationships throughout school districts and yielding entirely new organizations with which districts interact in complex ways. As we show through a series of case studies, this amounts to a major re-structuring of the American school system, and one that has been largely accidental.

This restructuring is a natural consequence of the fundamental rethinking of the education system engendered by the shift from school-level to teacher accountability. With school-level accountability, individual teachers are largely insulated from accountability. The worst case— and very rare—scenario with school accountability is that a school will be closed and the

teachers dismissed. Teacher accountability in contrast reaches every teacher in every school, as well as a host of indirect and secondary influences that we explore through a series of case studies.

When the stakes were low, test designers (mostly for-profit companies) received little scrutiny and their decisions had relatively little influence on the working lives of individual teachers. Our discussion of Hillsborough County shows that the testing companies do not yet have the capacity to meet the demands of these new systems and this means that, for now, assessments will have to be designed in different ways, by other groups. When data were only available at the school-level, a simple spreadsheet was all that was needed for data management. Our discussion of Hillsborough County and Center for Educational Innovation - Public Education Association show that for-profit database management and/or data warehouse companies now play an integral role. This has been driven, indirectly and apparently unintentionally, by educator accountability. Without it, there would be little impetus to change.

Educator accountability has generated a need to *coordinate*. In the past, the assessment director might rarely have spoken with, let alone coordinated with, the director of professional development or human resources. Another recent federal accountability program, the *Teacher Incentive Fund*, requires that teacher and principal professional development be explicitly targeted at weaknesses identified by individual teacher evaluations. To do that, different district administrators have to talk with one another.   The linkage between evaluation results, selection and assignment of appropriate professional support resources, the provision of those services to those individuals across many schools requires a level of coordination and cooperation uncommon to the traditional American school district. Educator accountability is leading to the integration of data from assessments, human resources, professional development, and

curriculum departments, making it possible for them to coordinate. This pressure has spawned a breakdown in data silos and management isolation so that districts can succeed in meeting state and federal mandates and to ensure that individual teacher performance measures can be fair and accurate.

When teacher evaluation lacked teeth, it mattered relatively little what principals thought of their teachers. New systems of evaluation not only require formal evaluation of individual teachers, but that action be taken to address shortcomings and principals necessarily play an important role in that. Many *Teacher Incentive Fund* programs also create new roles for teacher leaders – master or mentor teachers, peer evaluators, etc. The organizational changes at the school level required to distribute leadership tasks and embed professional development in the working day are substantial. Our discussion of the three *Teacher Incentive Fund* projects, which all include teacher and principal evaluation components, suggest that times are changing.

To make teacher accountability a reality, many state and federal policymakers have decided to usurp responsibilities for evaluation, compensation, and dismissal from teacher unions. Once these big leaps are taken, it is only a short step to accountability for the programs and institutions that prepare teachers. We discuss the case of the Network of Effective Teaching (NExT) in which participating schools of education (SOEs) provide guarantees of their graduates' effectiveness. In the past, schools of education could largely do what they wished. Graduates enter teaching, or not, teach well, or not. Perhaps the most positive role of the SOEs was to identify teachers who were committed enough to jump through all the hoops the programs required, even if those hoops did little to prepare teachers for the classroom. Since schools themselves were not focused on effectiveness in the hiring process (Harris, Rutledge, Ingle, & Thompson, 2010), graduates could find jobs in any event. But once we can evaluate individual

teachers, it takes little imagination, and almost no statistical knowledge, to place teachers into groups based on where they were prepared and to compare average performance across SOEs and other training institutions.

While little research and evaluation on these reforms has been published, we discuss ongoing cases of reform with which we have had direct experience. One of us (Thorn) works for a non-profit organization to which power has shifted: the Value-Added Research Center (VARC) at the University of Wisconsin at Madison. VARC is a supplier of value-added measures and professional development to school districts seeking to use it. The other (Harris) works with states and school districts across the country in an advisory capacity as they implement value-added-based accountability. Drawing on these experiences allows us to provide an up-to-date analysis of what is happening in this fast-moving environment. We also keep our eyes on what these experiences mean for fundamental questions of school reform, governance, and management. Test-based accountability is part of a larger, long-term trend toward performance management in the public sector more broadly (Heinrich & Marschke, 2010; Moynihan, 2010).

We continue in the next section by discussing the one element that all the cases have in common: teacher value-added estimates. Because we are interested in how accountability is changing the ways in which schools and districts operate, we leave it to others who have discussed the technical properties of the measures (e.g., Harris & Sass, 2006; Harris, 2009, 2011; Glazerman et al., 2011). We then discuss four cases that have teacher accountability based on value-added at their core. Many of these cases are in their early stages and, while there is little formal research, we draw on first-hand experiences and anecdotes that lead to a set of observations and questions, all of which we believe have important implications for the future of education accountability as well as its effects on the daily lives of educators and student

outcomes. It appears to us that the pressure for individual accountability is leading to a possibly widespread and permanent restructuring at every level of the American school system.

Value-Added and Observational Measures of Educator Performance

At the heart of the teacher accountability movement, and our case studies, are value-added measures of performance. While the statistical models differ from site to site, the basic concept is the same. Rather than rely on snapshots of achievement at a point in time or on changes in school- or grade-level average test score, value-added models take into consideration the prior level of achievement of each child. Children come into classrooms at different starting points. Value-added models capture the progress students make, given where they start. There are a number of different modeling choices that one might consider – the inclusion of student demographics, multiple years of pre-tests, classroom characteristics, etc., but these are extensions of the basic notion of capturing the contribution of the teacher.[1]

Aside from any modeling challenges (deciding what features are appropriate for a particular use), there are substantial technical difficulties associated with implementing teacher-level value-added models. The biggest problem is having high quality attribution (linkages) between teachers and students by subject taught. This may sound trivial, but most student information systems (SIS) are designed to manage scheduling and basic record keeping. They are not designed to manage the complexity of how learning is delivered to individual students. For example, most SISs indicate only the teacher of record for any given room or course section. This does not include formal or informal team teaching arrangements, ability grouping, ELL or special education aides. If one were to use simple administrative data from the SIS, the resulting value-added results would contain a number of unknown attribution errors that could seriously

---

[1] See (Harris, 2010; Kane & Staiger, 2012) for more.

impact the effectiveness scores of the teachers involved. Another problem associated with student linkage is detailed information on *dosage* – or the amount of instructional time received by each student preparing for any given assessment. Many students receive after school tutoring or other services that are not traditionally captured in SIS records.[2]

There are several groups working at the state, district, and school level to provide tools to overcome the limitations of SIS data. Battelle for Kids provides support for this at the state and district level through a linkage tool they have developed called BFK•Link®. The National Institute for Excellence in Teaching also provides a comprehensive data collection and linkage system that supports called CODE that supports schools using the Teacher Advancement Program (TAP). Many TIF districts have developed simple online or manual student roster verifications without any intelligent automated processes around it to lessen the burden on users. Imagine a solution that provides teachers with pre-populated spreadsheets that contain course rosters based on the information in the student information system. If a student is missing from the roster, there is no automatic feature search function to help find that student in a building-level list. The verifying teacher must hand enter the information with the resulting chance of error when that data is compared to online data at a later point – spelling differences, typos, etc. all introduce error. At the building level, one of the important roles of the principal is to make sure that no students are "unclaimed" by teachers or that all teachers have been linked to classrooms to reflect team teaching or other more complex organizational models. These tasks are relatively simple tasks in automated linkage systems. The same tasks are quite labor intensive in a manual system. The other major difference in automated systems is the ability to capture the linkage of support services to individual students. Many students are served by a number of other

---

[2] See (Battelle for Kids, 2011; Watson, Witham, & St. Louis, 2010) for more.

adults. Dedicated linkage systems can more easily capture these more complex relationships – things that often don't fit in an SIS.

In addition to data quality problems, there are concerns about the quality of the assessments available. Most state-mandated assessments were created to measure particular proficiency cut points at each grade. They are less accurate measuring students who are performing above or below grade level. There are also concerns about the level and types of mastery that can be measured with traditional fill-in-the-bubble assessment formats. Many curricular standards direct teachers to bring students to a level of mastery that allows them to apply concepts learned during the year to novel problems – not just recitation of facts or formulae. There can be a significant mismatch between the level of mastery that can be measured by the available assessment and the educational goals of the school, district, or state.[3] Those assessments are most accurate for measuring the performance of student at or near those cut points.

The federal assessment consortia – Partnership for Assessment of Readiness for College and Careers (PARCC) and the SMARTER Balanced Assessment Consortium (SBAC) – are intended to address some of the concerns about test quality in the core grades and subjects. The PARCC group will provide both interim and summative, computer-based, performance based assessments in grades 3-8 in mathematics and English language arts based on the Common Core Standards. There will also be grade-based high school assessments in the same subject areas and formative, vertically linked items for grades K-2.[4] The SBAC will provide online, computer adaptive summative assessments math and English language arts in grades 3-8 and 11 based on the Common Core Standards. The group will also provide optional interim and formative

---

[3] See (Baker et al., 2010; Koretz, 2008) for more.
[4] http://www.parcconline.org/parcc-assessment-design

assessment for capturing student progress through the year.[5] Both groups will be using common interim assessments.

Because almost all states will be implementing these assessments, it will be possible for the first time to compare educator performance across states. This could open up new areas of institutional cooperation and collaboration. As others have pointed out regarding the benefits of common standards and assessments (e.g., Harris & Goertz, 2008), this will facilitate professional development and other bases for school improvement that is both curriculum specific and cuts across states. This could increase the roles of national companies and non-profits who could take advantage of economies of scale (Harris & Taylor, YEAR).

Under NCLB, which covers only Math & English/Language Arts in grades 3-8), educational agencies have already begun purchasing commercial assessments in other grades and subjects. In other locations, the creation of new assessments has been an explicit partnership between leadership and teachers and curriculum experts. It may be that localized test design becomes an important trend over the next several years, but it is difficult to see how this will stand up to political demands for standardization and economies of scale – not to mention potential profits that could be leveraged by the large testing companies.

Moreover, in our recent experiences working with school districts that are implementing value-added measures for accountability, we also observe a general crumbling of the walls and silos that separate school district administrative activities. While in many cases this has positive implications since many silos have resisted evaluating their own practices or using data at all to make decisions, it is often done with little planning or attention to the staff development needs of the actors involved. Informal discussions with leadership teams across the country suggest that basic assessment and data literacy are critical problems for district staff. Grant requirements for

---

[5] http://www.k12.wa.us/SMARTER/FAQ.aspx

Teacher Incentive Fund recipients to link professional development to adult needs identified by evaluation results provide a unique opportunity to link this important source of "inputs" to outputs. The data linkages and data quality needed to routinely evaluate educational interventions are actually being put into place. What is missing in most schools and districts are staff members with the analytical training to design and implement robust evaluation designs. This skill gap stands in the way of schools and districts doing their own research on "what works".

This shift in large scale accountability is based on the assumption that teacher effectiveness measures, both value-added results and observational measures of teacher practice, meet data quality requirements necessary to accurately attribute connections between adults in schools to students by subject or program area. One of the challenges is that the assumptions about validity and reliability at the macro level are likely do not recognize the variability of local programs or practices. The challenge will be to ascertain which validity assumptions are met and what the implications of misfits are for using the accountability data.

In addition to the use of measures of student growth on assessments, all of the examples cited here rely on observation as another critical element of understanding effective teaching. There is a strong sense across most scholars who study teacher effectiveness that observational measures of teacher practice are important to understand which practices are most effective at delivering growth in student learning. In untested grades and subjects, observational measures of practice are often the only objective measure available that attributable to the individual teacher. Across TIF sites, adaptations of the Danielson Framework[6] are very common. Other frameworks in use include Pianta's CLASS[7] or Marzano's Teacher Effectiveness Framework[8]. All TIF grants

---

[6] http://charlottedanielson.com
[7] http://curry.virginia.edu/research/centers/castl/class

are required to observe teachers and principals at least twice a year and demonstrate procedures for achieving high inter-rater reliability between scorers. Many larger districts have struggled to scale these systems to robustly collect results, share those results with leadership, and support targeted responses to identified teacher or principal needs. Requirements to isolate human resource records from other data have induced many TIF districts to create standalone systems to support these needs.

While our focus, and that of reformers, is on teacher accountability, the *Race to the Top* and related reforms also involve revamped evaluations of administrators. These, too, are based on a combination of student learning and measures of practice. This will no doubt have significant implications for teacher-leader relations and perhaps school organization, although we focus here primarily on higher bureaucratic levels—districts, unions, non-profits, and companies.

The success of the current reform movement will depend substantially on the quality of data collection and validity and reliability of all the associated performance measures. This part of the process itself—regardless of the accountability mechanisms attached to performance—will likely influence the structure of schooling systems. We are uncertain about how these organizational shifts will play out. There seems to be a growing appetite across the country for value-added measures as evidenced by numerous *Race to the Top* and *Teacher Incentive Fund* grant applications and related legislation. While this analysis is only possible in the 30% of grades and subjects covered by mandated state testing, there is pressure to expand the assessment environment to other grades and subjects to provide objective measures of educator contributions to student learning in those areas. At the same time, states are struggling to maintain existing testing regimes in the face of shrinking departmental budgets. There is also a tension between the

[8] http://www.marzanoevaluation.com/

design to measure productivity – which give high level feedback about whether educational systems are working – and detailed, timely feedback in the classroom that can be used to improve day-to-day instruction. Most recently, an education blogger noted the evolution in Education Secretary Duncan's public statements about value-added measures. He roughly equated teacher value-added scores with educator effectiveness in August 2010. In May 2011, he spoke out in favor of the combination of value-added and observational measures. On February 6th, 2012 he is quoted as stating that test scores should not be used as a part of teacher evaluation.[9] On January 31st, 2012, Texas Education Agency Commissioner Robert Scott spoke to a large crowd of Texas educators stating that "testing was 'good for some things,' but that in Texas it has gone too far" and would probably experience a backlash in the next legislative session (Smith, 2012). It is in the light of this rapidly evolving policy environment that we consider 4 short case studies of reform efforts in the field.

Cases Studies in Teacher Accountability and System Restructuring

While all of the cases we consider include teacher value-added measures as a core component, they also differ in important respects. Even the three cases that all have Teacher Incentive Fund (TIF) funds, which have common federal requirements, have played out differently across the three contexts. We start, however, with a distinctive case focused on teacher preparation.

Case #1: Accountability and Teacher Preparation in the Three Midwest States[10]

The Network for Excellence in Teaching (NExT) is a 10-year partnership of 14 teacher-preparation programs and the Bush Foundation. The NExT project is focused on increasing the

---

[9] http://www.schoolsmatter.info/2012/02/evolution-of-arne-duncans-position-on.html
[10] http://www.bushfoundation.org/education/network-excellence-teaching-next

overall quality of teachers in the three partnering states – Minnesota, North Dakota, and South Dakota – over the next decade. The project will do this by marketing to young adults most likely to be effective teachers, improving teacher preparation programs, creating a more systematic hiring process, and providing adequate support to new teachers to move them from novice to mastery in their first three years on the job (Mohr, Broton, & Muelle, 2010).

This four-part strategy addresses what the Foundation and participants agree are the major levers for improving teacher quality and student learning. The Foundation has supported a general marketing campaign targeting high-achieving high school students and undergraduates to encourage them to consider teaching as an exciting and rewarding career. Individual IHEs have created targeted campaigns that focus on particular communities or program strengths. Some programs have struggled to recruit high quality undergraduates to their programs. One other important element of the strategy has been the engagement of both state-level and local union representation. The Foundation has supported outreach efforts to the state chapters and has included union leaders at many of the NExT meetings between IHEs and partner districts.

The Foundation has also examined school district hiring practices. One of the promises of the value-added analysis is to working with all IHEs and their partnering districts to better understand how schools make hiring decisions and how well they are able to select teachers who turn out to be highly effective. Prior research suggests that schools select teachers as much on their affective traits and how well they fit in with school culture as their teaching ability (Harris, Rutledge, Ingle, & Thompson, 2010).

The most striking aspect of NExT is that participating IHEs will provide a guarantee of effectiveness for each program graduate (Hakes, 2011). The guarantee is made to each student, but, of course, it plays into the hiring process since the guarantee is met by meeting specific

performance requirements. The language used by the Foundation in its original organizing materials was that teachers from NExT IHEs should *deliver a year of growth*. Operationalizing the intent of the project is more complex. The policy goal is to shift the distribution of new teacher effectiveness *to the right.* While this is policy intent, we also know that most teachers are not teaching in grades or subjects with mandatory state tests. We also know that IHE programs differ, their student populations differ, and they have different program goals that make comparison difficult. One of the most important aspects of the program design and the commitment to a guarantee remains the focus on improving the performance of every student who enters a teacher preparation program.

On the part of IHEs, that has led to substantial program redesign[11]. Pre-service students, in most cases, engage in practicum experiences earlier in their preparation – primarily co-teaching, receive support and performance feedback during that period and each of the first three years after their placement as a teacher. There is no single model, but the IHE program leaders spent substantial time together in planning sessions sharing insights from their own programs, discussing research on effective teacher preparation programs, and working with the same set of external consultants who helped all participants prepare the proposals. The differences in the programs reflect rural or urban locations, religious affiliation or public status, undergraduate versus graduate programs, etc. What is common to all is that the graduating institution is responsible for providing professional development to all of its teachers – *post placement* – to bring them to mastery and satisfy the guarantee. The guarantee differs by institution, but it includes ongoing support and training.

How is it that 14 IHEs have come to participate? Voluntary participation and the collaborative process are two key reasons. One of the major cooperative activities between the

---

[11]

14 participating IHEs has been a series of meetings to suggest, craft, and adopt *Common Metrics*[12] that will allow programs to gather consistent information on participations and program features through the entire life-cycle of the program. Since this large scale data collection and analysis effort extends across all partners and multiple states, strategies that would leverage the licensure or certification systems of individual states would not be sufficient.

Given the high-stakes involved for the schools of education, the program requires detailed and accurate teacher performance measures. The NExT initiative requires a series of data collection, data management, and analytical innovations that go beyond what is currently fielded by either district or state agencies. It is not that the analytics themselves are that different. The novelty is the self-imposed accountability and the required buy-in from both the IHE and hiring schools to share data to evaluate teacher effectiveness. This partnership will require that IHEs and districts are more transparent and collaborative in structuring their pre-services placements and mentoring services after initial hiring, if the districts are to take advantage of the support services provided by the IHEs for their graduates. Districts may also chose to use the value-added results they receive for other purposes, such as program or project evaluation. The Bush Foundation acquired the services of a commercial survey vendor to assist the participants in the collection, organization, and analysis of survey results for each of their programs – with the ability to compare their own results to the aggregate results of the entire 14 member initiative.

The other primary challenge of the NExT initiative is to provide analytical support for creating and using effective measures for the IHE partners. This includes calculating teacher-level value-added results for all program participants using state-mandated and locally adopted assessments. The state has provided statewide assessment data that allows the project team to

---

[12] http://blog.lib.umn.edu/cehd/teri/docs/Common%20Metrics.pdf

calculate school and grade level value-added. For most of Minnesota and North Dakota,

NWEA's Measures of Academic Progress is being used as a local assessment. The majority of

those districts have allowed VARC to have access to these data as well to provide school and

grade level value-added for this local assessment. The details of this aspect of the project are still

emerging as IHEs finalize their program designs and begin to put the first cohorts through there

redesign programs. The original intent was to create teams across IHE and partner districts to

share what assessments they currently have in place to encourage some collective action. New

teacher evaluation legislation in Minnesota has led to the creation of state-level working groups

to explore how the requirement for measures of student growth will be implemented. In this case,

the original program design has been overtaken by events. The original intent of the NExT

project was to move slowly towards building consensus around a well-understood model of

teacher effectiveness. None of the original members of the initiative considered it likely that

many states would be moving to implement educator effectiveness with test and observational

measures. Very shortly, we will have the decisions of the working groups in Minnesota charged

with recommending how these measures should be calculated. It is hoped that the NExT project

will be able to use these new assessment as the core assessments for the IHE guarantees. Where

possible the guarantee of new teacher effectiveness will be in comparison to average value-added

in a set of baseline years. VARC is the Foundation's partner is this work. While school and

grade-level value-added results can be calculated using state-level student administrative and

assessment data, the ability to calculate classroom and teacher value-added scores requires

access to high-quality student-teacher-course linkage data that is currently only available at the

school level and requires engagement and verification by individual teachers. The social,

technical, and political challenges of implementing these data collection efforts and engaging

local staff to expend the effort to achieve the required high quality data are not trivial.

Case #2: Merit Pay in Traditional Public Schools [13]

The Teacher Incentive Fund (TIF) program was created in 2006 (U.S. Department of

Education, November 14, 2006) as a competitive grant program administered by the U.S.

Department of Education that was intended to allow schools, school districts, state agencies, and

non-governmental partners to design and implement performance-based compensation programs.

There were two rounds of grants made under the original funding. The appropriation was

allocated in two rounds of grants. The first, referred to the by the Department of Education as

Cohort 1 received a total of $42 million award across 15 grantees in November 2006. The second

group of 18 recipients – referred to as Cohort 2 – received $38 million in June 2007 There was a

third round funding announced in May 2010 (U.S. Department of Education, 2010) and awards

made in September 2010. The third round of TIF grants - Cohort 3 – were made from an

appropriation of $437 million. The grants were awarded to 62 recipients – 50 grantees in the

general program and 12 grantees who also agreed to participate in a randomized, controlled trial

of their designed program against a simple whole school bonus treatment. The focus of this

program is stated in the press release announcing the most recent call for proposals as follows;

"TIF grants support local projects that reward teachers, principals and other school personnel

who improve student achievement."[14] This basic notion has been operationalized differently

across the three rounds of funding. Changes to the program requirements for Cohort 3 reflect a

substantial shift from a reliance on compensation alone to a set of measures (growth on student

assessments, proficiency for teachers and leaders on observational rubrics) and interventions

---

[13] http://www.maricopa.gov/schools/service-home.aspx?sid=1&cid=1
[14] http://www2.ed.gov/news/pressreleases/2010/05/05202010.html

(performance pay, new roles for teachers, and targeted professional development for identified needs) designed to achieve the desired core policy goal of improving outcomes for students.

The Rewarding Excellence in Instruction and Leadership (REIL) initiative is a multi-district TIF Project that has a regional Education Service Agency as its fiduciary agent and coordinating partner. The project team at the Maricopa County Educational Services Agency (MCESA), in addition to supporting the development of quality student teacher linkage data for all grades and subjects, is also developing a system for the collection and reporting of teacher observation results, building tools for analyzing the budget impact of different teacher compensation scenarios, and assisting in the pilot testing of the state's newly purchased system for linking all secondary education courses to a standard database of NCES course codes linked to common core standards. Arizona is a *right to work* state, but has an active teacher representation group, the Arizona Education Association, which explicitly supports the goals of the project and participates in the work through its members.

The other important area of work at the regional level is a pilot effort that, while not as extensive or mature as the Hillsborough case we discuss later, uses local curriculum experts from across the regional service area to create high quality items to build local assessments that could be used in grades and subjects that are not covered by state-mandated tests. Unlike Hillsborough, MCESA has pursued a mixed strategy of working with an existing state assessment vendor to partner with them to help them accelerate their expansion into other core subjects as well as to expand to other grades. MCESA has also recruited test developers from WestEd provide technical guidance and quality assurance to their efforts to create assessment in traditionally untested such as fine and performing arts. The project was expanded beyond the original TIF partner districts to get enough experts across the many curricular areas and levels to provide the

required level of expertise and diversity of experience. It may be that MCESA's responsibilities as a regional service agency provide them with economies of scale and access to more partners to support a more complex approach to filling in the gaps in traditionally non-tested grades and subjects. The requirements of creating high quality assessments as well as the required software for item collection, editing, refinement, testing, and publishing are all included in this ambitions effort.

The REIL project, like all TIF projects, also include a observational framework for teachers[15]. They regional service agency and the participating districts have collaboratively developed an evaluation rubric and are field testing it during the 2011-2012 academic year[16]. The district is also developing supporting online tools to collect individual observation results across all schools and provide aggregate reports for school leaders and individual reports back to individual teachers and mentor teachers to provide individualized feedback and a basis for the allocation of additional professional support, if needed.

The REIL project, like many other TIF projects, is phasing in its compensation system. The REIL project targets only high poverty, hard to staff schools. The pilot year focused on the development and field-testing of the observational framework and exploration of value-added approaches for the current assessment system. Year two will see the first implementation of the observational system. Highly effective teachers will initially be identified using the observational measures. In year three, teacher effectiveness scores will be a combination of observational results and either classroom or school value-added scores.

---

[15] http://www.maricopa.gov/schools/webcontent/docs/REIL_Learning_Obs_Instr_7-11-11_lr_51.pdf
[16] See the rollout plan at
http://www.maricopa.gov/schools/webcontent/docs/REIL_WebcastLearningObservationInstrument_51.pdf

...

-- DRAFT --

Case #3: Accountability and School-Level Decisions in Hillsborough County, Florida[17]

The Hillsborough project is the result of a Bill and Melinda Gates Foundation Intensive Partnerships for Effective Teaching grant, but also includes a combination of other funding sources (including a U.S. Department of Education Teacher Incentive Fund grant) and leverages strong labor-management engagement, leadership stability, and a 30 year tradition of developing local assessments.

The district has implemented teacher-level value-added measures based on either state-mandated end of grade or locally developed (or purchased) end of course assessments. This means that while most states assess students in grades 3 through 8 in mathematics and reading/language arts, 2 times in science and math during that period, and then again once high school in those subjects, Hillsborough assesses student performance in nearly all grades and subjects. The district fields around 700 different tests each year to cover all subjects and student electives.[18] To our knowledge, Hillsborough has gone much further in creating their own assessments than any other district in the country. This also means that the district has invested substantial efforts in training teachers and administrators to understand and apply the skills required to develop high quality test items.

Indeed, the data requirements for this work have induced the district to partner with an outside data warehouse and data analytics partner (Convergence Consulting Group) to support this complex, ongoing work[19]. In the past year, the district has also included support for improving their student-teacher-course linkages and their ability to deliver reports and embedded professional develop around the reports to support their interpretation. The district also provides

---

[17] http://communication.sdhc.k12.fl.us/eethome/
[18] http://www.sdhc.k12.fl.us/eet/v6/
[19] http://www.convergenceconsultinggroup.com/convergence-consulting-group-recognized-as-premier-small-business/

a separate website with extensive information about the initiative with newsletters, videos, and

answers to over 230 *Frequently Asked Questions*[20]. This site represents an aggregated approach

to dissemination that has material for teachers, parents, and the community at large. Resources

on the site include technical documentation on implementing the rubric for observing effective

teaching as well as the entire history of newsletters, press releases, and media coverage of the

project.

Teacher value-added measures represent 40% of the overall measure of an individual

teacher's effectiveness. The value added result is combined with rubric-based classroom

observation scores from principals (30% of total) and peer and mentor teachers (the remaining

30%) to create the measure of overall effectiveness. The supporting website for the program

provides the actual rubrics for teachers serving in various roles[21] and extensive questions and

answers about the details of the evaluation process, the role and training of peer and mentor

observers[22], and other critical elements of the implementation of the observational system.

Case #4: Merit Pay in Charter Schools[23]

The Partnership for Innovation in Compensation for Charter Schools (PICCS) project is

in its second iteration. The first effort was in the second round of TIF grants and is now winding

down. The new project leverages what was learned from that work as well as the technical

infrastructure. This TIF round 3 project includes another group of independent charter schools in

New York City and Buffalo, New York.[24] The non-profit Center for Educational Innovation -

Public Education Association (CEI-PEA) works with non-affiliated charter schools and provides

much of the technical support that would normally come from a district office – basic back office

---

[20] http://communication.sdhc.k12.fl.us/empoweringteachers/
[21] http://communication.sdhc.k12.fl.us/empoweringteachers/?page_id=706
[22] http://communication.sdhc.k12.fl.us/empoweringteachers/?page_id=313
[23] http://www.piccs.org/
[24] http://www.piccs.org/participatingschools.php

support and access to professional development resources. It also provides a social networking

platform and data warehouse that supports the collection of operational data, ongoing

professional engagement, and data use for improvement and program evaluation purposes. The

PICCS project includes several partner organizations that provide core technology services for

this distributed set of organizations – including SunGard[25] and Follett/TetraData[26] - for data

collection and data warehouse services respectively.

Since many of the schools have substantially different organizational principles and

curricular approaches, the infrastructure required to support them is not explicitly tailored for

charter schools or for any particular school design framework. The data warehouse collects data

that schools are required to report the city, state, and federal agencies. It also includes site

specific flexibility that allows each school to track and monitor data that peculiar to their

individual programs. On the other hand, the teacher and leader effectiveness frameworks are the

same across all sites. One of the lessons learned from the first project was that staff time required

to create a rigorous and fair framework from scratch was far too great. The current project adopts

a widely accepted framework - the Danielson Framework for Teaching and the ValED

assessment for measuring principal performance. The CEI-PEA team then works to engage and

train teachers and building leaders to understand and implement that framework.

The PICCS project will use the growth measures created by the newly adopted teacher

effectiveness standards for the state of New York. The data quality of those measures has not yet

been tested. The CEI-PEA leadership team has processes in place to supplement that work if the

data is not good enough to support high stakes use for the payment of productivity bonuses.

---

[25] http://www.sungardps.com/
[26] http://www.tetradata.com/

Case Themes

While each case is different, there are several commonalities in their implications for the structure of schooling. Clearly, the established institutions—districts, unions, and state and federal agencies—are playing new roles. Silos are breaking down and collaboration among district units is on the rise. In the case of teacher unions, participation is somewhat forced because of the strong political pressure to alter long-established traditions embodied in negotiated contracts. However, the Hillsborough case is distinctive for close cooperation between the district and union—including daily conversations between the union president and assistant superintendent leading the project (Eckert, 2011).

New organizations, both non-profit and for-profit, are also seeing increasing roles. Some organizations, such as VARC, are creating new measures and providing targeted professional development along with them. Other organizations (usually for-profits) are helping create and manage the data that feed into the performance management systems. Of course, foundations (Bush, Gates, and others) are playing an influential role in certain districts, providing both funding and rules and strongly guide district reform efforts.

As charter schools themselves are often new organizations, it is not surprising that the combination of market- and test-based accountability is leading them to take a different structural path. Rather than building a new system from scratch in each school, each participating organization in the PICCS project is trained on the implementation and use of a common evaluation framework. While this might first seem at odds with the typical independence sought by individual charter schools, the focus of the project is not is on the novelty of the program being provided to the students, but on shared notions of effective

teaching. The group uses both a formal peer review process for a curricular unit[27] as well as the Danielson Framework and VAL-ED as the primary measures of effective practice.[28]

In the coming years, these structures are likely to evolve further: each organization will learn through trial and error, common assessments will make it possible to create national systems of teacher profession development and decision support, and the market will evolve from one in which individual districts innovate to one in which many states and districts coalesce around common reforms and seek common—and inexpensive—solutions.

School management and leadership in this new world is likely to be characterized as a complex network than a bureaucracy. While federalism has always been a part of the U.S. education system, the roles of local, state, and federal agencies increasingly being equalized. Combined with the new organizations, and increasing collaboration within districts, these interactions will be inherently more complex. This makes it even more difficult to predict how the reform movement will evolve.

Questions Going Forward

Since we are discussing cases that are just beginning, and there is unlikely to be any evidence about these programs for several years at the earliest, we instead use these cases to motivate a set of critical questions about how teacher accountability proceeds. These include policy questions—questions that policymakers have to make—as well as empirical questions that will hopefully engage researchers in the coming years, and help policymakers make wise decisions.

---

[27] http://www.piccs.org/peerreview.php
[28] http://www.piccs.org/teacher_eval.php

Policy Questions

(1) Will the pressure to measure student outcomes in every grade and subject continue or will states decide that classroom observations alone are more appropriate in some subjects?

If these cases are any indication, it looks like the student testing will become ubiquitous across essentially all grades and subjects. It is possible, however, that these early adopters are simply those most supportive of this approach and others may move more cautiously in this direction or rely more heavily on classroom observations. Given the need for formative assessment of teaching, classroom observations are arguably more necessary than value-added measures, but they are also more costly. Interestingly, it is in some cases the educators themselves in these non-tested grades and subjects who are making the push because they believe that, without testing, their work will be further marginalized.

While apparently in the interest of fairness, one can imagine increased support for assessments, there are most certainly critics who would question the availability and quality of standards in non-core subjects (and across all grades). There are also concerns that low numbers of students in many courses and low numbers of teachers in many subject areas will make it very difficult to make value-added comparisons across the majority of courses and sections. There are some indications that the pendulum might actually be swinging away from ubiquitous testing to some mix of observation and stronger forms of goal setting at the classroom and student level.

Under No Child Left Behind about 30% of student-teacher engagement is covered under the required grades levels and subject assessment. There is a tension across many of these projects related to assessments. There is a strong sentiment that too much time is devoted to testing. At the same time, educators in non-tested grades and subjects generally do not want to be held accountable for student performance in areas unrelated to their instructional expertise. In

many places, teachers in music, art, social studies, vocational arts, etc. are looking to either begin administering nationally available assessments or to begin developing local assessments that could be used to measure growth for their students. This shifts the notion of who are test developers. It is clear that power will shift to test developers under models of accountability that include assessment-based measures of teacher and school productivity. Currently the major test vendors and the purchasing agents for the most populous U.S. states are the actors who will have the most power in the emerging environment if commercial tests remain as the core solution. The countervailing force here would be strong national standards (not federal mandates, but consensus standards such as Common Core or the Arts Education Partnership) and efforts like the work Hillsborough or MCESA are pursuing to leverage collective action to create not just the assessments but to build substantial assessment literacy at the local level.

(2) Assuming schools choose to use student outcomes measures beyond those mandated and provided by state and federal governments, who will create these alternative measures?

Districts such as Hillsborough have approached this for years by creating their own assessments, but this is a luxury that only a very large district can afford. Going forward, others will no doubt seek some more centralized approach, which will reduce costs dramatically, while probably also increasing measurement validity and reliability. But will this centralized solution be within the sphere of state government? Non-profits? For-profits? Traditional testing companies? Or perhaps Hillsborough and other will decide to sell their wares themselves and generate revenue? The development of testing instruments is the backbone of the current reform movement, so the choice of suppliers is obviously very important.

(3) How will the tension between cost and customization be resolved? To what degree will state and federal governments regulate and/or take over data collection, student assessment design, teacher evaluation, and data analysis tools?

With advancements in data management and analytic technology, we are seeing vast improvements in the ability of suppliers to customize at a reasonable cost, but there will always be a trade-off here. Some states, such as Wisconsin, are moving toward statewide data warehouses that allow school districts to create and store their own locally determined content, in addition to data required by state and federal governments. These efforts will eventually reduce technology costs at the local level (e.g, each district will no longer have to hire its own technology vender without leveraging economies of scale), which may facilitate the customization districts and educators will undoubtedly prefer.

(4) What will happen if and when federal funding for these initiatives dries up? Will states and districts continue to invest in accurate measures as well as in the professional development necessary for educators to understand and appropriately use them?

Many of the activities and costs we have discussed are upfront development costs, which are not threatened when external funds are pulled away. But not all costs fall into this category. The costs of classroom observations can be quite high, especially when the system involves evaluators other than school principals. Also, ongoing professional development will be required for teachers to understand the measures and make full use of them, both for understanding their own strengths and weaknesses and those of their students. There is some hope that technology can mitigate some of the costs. The Measures of Effective Teacher study tested the ability to gather classroom video for evaluation purposes at scale. This both reduced the burden on leaders and allowed for the centralized management of observational data and the work of scoring. All of

this translates into cost savings and improvements in system quality. It is very likely that state education agencies will need to make strong "return on investment" arguments for the infrastructure required to support state-wide systems. It is only at scale that the cost of these support systems can be reduced significantly. In the case of Hillsborough, the reforms have already been going on for many years, facilitated by unusually low turnover among district leadership and close ties with the union.[29] Long-term, stable leadership allows groups to build and maintain trust. It also allows actors to agree on longer-term payoffs with relatively high certainty that commitment to shared goals will remain high.

(5) As these reforms are fundamentally about performance measurement and accountability, what stakes will be attached to the measures?

We obviously cannot answer this, but the implications are noteworthy. As Harris (2011) writes, we should "attach stakes to performance measures that are proportional to the quality of those measures." So, as the stakes rise, so should the willingness to invest in data collection, analytic tools, and professional development. Whether policymakers will follow this rule is a political prediction beyond our scope (or ability).

Empirical Questions

Researchers can help policymakers make better decisions. Given how little we know at this point, one could begin almost anywhere and make a useful contribution, but below we focus on what we see as some of the key questions deserving particular attention.

(1) While other sites will no doubt be keeping track of the decisions and activities of the early adopters described in these cases, will they make the same decisions and with the same effect if and when the programs diffuse and/or become required? Unlikely. Based on our

---

[29] Average tenure in the largest districts in 2010 was 3.6 years. The average for Hillsborough over the past 15 years is 8.1 years. See http://www.districtadministration.com/article/superintendent-staying-power for more details.

experiences, the early movers are those with long-term commitment and support for teacher accountability. As the recent debates about *Race to the Top* and *No Child Left Behind* make clear, the appetite for these approaches is far from universal and, without widespread support, the move toward teacher accountability may amount to little more than a full employment act for purveyors of large computers and standardized tests.

While there is much to applaud across the four examples presented here, they may represent the very highest capacity actors in this policy space.

(2) How do the new educational organizations fit into the educational governance and management mix? How will they use their newfound power?

Labor unions may take ownership of some elements of accountability and start to self police. That seems to be what is at the core of the "new unionism" being promoted by AFT and NEA (American Federation of Teachers, 2011; Commission on Effective Teachers and Teaching, 2011). What is less clear is how networks formed through collective action will impact the governance and management mix. Partnership with private funding agencies and commercial solution providers create linkages that cross state lines and can provide access to policy options or technical solutions that are at odds with (or at least quite different than) official offerings. Indeed, some solution, such as approaches to educator effectiveness being tested by the Measures of Effective Teacher study might actually call the legitimacy of some state-mandated approaches into question..

(3) How, and how well, will the increasingly complex governance and management work?

Tight budgets and new requirements for accountability are part of the motivation behind new forms of collective action and shared governance. Will these forms survive a return to

economic health? One might see some elements of corporatism in these new constellations of actors – with close collaboration between labor and management and the active participation of state actors. .

For example, the evolutionary elements and long range planning of the leadership group in Hillsborough have allowed for the gradual scaling up of the systems for supporting the collection and analysis of data. All elements of the EET project are also piloted for one or more years before they go into production – allowing for tweaks of the infrastructure, the measurement tools, and the professional development supporting the evaluation tools and the use of the outcome. This strong sense of joint action, shared leadership, and incremental improvement provides both management and labor with the tools and the trust to maintain sense of shared ownership.

The outcome for charters might be quite different, given the need for collective action on start-up. CEI-PEA serves some of the functions of a district in that it provides access to technology, data systems, professional development, etc. much like what a district would provide. At the same time, the schools are each independent legal entities that have agreed to participate in this network for mutual benefit. The power relationships are complex and all of the actors have shared interests around collaboration – for funds, shared services, access to expertise, etc. – rather than the more formal structures in an actual school district.

(4) Will school districts maintain their power but simply exercise it in different ways? Will the breaking down of silos within school districts fundamentally change how they work?

The increasing engagement of the state and regional actors in the provisioning of services and the support of common standards certainly erodes the important of local districts. One theme that runs to all of these cases is a growing need for collaboration between departments at all

levels – in IHEs as well as at the state, district, local levels. Out in the field doing technical assistance with districts and state agencies, it is increasingly common to hear from human resource, research, and assessment departments staff members that they are really working together for the first time in their careers. In the past, they routinely regarded curriculum and instructional support services as the slowest movers and likely to adhere to internal agendas and products development plans known to no one else. Even they are being drawn into the complex evaluation and service provision discussions around measuring teacher effectiveness and what to do when a teacher or principal is identified as need some specific support. . This pressure to provide comprehensive support to around educator effectiveness is often combined with cost concerns as budgets remain flat or decline. Shared interests and external political and financial threats are providing incentives to engage in meaningful change.

The REIL project provides an example of the sort of complex collective action seen in some emerging partnerships. REIL is a multi-district partnership, but all of the partners are part of the Maricopa County Education Service Agency (MCESA), a regional educational service agency. The superintendent of MCESA is an elected county official and is responsible substantial back office and educational support services for the 58 school districts in the county – 6 of which are members of the REIL project. [30] MCESA has as part of its normal mandate providing support for school improvement, educational program development, financial and human resources support, etc. The leadership team was able to recruit districts and, more importantly, hire staff to begin designing the teacher evaluation program, co-designing detailed, district- and role-specific communication plans before the grant was won. The MCESA leadership was able to articulate a school improvement and compensation reform strategy that fit within its existing mandate for reform and would provide design feedback for the improvement

---

[30] http://www.maricopa.gov/Schools/bio.aspx

of most major technology systems across the county agency. The TIF grant provided resources to support activities that were in line with the existing political structures and plans for improvement.

The other more recent work on this project has focused on choosing or developing new assessments in grades and subjects not covered by the mandated state assessments. MCESA staff members were able to leverage their existing mandate to provide support for local requests for high quality assessments. In particular, the project hosted a series of meetings open to all 58 districts in the county to build support for measuring student learning across the curriculum. They also identified technical experts outside the group to provide the expertise necessary to create the sort of high quality tests necessary for high stakes use. The roll of coordinator and convener is part of the MCESA mandate. In this way, the role of TIF fiscal agent is consistent with the service agency's role in the state system. This consistency of roles helps avoid political challenges or power imbalances within or between the partnering districts. The distributed accountability of the MCESA leadership accountable to country voters and the local districts accountable to their local school governing boards also provided checks against more extreme positions being taken by any partner.

(5) What will be the indirect and unanticipated effects of the revised mixture of governance and management?

The direct and obvious implication is that state and/or federal governments will have more power through common curricular standards, teaching standards, and shared information technology infrastructure. The fact that many of these ideas are not required by the federal government (see *Race to the Top*), and are likely to remain outside the realm of ESEA after the re-authorization, does not mean the federal government lacks power. On the contrary, the

incredible spread of teacher accountability in the past few years shows just how powerful the federal government has been. This approach has been so successful in spreading reforms that almost every major policy initiative since *Race to the Top* has adopted the same competitive grant approach (and even the same rhetoric).

Partnerships are a part of all modern organizations. In the case of the partnership cited here there are a mix of groups that provide analytical support, technical solutions, professional development, and communication strategies. In no case, are any of these groups presented here calculating student growth measures on their own. Neither are they trying to do student-teacher-subject linkage completely on their own. Every group is partnering with state or research partners to develop and test their solutions to this challenging problem. One of the lessons we have taken away from reviewing the challenges of implementing new teacher and principal evaluation systems is that even the most capable actors require outside expertise – to the extent that some major components, such as value-added measures calculation or data warehousing, may be completely outsourced.

This may change as Race to the Top (and similarly motivated) evaluation systems role out to districts. Home grown solutions may be superseded by state mandates. There will also be opportunities for some of these local providers to move up and provide statewide solutions. This emerging market will likely show up first in Race to the Top states. While districts will, in some sense lose power, they also gain the power to dismiss low-performing teachers. On the other hand, teachers and schools will become more accountable locally. They can also use performance feedback to make a case for tax levies, etc. It is difficult to anticipate how the participation of large assessment and curriculum vendors make take advantage of (and thereby shape) this newly forming market for educator effective services. The market opportunities

around the Race to the Top projects alone have brought all of the major vendors of educational

systems, materials, and assessments into the mix. Collective action around these policy reforms

are providing the specifications for new products and services.

(6) How will other school reforms (especially school choice and market competition)
influence the development of individual teacher accountability?

Common standards and effectiveness measures are likely to expose the highly variable

quality of some of the market options - a fact masked by many current exemptions from

accountability. It seems likely that a shift in focus to effectiveness might undermine the claims of

market competition folks that only the market can do this. If effectiveness information is

routinely available and there are political levers to better allocate resources, the provision of

public goods might make a resurgence.

The most important empirical question in all of this is, will the reforms "work" in any

meaningful sense? Student test scores are an obvious first place to look, but even more

important may be whether educators sense that the reforms are improving instruction while not

stifling creativity, creating undue burdens, and undermining their work lives. Given the obvious

importance of teacher effectiveness in student outcomes, the approach has promise, but it is also

possible that policymakers will over-shoot the mark, try to "measure everything," attach high-

stakes to low-quality measures, and generally implement a good idea poorly. In this sense, these

early adopters we have described in these cases will play a critical role. Given the reluctance of

educators to move in this direction, it is unlikely that they will be implemented as intended

unless they have some confidence that, in their own eyes, the approach will work.

There is one looming problem for all of these projects. PICCS and MCESA have the

mandated external evaluations of their programs in place. What we don't have, however, are

large scale evaluation that will tell us with any certainty whether these programs are having the desired effects. There will be some limited before and after analysis, but most of this work is being doing using best guesses with what will be effective and assuming that feedback in the form of growth measures, observational ratings, and embedded professional development will be strong enough to produce the desired improvements in student achievement gaps, teacher recruitment and retention, etc. Without a more strategic focus on evaluation, much of what could have been learned from these novel implementations will not be available use elsewhere with any level of rigor.

References

American Federation of Teachers. (2011). *Weingarten proposes aligning evaluation and due process* Retrieved 3/18/2011, 2011, from
http://www.aft.org/newspubs/news/2011/022411weingarten.cfm

Baker, E., L.Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R., et al. (2010). *Problems with the use of student test scores to evaluate teachers* No. 278). Boulder, CO: The Economic Policy Institute: University of Colorado-Boulder.

Battelle for Kids. (2011). *The importance of accurately linking instruction to students to determine teacher effectiveness.*
. Columbus, OH: Battelle for Kids.

Commission on Effective Teachers and Teaching. (2011). *Transforming teaching: Connecting professional responsibility with student learning*. Washington, DC: National Education Association.

Hakes, J. (2011, April 23, 2011). Bush foundation to provide unprecedented support through teacher effectiveness initiative. Message posted to
http://theacademicvillage.blogspot.com/2011/04/bush-foundation-to-provide.html

Harris, D. N. (2010). *Value-added measures of education performance: Clearing away the smoke and mirrors* No. 10-4). Stanford University: Policy Analysis for California Education, Stanford University.

Harris, D. N., Rutledge, S. A., Ingle, W. K., & Thompson, C. C. (2010). Mix and match: What principals really look for when hiring teachers. *Education Finance and Policy, 5*(2), 228-246.

Heinrich, C. J., & Marschke, G. (2010). Incentives and their dynamics in public sector performance management systems. *Journal of Policy Analysis and Management, 29*(1), 183-208.

Kane, T., & Staiger, D. (2012). *Gathering feedback for teaching: Combining high-quality observations with student surveys and achievement gains*. Seattle, WA: Bill and Melinda Gates Foundation.

Koretz, D. (2008). A measured approach: Value-added models are a promising improvement, but no one measure can evaluate teacher performance. *American Educator, Fall*, 18-39.

Mohr, C., Broton, K., & Muelle, D. (2010). *Bush foundation teacher effectiveness initiative: Phase II feedback*. St. Paul, MN: The Archibald Bush Foundation. Retrieved from
http://www.bushfoundation.org/sites/default/files/Web_Content/Education/PDF_Files/Phase_2_Feedback_WR.pdf

Moynihan, D. P. (2010). Performance management. In R. Durant (Ed.), *Oxford handbook of american bureaucracy* (pp. 278-302). Oxford, UK: Oxford University Press.

Smith, M. (2012, 1/31/2012). Texas schools chief: Testing has gone too far. *The Texas Tribune,*

U.S. Department of Education. (2010). Office of elementary and secondary education: Overview information; teacher incentive fund: Notice inviting applications for new awards for fiscal year (FY) 2010. *Federal Register, 75*(98), January 12, 2012-28740. Retrieved from http://www2.ed.gov/legislation/FedRegister/announcements/2010-2/052110c.html

U.S. Department of Education. (November 14, 2006). Office of elementary and secondary education; overview information; teacher incentive fund; notice inviting applications for new awards for fiscal year (FY) 2006 *Federal Register, 71*(219), January 12, 2012. Retrieved from http://www.gpo.gov/fdsys/pkg/FR-2006-11-14/pdf/E6-19193.pdf

Watson, J., Witham, P., & St. Louis, T. (2010). *Evaluating student-teacher linkage data in teacher incentive fund (TIF) sites: Acquisition, verification, and system development.* Washington, DC: U.S. Department of Education, Office of Elementary and Secondary Education.