Harvesting Lessons on Educator Incentive Plan Design from Technical Assistance Provided to Teacher Incentive Fund Grants

Anthony Milanowski
University of Wisconsin-Madison
1025 W. Johnson St.
Madison, WI 53706
amilanow@wisc.edu

Peter Witham, University of Wisconsin-Madison

Patrick Schuermann, Vanderbilt University

Steven Kimball, University of Wisconsin-Madison
.
Darcy Pietryka, Westat

The work described in this paper was supported by the U.S. Department of Education through the Center for Educator Compensation Reform. The opinions expressed are those of the authors and do not necessarily reflect the view of the U.S. Department of Education, the Center for Educator Compensation Reform, or the institutional partners of the Center. This is a working paper describing initial results of an ongoing project. Comments and suggestions are welcome.

One of the most remarkable phenomena on the US K-12 educational reform scene has been the rapid rise of new forms of teacher compensation, especially those that provide financial incentives for teachers linked to student achievement. While still not common, there are dozens of school districts in the US that are experimenting with performance incentives. (See http://cecr.ed.gov/initiatives/maps/ for a partial list of such initiatives.) The allocation of additional federal dollars in 2010 for an expansion of TIF, and the inclusion of performance incentives in many Race to the Top proposals will substantially increase the number of states and districts with these programs.

As more states and districts think about, or experiment with, performance incentives, they face a capacity problem: relatively few educational administrators have training or experience that would help them design and administer performance pay systems for teachers or administrators. The simplicity and longevity of the traditional salary schedule has made it unnecessary to do much thinking about alternative compensation approaches. The training of most educational administrators has probably not included courses in compensation, or even human resource management. Thus it is likely that many states and districts will struggle to design effective plans, and then to implement them in a sustainable manner. Because we believe that plan success depends on the details of design and implementation, we think it is important to study and publicize how incentive programs have addressed common design and implementation issues. By doing so, we can contribute to building the capacity of the K-12 system to develop viable programs. In this paper, we try to contribute by describing the design and implementation issues we encountered providing technical assistance to five Teacher Incentive Fund grantees, and the solutions that were developed. The paper is not intended to be a comprehensive assessment of the success of the incentive programs, nor even an evaluation of our technical assistance efforts. Rather, it represents an attempt to harvest what we have learned, for the benefit of those who want to do a better job of designing and implementing incentive programs.

This paper is based primarily on the experience of technical assistance providers working with the Center for Educator Compensation Reform (CECR), a US Department of Education-funded center that is responsible for providing grant administration assistance to Teacher Incentive Fund program staff and technical assistance to Teacher Incentive Fund grantees. The Teacher Incentive Fund (TIF) is a Department of Education program that provides nearly $99 million to a variety of grantees, including states, school districts, charter school consortia, district consortia, and individual charter schools to assist them in developing and implementing performance-based teacher and principal compensation systems in high-poverty schools. The program began in 2006 with 14 grantees competing successfully for the first round of funding. Twenty more grantees were added in a second round in 2007. The CECR provides technical assistance to these grantees through a combination of resources published on the internet (http://www.cecr.ed.gov/guides/), conferences, individual consulting, and referrals to persons or organizations with appropriate expertise.

Method

The Teacher Incentive Fund program currently includes 33 grantees. (A list of the grantees can be found at http://cecr.ed.gov/initiatives/grantees/profiles.cfm.) At the time data was collected, 14 of these were in their third year of the grant program, and the rest in the second year. Thus

many have accumulated a considerable experience and most are well past the start-up stage. Most have made at least one payout under their incentive plan design.

We selected five TIF grantees to study. Our selection was purposive, guided by the desire to represent both large and small grantees, conventional school districts and charter schools, and programs in both the second and third years of the grant. We also had to consider the amount of technical assistance that we have provided to the grantee, because that technical assistance experience was the primary "data base" we had available.[1] Based on a thorough discussion of these considerations, we agreed on the five grantees to use as the cases for this study. These are profiled in Table 1 (See Appendix A). Actual grantee names are not used because we guaranteed grantee program coordinators confidentiality to encourage candor about the problems they had encountered.

Each of the authors developed one or more case study narratives. These narratives were written to a standard format, so that similar information would be available for each case. The format required case writers to describe the grantee's context and the incentives provided, summarize the evolution of the incentive program, summarize the chronology of technical assistance provided, identify the most important technical assistance issues, and assess the effectiveness of the solutions developed to address these issues. The case writers developed the cases by reviewing documents such as grant proposals, progress reports, memos, and emails. In addition, they interviewed CECR staff who provided technical assistance, and in some cases technical assistance providers from outside the CECR technical assistance group. CECR staff assigned to monitor grantee compliance with the Federal contract were also interviewed. Lastly, one or two grantee staff were interviewed, including the TIF program coordinator. Interviews were done using a semi-structured protocol, but due to the variety of issues that could come up in the cases, we encouraged follow up questions to bring out the details in each case.

When case drafts were completed, each author read all of the cases and noted where additional information was needed. Authors revised cases in response, occasionally re-contacting sources to obtain additional information. Cases were discussed at a group meeting and a set of cross-case themes related to issues that the grantees had to address was developed. The first two authors then drafted a summary of the case material relevant to each theme. The second two authors reviewed and made additions or corrections to this material. The result was a summary of our conclusions about each theme and a short description of how the issues underlying the themes were addressed in each case.

Results

This section presents our conclusions about each theme, a summary of how it is illustrated in each case, and what we take to be the implications for incentive program designers.

Knowledge/Capacity and External Partners

[1] Because of concerns about burdening grantees and getting in the way of the national evaluation of the TIF Fund, we had to limit our contact with grantees themselves. We therefore did not choose grantees with which we had few technical assistance interactions.

The first theme we found important in all five cases was that grantees initially lacked knowledge about what would be required to actually implement a performance pay program, as well as certain important implementation capacities, and that experienced external partners were very helpful in supplying these capacities. State, district, and/or school staff themselves typically knew little about how to design and especially how to implement performance pay systems. Even if the plan looked good on paper (i.e., on the application for TIF funding), the details of how to communicate with stakeholders (especially teachers), measure performance, and make accurate payouts were often not fleshed out and systems needed to be constructed. Grantees were at times unaware of important implementation issues such as the need to have data systems that could link students and teachers, or tests that had credibility with teachers. We found that the grantees that began with or quickly added external partners with experience in these areas had more knowledge and capacity at their disposal and this knowledge/capacity also transferred over to them.

Grantee A's internal staff had limited knowledge and capacity, but this was offset by the capacity and knowledge provided by external partners, who were brought in during the grant development process or early in the first year. The grantee began by partnering with a national organization with a tested performance pay model that included many supporting elements such as communication with teachers. Their "off the shelf" model required stakeholder buy in and lowered the potential for confusion among partners and stakeholders about what the incentive system would entail. The program was designed to scale up from a small subset of schools that volunteered to participate. The program was presented as part of a broader set of human capital initiatives, which helped to attract some additional foundation support which facilitated bringing in additional partners and experts when problems arose. Foundation resources allowed the grantee to bring in a national research organization to evaluate the program and to help in initial school selection. As part of another initiative, the district also contracted with an experienced vendor to design and implement a school and classroom value-added system. This grantee basically contracted for most of the key elements of its program. Though there were some initial implementation difficulties, they did not prevent a successful roll out of the program as designed. After the initial year, the capacities of the external stakeholders limited the need for CECR technical assistance.

Grantee B, as a fairly large district, had an ample administrative infrastructure, but lacked some key knowledge and capacities. In particular, capacities for communication with stakeholders (especially teachers and the press), data systems (especially linking students with teachers to create valid value-added estimates), and doing value-added modeling were limited. This grantee initially relied on its own resources, rather than bringing in external partners, and moved quickly into full implementation without a planning year or pilot. Two problems threatened to discredit the TIF program after the initial year: errors in the value-added estimates of classroom achievement and an initial highly negative response from teachers due to errors in teacher-student linkage and limited initial communication. Press coverage was also critical. However, this grantee quickly mobilized additional financial support from foundations to bring on board an experienced value-added vendor, a public relations firm, and an organization with expertise in developing systems to help teachers use student achievement. These external partners worked closely with district-level administrators to improve and refine the value-added-based performance measurement model, communicate about the program to faculty and the public,

design implementation systems, and encourage support and buy-in from students, their parents, and community and business leaders. The second year of the program was much less troubled, with accurate payout delivered on time, though performance incentives remain controversial with some stakeholders. The presence of these strong external partners limited the need for CECR technical assistance.

Grantee C had no prior experience with performance incentives. The small external partner organizations had significant experience working with each other in the design, implementation, and evaluation of education initiatives but no experience with performance incentives. Facing considerable geographic, infrastructure, and stakeholder engagement challenges, the lack of resources and capacity of these partners resulted in the lack of a systematic and cohesive communications plan, a systematic approach to data quality, the initial use of relatively simple performance measures, and slow development and implementation of program improvements. This grantee was not able to find additional external resources to support the program. The external evaluator did step in at some points to provide technical assistance and keep the program on track. CECR technical assistance has also been important in moving toward a statewide growth model for use in school-based incentives. The lack of resources, capacity, and knowledge slowed development of the incentive program..

Grantee D also had limited knowledge and capacity, due in part to its small size and rural location. Though it had hired a coordinator with experience with a larger district's performance pay pilot project, no one else in this small (3 school) district had experience with performance incentives. The external partners were mostly small organizations with no experience in performance measurement, incentive design, or communications. CECR technical assistance has therefore been important in helping this make use of the state growth model.

Grantee E staff had substantial experience with a variety of educational innovations, though little with performance pay. Program administrative staffing was very lean, Individual charter schools had no experience and very little infrastructure to administer a program. The initial plan was to use a variety of external partners to supplement the limited capacity of the schools and the central consortium staff. Partners for communication, data system development, and professional development on data use were lined up early in the project. As the project progressed, additional assistance on plan design and teacher evaluation was acquired from smaller but quite competent vendors. Without these partners, there would likely have been little progress toward implementing most of the program's relatively complex incentive model. Also, this organization wisely used their first grant year for planning, which helped develop capacity. During the first year, consortium staff worked with school based teams and charter school governing boards to design specifics of the incentive program based on a framework that was to be adapted to all schools. Overall, the knowledge and capacity of external partners didn't transfer as well as it could have because of the general nature of the framework and lack of accountability across the consortium for implementation of the framework. Resources available but not utilized consistently due to lack of accountability.

Our conclusions around this theme are that most states and districts are not likely to have sufficient knowledge and capacity on their own to design and initially implement an effective, sustainable performance pay program. Bringing in well-resourced and experienced external

partners seems to be the easiest way to add the needed knowledge and capacity. The larger urban grantees had the resources, including extra foundation support, to bring in such external partners. The other three grantees tended to have fewer internal resources and smaller partners with less capacity to overcome the many obstacles facing a complex and ambitious initiative. Having strong external partners with expertise in program design, data management, and developing and implementing performance metrics appears to contribute to getting a viable program up and running more quickly. Unless a state or district has substantial prior experience in performance pay system design, performance measurement, communications, and data system design, beginning with capable external partners lowers the odds of design and implementation problems. An implication for governments or foundations seeking to fund performance pay initiatives is that proposals relying mostly on in-house resources should be scrutinized closely. It may also be useful to develop a clearinghouse that would list potential external partners with a track record of helping districts design and implement these initiatives.

Leadership

As might be expected, leadership played a large role in the successful design and implementation of these grantees' performance pay systems. Overall when leadership made the program a significant district priority and aligned the program with larger district goals, resources and capacity were galvanized for successful design and implementation of the program. Where leadership was able to situate performance incentives within the districts larger vision, and gather key partners to increases the resources and capacity of the project, there was more momentum behind the project.

Grantee A's top leader made the incentive program part of the district's human capital initiative, which signaled to the district the importance of the TIF grant initiative. This also helped recruit key partners and attract additional foundation funding. Leadership took steps to engage multiple stakeholders in the design and implementation of the program, including both teacher association and district central office staff. However, there was considerable churn and turnover in the grantee's TIF project team. This and the low hierarchical status of the project director limited the integration of the TIF project with other district initiatives. Fortunately, the grantee was able to rely on the external partners to implement the project. The main role of leadership was to get the right players on the field, emphasize the importance of the project, and defuse potential initial opposition from the teachers' association. This was done by bringing in national association leaders and value-added experts to help explain program features and reassure local association leaders, and by establishing an oversight group including the association president, along with members appointed by the school board, the association, and a local foundation that had sponsored the program.

Grantee B's top leadership championed the performance pay program from the start, and pushed it ahead before all the supporting systems were in place. This got the program up and running quickly but resulted in the first year problems discussed above. However, Grantee B leaders quickly realized there was need for skill as well as will, and moved to develop a representative advisory committee that would determine the best way forward based on thorough examination of national landscape and counsel of national education experts. They also moved the project forward by bringing in external partners with significant capacity in the areas most in need

(performance measurement/value-added, professional development for teachers and administrators on the program and specifically value-added, and communication with stakeholders. By recognizing the need for improvement and finding the necessary resources for success, leadership signaled to the district and stakeholders the importance of the project and its place in the larger district reform initiative.

Grantee C's leadership was not aggressive about making the performance pay program central to the state's education initiatives, rather rolling it out as more of an experiment or pilot for certain districts. The largest district in the state dropped out of the program at its inception, which was a blow to the project's credibility. Project leaders did not respond to capacity limits by seeking external key partners. Though it involved stakeholders via a representative and collaborative administrative and governance body, this group was not able to energize the program and failed to even meet as often as originally planned. It appears that no strong champion of the program has yet emerged. Grantee leadership did not put high priority on some of the foundations of performance pay, including communications, data quality, and performance measurement.

Grantee D's leadership also did not make the performance pay program a central part of the district's reform strategy. However, the project director and other leaders made continuous efforts to keep the project going in the face of misunderstandings, changes in and shifts in the philosophy of the governing committee. This grantee experienced two substantial shifts in its incentive program design. The leadership was able to change the design to respond to problems and objections from stakeholders. The leadership was also proactive in finding external expertise (largely CECR technical assistance providers) to improve performance measurement techniques, data quality, and communication strategies.

Grantee E had a strong champion in the project director, who made considerable effort to integrate performance pay into a larger vision for improving instruction, communicate this vision to the schools in the consortium, and to find additional external partners and resources when needed. However, the director and the organization that runs the project did not appear to have enough authority to compel the ten independent charter schools to follow a uniform model or implement them with care. Individual schools and their governing boards were ultimately responsible to develop the program with the resources that were made available. Limited accountability of the schools to the project created variance in program design, implementation, and likely sustainability.

Implication: A clear champion seems important to ensure performance pay programs move from design to implementation, evaluation, and program improvement. Because strong leadership will always be in short supply, there will likely be a substantial number of underperforming and even failed performance pay projects. At some point, however, a model of performance pay may become dominant, and this may reduce the demands on leaders. Perhaps this will happen via the development of a market for external vendors of turnkey systems. (Grantee A illustrates how an external vendor can reduce demands on local leaders.) External partners can also provide some continuity of focus to offset the churn in TIF program directors experienced by some grantees. Government or foundation funders may also want to look at the track records of leaders in implementing similar initiatives before committing funds. The US Department of Education appears to be taking a similar approach in considering states' Race to the Top applications.

Design Evolution

In four of the five cases, the performance pay designs changed substantially from that described in the grant application. Much of this was due to grantees finding out they didn't have the data, systems, or support needed. While many intended to use some form of classroom value-added as a basis for teacher bonuses, all but one found they did not have the requisite data or data systems in place. One common response was to begin with school-level performance such as change in attainment as the basis for the incentive, until measurement systems would catch up enough to allow value-added indicators and classroom-level performance to be used. Some design evolution has also been toward less differentiation among educators in the payout distribution than expected based on original proposals.

Grantee A had to forgo individual teacher incentives based on classroom value added in the first payout year, because classroom value added data was not available. The money was used for school-wide achievement bonuses instead. Data quality issues also made payouts based on value-added problematic at the high school level, forcing the district to develop a novel approach different from the initial design (i.e., using average value-added results from the elementary level for payouts to high school teachers, which led to all high school teachers receiving a bonus). While the intention is to use classroom value-added as a basis for payout in the second year, the status of classroom level data is still in doubt. Given on-going concerns with the quality of data at the elementary level, there is now consideration of using grade level value-added results as a substitute for classroom-level until data quality improves.

Grantee B had to make the least change in its incentive design. Its initial design called for bonuses to be based on school and classroom value-added, and it has stayed with these measures of performance. Instead of reacting to first year problems with its value-added estimates by changing the design, this grantee brought in the resources needed to correct the problems and move forward with the initial design. The major refinement was a more sophisticated value-added model. The "tournament" nature of the design (only the teachers and schools in the top quartiles of value-added receive awards) preserves incentive differentiation.

Grantee C began by relying on state school-wide performance measures based on change attainment, such as making AYP. They also intended to include a classroom-level measure, but this was not implemented. The exact basis for the proposed individual teacher incentive was not clearly specified in the grant proposal. It appears this was to be some combination of a student growth measure, a practice evaluation, professional development, and leadership roles. The original principal and teacher evaluation measures were also changed from being based on rated performance to being based on simply participating. This likely limits payout differentiation. There has been some interest in a student growth (value-added) model, but implementing this would require more state investment than is likely to be made.

Grantee D made a number of changes to its design to respond to both measurement and stakeholder issues. The initial first payout was based on a formula that rewarded school performance with respect to state accreditation standards (effectively change in attainment). A second payout in the first year was instituted, based on school attainment relative to state

averages, to respond to stakeholder concerns. One result was low differentiation. Then a third formula was developed based on growth as defined by a state school value-added model. The basis for incentives was still school performance. For this grantee, the development of the state model was important because it did not have the resources to develop its own growth model.

Grantee E began with an ambitious three-part incentive model based on student, classroom, and school achievement, teachers sharing best practices, and teachers taking on additional roles. The plans implemented by the 10 independent charter schools in the first payout year varied considerably in how many of these elements were implemented. Data availability limited the use of student and classroom performance indicators for incentives, and the grantee has had difficulty deciding on a growth or value-added model that could be used by all schools. Neither grantee staff nor most schools initially had the time or capacity to implement the teaching best practices component. There was also a tension between the consortium-wide model, which emphasized growth, and schools' charter goals, that emphasized attainment or attainment change, and the latter were the main element of many first year plans. Some participating schools also appeared to consciously chose performance measures that led to minimal differentiation in payouts.

One implication of the tendency of incentive designs to change is that program designers may want to consider a pilot year in which they implement on a small scale in order to see if they actually have the data to implement the design, how stakeholders will react to it, and to work out glitches in their systems. Even at the planning stage, it may be useful to have the plan reviewed for implementability and "signed off" on by administrators in charge of data systems, human resources and payroll, or other relevant functional units. Government and foundation funders of such programs may also want to be cautious about approving a proposal based on the details of the proposed design, since it may not be implementable and will likely be modified. They may want to require a planning year in initiatives they fund. The US Department of Education is doing this for the third round of Teacher Incentive Fund grants.

Communication

Communication of any performance incentive plan is a critical part of its implementation. First, assuming that the point of the plan is to motivate educators to expend effort to improve performance, the target educators will need to understand key plan features such as the size of the incentives (e.g., how large are the promised bonuses), how performance is defined, and what levels of performance are required to receive the incentives. Second, the support of key external stakeholders will be affected by efforts to communicate the purpose, scope, and potential benefits of the plan. In particular, the local media will likely be interested in aspects of the plan such as the size and costs of the incentives and educator reactions. In this paper, we are largely concerned with communication with internal stakeholders. Our cases showed a lot of variation in communication approaches. While some grantees attended to communication strategies from the beginning of the project, most did not realize the extent of communication that would be needed until their projects were well underway. Most added to their communication efforts on an ad hoc basis in response to implementation challenges and stakeholder feedback. Some really struggled to develop effective communication strategies for their particular context.

Grantee A began with a national model that included strategies for communicating with stakeholders about this program. The model provider required that participating schools take a vote on participation, and that a majority at each school agree to participate. This approach encouraged the educators that were potentially affected to learn about the model, and it appeared that principals and teachers knew prior to the program's roll out what its core components were. In addition, this grantee hired an external vendor early on in the process to develop a district-wide communications plan about the program. During the initial year of implementation, feedback from the model provider, school stakeholders, and the communications vendor identified areas that required more communication, such as the hiring process for teacher leaders. One issue that arose at the very beginning of this grant was potential opposition from the local teachers' association. During the development of the proposal and initial implementation, there were strong concerns about how performance would be measured and how school principal evaluation of teaching would affect awards. The district and especially its foundation supporters responded by bringing in a variety of external experts, including some from the national association and the vendor of the value-added performance measures, to engage in intensive discussions with local association leadership. These external experts were able to address the concerns of the association and win its tentative support. A memorandum of understanding was developed that spelled out how the association would be involved in implementing and overseeing the program with the district, including the designation of an association staff person to act as a liaison and serve on an association-management joint committee concerned with the program.

Grantee B is located in a state where several other districts had experimented with performance pay programs, and so district educators had some preconceived opinions about performance pay which were not necessarily positive. The district at first did little to explain its program to target educators, or the local media. When the first payouts were made, the grantee was surprised when a local media requested a list of all educators who received an incentive for publication. Worse, there were a substantial number of errors in first round payouts. This lowered the credibility of the program with teachers, their representatives, and the public. At this point the district realized a comprehensive and systematic communications strategy was essential for the program's survival, and contracted with two external providers using external foundation funds. These groups worked collaboratively with district staff to develop strategies for communicating with the media and target educators. Notably, the district admitted that errors were made but also made clear the extent of its efforts to correct them and ensure they were not repeated. Because this grantee did not have to work with a teachers' association, it targeted its efforts toward direct communication with educators, creating dedicated web portal where all information about the program could be accessed, developing materials to explain how the value-added classroom and school performance measures worked, and providing opportunities for stakeholder feedback. The grantee's comprehensive re-branding of the program also stressed other aspects besides the financial incentives including data use, supports for teachers, and the potential benefits to students and the community of improved student achievement. As a result, the program survived its first year. Though teachers' representatives continue to express opposition, the program appears to have support from the local community.

Grantee C faced several barriers to effective communication from the beginning. These included the geographic isolation and lack of resources of the participating school districts, and a lack of

fiscal resources to partner with an experienced external communications vendor. The original communications strategy was primarily comprised of periodic meetings with district leadership, field teams and exposing participating district coordinators to annual summer institutes. The ineffectiveness of these strategies was illustrated by one educator's comment, "We got a check for something, but don't even know what it is for." The lack of a defined mechanism for stakeholder feedback placed the burden on principals, superintendents, Field Teams, and the external evaluator to provide feedback to the district. The external evaluator found relatively low understanding of the program by target educators, and recommended that the grantee find better ways to engage stakeholders and participants in the project. This feedback did lead to a revised communication plan that took additional measures to improve clarification and consistency of messaging. One of the most significant changes to the communication plan was that it used email as a way to bridge the gap between the state central authority and teacher participants. The previous dependency on superintendents and principals led to a great deal of variance in teacher knowledge about and understanding of the program. Additionally, the project improved its external communication strategies so that the state level stakeholders had a deeper understanding on the state of the project.

Grantee D, a single small district, had fewer barriers to overcome in communicating about its program. From the beginning it had a coherent communication strategy that provided stakeholders with ongoing and targeted information about the grant. The small size and use of a representative design committee allowed for quick communication and feedback from internal stakeholders. The grantee's communication strategy was tested by problems with the initial design of the incentives. The small size and cooperative oversight committee process allowed the grantee to work out these problems without threatening stakeholder support. Once the committee decided to use the state growth model as the basis for performance measurement, it was then able to communicate about why this was a fairer measure.

Grantee E also required school staff to vote on school participation in the grant. This created at least a minimal understanding by internal stakeholders of the basics of performance pay. Grantee central staff also developed a framework for incentive plan design and used the first grant year as a planning period for the independent charter schools that decided to participate to develop their own adaptations of this framework. During this period, the central staff also fed information about plan design to the schools and hired an external consultant to provide training to schools. The central staff also contracted with an external communications group which developed a communication plan that used multiple tools, including a web portal which was designed to engage educators in the schools. The incentive plan was also presented as part of a broader strategy to improve instruction and student achievement. The portal included links to tools for these uses. However, though these efforts were quite sophisticated, the capacity of the individual schools to respond was often low, due to relatively small sizes and limited administrative infrastructures. Also, the central consortium staff (as in grantee C) did not have the authority to require engagement with the resources provided. The variation in the quality of the initial incentive plan developed during the planning year suggested that several schools were unable to make use of the information they received.

We conclude that performance incentives are likely to require a lot more effort toward communication than program designers might think. As designers of incentive programs in the

private sector emphasize, there cannot be too much communication (Hale & Bailey, 1998; Flannery, Hofrichter, & Platten, 1996). Districts seeking a successful initial roll out of their performance incentives should probably seek technical assistance from an organization with expertise early in the planning process. There are some promising communication ideas exemplified in other cases, notably Grantee B's extensive efforts to communicate directly with teachers, its rebranding of incentives as part of a broader district improvement program, and its use of experienced external partners to develop web-based support. These ideas are likely applicable in larger organizations. In small districts, structures like Grantee D's representative committees, combined with an openness to listen to feedback, would seem promising. It is likely that consortia of independent districts or schools will present a communications challenge to a central directing authority that has only limited authority. Government and foundation funders may want to require an explicit plan for communicating with target staff and external stakeholders as a condition for funding.

Use of Value Added Performance Measures

While value-added measures of school or teacher performance are the current state of the art, actually implementing them has proven more difficult than might have been expected, especially at the classroom level. Though all of our cases illustrate some intention to use value-added type performance measures, making valid student-teacher links has been a common problem limiting the use of classroom level value added. Some grantees lacked understanding of what value-added measurement required both in terms of data and of how to explain the concept and the results to stakeholders. In some of the cases, inability to get planned value-added type systems up and running resulted in use of attainment change or very simple growth models as the bases for incentives.

Grantee A's turnkey model included the use of both school and classroom value-added measures as two of the three bases for performance bonuses. Even this relatively large district did not, however, have a value-added measurement system in place, nor did it have the internal capacity to build one quickly. Fortunately, as a part of another initiative, it had contracted with an experienced external partner which worked to develop school and classroom measures for the entire district, as well as the TIF schools. Unfortunately, as described in the Design Evolution section above, the district's data systems were not initially adequate to support valid classroom value-added in the first year. This required using only school value-added and teachers' performance evaluation ratings as the bases for the first incentive payout. For the second payout, grade level value-added was used instead of individual level, as the data system issues began to get resolved. The district and value-added vendor have continued to expend considerable time and effort on teacher-student link issues.

Grantee B's initial foray into classroom value-added was plagued by teacher-student linkage problems due to data system inadequacies. As discussed above, these led to errors in bonus calculations and threatened the credibility of the project as a substantial number of teachers found they were being credited or discredited based on students they had not taught. Teacher understanding of value-added also appeared to be limited, and the district had difficulty explaining the value-added model. The district threw a lot of resources into correcting these problems, and appears to have succeeded, at least as indicated but many fewer problems with

bonus calculations in the second round of payouts. An external vendor with considerable experience brought in a model that had the credibility of use in other jurisdictions, and provided training to school teams of principals and teachers on the value-added data, its uses in the incentive model, and how it could be used in the classroom. This vendor was supported by another partner specializing in assistance in the areas of professional development, creating data files for teacher-student linkages, dissemination of teacher profiles for value-added results and payout information, staffing and maintaining help-desk phone lines, and conducting trainings for core teams, teachers, and principals.

Grantee C did not include value-added measures in its initial incentive design. Student achievement entered into the incentive payout based on the state's AYP criteria. The data systems needed to support a classroom value-added system did not exist at either the state or individual district level. The grantee also had limited technical capacity to develop new data systems or to develop a school value-added model, and did not have the resources to bring in an external vendor. There was also limited capacity to communicate about value-added to the participating districts. Not surprisingly, the there was not a lot of buy-in for teacher level value-added. Another challenge faced by this grantee is the small number of schools and even teachers in some districts, making district-specific value-added models problematic. The grantee basically has to wait until the state can deploy its limited resources to develop a statewide model that the participating districts could use. Changes in the state's achievement tests further complicate the development of value-added models due to the need to equate the new and older tests.

Grantee D began with performance measures that did not include value-added, but were based on the state accountability system, which emphasized attainment and improvement in attainment. It appears that grantee staff did not initially believe they had the knowledge, capacity, or need to use value-added performance measures. In the absence of a strong external partner with expertise in this area, it is easy to understand why this small district initially shied away from value-add. However, dissatisfaction with the perceived unfairness and lack of differentiation in the first year results led this grantee to look for a way to use value-added. CECR technical assistance staff helped grantee staff understand how value-added measures could address their concerns, and also how to incorporate the state's new growth model in their incentive design. This growth model was limited, however, to measuring school growth. The state model does not currently provide classroom value added estimates, and it is unlikely that this grantee has the resources to develop its own classroom model, even using state data. Because of its small size, the district would also have relatively few teachers on which to base a model, limited reliability of the results.

Grantee E began with a vision that included focusing teachers on student growth, and their incentive model was developed to do this. The model included an incentive for having students meet individually-specific goals set at the beginning of the year based on an unspecified growth model. Classroom and school achievement could also be measured in terms of value-added, but attainment or change in average attainment were also options. Due to the small school size and small number of schools, a value-added model specific to this grantee was not technically feasible. Teacher-student links, data management, and test credibility issues also worked against the development of a value-added model. Thus value-added was not used in any of the first year plans. While the grantee has put substantial effort into developing a data warehouse with

teacher-student linkages, and has developed school-specific systems to maintain linkages, the grantee continues to have concerns about value-added, in part due to lack of consensus in the field about what the most valid model would be, and in part due to lack of understanding of the technology. It looks as if none of the second year plans will use value-added measures. The student growth idea is being implemented in terms of changes in proficiency level or an increase in NCE scores rather than value-added as implemented by Grantees A and B.

The experiences of these grantees suggest a couple of conclusions about the use of value-added measures of performance. The first is that it is harder to do value-added well enough for pay purposes than many grantees expected, and than many policy advocates would like. The development of valid system of measurement is likely to require the involvement of external actors with more expertise that most districts currently possess. The good news is that such expertise is being developed and can be brought to bear if resources allow. The larger, more resourced grantees (A and B) were better able to utilize value-added results for the student achievement portion of the performance incentive and to provide professional development about the measures. Second, the best route to using value added for small programs is to piggy back on larger state or regional efforts. Small districts don't have the resources to develop valid measures on their own or to bring in an external vendor, and may have trouble getting reliable measures when schools are few and school sizes small. Race to the Top requirements will provide additional incentives for states to develop value-added and supporting data systems, which would make it much easier for small grantees to use value-added in performance pay systems. Government and foundations seeking to fund initiatives that use value-added may want to encourage program leaders to partner with proven providers of value-added measures, or to require programs to have some experience with producing value-added estimates.

Verification Mechanisms

One aspect of implementing performance pay systems that has been underappreciated by both grantees and policy advocates is the need to ensure that incentive checks for the correct amounts are sent to the right people. Three of the grantees we studied had to develop verification mechanisms from scratch to ensure that the right educators received the right payout. The experience of some of the grantees showed how important this was. Developing and implementing such a process appears to require a substantial commitment of time and effort.

As discussed above, Grantee B had major difficulties in its first year when payouts to individual teachers were made based on errors in linking students to teachers. The district invested substantial resources, obtained assistance from an external partner, and developed verification systems including a web-based tool for teachers to confirm the students they were responsible for instructing.

Grantee A's phase in of the incentive program and use of an external model has made verification less of an issue compared to Grantee B. For example, with only thirty initial schools it was possible to require the schools to audit classroom rosters to verify and correct student-teacher links. However, cooperation from schools varied, and district data systems continued to have weaknesses that require manual review. While few if any payout errors were reported, scale-up to the entire district will require a more sophisticated effort. Initial work on district-wide

value-added by the external vendor suggests that without attention to a verification process, problems with teacher-student linkages could lead to a smaller scale version of Grantee B's initial experience.

Grantee E took as proactive stance to verification using a combination of web-based tools and manual review to pre-audit all payouts. This process brought to light a substantial number of errors in payouts and it took several weeks working with schools to resolve them. However, this process minimized later complaints from educators and avoided the negative media attention experienced by Grantee B. As its information technology capacities increase, this grantee will be trying to automate more of this process.

Grantees C and D have used relatively simple incentive designs that have not presented major challenges of verification. Existing state systems are used for school-based student achievement measures, avoiding teacher-student link issues. Other measures are based on participation, which can be handled using simple additions to existing systems. These grantees would, however, need to make major efforts to ensure data quality if they moved to more complex incentive designs.

The major implication of these grantee's experience is that those designing performance incentive systems need to design and implement data quality measures (especially for teacher-student links if classroom value-added is to be used) and develop verification procedures to head off potential errors that could lower the credibility of and support for the program. As experience accumulates, we will be learning more about typical data quality problems and how to address them, and hopefully model process will become part of the knowledge base. External funders may want to require initiatives to be funded to develop data quality plans and verification systems.

Teacher & Principal Evaluation

While the TIF Request for Proposals included the extent to which the grantee "proposes to develop and implement a fair, rigorous, and objective process to evaluate teacher and principal performance multiple times throughout the school year" as a selection criterion, this has not appeared to have driven much innovation or use of performance evaluation as a basis for incentives. Across the five cases, there is considerable variation in the emphasis placed on teacher and principal evaluation.

Grantee A's use of an external model brought with it a well-developed teacher evaluation model that was integrated with the vendor's vision of instruction and supplemented a weak district process in the participating schools. Teachers in these schools were also required to get a rating on the district's system, but with the vendor's model given more emphasis (the ratings on this model were used as a criterion for the incentive amount). In the second year, the district decided to improve its own evaluation process, causing a conflict as to whether the vendor's or the districts system would be used in participating schools. Principal evaluation in participating schools is similarly bifurcated. The vendor's program implementation standards are part of the evaluation process and used as a basis for principal incentives. This is intended to focus principals' efforts on program implementation. The principals appear to be evaluated on the

district-wide system as well, and their schools' performance is also tracked by the district accountability system.

Grantee B uses a state-designed evaluation process, which is rather cumbersome and was not designed to align with district instructional strategies. Though teachers must be evaluated using the state system to be eligible for an incentive based on student achievement, there are no other links to pay. It is clear that Grantee B's the theory of action does not include teacher evaluation, nor was a process for classroom observations addressed in the original proposal. Principal evaluation is based on a 12 dimension evaluation instrument that includes one outcome-based dimension, but this is not integrated with the performance incentive which is based on different measures of school-wide student achievement. The district appears to be simply continuing the pre-TIF process which was designed to meet state rules for principal evaluation.

Grantee C's thinking about teacher evaluation has evolved over the course of the project. In the context of a cross-district project, one issue was how to unify the several district approaches. The original proposal was to use a research-oriented classroom observation protocol. After much deliberation, this grantee moved towards using the Take One Program of the National Board. The advantages are that this is pre-developed, credible, and does not require participating districts to change their evaluation systems. While this does not appear to fit with the idea of multiple evaluations during the school year, the grantee has based part of the teacher incentive on participating in the Take One process. Grantee C originally planned to adapt a national school audit tool for principal evaluation. Support would be provided around the 9 dimensions of this tool through a leadership academy and other training. It appears that this tool was never used. There are no incentives based on assessment results from this or any principal evaluation tool. There is an indirect link, however, in that some principal incentives are based on participation in leadership academy training and building leadership teams.

Grantee D began with an intention to change teacher evaluation. The proposal mentioned a model that included quarterly review of grade books, lesson logs, and student work as well as observation based on standards from the Center for Research on Education, Diversity, and Excellence. (CREDE). Efforts were made to revise the evaluation system, but it appears that what was implemented was a streamlined version of the system used prior to the TIF project. This grantee does not link incentives to evaluation ratings. For principals, this grantee initially proposed to build a "research based" principal evaluation system based on North Carolina Center for School Leadership Development materials. This development is ongoing. Evaluation results are not linked with the performance incentives provided to principals.

Grantee E's member schools appear to have started with either no or weak evaluation processes in place, and the project coordinator recognized early that a better model was needed. The project theory of action required a better evaluation approach to contribute to the development of a common language about instruction. External partners were used to introduce a new approach, but implementation has been slow and encountered problems typical of implementing a more rigorous approach. However, some schools have used teacher evaluation ratings as one criterion in determining the performance pay amount, and others have required that an evaluation be made for a teacher to be eligible for bonuses based on student achievement. With respect to principals, participating schools also appear to have had at best weak evaluation processes in place. Because

these schools are charters, the situation is complicated by the role of individual school boards and charter goals in assessing principal performance. Each principal thus faces a unique set of evaluation criteria. The grantee's central staff has encouraged schools to experiment with the Val-Ed 360 degree evaluation system, by has encountered resistance from some schools. This grantee is still working to get a uniform principal evaluation model in place.

The major implication for incentive plan designers is that teacher and principal evaluation may be difficult to add into an incentive model. There are a number of reasons for this. First, districts may have little faith in either the validity or credibility of current evaluation systems. There is reason to doubt whether such evaluations distinguish among teachers (Weisberg, Sexton, Mulhern, & Keeling, 2009) or influence teachers to improve practice (Peterson, 2000). Second, more effective evaluation, especially if the resulting ratings are to be linked to pay, is likely to require more resources (Heneman, Milanowski, & Kimball, & Odden, 2006). Given the emphasis of most performance incentive programs, including the Teacher Incentive Fund, on outcomes, it may not be perceived as worthwhile to invest in better teacher and principal evaluation. In two cases (Grantees A and E), teacher performance evaluation was intended to support a change in teaching practice that is in turn intended to improve learning outcomes. In both cases new evaluation systems had to be brought in and more resources provided to make them work. In considering whether to invest in better teacher evaluation, a district needs to think through how evaluation will fit into its overall strategy for improving teacher performance.

Principal evaluation is even more problematic. With the possible exception of Grantee B, there appears to be little integration of principal evaluation with outcome-based performance incentives into a coherent performance management system. Grantees appear to have difficulty developing and implementing principal evaluation, likely because the state of the art is less well developed than teacher evaluation, with fewer models and less supporting research. Part of the problem may also be that grantees lack the capacity to make major changes in both principal and teacher evaluation. It is notable that one of the grantees (though not one in our case sample) that has made progress in principal evaluation has a program that includes only principals.

Performance Pay Program Evaluation

Evaluation designs varied across the five cases. The strength of the evaluation was dependent upon the resources, knowledge, and capacity of the evaluator as well as the evaluator's role in designing the performance incentive system. The two larger grantees put more emphasis on impact evaluation (e.g., impacts on teacher retention or student achievement), while the three smaller grantees' evaluations tended to focus mostly on implementation issues. Implementation evaluations appeared to be generally useful to program managers in making mid-course corrections. Most grantees' evaluations did not appear to be guided by an explicit theory of action or logic model. The majority of evaluations did address some basic parts of standard evaluation models, such as describing the context of the program, documenting program components, and measuring the fidelity of program implementation. When outcomes were assessed, this was typically via changes over time without using experimental or strong quasi-experimental designs. Only one of the grantees will be able to provide a reasonable estimate of a causal effect on student achievement outcomes.

Grantee A's program evaluation was the most sophisticated. Using external funding, both an internally-conducted process evaluation and an impact evaluation conducted by a major national evaluator are continuing. The impact evaluation is using an experimental design with volunteer schools assigned at random to implement the model in the first and third year of the grant.

Grantee B, with significant internal evaluation resources, was able to evaluate both program operation and impacts. The program operation evaluation provided district leaders and the external partners with information on what was working well in the district and what needed improvement. Findings for program improvement included 1) eligibility issues, 2) model fairness, 3) streamlining verification process, and 4) requests for increased compensation. These recommendations allowed the district to consider strategies for mid-course correction.

Grantee C employed a local independent evaluator, with significant experience evaluating statewide policy initiatives. The evaluation is covering both program operation and impacts, but the emphasis has been formative. Partly because of limited institutional capacity, the evaluator has also played the role of technical assistance provider, responding to the grantee's need for more help in modifying the incentive design and performance measurements, as well as providing information to the grantee leadership on the readiness of participating districts for performance incentives, and challenges and obstacles to implementation. While the evaluator role provided the project with substantive information to improve program design and implementation, this was not always translated into substantive project improvements.

Grantee D contracted with an external evaluator with experience in evaluation of educational policy initiatives in the state. The annual evaluations are largely formative. The evaluation was particularly strong in its documentation of challenges and obstacles to the program, as well as strengths and weaknesses in fidelity of implementation. The evaluator worked closely with the district and provided ongoing feedback for program improvement.

Grantee E contracted with a regional external evaluator with experience in evaluating educational initiatives. The evaluator initially focused on assessing participating schools' readiness for a performance compensation system. The evaluator also used surveys, interviews, and observations to establish a baseline for some of the intermediate outcomes such as changes in professional development or school culture. In the second year, the evaluation cited some improvements in some student achievement measures, but stopped short of attributing them to the program. The evaluation design would not support such an attribution.

The usefulness of implementation/formative evaluations in helping modify programs or focus implementation resources suggests that program designers should include a strong evaluation component in their project. The usefulness of evaluation might be improved, however, if designers worked with evaluators to develop an explicit theory of action or logic model to guide formative evaluation. Since smaller districts are less likely to have the resources or the numbers of schools/teachers to allow a rigorous impact evaluation, there may be a role for states or consortia of districts in developing impact evaluations. It may be most efficient to do the formative/implementation evaluation internally and let a consortium do the impact evaluation. Program funders may want to get more mileage out of evaluation by requiring the development of an explicit theory of action or logic model to guide evaluation efforts, and to require more

rigorous impact evaluation. This would likely require that program designers bring in more capable external evaluation providers and that funders make more resources available for evaluation.

References

Flannery, T.P., Hofrichter, D.A., and Platten, P.E. (1996). *People, Performance, & Pay: Dynamic Compensation for Changing Organizations*. NY: The Free Press. Chapter 10.

Hale, J, & Bailey, G. (1998). Seven dimensions of successful reward plans. *Compensation & Benefits Review*, 30,4, 71-77.

Heneman, H.G. III, Milanowski, A., Kimball, S.M., and Odden, A. (2006). Standards-Based Teacher Evaluation as a Foundation for Knowledge- and Skill-Based Pay. CPRE Policy Brief RB-45. Philadelphia, PA. Consortium for Policy Research In Education.

Peterson, K.D. (2000). *Teacher Evaluation: A Comprehensive Guide to New Directions and Practices*. Thousand Oaks, CA: Corwin Press.

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. The New Teacher Project. Available at: http://widgeteffect.org/downloads/TheWidgetEffect.pdf

Table 1
Profile of Five TIF Grantees

Grantee	Type/Location	Year of Implementation	Number of Participating Schools	Incentive Types	Other
A	Large Midwest District	3	30	Bonuses for teachers & principals based on school & classroom value-added (when available). Also bonuses based on practice assessment.	First year was used for planning
B	Large Southwestern District	3	104	Bonuses for teachers & principals based on school or classroom value-added derived from state tests.	
C	State-led Consortium of Rural & Small City Districts	2	40, over 10 districts	Bonuses for teachers & principals based on meeting school-wide state test attainment goals. Additional bonuses based on professional development and practice assessment.	
D	Small Western District	3	4	Bonuses for teachers & principals based on school-wide improvement in attainment on state tests.	Has changed incentive model during course of program.
E	Consortium of Urban Charter Schools	2	10	Bonuses for teachers & principals based on a variety of indicators including test score attainment, improvement in attainment, and student growth. Also bonuses for filling teacher leadership roles & sharing best instructional practice.	First year was used for planning