UNLOCKING THE BLACK BOX: HOW DELAWARE ELEMENTARY SCHOOLS FULFILLED THE EXPECTATIONS OF PERFORMANCE MEASURES AND ACCOUNTABILITY

by

Glenn L. Silverstein

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Public Policy and Administration

Spring 2011

Copyright ©2011 Glenn L. Silverstein
All Rights Reserved
Chapter 7

CONCLUSION

This country began as an experiment in balancing the competing rights, values, and responsibilities of the federal, state, and local governments with those of the individual. Perhaps, nowhere is this competition more clearly and passionately fought than the recent battles to reform education. In the latter half of the twentieth century, the battle of federal education policy centered on the rights and values of access and equity; the new century brought with it new concerns over school performance and accountability. These new ideals, however, are in conflict with traditional values of local control of public schools. Since education is a ubiquitous methodology by which society passes on to its children its values, its history, and its future, the victor in this conflict will have consequences long into the future. It is unclear which system will ultimately produce a “better” citizenry, but the conclusion of this conflict will indicate which direction this country is moving. Does the country still maintain its confidence in localism, in regional diversity, in its teachers and in its schools, or for the sake of potential effectiveness and control, increase the uniformity of the educational system and move the locus of control away from schools and school districts and towards state and federal governments. This dissertation does not take sides in this conflict; it does, however, mark its effects.
The conflict surrounding accountability and testing is really involves questions about school and, to a lesser degree, effectiveness. Improvement to effectiveness once primarily involved discussions surrounding pedagogy and learning styles; and while these subjects are still important to teachers and other education professionals, discussions addressing broader student achievement have moved beyond events in the classroom to the effectiveness of the school itself. NCLB is the most recent manifestation of these discussions, and while it is certainly not the last word in education reform, it is such an important step in the process that it was imperative that an understanding of its effects upon the schools be developed. To fulfill the mandates of NCLB, Delaware developed the DSTP which required school to attain specific achievement benchmarks. To achieve these benchmarks, most Delaware schools changed many aspects of their respective organizations. To assist in achieving a broader understanding of these changes, this dissertation had three goals: (i) to explore the organizational changes that occurred in Delaware elementary schools, (ii) to determine if the organizational changes identified had any relationship to educational achievement, and (iii) to determine to what degree the DSTP fulfilled the requirements of a proper performance measurement system.

Since this dissertation examines education policy, it seems logical that one way to determine the level of achievement of the DSTP as a performance measurement system is to review the achievement levels it earned against Hatry’s rubric. By this measure, the DSTP “Exceeded Expectations” on eight key elements, “Met
Expectations” on eleven key elements and fell “Below Expectations” on five key elements. This would suggest that the DSTP is an adequate performance measurement system with several design and implementation aspects that are very commendable, but it also has some serious flaws that must also be addressed.

Some of the commendable aspects include the design of the DSTP as well as the dissemination of the results. The DSTP was designed with input from many groups of people with a wide range of interests, experiences, and perspectives. By getting input and commitment from many different stakeholders the end product was improved, implementation resistance was minimized, and support, particularly among district and State leaders, was strengthened. Administrative support for the DSTP is evident by the increase in data and analytical capacity of the State.

Some of the requirements mandated by the federal NCLB legislation also contained some good ideas. The requirement that the test data be disaggregated allows communities and individuals with particular needs to be more easily identified and the 95% participation means that it is more difficult to game the test by hiding the results. The requirement that test data be widely reported ensures that key information reaches everyone. And while the released test data may not provide enough information to confidently base school-wide decisions upon, it may provide enough information to identify some individual student grade, and school weaknesses as well as indicate individual teacher development needs, teacher and school professional development needs, and budgeting decisions. While the information does not disclose
why the test results occurred, it does provide information on what performance level indicators were not achieved.

Another area where the DSTP excelled, however unintended and undirected, was the effects that testing and AYP has had upon the educational organization itself. One of the most interesting aspects of these changes is the effect that the DSTP has had on school principals. The expectation level of AYP is relentless. The goal is unaffected by any variable, endogenous or exogenous, including school population, teacher education or teacher experience. Moreover, the penalties for failure are significant, public, and escalatory. Since school principals are not protected by union contracts and are the chief executives of the school, they are most exposed to individual sanctions for failing to meet AYP goals. School principals, however, have not become more desirous of control in response to AYP mandates; instead, they have embraced new ideas about leadership and decision making. School principals in nearly every achievement category have become more instructional in their leadership approach. While accepting that school principals cannot be the sole repository of knowledge, they have nonetheless broadened their understanding of the curriculum, teaching methodologies, and educational expertise. In addition, rather than centralizing the decision making process, school principals have empowered and encouraged shared decision making with various teams within their own buildings. The degree to which decisions are shared is more varied than the idea of instructional
leadership, however, even schools that only have the “district-mandated” team in their building consult with them and are accepting of differing points of view.

The change in leadership culture has also fostered a change in staff culture. The idea of performance measurement is one of competition, but within the schools teachers have become more cooperative and cohesive in response to the challenges posed by AYP. Nearly all of the schools included in this dissertation have reported improvements in the climate of the school, increased desire to succeed, and greater usage of data-driven education. These changes are no mirage, they are real; this is somewhat surprising, and reflects a “rise to the challenge” attitude. This is not to say that as AYP becomes more difficult to achieve and the bar is raised once again that school cultures will not suffer the malaise or frustration that failure often brings, but for now, these cultural changes are a positive result of the DSTP.

Just as the positive contributions made by the DSTP are impressive, the severe limitations of the DSTP are equally problematic. The limitations include: questionable definition of student achievement; problematic correlation between standards, curriculum, course arrangement, and the performance measure; the measure itself; testing effectiveness without regard to efficiency; and the increasing distance between principal and agent. For each of these limitations there is a cadre of educational researchers that have voiced their objections. Even if these limitation did not exist, there is still the larger questions regarding whether high-stakes testing can
achieve universal proficiency and if so, is it economically effective, efficient, and fair to do so.

For the performance measure to accurately reflect the knowledge of the students that the DSTP is supposed to judge, there must exist a very strong correlation among the standards, the curriculum, and the assessment. As shown in this dissertation, such a strong correlation is suspect. The assessment is reviewed often and individual items are replaced each year, but the level of alignment between the DSTP and the Delaware content standards and grade level expectations required to confidently impose the significant punishments upon a school is not fully present. Perhaps, statistical evaluation will show an improving situation in the future, but until the measure has been shown to produce the advertised information, it is unfair to hold schools accountable for the results.

Student achievement is a growth process not a single destination; however, using a single measure suggests that everything about a student’s educational experience can be discerned by how well they perform on the test. A single measure provides information about the condition of knowledge at a discrete point in time; it does not indicate the growth in knowledge over the course of the school year or even the effect summer vacation has on knowledge retention. The State of Delaware appears to recognize this since it collects a plethora of data about the process of education, but while clearly the State thought this information essential enough to collect, it cannot be used for accountability purposes under the DSTP. By focusing
solely on the score of the test, it is possible that the skills required for learning beyond the curriculum will be lost. Developing good study habits and research skills, promoting outside reading, and encouraging curiosity are essential to producing lifelong learners and not simply test takers.

Without adding other performance measures to the DSTP it is unlikely that it can ever become more than an adequate performance measurement system. One method to improve the DSTP is, of course, to broaden the definition of student achievement and collect additional data. As stated above, Delaware already collects additional information to improve the DSTP system and even if those variables are not acceptable to the federal government the DDOE could still use the information as part of a supplemental system to improve student achievement, but, to date, it has declined to do so. A second method to improve the DSTP is to keep the definition of student achievement but change it from a single cross-sectional measurement into a longitudinal growth-based measurement. To implement this, the NCLB or its companion regulation must be altered\(^{14}\).

\(^{14}\) Since the writing of this dissertation, Delaware has implemented the Delaware Comprehensive Assessment System (DCAS) which has changed student testing methodology and the performance measure from a single cross-sectional measure to a computerized measure collected three times a year. Further research is needed to evaluate the results of this change and whether the measure has been improved by the implementation of DCAS.
Earlier this year, the Obama administration released its first blueprint for changes to the Elementary and Secondary Education Act. Although this plan did not directly alter the present definition of student achievement defined by NCLB, it did include a program called *Race to the Top* (Department of Education, 2009) which included a third method for improving the DSTP which is to control the variability in policy implementation and data analysis. In other words, *Race to the Top* adds greater control over the endogenous variation surrounding the collected measurements by shortening the distance between the principal, which is the federal government, and the agent, which have always been the schools.

*Race to the Top* has some admirable stated goals (Department of Education, 2009):

(a) Adopting internationally benchmarked standards and assessments that prepare students for success in college and the workplace

(b) Building data systems that measure student success and inform teachers and principals about how they can improve their practices

(c) Increasing teacher and principal effectiveness and achieving equity in their distribution

(d) Turning around our lowest-achieving schools

The underlying methodology that it uses, however, outright threatens the idea of local control of schools by moving curriculum development and testing to the national level and by moving implementation of *Race to the Top* to the state.
In March 2010, Delaware became only one of two states to be accepted into the federal government’s Race to the Top education plan. While admission into this program offers the opportunity to feed at the federal trough, the requirements make it clear that there is no free lunch.

To satisfy the goal of Race to the Top’s benchmarked standards, Delaware has decided to adopt the “Common Core Standards developed by the Council of Chief State School Officers (CCSSO)/National Governors’ Association (NGA)” (The Delaware Department of Education, 2010). As shown in this dissertation, Delaware spent years developing their own standards in cooperation with many different stakeholders. By agreeing to adopt these uniform standards, Delaware has now limited the ability of these stakeholders to comment on or modify the foundation of their respective schools’ curriculum. On the other hand, Delaware also plans to implement the Delaware Comprehensive Assessment System (DCAS) which can provide “up to three computer-adaptive formative assessments and one summative assessment per year, including end-of-course exam in high school” which will enable Delaware to add a longitudinal aspect to the performance measure it now collects (Department of Education, 2009). In order to implement this system, Delaware plans to use a significant portion of the federal grant money to upgrade its computer system.

Delaware receives other benefit as well, particularly in teacher training and professional development, but school principals will no longer be in charge of policy implementation. The state has moved implementation of the reform plan and
responsibility for the outcomes out of the schools and into DDOE’s own offices by 
establishing the Project Management Office (PMO). “The PMO’s Performance 
Management Team (PMT) will monitor performance against goals, identify when 
performance is off-track, and intervene early to ensure that goals are met on budget 
and on time. The PMO’s Teacher and Leader Effectiveness Unit (TLEU) will be a 
key implementation resource, managing reform program in professional development, 
evaluation, career paths, recruitment, and retention…In short, these teams will be 
According to the Race to the Top application, not only are school principals removed 
from a large part of present school governance, the shared decision making that has 
been developed by the initial implementation of the DSTP may be laid waste as well. 

One of the purposes of this dissertation was to explore what happens to the 
educational organization when it is forced to achieve goals not of their own choosing, 
as related by several principals and in response to the exogenous stimulus. Teachers 
modified their approach to teaching, shared decisions, shared information, became 
more cohesive, and mentored each other so that the school was in a better position to 
reach the required level of achievement. School principals were also part of the team. 
School principals embraced new leadership approaches and roles and adjusted their 
own approach to success to the new reality of accountability. They served as leaders, 
resident experts, cheerleaders, and primary authority figures, and according to the 
surveys, they were not only viewed positively, they contained the multitude of skills
necessary to achieving the quantitative goals established by the state and the qualitative goals established by the students, parents, school staff, and other stakeholders. It is not known whether the members of Delaware’s new PMT or TLEU will garner the same respect or be able have the same positive effect upon the school as its principal. It is also unclear what role the school principal will have in this new educational regime and how s/he will interact and react to the new *Race to the Top* reality, but it seems possible that dividing the leadership and responsibility between the school principals and state officials could have detrimental effects and will require a similar study to this dissertation.
Future Research

Dissertations serve several purposes. Certainly, it stands as the final piece of evidence that a student has the mind and ability to command the title Doctor of Philosophy, but for those students that view the dissertation as a beginning rather than an end, a dissertation identifies future research opportunities. This dissertation certainly identifies several likely prospects.

One of the first and easiest research projects that should be undertaken is testing the regression equation against a broader population of schools. If test scores can be reliably predicted or influenced by so few variables, than a case can be made that the most efficient method for improving student achievement is by directly addressing those particular variables. The two variables in this study, percentage of teachers with five or more years of experience and percentage of non-white students, explain approximately fifty percent of the variation in test scores. That still leaves a considerable amount of variation unexplained, but eliminating the lion share of uncertainty is a good start.

Another area of research is provided by the organizational findings in this dissertation. Having found the widespread application of instructional leadership and shared decision-making concepts among all elementary schools in the study, a more refined protocol may identify the details behind these organizational changes and to uncover the actual causation of their widespread implementation. Additional
comparisons of organizational changes can be explored to discover similarities and
differences based on demographic, geographic, or achievement categories.

Improvement to school culture is another promising area for future research. Most school principals reported that the DSTP has made significant improvements to the culture and climate of their school, but it is unclear what specific aspect of the implementation has been so beneficial. This is particularly important given the general negative attitude towards the teaching profession expressed recently by Governor Chris Christie and the negative view of teaching expressed in No Child Left Behind itself.

Finally, this dissertation performed a performance measurement analysis of the DSTP and found it to be an adequate performance measurement system. Race to the Top, however, makes substantial modifications to the current system. It is logical, therefore, that before a new system is implemented, that a full program evaluation be conducted on the current DSTP to gain a better understanding of the reasons underlying the performance measures. What did the DSTP perform well, what were the problems? What system gaming, if any, took place? Were the previous underserved and minority communities better served by the DSTP than previous education reform regimes, and if not, why not? What specific policy or program had the greatest impact on the performance measures? These questions can be answered only by a full program evaluation and these questions should be answered before a new education reform regime is installed because once the new system is
implemented, it will be far more difficult to find answers free from confounding interactions.
Afterwords

The goal of the DSTP is to improve student achievement, but like many implemented policies, there are effects that extend far beyond the policy's intention. This section has highlighted many of the organizational effects initiated as a consequence of the implementation of the DSTP. Many of them seem to be beneficial. Broadening school principals' notion of leadership can be beneficial if it provides them more options to lead and motivate their staff. Shared decision making can be beneficial if for no other reason than it brings different ideas to the table. Since teaching is at least as much an art as it is a science, additional voices will add more perspectives to assist in achieving AYP, and it may help younger teachers develop by providing better avenues for communication. The changes in school culture described by the school principals seem to indicate that this type of information sharing has already begun to have positive effects.

Not all the consequences, however, have been positive.

There is less joy...Yes, it's more of a business of education rather than an experience...I find I am equally as motivated now as I was before, I find it more challenging because I have less control. I can't [decide] who comes in here; I can't decide who I hire; I can't decide lots of things, but yet I am still responsible for everything. So while I am still motivated as the building administrator, I find it
increasingly challenging to do what I know I need to do and what I want to do with having my hands somewhat tied behind my back... (Supreme, 2007).

This quote from the principal of the most accomplished school dramatically highlights a difficult truth about performance measures and accountability. To measure performance you need to eliminate variation so the results have validity and reliability. In education, accountability means creating a single curriculum, using a single text, and promoting similar teaching styles. It runs the risk of turning the teaching profession into an assembly line job where every third grade English teacher is doing the same thing on November 22 as every other third grade English teacher in the state. It also means that principals are constrained to innovate and are reduced to being scientific managers simply ensuring compliance. It is no wonder that this particular principal, one of the longest serving principals and one of the most successful, finds the new accountability restraining and is concerned that the result of accountability is less joy. No Child Left Behind implemented a system where curriculum is unified and the only measure of consequence is a single test; the results of which contains only sticks for failure, but no carrots for success. In addition, just as NCLB views teachers as unprofessional and lazy, it also views students as passive non-contributors in the process of their own education (Ravitch, 2010). Families are also a non-consideration in the education process, which seems more than a bit disingenuous since “parent involvement” has been such a school improvement indicator for government officials since before NCLB. Accountability may very well
improve student performance, but over time, there is a danger that the teaching profession may be sacrificed in the wake of performance measurement.