The Role of Accountability Systems in Driving School Improvement: A Case Study

Michael Flicek
Natrona County Schools, Casper, WY

Since the passage of the Elementary and Secondary Education Act (ESEA) in 1965 through its current incarnation as No Child Left Behind (NCLB) the United States has had a public policy designed to encourage equity in educational opportunity for disadvantaged students in kindergarten through grade 12 settings. Substantial federal resources have been directed toward this effort over the years. Progress toward this goal has been measured by monitoring the gap in academic skills of students from disadvantaged backgrounds versus those from more advantaged backgrounds. One indicator of disadvantage is family income. Specifically, the gap in academic skills of students eligible for free or reduced price lunches versus those not eligible for free or reduced price lunches has been monitored. Despite the 40+ year public policy history of addressing this problem, considerable inequity in outcomes persist for students on free or reduced price lunches (e.g., see Lee, Grigg, & Donahue, 2007).

The support for public policy in the United States around closing this achievement gap is not surprising in a country that sees itself as the “land of opportunity” (Sawhill & McLanahan, 2006; Sawhill & Morton, 2007). Americans have faith that hard work and merit will, and should, be rewarded. As such, Americans tend to support “equality of opportunity” while they are less compelled to support “equality of outcomes” (Aaberge, Colombino, & Roemer, 2001; Benabou, 2000), particularly if such equality were to be achieved in the absence of merit. Equality of outcomes unrelated to merit is viewed, with justification, as patently unfair. Outcomes tend to be easily measured while the measurement of opportunity is more elusive. Without inequality in outcomes there would be little reason to be concerned about inequality of opportunity. Inequality in outcomes, therefore, constitutes necessary evidence that inequality in opportunity may be present. A search for potential instances where opportunity is not equitable is appropriate when gaps in outcomes are present.

The achievement gap is an easily measured outcome. The National Center for Public Policy and Higher Education (2005) offers a powerful economic argument supporting efforts to narrow the achievement gap. They concluded that if the current gap persists, a per capita decline in income in the U.S. will result. Conversely, their report concluded that narrowing the gap would be associated with a gain of as much as $425 billion in total U.S. income. A more recent study by the Organisation for Economic Co-operation and Development (OECD, 2010) demonstrated the importance of improved test scores on the Programme for International Students Assessment (PISA) to gains in GDP. Reasonable gains in PISA scores were projected to be associated with a $41 trillion gain in GDP over the next 80 years. These economic models...
suggestion that goal of narrowing the achievement gap while at the same time lifting the achievement of all students would have clear economic benefits to a society.

Growth, versus decline, in economic output is clearly in the interest of all parents. Our children have the best prospect of attaining the American dream (i.e., a higher standard of living than their parents) in a growing economy (see Sawhill & Morton, 2007). Sawhill and Morton put forward the following argument. In a growing economy, children with the same education level and occupational level as their parents tend to have a better standard of living than their parents as a result of the growth in income. *A rising tide lifts all boats.* In contrast, when the economy is stagnant or in decline, for every individual that gets ahead of their parents, another individual must fall behind. *The boats change places.* The success of schools in narrowing the achievement gap is fundamental if we are to have a growing economy in which all boats have the opportunity to rise (i.e., for the American dream to be realized).

Berliner (2005) advanced an argument challenging America’s will to live up to the ideals of meritocracy that are outlined in the above paragraphs. He used comparisons on international indicators to advance a compelling argument that responsibility for the achievement gap in U.S. schools does not reside with schools alone. Berliner presented a convincing case that, as a nation, we (i.e., the U.S.) have accepted conditions such as unacceptably high levels of childhood poverty and inadequate health care for children. These conditions result in children coming to school not ready to learn. This is not to say that schools should not be held accountable, rather, he argues that accountability should be reciprocal. Both schools and communities have some responsibility. Our communities should be held accountable for delivering children to schools healthy and ready to learn. He concluded that many of the advantaged in our society “vote to keep income inequality as it is” (p. 989). That the advantaged would favor conditions that, in their view, would preserve advantage for them and their children is understandable, however, it may well be short sighted given the economic theories put forth above. Nevertheless, such actions do work against the American value of meritocracy and, also, as outlined in the above paragraph, to the extent that these views result in preserving an achievement gap, they ultimately work against the long term interest of all members of our society. All of that aside, Berliner is certainly correct that schools would be more successful in efforts to reduce the achievement gap if the outside of school social conditions for disadvantaged students were to improve.

While it may be true, as Berliner suggested, that schools are being scapegoated, my belief is that there remain things that many, and perhaps most, schools could do that would increase equity in opportunities for disadvantaged students and thereby improve equity in outcomes as well. My experience suggests that some and perhaps substantial narrowing of the achievement gap is within the control of schools. To the extent that this is true, schools can play a meaningful role in supporting our meritocracy and assuring that there is within our economy a rising tide that will lift all boats. This paper will address the role that accountability systems play in supporting
efforts to increase educational outcomes and increasing equity by diminishing the achievement gap.

Test-Based Accountability: A Theory of Action

Regardless of whether formal test-based accountability systems are in place, Elmore (2004) has properly observed that there are “stakes” for students associated with “being poorly educated” (p. 282). Those stakes include a range of well known negative outcomes like dropping out of school and diminished lifetime earnings. Schools with large proportions of students from economically disadvantaged backgrounds are most likely to be underperforming. Hence, it is economically disadvantaged students who are most likely to pay the price of “being poorly educated” due to inequality in opportunities. A foundational rationale for the implementation of school accountability systems that rely upon the test performance of students is the belief that transparency around school performance might serve to motivate school improvement efforts (Baker & Linn, 2004). This sentiment is clearly evident in Bryk, Sebring, Allensworth, Luppescu, and Easton’s (2010) observation from their work in Chicago public schools when they noted that a centralized accountability system that identified underperforming schools was needed to “jump start” school improvement efforts in some schools.

The experience in Chicago (Bryk, et al., 2010) demonstrated that some underperforming schools were able to improve following increased transparency around performance. This suggests that some schools have the capacity needed to improve and that transparency around test performance alone might be sufficient to motivate changes that lead to improvement. To the extent that transparency derived from an accountability system motivates some schools to improve student achievement, some students will have avoided being poorly educated. This fact alone supports the call for transparency around student achievement at schools unless, of course, there is evidence that increased transparency of test results contributes to a decline in student achievement at other schools. It is, however, difficult to envision a mechanism by which increased transparency around performance would lead to a decline in performance.

When individual schools lack the capacity to improve, more than additional motivation from an accountability system will be needed to produce better results. Helping these schools to develop the needed capacity to improve is first and foremost the responsibility of the district’s central office since the central office shares with the school the responsibility for the stakes to the students that are the result of poor achievement. Bryk, et al. (2010) called into question the capacity of central offices to assist schools in developing capacity. They indicated that the capacity of Chicago’s central system to support school improvement or to “getting it right” (p. 218) even if they tried was quite limited. A potential obstacle for direct involvement of central office staff in a school’s improvement efforts involves relational trust. Bryk, et al., found relational trust to be an essential ingredient for school improvement to occur. Relational trust among school staff and central office leaders can be a challenge. In fact, one argument that Bryk et al., made in favor of decentralization was that energy that school staff use to “resist and overturn” (p. 219) reforms initiated by the central office could be channeled into energy to improve. State departments of education would likely have all of the challenges that district central offices had when it comes to helping schools build needed capacity for improvement.

State departments could be expected to have all of the issues of capacity and relational trust that a central office would have and they have the added burden of being even further removed from the scene of the action.

On a more positive note, Datnow and Strigfield (2000) reported that districts have an important role in supporting schools as the schools build capacity for improvement. Furthermore, both district offices and state departments can help to “protect” the efforts of schools to reform. In addition, district offices and state departments can provide schools with the needed resources for improvement. Fullen (2003) also reported that the integration of accountability with capacity building by districts and states can contribute positively to large improvements in student learning. He cited examples in two large U. S. school districts and in England. An optimal situation might exist when the level of the system that administers the accountability system (i.e., the district or the state) also has the capacity to assist underperforming schools in building the capacity that they need to improve.

Research provides ample guidance to states and districts that are interested in supporting the effort of schools to improve. There is a growing body of research documenting the conditions that are associated with school improvement (e.g., Fullen, 2001a, 2001b, & 2003). For example, Newmann, King, & Young (2001) indicated that capacity to improve consisted of (a) knowledge, skills and dispositions of individuals, (b) professional community, (c) program coherence, (d) technical resources, and (e) principal leadership. A more recent very rigorous empirical investigation of school improvement in Chicago (Bryk et al., 2010) identified a remarkably similar set of conditions as necessary for schools to improve. The conditions identified by the Chicago researchers were (a) school leadership, (b) professional capacity of staff, (c) an instructional guidance system, (d) parent community ties, (e) a student-centered climate, and (f) relational trust. A careful study of this literature can provide district or state leaders with a clear road map to follow as they endeavor to support the efforts of schools to work to develop the capacity needed to improve.

A caveat involves those schools that serve large concentrations of students who are living under what Bryk et al. (2010) called “extraordinary circumstances” and schools that serve students coming from neighborhoods with extremely low social capital. Homeless students, students who are abused or neglected and students living in foster care are examples of students in extraordinary circumstances. Neighborhoods with extremely low social capital are those with high crime rates and ineffective social organizations. Children in these neighborhoods have unmet needs for physical and emotional safety that push educational concerns into the background. David Berliner’s (2005) call for neighborhood’s to send students to school ready to learn really resonates in schools serving large concentrations of students living in either of these conditions. School improvement under these conditions is but one part of a larger societal challenge of improving conditions for all citizens in these neighborhoods who are under these multiple sources of duress. While schools in my district are dealing with individual students living in these extreme conditions, I do not believe that these conditions are present in sufficient concentrations in any of our schools that they would prevent improvement from occurring.

How This Works in Natrona County School District (NCSD)
NCSD has about 11,800 students. I presented a detailed description about NCSD’s path to the development of a district accountability system at last year’s annual meeting (Flicek, 2009). The 2009 paper also described the accountability model in considerable detail along with some findings from the first couple of years of using the system. I will not repeat all of those details here. Instead, I will provide a brief description in order to show how a credible test-based accountability system is playing a role in the improvement of schools in NCSD. We wanted an accountability system that would produce conclusions about school performance that were credible to the staff working in the schools. We chose not to use the NCLB accountability mechanism that was being administered by the Wyoming Department of Education because of our belief that it was not a valid system. Small schools with poor performance met adequate yearly progress (AYP; i.e., “met AYP” is equivalent to a grade of “pass” in a pass/fail system) while large schools with much better results did not meet AYP (i.e., “failed”). More to the point AYP was not viewed as credible by the staff working in the schools. Logic suggests that results from an accountability system that are viewed by staff in the schools as credible would be more likely to motivate improved effort.

Adequate Yearly Progress (AYP)

This section briefly describes some reasons for developing a district accountability system rather than relying upon AYP. First, we wanted to celebrate schools that were excelling. AYP does not have a category for excelling schools. Our district system calls schools that are identified as underperforming focus schools. Our district accountability system identified eight of the 22 elementary schools as focus schools during the first year of implementation of the system. This occurred despite the fact that all 22 of the schools met AYP. Our judgment was based upon available data that clearly showed underperformance at these schools relative to all other schools in our system (e.g., the focus schools had low status, they were not improving, and they had students showing low growth over time). Because of the problems evident in AYP accountability that were associated with holding schools accountable for subgroup performance, we chose to focus on subgroup performance at the district level rather than the school level.

In addition, AYP was based almost exclusively on the percent of students who were proficient or above on one test (i.e., the state test) in two content areas. Using this approach scale scores are reduced to a dichotomous score and valuable information is lost. More concerning, however, this approach encourages schools to give more attention to students near the proficient cut-score to the neglect of students above or below that cut-score. Schools earned credit or blame only when students move up or down across the proficient cut score only. Up or down movement away from the proficient cut-score was not recognized. We wanted to encourage up movement and discourage down movement for all students, not just those near the proficient cut-score.

Measures of Student Achievement

We embarked upon implementation of a district accountability process with the full knowledge that no tests and no accountability system would be perfect. We intentionally designed an accountability system for purpose of motivating improved reading, math, and
writing skills. We had good measures in these areas and the skills in these areas were seen as foundational to success in other content areas. Baker and Linn (2004) have argued that accountability systems should use multiple sources of data. We had two important sources of test scores in the reading, math, and writing/language usage. Our district had been using achievement measures from Northwest Evaluation Association (NWEA) for nearly a decade at the time that we began developing our district accountability system. The current version of the test is a computer adaptive test called the Measures of Academic Performance (MAP). The test locates each student’s position at any point in time along a vertical learning progression of basic skills. The MAP encouraged the differentiation needed to move students from their current location on the learning progression to higher levels regardless of their grade in school.

In contrast, the state tests in these content areas, and the standards upon which they were built, tend to promote common targets for all students based upon their grade in school. In Wyoming, the state test was built to the specifications recommended by the Commission on Instructionally Supportive Assessments (2001). In reading, for example, there are just eight important reading comprehension skills that are measured on Wyoming’s test (e.g., plot in narrative reading, major points and supporting details in expository reading). This test assumes students reading skills are at a grade appropriate level on a learning progression of reading skills. Grade level appropriate reading skills are necessary but not sufficient for success on the state test. Mastery of the eight important reading comprehension skills is also necessary for success. The MAP tests allow us to monitor student progress along the progression of reading skills (i.e., where student skills are in relation to grade level expectations) and the state test allows us to determine if students have the important comprehension skills. The two tests were similarly complementary in the other content areas. Given these considerations, it was important for our school accountability system to encourage schools to have students perform well on both of these measures.

Questions about Test Performance

Carlson (2002) demonstrated that test results can be used to answer a variety of questions about a school’s performance. Different conclusions about a school’s effectiveness are suggested depending upon which question is being asked (Carlson, 2002; Flicek, 2004). The first question involves the absolute achievement level at a school. Instead of measuring a school’s status against an arbitrary standard of proficiency, we chose to monitor our progress relative to normative benchmarks drawn from beyond our district. Because the benchmarks were from outside of our district (e.g., the state mean versus the district mean) all of our schools had an opportunity to meet the benchmark. While normative comparisons were out of favor not so long ago, Linn (2004) reported that there was a trend developing among states of incorporating norms into their accountability models. Specifically, for each grade at each school we were interested in the extent that the mean scale score at the school differed from the state-wide mean. On the MAP test we compared grade level mean scores from the schools to the norm sample mean scale score (i.e., which is referred to as a RIT score).

A second question about a school’s performance that interested us was referred to by Linn (2004) as a successive groups model. We refer to this indicator as improvement. The focus

here is upon changes to mean scores at each grade over time within each school. We acknowledge that reliance upon these scores in the absence of status scores and the next score of interest, growth scores, would be problematic. Specifically, these change scores contain enough error variance (Kane & Sanger, 2002) so that not all of the change in improvement scores over time can be attributed to true changes in the school’s effect on student achievement. Nevertheless, when used in combination with status scores and the next scores of interest, growth scores, we feel that change scores contribute to the robustness of our understanding about the effectiveness of our schools. An appeal of this method is that progress can be demonstrated even when schools serve students who start in different places. This method assumes that students at a school tend to be similar across school years. Unless external variables are causing sizable demographic shifts at a particular school, Linn (2004) suggested that it is “reasonable” to assume that the students schools serve tend to be about the same from year to year.

Finally, we were interested in the extent that individual students were progressing over time at a school. We refer to this as achievement growth. The state test, being a grade specific test, was not well suited to the measurement of growth. The MAP test, on the other hand, was specifically designed for the measurement of achievement growth. Changes to a student’s position on the vertical learning progression from one spring to the next were an indication of the student’s growth in that content. NWEA has growth norms that include more than 1.5 million students. Within the norm sample there is a sample for each grade-by-pretest RIT score. Using this norm sample it is possible to compute spring-to-spring $z$ scores for growth for each student (see Flicek, 2004). This pretest-posttest approach to the measurement of growth is problematic in that it is difficult to separate true growth from measurement error (Singer & Willett, 2003). We mitigate this problem by averaging individual student growth $z$ scores across two school years with three waves of spring test scores involved. The result of this procedure is similar to what would be achieved using a two level hierarchical linear model with three waves of data. It differs in that the growth comparison is against the NWEA norm group of more than 1.5 million students rather than a comparison within a local sample. This gives us an external benchmark for growth so that, once again, all schools within our system have a chance to succeed.

To summarize, there are five indicators of performance at each school in each of three content areas. On the state test both status and improvement are considered. On the MAP test three indicators are considered; status, improvement and growth. More specifics about these indicators are available in Flicek (2009). As mentioned above, we decided to address subgroup performance at a district rather than the school level. Many of the issues around poverty and disabilities involve policy and practice issues that are centrally administered through Title I services and special education.

Professional Judgment Panel Process

With three different ways to look at the data from one test and two ways to look at the data from another test we were confronted with the problem of how to aggregate results across the different methods and tests within a content area to reach an appropriate classification of focus (i.e., underperforming), excelling, or unclassified (i.e., neither underperforming nor excelling) for a school. We concluded that a judgment panel approach might have certain
advantages over a formulaic approach. Specifically, if a panel of professional educator judges could reach agreement about a schools’ classification by inspecting data from the two measures in the multiple ways, then we’d have some evidence that they understood the indicators and could explain their reasoning to others. The ability for this to occur would lend a certain credibility to the process that might have been lost had a formulaic approach been used. Furthermore, weighting of the different indicators was no simple matter. For example, a high status school might have little room for improvement. Improvement should be given less weight in the classification of this school. Another school might have low status with little evidence of improvement. Improvement should probably be given more weight at this school. Finally, growth might be weighted as important regardless of status or improvement. This issue of weighting is routinely discussed during training of panel members. Baker and Linn (2004) argued that any weighting of different indicators in an accountability system should be made explicit. Nevertheless, we have so far avoided hard and fast rules around the weighting of the various indicators. As the process matures we might be able to begin providing more specific guidance based upon the history of panel decisions.

The panels have typically included three or four district curriculum leaders, teacher association leaders, principal association representatives, special education representatives, and Title I representatives. Panel membership has ranged from ten to 16 individuals. The professional judgment panels were facilitated by staff members from the assessment and research office. The panels were convened separately for each of the three content areas of reading, mathematics, and writing/language usage. There was a packet of data for each of the 30 schools and the schools were identified by random codes so that the school identity was not readily apparent to the raters. Training of panel members involved going through the data from a randomly selected school and discussing each indicator and then what rating might be appropriate and why. There was considerable discussion during this process. Once panel members began making independent ratings they were free to ask the assessment office facilitators for clarification or to check for understanding. A form was provided for each rater to rate the school on each of the five indicators. Ratings have been used to help panel members organize their judgments about each school and they have not been collected. Next year we plan to collect these ratings to run generalizability theory analysis to check on the consistency of ratings at this level of the process. For each of the five indicators for each school the panel members rated the school on a continuum from “clearly negative”, to “generally negative”, to “neutral or mixed”, to “generally positive”, to “clearly positive”.

After the initial independent ratings of all schools were completed, the raters categorized each school as focus, unclassified, or excelling using a form provided for this purpose. Once each panel member had independently completed this form, their categorizations of the schools were tallied by the assessment and research office staff. Beginning next year we will also collect these forms and use the ratings data for generalizability theory analysis of this level of decision making. For some schools there was unanimous agreement from the onset. In other instances, there was disagreement. A consensus process was used to resolve school classifications when there were disagreements. In these cases the panel members reviewed the reasons for their decision and openly discussed the indicator data that persuaded them to support one category or another. In every case over the first two years, a consensus was reached by the panel members.
about the every school’s category. The third year, there was one panel member who disagreed with an otherwise consensus decision about a school in one content area.

**District and School Response**

Prior to implementation of this accountability model there had been no formal district level review of school performance on measures. Implementation of this process represented a substantial change in practice and culture within the district. The accountability model was piloted one year and school leaders were informed that the system would go live the following year. All elementary schools in the system met AYP requirements during the first year that the district accountability system went live. For some of the eight elementary schools that were identified as focus schools that year the discrepancy between the two accountability conclusions fueled some anger at the district process. Each of these schools were visited by district leaders who took the data used by the judgment panel and walked the school’s staff through the process of judging data from three schools, one of which was their school. The data from the other schools was anonymous but it came from an excelling school and an unclassified school. The focus school teachers were given the forms used by panel members and asked to complete ratings from clearly negative to clearly positive on each of the five data indicators for each school. This process gave the staff at the focus some comparisons from other schools in the district that were performing better and helped them understand the data that led to the judgment panel classifying their school as a focus school. At most of the focus schools district staff then worked with the school’s staff to drill down to the class and student level on each of the data indicators. This process was very helpful in solidifying the credibility of the district accountability process within the district since it became obvious to the teachers and leaders at the focus schools that the data supported the conclusion of the judgment panel about their school.

Beyond our intention of improving district achievement by celebrating excelling schools and working to improve focus schools we did not have a systematic plan in place for accomplishing either of these objectives until half way through the second school year during which our accountability model was operational. Central office leaders were in agreement, however, that there was no point in classifying any school as underperforming unless the central office was committed providing the schools with the support they needed to increase their capacity to get better results.

The district now has a well developed plan in place for celebrating excelling and improving schools and for providing support to focus schools. The plan calls for an increased level of transparency around achievement data within all schools but especially within focus schools. In addition, external partners with a track record of helping schools improve were contracted to begin working with some focus schools. The primary emphasis of this work is system improvement with goal setting at every system level at the school (i.e., school, classroom, teacher, and student). From the onset of the operational implementation of the district accountability model, however, some focus schools have called upon central office staff to assist them in working to better understanding their data and developing strategies for improving results. This resulted in collaborative work between some central office staff and the staff at those schools.

One School’s Response

The author was invited by one focus school principal to participate in some of the improvement efforts that were underway at the school. In recent years this school (i.e., hereafter referred to as School A), where about 75% of students were on free/reduced lunch, had been experiencing high staff turn-over. Test results at School A were consistent in suggesting that underperformance had been present for many years prior to the district’s implementation of an accountability system. Work on improvement at School A that followed it’s identification as a focus school was wide-spread and the author was only involved in a part of that work. What follows relates only to those initiatives at the school about which the author has direct knowledge.

The opportunity to participate with staff at School A proved to be a valuable learning experience for the author. The school initially worked to develop a deep understanding of the test results at the school, grade, classroom, and student level. This work was very transparent and it involved each teacher or teacher team working with a leadership team that included the principal, an instructional coach, and tutors. Over time these staff members have developed a deeper understanding of the meaning of test results for their students. Teachers learned to set high but achievable goals for each student by using their recent and historical assessment information. These goals are adjusted as a result of student progress and new evidence. The goals were communicated to students and parents. This helped students develop ownership for their own learning.

Next, School A’s staff worked to develop a deeper understanding of the skills that were measured on the tests. Having a deep understanding of the tested curriculum empowered the teachers better prioritize so that planned units and lessons assured that the tested curriculum was delivered to students prior to testing. Another important change involved creating a testing environment that encouraged optimal effort by students on the tests. This began with the goal setting that (a) communicated very clearly to the students that teachers at the school valued the results from the tests and (b) encouraged students to work toward a personal goal. Having a personal goal for the test contributed to improved engagement during testing for many of the students. During the actual testing, more adult supervision than in the past was provided and students who appeared to be disengaging were identified. Physical proximity of adults and occasional redirection to read the items carefully occurred.

A year and a half after this work began, student test scores at School A showed dramatic improvement. School A improved from focus status to unclassified status following that test season (spring 2009). Test results suggested that the proportion of students at School A that were paying the price of being poorly educated had dramatically fallen compared to past years. Figure 1 shows box and whisker plots for four years of reading test scores from grade 3 at School A. The dramatic improvement in reading test scores in grade 3 during the spring of 2009 at School A is clearly visible in Figure 1. The top whisker was higher in 2009 than previous years and the bottom whisker was dramatically higher and closer to the top whisker than in past years. The mean had risen as well and the box was smaller. This pattern indicates that the distance
between the high and low performing students had narrowed. The distribution shows that improvement of low performing students (i.e., those lowest on the distribution of scores) was not obtained at the expense of high performing students. All students were benefitting from improved performance and the achievement gap had narrowed. Our objective when implementing the district accountability system was to realize precisely this type of change in the distributions of scores at the district level. Figure 2 provides a box and whisker plot for grade 3 reading in the entire district. The results at the district level were not as dramatic as those at School A. Nevertheless, early evidence of improvement can be seen at the district level. The mean is at a high point and, as importantly, the distribution has narrowed primarily because the bottom of the distribution has risen toward the mean. As such, in grade 3, there is evidence that equity in reading is improving within the district. The improvement for the overall district is being driven by the improvement of schools like School A.

My observations at School A suggest that the staff became stronger on the five conditions that Bryk et al. (2010) identified as necessary conditions for school improvement to occur. These five conditions include school leadership, the professional capacity of staff (i.e., a dramatically improved sense of efficacy), an instructional alignment system (i.e., alignment with measured skills), a student centered climate (i.e., one that features student goal setting and ownership of learning), and relational trust among staff. On the variable of parent community ties, I did not have relevant observations to report.

Summary

Economist have reported that narrowing the achievement gap would contribute to higher economic growth which would create the condition for a rising tide to lift all boats (i.e., for children to have the opportunity to have a better standard of living than their parents). For this reason, narrowing the achievement gap is in a society’s interest. School improvement is necessary to narrow the achievement gap. Evidence that school improvement is happening occurs when the distribution of test scores within a school, district, state, or country moves higher and the range of the test scores in the distribution narrows. These changes to test score distributions signal that outcomes are becoming more equitable and the achievement gap is narrowing.

Much is known about the conditions that are necessary in order for a school to improve. These conditions provide a roadmap to states, districts, and schools about the conditions that must be created in order for improvement to occur. Credible test-based accountability systems are enough to motivate some schools to develop the capacity needed to improve (e.g., see Bryk et al., 2010). State departments and district offices have potential, and I personally believe, responsibility, to make positive contributions to the capacity to improve at all underperforming schools. Credible test-based accountability systems can help state departments and district offices appropriately identify the schools that might need extra support.

Whether or not a test-based accountability system is in place, there is a cost (i.e., stakes) to students associated with low achievement. The cost is an increased likelihood of a range of negative outcomes. When school improvement happens, fewer students experience the negative
costs associated with low achievement. The implementation of a credible accountability system by itself, in the absence of additional resources or external support, has been shown to drive improvement at some schools. Provided that the implementation of a test-based accountability system is not associated with a corresponding decrease in performance at other schools, then implementation a test-based accountability system would be in the interest of students and society at large. Specifically, more students would avoid the costs associated with low achievement and society would ultimately experience the economic growth associated with a reduction in the achievement gap. Perhaps an optimal situation exists when district and/or state leaders have the capacity to support the efforts of properly identified underperforming schools as they work to build the capacity needed to improve.

There are some schools where improvement is unlikely to happen as a result of the efforts of the schools, districts and/or state departments of education alone. These schools serve students who live in communities where threats to their personal safety are routinely present due largely to an absence of functioning social institutions and organizations. It is up to the larger society to provide resources, policies and actions aimed at securing the necessary community-wide improved conditions. As improvement occurs within these communities students living in them will have a better chance of arriving at school “ready to learn” which helps to create a condition where school improvement is possible.

References


Figure 1. School A Grade 3 Reading Scores from the MAP.
Figure 2. NCSD Grade 3 Reading Scores from the MAP.