Confronting Accountability:
A Call for Achievable Alternatives to No Child Left Behind
CONTENTS

I. Introduction 3
II. Literature Review 7
III. Historical Background 25
IV. Research Design 38
V. Results 45
VI. Discussion 64
VII. Conclusion 72
VIII. Works Cited 74
I. Introduction

47 million young people attend public schools in the United States each day. Today, the quality of their schools is subject to the accountability measures in a single piece of legislation: No Child Left Behind. It is time to reevaluate No Child Left Behind both in theory and in practice because of the perceived inadequacies and the pending reauthorization of the law. In their proposals for the reauthorization, both conservative and progressive education reformers have been circumventing the specific questions of accountability. The goal of this project is, therefore, to confront the question of accountability directly. This thesis will engage in an explicit conversation asking why it is important, how it is being used, and what futures are possible.

Working Framework for Educational Accountability

All accountability systems are comprised of three main criteria. (1): to whom should schools be held accountable? (2): for what should schools be held accountable? and (3): how should schools be held accountable?

Within the to whom category a number of subquestions arise such as: should the federal, state, or local government be primarily responsible? What role should parents and the local community play? Similarly how should the business-community and policy makers be involved? Depending on who schools are accountable to, who should become the architect of the system and how should it be funded? When thinking about political decision making around educational issues, specifically assessment policies, there is a normative argument over whether educators,
because of their positioning in schools or policy makers, because of their responsibility as elected officials, should have more of a say (McDonnell 2005, 36).

The second component of an accountability system is “for what purpose should we be holding schools accountable?” What do we hope our schools will accomplish? The answer to this question is perhaps the most important question that has been asked since the idea of public education began to take root. The answer to this question wholly influences the other factors that go into creating an accountability system. Biesta (2009) writes, “when we are engaged in decision making about the direction of education we are always and necessarily engaged in value judgments - judgments about what is educationally desirable.” It is useful here to present briefly the overarching debate of the purpose of school. Put simply, there are two schools of thought when it comes to the role of public education. The first is driven by a desire to increase economic productivity and international competitiveness. The second is driven by a commitment to fulfill and strengthen our democratic society and individual communities. While the approaches to accountability presented throughout this thesis are not specifically broken down into these two categories, it is a critical and divisive overarching framework that guides the development of all educational reform efforts.

The third criteria used for assessing accountability systems, and perhaps the most widely discussed at present, is the specific “how to:” what measures will we use to hold schools accountable—to judge their successes and failures. This last section, on measures or methods, is influenced greatly by who is holding schools accountable and what they are holding schools accountable for. Answers to these specific questions span a wide range, from standardized testing to performance-based assessments, both of which will be defined in more detail throughout.
Overview

This thesis begins with a Literature Review where I examine a range of answers to my specific research question in the broadest sense: *how should we hold schools accountable?* The Literature Review is divided into three subsections: test-based accountability, standards-based accountability, and performance-based accountability. I then move into a Historical Background section where I present very early cases of accountability in education. I then chronicle the evolution of philosophical, social, economic, and political developments in twentieth century United States history that have directly influenced how we understand and implement accountability systems in public education. The appreciable part of this section refers to the development of and nuances in the modern “standards and accountability” movement as well as the development of and nuances within the No Child Left Behind Act. I also describe what Obama’s Blueprint for the reauthorization of the Elementary and Secondary Education Act proposes for the future of federal accountability policy.

The third section of this thesis is a Research Design, where I layout the methodology that I used in order to investigate an alternative system of accountability that endorses standards and performance-based accountability rather than our current system of test-based accountability. I Science Leadership Academy (SLA), a new small, progressive magnet school in Philadelphia, PA that is attempting to pilot a system of Standards-Based Reporting (SBR) to study. I collected data on SLA through interviews, observations, journal articles, education blogs and school report cards. After the Research Design, I present the results from my empirical research. This Results section displays information on how SLA is held accountable by the state, city, and itself, and documents the staff’s hopes for SBR. The last section of this thesis discusses the findings from
my research in light of how schools should be held accountable for student achievement and separately for school quality. Additionally, in order to inform my discussion on SLA I draw upon two case studies, The New York Performance Standards Consortium and the Kentucky Education Reform Act, both of which illustrate places of tension between test based accountability and performance based accountability. I conclude with suggestions for future research.
II. Literature Review

Current Approaches to Accountability in Education

In this section, I will attempt to present a coherent map of the various ways in which educators and educational researchers, the business community, as well as those working in public policy have answered the question, how should we hold schools accountable? Their answers can be broken down into three schools of thought: test-based accountability, standards-based accountability, and performance-based accountability. During my analysis of each, the three aspects of accountability systems, as discussed in the introduction, will be addressed: accountable to who, for what, and how.

Test-Based Accountability

Test-based accountability is defined as using large-scale, high-stakes tests to identify schools that need assistance. Such tests have uniform tasks, are developed by external parties and administered under standardized conditions. The results of these tests are used to determine rewards and sanctions for individual schools (Hamilton et. al 2002). Test-based accountability "is premised on the notion that standardized tests can and do measure an important dimension of..."

---

1 Given that NCLB is our most contemporary and omnipresent model of accountability, most authors writing about accountability today have located their discussion in terms of NCLB. The No Child Left Behind Act, however, as discussed in the Background section, is far more than an accountability system. In light of this I want to stress that my question was not, "what are the weaknesses of NCLB" or "how can we improve upon NCLB." In order to analyze a wide range of responses to my research question, which specifically focuses on how we should hold schools accountable, I will treat NCLB as one particular model of accountability, not the starting place for conversation. I think there is value in trying to position this conversation as one that exists outside of the current framework of No Child Left Behind.
educational quality" (Peterson and West 2003, 3). Such systems were established in reaction to mounting data that uncovered the inequality between and poor quality of American schools. Test-based accountability systems, are necessary when “school improvement no longer rests primarily upon individual volition or intrinsic motivation” (Hess 2003, 57). The increase transparency of school quality and encourage parents to express concerns about low or failing test scores (Peterson and West 2003).

No Child Left Behind was the first federal education policy that mandated test-based accountability. NCLB makes schools primarily accountable to the state and federal government for student’s math, reading, and science achievement scores, using high-stakes standardized tests. Specifically, accountability under NCLB relies on the testing of students in math and reading, annually in grades 3 through 8 and once during high school. Schools are expected to make Adequate Yearly Progress (AYP) toward the goal of universal student proficiency by the year 2014. Central to NCLB is the fact that test results are disaggregated to ensure that students from all racial backgrounds and those on free/reduced lunch are progressing. The results of the annual test scores and high schools graduation rates are published on school-wide Report Cards that aim to make a school’s quality known to parents and the community. From the second year to the sixth year that a school does not meet AYP an increasing ladder of sanctions are employed from providing supplemental educational services to complete school take-over or restructuring. A remaining distinctive feature of NCLB is the school choice option which allows for parents to transfer their students out of a school which has not met AYP to a better performing school in their same district (United States 2002).
Margaret Spellings, Secretary of Education for George W. Bush, and one of the architects of NCLB remains an advocate of this type of accountability. One of her key reasons is that the structure of test-based accountability gives a "sense of urgency" to the problems of public education (Margaret Spellings 2007). Given her position in the executive branch and role in crafting NCLB, Spellings is a clear advocate for the federal government prescribing the way in which schools be held accountable. She argues strongly for the 2014 deadline by which all students must be proficient in math and reading as well as science on standardized test scores. In response to the argument that schools should be held accountable to measures in addition to test score results, which I will present in detail below, Spellings said, "I have grave reservations about the capacity and capability of States and districts to administer... complex levels of accountability." She then added that including "multiple measures" in addition to test scores would be "confusing and burdensome" (Margaret Spellings 2007). Her vision of accountability, however, strays from the original formula prescribed by NCLB in that she recognizes the importance of incorporating growth models of student achievement as well as creating a system to differentiate the sanctions of failing schools in order to prioritize schools most in need of improvement.

John Castellani, President of the Business Roundtable, joins Spellings in her belief that the crux of holding schools accountable should be focused on academic achievement (in math, reading, and science). He supports the goal of 100% student proficiency by 2014 and the federal government taking responsibility to ensure that this goal is met. Likewise, he agrees that too complex of measures to hold schools accountable would render accountability systems meaningless. Given that he is a representative of the business community and has a direct interest
in postsecondary and workplace readiness, Castellani, however, thinks that graduation and college enrollment rates should be included in addition to academic test results (Reauthorization Castellani 2007).

The Hoover Institution’s Koret Task Force on K-12 education claimed, in 2005, that the goal of student proficiency in math and reading only, by 2014, was fundamental to the success of the accountability system that NCLB laid out and that standardized tests should continue to be the measure by which we judge success in those two areas. In order to encourage high standards and a more unified definition of “proficient,” the Task Force called for states to align their standards with those on the NAEP test. To encourage the success of this high goal they endorsed the strict sanctions that schools would face for not meeting AYP and the increased visibility of student achievement data to call attention to the success and failures of various schools. This would encourage families to transfer out of failing schools and into other public, charter, or private schools that accepted district money: another critical piece of “strong” accountability. Within trying to find a balance between federal goals and local control, which they acknowledged as being fundamental to the structure of governance in the United States, they recommended that accountability provisions be taken away from the school district and given to the state education agency. In addition the Task Force clearly stated that the federal government should not be responsible for giving more funding for what states should already have been providing. After describing the various ways that NCLB should be tightened to truly hold schools accountable Chubb, the author of the Task Force, wrote, “[our recommendations] must not be confused with all-too-frequent calls for ‘midcourse corrections’ by interest groups and policy analysts that would effectively gut NCLB in the name of ‘feasibility’ and ‘fairness’” (xii).
In 2009, however, Chubb, on behalf of the Task Force, wrote another book *Learning from No Child Left Behind*, where he strayed slightly from some of the principal claims they made in 2005. In 2009, Chubb cited student growth on proficiency tests as something that should be included in how schools are held accountable. He also proposed a “simple and differentiated system of incentives and sanctions.” To this he added that schools with the highest levels of achievement should be rewarded financially. Chubb also recommended that schools be held accountable for their teaching of history and science and in general supports national voluntary standards. By adopting national standards, Chubb proposes that states should have an extra six years to meet proficiency. Instead of having all students be proficient by 2014, schools will have until 2020 to meet this goal. After six years of not meeting AYP, schools should pick from three options: “(1) change the school’s governance by becoming a charter school; (2) delegate management of the school to an external manager, either not for profit or for profit; (3) or make a wholesale change of school staff, meaning the principal and 100% of the teachers – no exception” (Chubb 2009).

Test-based accountability systems defined by a strict adherence to the use of high-stakes standardized tests. Ravitch (2010) explains that the test-based accountability of NCLB prioritized “higher scores, without regard to whether students acquired any knowledge of history, science, literature, geography, the arts, and other subjects that were not important for accountability purposes” (Ravitch 2010, 30). While tests and ‘standards’ often overlap, one problem with NCLB, which only requires state testing in reading and math, is that Congress “left to the states the precise standards to be set, the specific design of their testing instruments, and the administration of their accountability systems...” (Peterson and West 2003, 8). As a result,
standards have been lowered in some states. The following subsection, standards-based accountability, identifies those who believe that the linchpin of an accountability systems must be high standards, not high-stakes tests.

Standards-Based Accountability

Kahlenberg (2008) and the contributing authors of Improving on No Child Left Behind hope that accountability can “recapture the early hopes for standards-based reform and promote both equity and excellence in K-12 schooling” (3,5). Within the book, Resnick et. al (2008) make the claim that we do not have a standards-based system but rather a test-based system and thus a false implementation of the standards-based reform effort. Hess (2003) outlines the distinction, “it is important to distinguish between high-stakes accountability systems that include sanctions for students, teachers, or both and those nonintrusive standards-based systems that do not,” he explains (57). He goes on to say,

Gentler, more suggestive standards-based approaches seek to improve schooling through informal social pressures, by using tests as a diagnostic device, by increasing coordination across schools and classrooms, and by using standardization to permit more efficient use of school resources. Suggestive accountability can produce educational benefits, but such changes tend to be modest and dependent on the ability and inclination of teachers to use the tests as pedagogical tools.” (Hess 2003, 58)

Ravitch is now the champion of this perspective, arguing that NCLB specifically allowed standards to be watered down and rendered effectively meaningless due to the false urgency with which tests-as-accountability policies were implemented (2010). Resnick et. al claim that “we need better standards and assessments, along with a redesigned accountability system that motivates positive effort rather than strategies for avoiding negative consequences” (Resnick et.
In order to achieve this, the standards and the tests must be linked together from the beginning by redesigning tests so that they specifically measure the learning standards and tests must be administered in more than the basic subjects of reading and math.

Resnick, et al offer an alternative to traditional standardized tests: assessing based on a *learning progression*. They explain, "there might be three to five major goals to meet in each subject each year. Students would not have to be tested on all of the yearly goals at one time. Instead, assessments on a particular goal could be administered after instruction on that goal had been completed" (Resnick et. al 133). An accountability system, they explain, should also have a graduated set of incentives such as public praise, financial rewards, and site visits. This last piece, on school evaluations, they suggest should only be pursued when schools are moving from "good-to-great," in order to keep the costs down and incentives up (Resnick et al 136). Resnick’s strict focus on carrots over sticks prohibited her from engaging with the discussion of schools that move from good-to-not good. It is unclear what she suggests should be done to schools that move in a negative direction.

The perspectives presented in this second approach to accountability span a relatively wide range. They share one common perspective which is that the primary purpose of accountability should be holding schools accountable for teaching students based on a set of common high-quality standards. While they all agree on having more coherent and strict standards, they waver on the role of high-stakes standardized testing as the method used to hold schools accountable.

E.D. Hirsch, founder of the Core Knowledge Foundation, writes that “curricular incoherencies” account for the failing of our K-12 education system (Hirsch 1996). Chester E.
Finn, also arguing on behalf of the Core Knowledge Foundation, proposes the creation of a voluntary standardized curriculum. His hope is that states will work together and form coalitions of curricular standards and tests that are appropriately aligned with these standards. He proposes this because of his belief that “having [standards] vary from state to state produces mediocrity, cacophony, waste, duplication, and confusion” (Meier and Finn 2009). Hirsch, forcefully claims that (standardized) tests are necessary to achieve excellence and fairness. Hirsch proposes widespread multiple choice testing with a performance component to assure the tests are sending out the “right educational message.”

Bob Wise, President of Alliance for Excellent Education, and Michael Cohen, President of Achieve Inc. are equally interested in holding schools accountable for their alignment with standards, but have a stronger interest in changing the means through which schools are held accountable. Wise testified in front of the Committee on Education and Labor and called for incentives for states to work together to establish and adopt common standards, and high quality assessments. Cohen’s commitment to standards comes from a desire for high school students to be taught exactly what they will need to know and do in post-secondary education and work. They are wary of including “multiple measures” such as college enrollment, dropout rates, and end-of-course testing. Wise and Cohen are both interested in the federal government creating a pilot program of how to learn from the efforts of states that are prepared to design such a system before giving all states the option as opposed to school of thought that advocates for test-based accountability (Reauthorization Wise 2007, Reauthorization Cohen 2007).

While tests and standards historically appear in pairs, test-based accountability allows standards to vary from state to state, be weakened or strengthened depending on political will,
and are often rendered meaningless. Standards-based accountability on the other hand does not prescribe how schools should be held accountable as long as they are held accountable for meeting high standards. Half of this school of thought is linked to the test-based accountability group but Ravitch is leading a new movement towards anti-testing standards-based accountability (Ravitch 2010).

Performance-Based Accountability

Deborah Meier is one the leading advocates for education that should be held accountable locally, in order to maximize student potential using a wide variety of authentic measurements. The work that she has done is work of an educator and activist not researcher or politician. While her arguments are grounded in education and political theory, many of her conclusions are based in her own practice of founding and directing the Central Park East Secondary School (CPESS) in New York City for over ten years. Meier’s work at CPESS has been showcased in over 20 articles and books. She is a proponent of maximizing a school’s autonomy and giving the power to make educational decisions and hold schools accountable to the local communities who know their schools and students well. Her ideals for education, and therefore accountability, cannot be separated from her vision that schools have a primary responsibility to uphold, embody, and educate towards democratic principles. She writes: “By shifting the locus of authority to outside bodies, it undermines the capacity of schools to instruct by example in the qualities of mind that schools in a democracy should be fostering in kids— responsibility for one’s own ideas, tolerance for the ideas of others, and a capacity to negotiate differences” (Meier 2000, 5). Because of this commitment, Meier believes that schools should develop their own standards with respect to what students will be asked to know and do upon leaving the school and should then measure
student success based on those standards. Student’s achievement should be measured using a
variety of methods, prioritizing, however, performance-based assessments.

Performance based assessments (PBAs) are assessments based on performance of a task
rather than on the memorization of knowledge. PBAs can include activities such as or similar to,
devising and conducting science experiments, the writing and revision process in the humanities,
or creating multidisciplinary projects. Performance based assessments involve multiple measures
of achievement which test for higher order skills, are aimed at inspiring student engagement, and
encompass the idea of transfer, that in-school learning will be applicable to out of school
situations (Archbald 1991). This is a fundamental justification for performance based
assessment’s combined focus on process and product. If the process of learning is valued in ways
such as questioning and research techniques, ability to develop and test hypotheses, make and
support arguments, join creative and academic work together, then educational spaces will create
opportunities for students to practice these skills and so, will be more able to use the same
strategies in new situations (Bruner 1966).

Because of Meier’s belief in local control over national standards and national,
accountability systems, she also believes that if schools are failing to provide well for students
the local community should be in charge of devising and enforcing sanctions. To close her essay
“Educating a Democracy” she writes, “All we need is a little more patient confidence in the good
sense of ‘the people’ - in short, a little more commitment to democracy” (Meier 2005, 31).

Ken Jones, the author of Democratic School Accountability, explains the construction of
accountability in his “learner-centered” model: "It must be centered on the needs of learners and
on the intentions of having high standards for all rather than the prerequisites of a bureaucratic
measurement system... Schools are complex and unique institutions that address multiple societal needs. A standardized approach towards school accountability cannot fit at all” (Jones 2007).

“Students, parents, and the local community are the primary clients for a school” says Jones. They should, therefore, be the ones primarily responsible for holding schools accountable for the physical and emotional well-being of students, which includes providing safe environments as well as the opportunity for positive relationships. In addition, Jones claims, schools should be responsible for student and teacher learning, equity and access, and improvement (Jones 2007). By laying out these goals, Jones' accountability system places value on a complex notion of student learning and well-being from specific discipline knowledge to affective needs. Student learning goals should be aligned with local curriculum and be assessed with performance-based assessments. The last piece that should be considered in holding schools accountable according to Jones is improvement. He claims that schools must be “continuously engaged in self-assessment and adjustment with respect to meeting the needs to their students” (Jones 2007).

Using the Balanced Scorecard, an accountability measure used in business, as an example, Jones suggests that the primary role of an accountability system should be giving schools information for improvement rather than rewards or punishment. The Balanced Scorecard model includes gathering information on 1) Leadership, 2) Strategic planning, 3) Student, stakeholder, and market focus, 4) Measurement, analysis, and knowledge management, 5) Faculty and staff focus, 6) Process management, and 7) Organizational performance results (Karathanos 2005). Jones also calls for accountability that includes the opportunity to learn
(OTL). Jones (2007) and Sloan (2008) define OTL as the various resources a school has to achieve its learning goals. They provide examples such as:

The number or percentage of fully and properly certified teachers at the school, the number or percentage of classrooms in which teachers are not fully and properly certified, the degree to which teachers of English language learners hold the appropriate credentials, the average number of years teachers have been teaching and have taught at the school and teacher turnover rates, teacher-to-student ratios, the availability of adequate and appropriate instructional materials such as books textbooks, and technology, the physical condition and maintenance of school facilities, and the extent to which the student population exceeds the capacity of a school, the availability and adequacy of high-quality college preparatory vocational education, honors courses for students in high school, the availability of instructional support from school librarians, nurses, and other school staff. (Sloan 2008, 122)

I included the entire list of examples to elucidate the depth to which opportunities to learn can influence every school's ability to meet its goals. Sloan (2008) also includes school violence rates in his definition (129).

Although I have titled this section performance-based accountability, it is important to note that in addition to focusing on the complex needs and possibilities of each student, they are also extremely focused on schools as learning communities. Another way to say this is that by recognizing the extent to which opportunities to learn play a role in achieving their goals, they believe that accountability systems must include room to assess school inputs as well as outputs.

Darling-Hammond proposes the use of assessments as indicators of progress and not to control rewards and punishments for schools. She claims that teachers should be involved in creating and scoring the assessments used for accountability systems and introduces the possibility of each state having a weighted set of indicators by which to hold schools accountable. The index would measure current status and growth at the district, school, and subgroup levels; a proficiency benchmark; and a means to set annual growth targets toward the
proficiency benchmark for each school. The proficiency index of a school's success would include: scores on student assessments, attendance, grade-to-grade progression, suspension rates, graduation rates, and additional quantitative measures of school learning conditions and effectiveness such as Advanced Placement courses (Darling-Hammond 2007). As opposed to the school report cards that are currently in existence, Darling-Hammond's proposal for a weighted set of indicators will force on conversation on which are most important. Currently while some of this data is made public, all except for test results are relatively ignored. Darling-Hammond also agrees that schools should be held accountable for the opportunities to learn.

Wood (1992) presents the idea of an Exhibition, one possible model for assessing students in his way and an alternative to exit exams to determine graduation eligibility. Exhibitions work backwards from desired skills to develop graduation standards that require students to prove their mastery of each standard. Exhibition can be made up of the presentation of a senior project, a graduation portfolio which is a collection of work from the high school career as well as course credits.

With the publication of his book, *Grading Education*, education columnist for *The New York Times*, Richard Rothstein joined Meier's ranks working towards progressive education reform. To begin, the accountability system proposed by Rothstein (2009) falls to the states to employ as he does not believe the federal government should manage individual schools. Rothstein (2009) argues for an accountability system that equally assesses basic academic knowledge, critical thinking and problem solving, appreciation of arts and literature, preparation for skilled employment, social skills and work ethic, citizenship and community responsibility, and physical and emotional health. In order to assess these eight tenets of a good education, he
suggests an accountability system made up of three fundamental pieces: performance-based assessment, school accreditation, and a return to the use of NAEP exams. Because performance-based assessments, used to measure student achievement, have been described elsewhere in this section, I will expand here on the two other components of his accountability proposal, accreditation and NAEP testing.

Accreditation systems, similar to quality school reviews, involve evaluators spending time in a school for a number of days and collecting qualitative data. Though public school accreditations have existed in the United States for over one hundred years, they have historically not been used to order rewards or sanctions to schools. The data accreditation teams find, however, is both reported publicly and used to make school improvement decisions. Accreditations can be used as a “visible credential validating school quality,” as a signal to parents and community members and colleges and universities that the school has met certain standards (Fairman et, al. 2009).

For Rothstein (2009), accreditation teams should look at basic academic knowledge, critical thinking and problem solving, appreciation of arts and literature, preparation for skilled employment, social skills and work ethic, citizenship and community responsibility, and physical and emotional health, the eight areas he designated, and should make their findings and clear recommendations public. Rothstein (2009) also argues for there to be consequences imposed directly by the state for schools that fail repeated accreditation reports but does not specify what those consequences should be (Rothstein 2009, 127). Rothstein (2009) based his recommendations for an accreditation system on examples from England, New Zealand and the Netherlands. In England, “Her Majesty’s Inspectors” (HMIs), has been in place since 1839.
Accreditations there are run through the Office of Standards in Education which subcontracts companies to hire and train retired principals and teachers as professional inspectors. Because this model has been in place for so long and because it is run through the national government, inspectors, and the process at large are well-respected.

The New England Association of Schools and Colleges, which is the oldest accreditation agency in the United States works on a ten year cycle with each school. The school does a self study for two years, is then visited for four days by a team of educators from other secondary schools who follow up with progress reports and monitoring. NAESC reviews schools based on seven standard areas: mission and expectations for student learning, curriculum, instruction, assessment, leadership and organization, school resources and community resources. During the visitation, teams meet with teachers, school-board members, administrators, students, and sometimes parents; observe classroom instruction; shadow and interview students; tour and examine physical space and resources, accessibility, maintenance, and safety (Fairman 2009).

In addition to being useful for validating school quality and holding schools accountable, accreditations serve to assist the school with improvements. As one principal of a school that was recently accredited in Maine said, “The thing we were looking for is what are we doing right, and what are we not doing. What’s working for us, what’s not, what holes are there. What are we doing that’s redundant. What are we doing that’s not effective...It’s a way for us to recenter ourselves. So for us it was an opportunity for growth and introspection” (principal, 9/09/09 qtd in Fairman 2009).

In 2006, a study was done evaluating perceptions of accreditation in the State of Maine. It investigated 102 public secondary schools that had been accredited and interviewed 30
superintendents, school board members, principals, teachers, and special education administrators from eleven high schools. Those interviewed responded positively to questions about improvements in academic environment, staff communication, improvement in organization and management, professional development, institutional leadership, and work environment for the school staff (Fairman et al. 2009).

Within the United States, Rothstein (2009) recommends:

To fulfill an accountability role, associations should become quasi-governmental agencies with tax support and budgets large enough to conduct school visits more frequently and to employ trained professional evaluators... For accountability reviews, volunteer observers could include parents and other members of the public, such as local business, civic, or political leaders. Exposing the process to public scrutiny in this way would give it much needed credibility. (Rothstein 2009, 127)

There are a number of constraints on accreditations, however. Few empirical studies have been done to document their successes and regardless, we already know it is a long and expensive process. Some of the schools studied in Maine reported that they had been hurt by budget constraints and had a hard time preparing the self-evaluation or enacting the recommendations made in their accreditation reports. Despite the constraints, the study on SEASC concluded that “School board members and superintendents emphasized the importance of the outcome of accreditation, in maintaining accreditation status to ensure public confidence in their local schools.” The potential for accreditations to improve schools and link accountability to school improvement has also been shown.

In addition to accreditation, Grading Education argues for a reinstitution of NEAP testing. NAEP, the National Assessment of Education Progress, sometimes referred to as the Nation’s Report Card, is the only nationally representative assessment of what students know and
are able to do. Pellegrino, in an article considering the role that NAEP should play in developing state standards today wrote: “NAEP is seen as a high-quality indicator of academic achievement, its performance standards are perceived to have greater rigor and validity than those set for many other assessments, including the achievement tests developed by individual states” (2007). When NAEP testing began in the 1960s it tested students in grades 4, 8, and 12 in math, reading, and science (Hoff 2001). Early NEAP testing was unique in that not every student took the NAEP test every year; NAEP aimed to gather a sample of the student body and so no one student took the entire test, and no testing period was longer than 45 minutes (Rothstein 2009, 99). This meant that individual students never received scores on NAEP tests and because the identical assessment was given across the country achievement of students in all states could be compared using a common metric. Likewise, the results of the NAEP tests have never had consequences attached to them, as they were purely informative (Hoff 2001).

This subsection on performance-based accountability discussed performance-based assessment, the balanced scorecard and weighted indexes, the possibilities for Exhibitions and graduation portfolios, accreditations, and the use of NAEP testing as a common metric. Those in this school of thought share the common belief that schools (and students within schools) should be held accountable for their results by demonstrating authentic performances. The Coalition of Essential Schools, a national network of schools and organizations that promote performance-based assessments have shown that their schools in New York have 17.7% higher college-going rate and a 9.7% lower dropout rate compared to the state average; in Boston, CES schools have a 26% higher college-going rate and 11% less of a drop-out rate than the city average (Measuring Up 2006).
The Connections

Regardless of their accountability allegiances, all the educators, scholars, and policy makers, described above and as far as I can tell everyone writing about accountability who was not surveyed in this paper agree that when holding schools accountable for student achievement we can not look at absolute proficiency goals alone. Many write about the idea of value-added accountability policies which measure the progress made by students year-to-year. In addition the vast majority of scholars agree with accountability policies disaggregating student achievement data by race and class except for Kane and Raymond (Peterson and West 2003).

One last method of holding schools accountable is the Hope Study, a survey created by a group of schools in Minnesota. “The Hope Survey measures student perceptions of autonomy, belongingness, goal orientation & rigor in the school environment, as well as students’ self-reported engagement in learning & their dispositional attitude toward achievement (i.e., their “Hope”) (Newell 2009). Admittedly this will not become national policy anytime soon, however, it is a case of accountability that values the emotional well-being of students. The data for the Hope Study is collected using a series of reliable, valid measurement instruments “borrowed directly from educational research, which are administered via the Internet with full security & confidentiality.” Schools who have participated in the Hope Study and use the results to think about improving “hope” within schools have found improved student behavior, attendance, and academic achievement.
III. Background

Educational accountability in the United States has had a long and serpentine history. In 1643 parents were fined when they did not teach their children to read and understand religious and secular law (Rothstein 2009). The first school superintendent of Portland, Oregon in 1874 preemptively intervened in schools by implementing a standardized curriculum across all elementary schools and developing exams to assess their compliance. When the exams were scored, he then published student data in the local newspaper (Jackson 2009). These two examples of early accountability systems differ in terms of who is held accountable: in the case of the 17th century system it was the parents and in the Portland case, the students. Horace Mann, a member of the Massachusetts House of Representatives and foundational educational reformer, implemented an accountability system in the 1830s that held schools accountable. Mann required schools to report data on their school’s operation. He included this information in “Annual Reports” where “[h]e publicized data showing that most school committees had not visited their schools in the past year, many had hired teachers without examining their knowledge of the subjects to be taught, and almost half of all students were absent on a given day” (Jackson 2009). These three cases of early accountability systems document the existence of accountability in education for over three centuries. In order to think clearly about how schools should be held accountable in our future, however, this section goes on to addresses why and how we have held schools accountable recently.

Modern accountability has become interconnected with notions of standardized testing (Darling-Hammond 2006). This formation and acceptance of this particular methodology has
been guided by two distinct social, political, and economic themes of the twentieth century: racial equality and globalization. In *Brown vs. Board of Education* (1954) the Supreme Court of the United States voted 9-0 that the having of separate facilities, such as schools, based on race was inherently unequal. This decision overturned that of *Plessy vs. Ferguson*, the 1896 case which legalized segregated schools. After *Brown vs. Board of Education*, the quality of schools and the inequality between schools came into the federal government's view and in 1965 President Lyndon B. Johnson signed the Elementary and Secondary Education Act (ESEA) as one way of confronting these issues. ESEA for the first time dedicated federal money to states in order to fund public education, now known as Title I funding. Under Title I the government gave $1.06 billion to states for local school district to assist "educationally deprived children." Also included in ESEA was Title II funding, which offered $100 million for textbooks, expanded libraries, and other instructional resources, Title III funding awarded $100 million for "supplemental services and centers" and lastly Title IV set aside $100 million to modernize and coordinate educational research (McAndrews 2006). Once President Johnson enlisted the federal government to contribute to public education, the message from the executive branch all the way to individual taxpayers has been that we must make schools accountable for their performance. The message went: if I am handing out money, I want to make sure that something productive is being done with that money (Sciara 1972). It took 47 years, however, for formal accountability measures to be added into federal education policy.

Nearly a decade after the *Brown vs. Board of Education* decision and the passage of ESEA, one of the largest sociological studies was done about race and schools. The Equality of Educational Opportunity, known more commonly as the Coleman Report, of 1966, concluded
that a student’s background and the background of the student’s classmates had greater influences of educational outcomes (student success as measured by test-results) than the quality of the school a student attended. This data was used as justification for desegregation, assuming that black children would achieve at higher levels if they went to school in the company of white students (Schneider and Keesler 2007). The report also showed that the differences in school funding between predominantly black and predominantly white schools were not as drastic as people had assumed, thereby shifting the public’s focus from school resources to educational outcomes (Sloan 2006, 4). This marks a major shift in education policy and encouraged the public, educational researchers, and politicians to collect data and make decisions about schools based school outputs; today often referred to as results.

The focus on results was magnified during the Cold War as the United States fought to maintain its position as the world’s dominant superpower. During this period, it was found that education in the United States was not only unequal at home but of poor quality compared to other countries around the world. International surveys and tests ranked American students behind their peers abroad. “The situation deteriorated the longer students remained in school. Among nine-year olds, U.S. Students performed in math and science among the top tier of nations...By age thirteen, U.S. Students had fallen below the international average in these subjects” (Peterson and West 2003, 5). The publication of data such as this promoted an attitude about education that situated the United States in comparison with other countries, explicitly drawing connections between our weaknesses in education and the health of our economy. It produced a sense of fear and urgency and instigated the business community’s involvement in education. A report of the National Alliance of Business affirmed that the “public education system will not respond to such calls for reform in the absence of the pressure to do so” (quoted
in Hamilton et. al 2002, 7). The business community’s involvement and inculcation of business principles (such as the notion of incentives) have been considerably responsible for the national shift towards educational accountability.

Even President Reagan who once called for the abolition of the Department of Education made education a priority after our global ranking was published and public outrage ensued (Peterson and West 2003, 6). During the Reagan presidency the Secretary of Education, Terrel H. Bell created the National Commission of Excellence in Education. It was this commission that published *A Nation at Risk* in 1983, which famously began with the line: “The educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people” (United States 1983). *A Nation at Risk* called specifically for six reforms in the education system:

1. More days and hours of schooling
2. More academic courses
3. More attention to “the basics”
4. More clearly written, measurable standards for evaluating and compensating teachers
5. More standardized testing to determine student achievement levels
6. More elaborate systems of reporting test results at the local state levels (United States 1983).

*A Nation at Risk* was not, however, education policy; it was never turned into law. The effect of *A Nation at Risk* most directly was on the quickly formed standards-based reform movement. “[The standards] movement’s core idea is that by expressing what students should know and be able to do, reformers could promote high-quality curricular frameworks, assessments tied to those standards, and professional development for teachers whose performance would be directed toward meeting a defined set of objectives (Schneider and Keesler 2007). As a result of A Nation at Risk a growing number of states began to create or to improve upon their learning standards. In the early 1990s, Ravitch worked on creating voluntary
national standards through the U.S. Department of Education (Ravitch 2010). Margaret Goertz, Professor of Education at the University of Pennsylvania explains the motivation for national standards: “The arguments in support of national standards today echo those of the past: they will promote democracy, equity and economic competitiveness. The arguments against national standards are also familiar: they will lead to the establishment of a national curriculum; one size does not fit all; and local communities, not the federal government know what is best for their students” (Goertz 2007).

Regardless of the question of standards, the amount of federal funding spent on education, inequalities between schools and our global academic ranking are three reasons why the United States continues to monitor school progress. How this is done and how it has been done in the past raise more interesting points.

It was the early 1900s, during the height of the industrial revolution, when centralized accountability truly began to take center stage in the education discourse, both theoretically because of a new mindset about production, and practically, because of the new methods of production. The industrial revolution was centered on the ideals of “scientific management.” Superintendents adopted this ideology and given the era’s booming new technologies, specifically the reduced cost of paper and ease of printing, jumped on the opportunity to put a standardized way of monitoring progress through paper and pencil exams into practice (Jackson 2009). President Johnson commissioned the National Assessment of Educational Progress (NAEP), the first periodically administered testing program to collect national samples of American students. NAEP has continued to be funded through the federal government and results from the tests are published annually as “The Nation’s Report Card.”
George H.W. Bush’s plan for education, *America 2000*, is eerily similar to the modern day education policy, No Child Left Behind. *America 2000* was an education policy built firmly on the basis of standards-based form. Using suggestive powers it encouraged states to produce high-quality standards and as Jackson (2009) explains, “intended to foster achievement... by preparing voluntary national tests in five core subjects at grades 4, 8, and 12 and preparing report cards at the school, district, and state level. The bill also would have amended ESEA to allow low-income children to use Title I funds as vouchers to attend private schools.” It was largely because of controversy over the voucher provision that the *America 2000* bill never passed (Jackson 2009). George H. W. Bush was also a catalyst for his successor’s education policy. Bush hosted the 1989 Education Summit where America’s governors worked to create a set of new goals for education. They decided upon six common goals related to inputs (students starting school ready to learn and safe learning environments) as well as a wide range of outputs (academic achievement in five core subjects, graduation rates, economic productivity and civic responsibility). President Clinton turned the six goals into *Goals 2000*, a federal grant project that went to states, districts, and schools to assist with reform initiatives (Jennings 1995). Under Clinton schools were asked to produce improvement plans that laid out how their state-wide assessment system was connected to the standards and then produce evidence every three years of making progress on that improvement plan. In addition to Goals 2000, Congress reauthorized ESEA as the Improving America’s Schools Act. IASA paid particular attention to educating students with disabilities as well as officially introduced the idea using test scores as a way to hold schools accountable into federal policy (Manna 2007, Schneider and Keesler 2007, 209).

Under Bush and Clinton’s leadership and *America 2000*, *Goals 2000*, and IASA initiatives, states across the country began to develop formalized accountability systems.
Governors responded quickly to the new attitudes about education. They instituted policies that allowed them to both spend more money on public education and mandate more-conservative requirements (Peterson and West 2003, 6). “While all states developed assessments, standards, performance reporting, and in most cases consequences for performances, states found different ways to define what it meant for schools to succeed, what indicators to include in their definition of success, and what the consequences would be” (Goertz 2007) Ross Perot, head of a state education commission in Texas was the leader in calling for testing procedures to monitor school progress annually (Peterson and West 2003, 7). The testing systems in Texas was received with national praise as test scores rose and “the achievement gap” closed. It did not take long, though, for these results to be dubbed the “Texas Miracle” and for studies to document that along with Texas’ increase in state test scores came a drastic increase in the dropout rate and no comparable success on other achievement tests such as NAEP or the SAT (Ravitch 2010, 96). The implications of dishonesty and coercion in Texas were not enough to deter Congress from being wooed by the appeal of an easy-to-administer, standardized, high-stakes test-as-accountability system.

George W. Bush’s reauthorization of ESEA in 2002, the No Child Left Behind Act, turned the suggestive accountability of IASA on its head. The No Child Left Behind Act, which was meant to promote sweeping education reform, is built on four core elements: “stronger accountability for results; greater flexibility for states; more choices for parents and children from disadvantaged backgrounds; and an emphasis on teaching methods that have been demonstrated to work” (United States 2002). The “landmark” education legislation passed during Bush’s first term with bi-partisan support in Congress. The law dictates that each individual state
annually test students in grades 3 through 8 in math and reading. Schools are held accountable for meeting “adequate yearly progress (AYP)” each year towards the goal of meeting 100% student proficiency by 2014. As discussed in detail in the Literature Review, under NCLB, sanctions are imposed when school’s fail to meet AYP. NCLB for the first time strictly enforced via sanctions what America 2000 and IASA had suggested: the adoption of state standards and testing systems.

While we are still situated within an era of test-driven accountability and scientific-evidence for data-driven decision making, which NCLB championed, the litany of complaints against the specifics of NCLB are paramount. In the last eight years, Schneider and Keesler (2007) have documented how the high-stakes annual tests have caused a widespread narrowing of curriculums in order for teachers to focus on test preparation. Ohanian (2002) has recorded the nationwide protests among parents when recess, art, and other non-basic education time is being cut in schools to also allow for test preparation. Rothstein (2008) notes that testing students with one test, one day per year is unreliable and due to the high-stakes nature of NCLB tests, schools and states are highly encouraged, and have been found, to game the system either by blatantly changing test results, creating such low standards that AYP will be easier to meet, or pushing low performing students out of their school (Darling-Hammond 2006). In response to the increased testing and narrowing of the curriculum, six mothers led 200, or 60% of eighth graders, in Scarsdale, NY on strike boycotting 13.5 hours of standardized testing they had in one school year. Melanie Spavin, one of the mothers was quoted as saying: “We’re law-abiding and we love process, believe me. But we’re a bunch of fed-up mothers!” Deborah Rapaport another of the leaders exclaimed, “What could be less productive academically than eighth graders spending
13.5 hours taking tests - especially when it’s not even clear just what these tests are designed to do?” (Ohanian 221). Even the public school choice pillar of NCLB has had deplorable results “In the 2006-2007 school year, only 45,000 students, less than one percent of the total eligible, used the choice option” (Derthick and Dunn 2009, 1022). Duncombe et. al, (2008) point out that states are unfairly punished by external sanctions being imposed on their schools, especially given states unequal resources and the fact that Title 1 funds do not account for all of the extra expenses schools and states incur when implementing the NCLB mandates. Nearly every educator/author regardless of their position on NCLB has expressed that a goal of universal student proficiency by the year 2014 is untenable (Rothstein 2008).

President Obama’s election in 2008 brought a sense of relief to many people in the country who did not agree with the policies of his predecessor. This included those who had been disappointed by No Child Left Behind. This said, it has also been pointed out that: “[a]lthough the mechanisms of NCLB may be perceived as too harsh and as placing too big a burden on the schools serving poor communities, it is doubtful that policy makers will move away from accountability and rigor” (Schneider and Keesler 2007, 213). At this point in time, with the acceptance and implementation of testing-as-accountability there are too many economic and political interests in perpetuating high stakes testing as the only form of accountability. Especially because it is not just assessments that have been influenced by the accountability mindset. “NCLB introduced a new definition of school reform that was applauded by Democrats and Republicans alike. In this new era, school reform was characterized as accountability, high-stakes testing, data-driven decision making, choice, charter schools, privatization, deregulation, merit pay, and competition among schools. Whatever could not be measured did not
count" (Ravitch 2010, 21). Organizations like Teach for America, a peace-corp like program which recruits high-performing college graduates to teach in low-performing public schools, and its spawns such as Superintendent of Washington D.C. Schools, Michelle Rhee embodied the state of education under this mentality. The landscape of education policy has changed remarkably in the last ten years as a result of NCLB and the commitment to data-driven decisions and high stakes standardized tests.

While President Obama campaigned on the platform of “change,” he now governs in an accountability driven political climate. The American Recovery and Reinvestment Act, more commonly known as the stimulus package, promised $787 billion to a variety of initiatives aimed at offsetting the effects of the economic recession and included $90.5 specifically for education. This was the largest and fastest increase in education spending to date (McNeil 2010). Within this, the administration set aside $4 billion for a competition it designed, now known as Race to the Top, to encourage reform in four main areas: implementing common academic standards and tests, improving data systems, turning around the lowest-performing schools, and improving the effectiveness and distribution of high-quality teachers, especially in high-poverty, high-minority schools (McNeil 2010). McNeil (2010) also reported that states can boost their chances of winning the grant if they embrace “aggressive action for low-performing schools such as firing most of the staff.” It is being reported that Race to the Top will be the model for the Obama administration’s proposal for the reauthorization of ESEA (Klein 2010, Sawchuk 2010). Secretary Duncan is particularly pleased with the Race to the Top plan because as he says, while it prescribes the goals states should be working towards it leaves the “detailed roadmap of reform up to the states” (McNeil).
The Blueprint for the reauthorization of ESEA was released in March 2010. It is a rather vague, 49-page document which covers a number of topics that will be addressed in a revision of the bill. From the outset, the blueprint separates itself from NCLB calling it a “failed” piece of legislation. Accountability in the blueprint has the following pillars: college and career ready standards, rewarding success, school turn around grants and a reliance on data (United States 2010). One of the largest themes throughout is a movement towards college and career ready standards and emphasizes that assessments should be aligned with those standards. Though the 2014 deadline of student proficiency was dropped, as were mentions of AYP, the blueprint lays out its own goal: ALL students graduating or on track to graduate from high school ready for college and career by 2020. Under NLCB making AYP was a way to check in on schools before the end date came upon them and sanctions were imposed when AYP was not met.

Obama’s proposal reads, “Instead of labeling failures, we will reward success. Instead of a single snapshot, we will recognize progress and growth.” This starting framework has places of distinction and overlap with the NCLB system. The Department of Education wants to separate schools into two groups: Reward Schools and Challenge Schools. Gaining flexibility in budget spending and becoming a model for Challenge Schools were the incentives for Reward Schools. Challenge Schools come in three tiers. One tier Challenge Schools will have to expand learning time, offer supplemental educational services, or public school choice. The next will implement locally developed strategies to improve. And the third must chose from one of four options:

1. **Transformation Model** - replace the principle, strengthen staff, implement research-based instructional program, provide extended learning time, implement new governance and flexibility.
2. **Turnaround Model** - replace principal, rehire no more than 50% of staff, implement research-based instructional program, provide extended learning time, implement new governance.
3. **Restart Model** - convert or close and reopen the school under the management of an effective charter operator, charter management organization, or education management organization.

4. **School Closure Model** - close the school and enroll students who attend it in other higher-performing schools in the district.

The similarities between becoming a Challenge School in Obama’s plan and facing sanctions under NCLB are paramount. One possibility for meaningful difference is that Challenge Schools have a chance to gain School Turnaround Grants from the federal government in order to specifically assist with schooling choosing one of the options above.

The focus for measuring student success, in addition to college and career ready standards will still be on math and reading test scores. Schools will also be required to collect graduation rates, college enrollment rates, and rates of college enrollment without the need for remediation—all desegregated by race, gender, ethnicity, disability status, English Learner status and family income. The blueprint also aims to legislate that data on school quality is collected, “States and districts will collect information on teaching and learning conditions - school climate such as student, teacher and school leader attendance, disciplinary incidents; or student, parent, or staff surveys about their school experience.” The last way the blueprint differentiates itself from NCLB is that it will reserve a portion of School Turn Around grants for additional activities such as investing in model school quality review teams to observe and assist in low performing schools. Within the blueprint for the reauthorization of NCLB, schools will be held accountable for student’s performance on reading and math standardized tests and college and career ready standards though it is not clear how the latter will be assed.

This historical background section aimed to show the evolution of accountability. While there are examples of local governments holding schools accountable, early accountability
systems reflected a desire to hold parties responsible when set educational goals were not met. As social and political changes occurred in the U.S. and more attention was paid to public schools because of issues of equality and economic competitiveness the federal government's role in public education grew. With the establishment of No Child Left Behind, however, the language of accountability and practice of standardized testing won permanent seats in conversations about education reform. Obama has the possibility to introduce a proposal for the reauthorization of ESEA that separates itself from the negative outcomes of NCLB but will be working within and against the powerful forces of political will and rhetoric that may limit the future of accountability.
Hess and Finn (2007) criticize No Child Left Behind by saying “NCLB’s remedies are more an assemblage of reform ideas than a coherent scheme. They were adopted with scant attention to how they would fit together, what resources or authority they would require, or whether they could be deployed sensibly through the available machinery” (9). NCLB, to put it kindly, has not lived up to its expectations. It has been shown that NLCB is too strict on schools, using all sticks and no carrots; has been underfunded; and has narrowed the curriculum (Derthick and Dunn 2009). Perhaps more importantly, Darling-Hammond explains that the problem with No Child Left Behind and a reliance on test-based accountability is that, “tests are asked to take on burdens of decision making and of instructional improvement, which they are not designed to carry and are not capable of accomplishing” (Darling Hammond 2006). Test-based accountability has proven at best ineffective and at worst deleterious.

Despite the criticisms, however, there is no doubt that accountability provisions will be placed in future education policy. Test-based accountability was in fact fashioned to call attention to the failures of public schools and with a sense of urgency support improvements. We must therefore, as supporters of this broad mission, follow Jackson’s (2009) claim that, “The order of the day will be engineering better accountability systems rather than abandoning them.”

I began this work by accepting the premise that has been laid out by progressive and traditional educators alike: NCLB has not had its desired effects of closing the achievement gap. Given that NCLB is federal education policy, if and where alternative systems of school
accountability do exist, they do so under the radar. Within that, the guiding question of this project is not what are other criticisms, or how can those criticism be demonstrated in a new way, rather, given the outcomes of NCLB, how should we hold schools accountable?

In the introduction of this paper I introduced three branches of accountability systems: (1) to whom are schools accountable; (2) for what are schools accountable; and (3) how are schools accountable. Scholars, educators, and policy makers have all weighed in on these questions by presenting historical, theoretical, or foreign examples of accountability systems that they claimed should be applied to the public school system in the United States. I previously grouped their contributions into three categories: test-based accountability, standards-based accountability, and performance-based accountability.

The distinctions between these three schools of thought are most clear when we think about their highest priority. For test-based accountability, the highest-priority is results on standardized achievement tests and its advocates Spellings (Margaret Spellings 2007) and Castellani (Reauthorization Castellan I 2007) contend that schools must be held accountable to the federal government. As we have seen under NCLB school structures including schedule and curriculum were modified in order to guarantee that all students took and were proficient on the tests. For standards-based accountability, Hirsch (1996) and Finn (2008) claim that states should be responsible for meeting state or consortia of state standards and their ultimate goal is a schools adoption of common standards. The assessment of those standards is more flexible than the teaching to the standards themselves. Performance-based accountability, championed by Meier, Jones and and Rothstein who argue that schools should be held accountable to their local communities and that social, emotional, and physical goals should be placed along side academic
achievement goals. This school of thought is flexible in terms of curriculum but stresses that assessments from the students up to the schools be conducted based on performance of real-life documentable work via performance-based assessments and some form of school accreditation.

**Philadelphia and the Science Leadership Academy**

For the purpose of this study, I set out to identify, define, and evaluate an alternative to test-based accountability. Given my location at Haverford College I began to investigate public high-schools within Philadelphia. Philadelphia in many ways is a model city of how to implement the NCLB accountability system. Its use of the state Pennsylvania System of School Assessment (PSSA) test scores to make decisions about school success is paramount and data on school are included in School Report Cards, which are publicly available. Within this system I found one school that actively modeled alternative practices: The Science Leadership Academy (SLA), which began just four years ago in September of 2006. SLA was founded by Chris Lehmann who is the author of popular education blog PracticalTheory and now one of the National School Board Association “20 [Administrators] to Watch.” SLA is a site-select, magnet school partnered with The Franklin Institute Science Museum that has 49.9% African American students, 35.6% White students, 6.6% Asian students, and 6.8% Latino Students. 48.5% of the student body is economically disadvantaged; 21.3% are mentally gifted whereas 4.4% are special education students. The current student enrollment is 489 students (Annual Report 2009).

SLA on one level exemplifies all three accountability-schools-of-thought. In addition to, or in spite of its 100% compliance with the PSSAs (test-based accountability) and other state and city accountability requirements (which will increasingly represent standards-based
accountability), SLA, internally, is a unique place school that is committed to project-based learning and performance-based assessments (performance-based accountability).

It was at SLA’s staff meeting that I heard a discussion about accountability take place in a way I had never heard before. The staff at SLA are discussing the creation and implementation of what they are calling Standards-Based Reporting. If approved by the staff, by consensus, this system of accountability/assessment would be created and implemented in addition to their traditional grades, benchmark projects, which students complete four times a year to replace the city-wide requirement for interim testing, senior capstone project (a requirement for graduation), as well as the PSSAs.

The idea of Standards-Based Reporting comes out of the school’s staff noticing that students were completing their benchmark projects and meeting proficient on the PSSAs without certain critical skills aligned with behavior such as being able to collaborate with others, as well as content related skills, for example being able to write a strong thesis statement. From them noticing this, teachers asked the question “are there other skills that students did not have that did not get noticed by the pre-existing performance based assessment or state wide accountability systems?”

I believe that within Philadelphia’s context, the Science Leadership Academy’s combination of performance based-assessments combined with a potential standards-based reporting system may present a widespread, compelling, and achievable alternative to NCLB. It is this question I hope to explore throughout my empirical research. The purpose of which is to gather data on the potential of an alternative to arrive at preliminary conclusions, which can then be tested in a more rigorous way.
Methodology

Through my perspective as a student teacher at the school I will have full access to staff meetings and professional development, as well as the staff's Internet forum where ongoing professional conversations take place. Within SLA I will go about my research in three ways. The first is through analysis of public data on school quality, the second is through interviews, and the third is through observations.

To begin researching I will study all data that is made public about school quality, including report card and progress report as well as newspaper and journal articles.

The second method of research within my case study will be four semi-structured interviews with (1) the principal of the school, Chris Lehmann, (2) a founding history and science teacher and active participant in the SBR discussion, Gamal Sheriff, (3) Matthew Kay, a fourth year English teacher who began his career at SLA and (4) Rosalind Echols, a graduate of the Teach for America program and third year science teacher. My interview questions are below:

Interview with Principal, Chris Lehmann:

On No Child Left Behind
- How is your school held accountable under No Child Left Behind?
- How does NCLB measure student achievement?
- What school quality data do you collect for the city and state?
- What school quality data does the city/state collect and give you?
  - How are you supposed to use that data? How do you use that data?
- Do you feel that your school report card score accurately represents your school quality?

On Pennsylvania and Philadelphia Standards
- How is your school held accountable for meeting the state and district standards?
- How do the state and district standards influence student achievement?

On Science Leadership Academy
- How do you hold your school accountable?
- How would you hold your school accountable? What do you wish for accountability?
How do you at SLA measure student achievement?
What extra, if any, data do you collect about your school quality?
   How do you use that data?
Is there any data you would like to collect but do not have the time or resources?

How would your answers to these questions be different if you did not have to take the PSSAs?
How do the PSSAs/NCLB affect your school?
   a. What are the positives?
   b. Negatives?

On Standards Based Reporting
   What is Standards-Based Reporting?
   What began the discussion of SBR?
   Are you in favor of it? Why or why not? What is your vision for it?
   How do you see it fitting into a conversation about accountability?
   Do you think the district/state will support Standards-Based Reporting, do they have to?

Interview with Teachers

On No Child Left Behind
   How is your school held accountable under No Child Left Behind?
   How does NCLB measure student achievement?
   What school quality data do you collect for the city and state?
   What school quality data does the city/state collect and give you?
   How are you supposed to use that data? How do you use that data?
   Do you feel that your school report card score accurately represents your school quality?

On Pennsylvania and Philadelphia Standards
   How is your school held accountable for meeting the state and district standards?
   How do the state and district standards influence student achievement?

On Science Leadership Academy
   How does SLA hold SLA accountable?
   How would you hold SLA accountable? What do you wish for accountability?
   How do you at SLA measure student achievement?
   What extra, if any, data do you collect about your school quality?
      How do you use that data?
   Is there any data you would like to collect but do not have the time or resources?

   How would your answers to these questions be different if you did not have to take the PSSAs?
How do the PSSAs/NCLB affect your school?
   c. What are the positives?
   d. Negatives?

On Standards Based Reporting
   What is Standards-Based Reporting?
   What began the discussion of SBR?
   Are you in favor of it? Why or why not? What is your vision for it?
   How do you see it fitting into a conversation about accountability?
   Do you think the district/state will support Standards-Based Reporting, do they have to?

**Observations**

In addition to interviews I will sit in on all staff professional development meetings for eight weeks and assess the language and attitudes that surround performance-based assessments, narratives, daily reports, standards-based reporting and other data collected by the state, city and school.

**Broader Context Analysis**

In order to help me frame the context of educational accountability in Philadelphia, my research for this project will also include an interview with David Weiner, the Chief of the Office of Accountability for the School District of Philadelphia. Given this interviews and continuing traditional research, I will evaluate how SLA differs from typical Philadelphia public high school, in its creation, assessment and accountability system, student body, and community partnerships. I will also compare SLA's attempt at creating an alternative system to two landmark cases of accountability systems that combine test and performance based criteria: The New York Standards Performance Consortium and the Kentucky Education Reform Act. In total, this research will situate SLA's proposed system of Standards-Based Reporting within an ongoing search for an new, achievable model of accountability.
V. Results

School Accountability in Philadelphia

To situate Science Leadership Academy within a context of educational accountability, I will begin this section by laying out how schools are held accountable in the School District of Philadelphia. The School District of Philadelphia (SPD) holds schools accountable in three main ways. The first is by measuring school’s AYP in order to follow state accountability regulations which come from No Child Left Behind. The second is through a system called Annual Reports, or school report cards, which publicize data on a school’s academic achievement, operations, and community involvement. The third is through the School Performance Index (SPI) which compares schools to each other based on similar data from the Annual Reports.

David Weiner, Chief Officer of Accountability at the School District of Philadelphia explained that the first stage of school accountability in SPD is inline with the states’s accordance with NCLB. In addition to AYP, he added “for high school they also include grad rates and K-8 they include attendance, small things, but largely the biggest bulk of it is through PSSA scores and when you talk about PSSA they look at it in three ways, improvements in reading, improvements in math, and participation rates” (Weiner).

Under NCLB, since 2006, four schools in the district have been selected as Blue Ribbon Schools, the reward for continuously making AYP. “The award distinguishes and honors schools for helping students achieve at very high levels and for making significant progress in closing the achievement gap” (The School District of Philadelphia 2008). On the other hand, at the beginning of the 2009 school year 16 SDP schools were categorized as Corrective Action II, 7th
Corrective Action II is one stage of the sanctions imposed on schools that fail to make AYP under NCLB. Interestingly, however, according to the law Corrective Action II is only supposed to be a two year process. One year for drawing up corrective action plans and the second year for implementing them. Socolar (2009) explained:

The fact that Philadelphia has dozens of schools that have been allowed to languish in Corrective Action for three to seven years suggests two things:
1. That NCLB has failed to provide the guidance and resources or focus the attention necessary to draw up and implement plans to actually take corrective action - to restructure and turn around persistently low-performing schools, and/or
2. That there has been a tacit consensus among decision-makers in the state and District that the unproven but still-fashionable restructuring options called for by NCLB (replacing the staff, charter conversion, privatization or state takeover) are not adequate to the challenge of turning around large numbers of persistently low-performing schools.

The city is currently (finally) managing the implementation of final stages of sanctions as imposed by the No Child Left Behind legislation in twelve of the Corrective Action Schools. The Renaissance schools, as they are now called, have all failed to meet AYP for the last six years will face complete school take over. While some Philadelphia school teachers understand this course of action, many feel that it will cause more unrest and instability in schools that were just beginning to make progress (Socolar 2007). Regarding the fear of teachers, this process in the city draws attention to one of the failures of NCLB that was previously discussed–measuring student achievement as compared to a static marker of proficient rather than measuring student achievement based on growth. For many of these schools, student achievement results, which failed to meet AYP, were plugged into the Pennsylvania Value-Added Assessment System (PVAAS) and at least 11 of the schools were found to be making growth at or above the growth standard (Socolar 2007).
Mezzacappa and Socolar (2009) reported that about half of SDP’s schools had to allow its students to transfer due to not meeting AYP. The problem with the school choice sanction in Philadelphia is that only a few schools are “designated as receiving schools, and so the number of slots limits the number of students who are approved for these transfers.”

The SPD, however, will presumably be moving away from measuring student achievement based on a static proficiency mark both because of the possibilities for the reauthorization of ESEA as well as Pennsylvania’s own newest initiative the Keystone exams. While the PSSAs test a student’s knowledge about a wide range of areas that students may have covered over a period of a few years, the Keystone exams are going to be slightly different. Weiner explains:

[The Keystone exams are] going to start, at least, in high school, and they essentially are like an end to the unit or an end of the school year test so like a keystone test will be algebra two, so whenever you take algebra two you take it in 9th grade I take it in 10th grade my sister takes it in 11th grade, whenever you take algebra two you're going to take a final exam... The order you take them will depend on your course load.

The Keystone exams will be introduced next year (2010-2011) for math, the following year for reading, which includes literature and compositions, and by 2016 will be available for all subject areas. Given their nature as end of course exams, the Keystone exams more than the PSSAs will need to be aligned with the state education standards—one of the suggestions of the standards-based accountability school of thought. The SDP has a core curriculum which nearly every school in the city follows. The core curriculum is based on Pennsylvania’s education standards and has regular standardized assessments to measure student learning. The “benchmark” exams are two day testing periods which occur four times a year for math, reading, and science. Weiner explained, however that though they are an example of linked standards and assessments they
will not be used as an accountability measure. He said, "[The SDP] wants teachers to use those [Benchmark exams] as instructional tools like, 'my kid is weak in double digit addition I need to do better lessons in double digit addition,' 'my kids are great at subtraction I need to skip subtraction, focus on more double digit addition.'" While the district may have the best intentions to provide useful resources to teachers, many see the increased testing, regardless of their link to the standards, as increased testing only. Regarding the announcement of the Keystone tests in addition to the benchmark system a teacher's op-ed piece in The Philadelphia Public School Notebook read:

The most distressing part of the plan was the announcement of increased testing for students - which I honestly did not know was possible. Now, in addition to the PSSA 12th grade retest, the ACCESS test, the PSSA Writing, Reading/Mathematics and Science tests, quarterly Benchmark exams, and (new this year) weekly standards-based assessments, students will be required to take an end-of-year assessment and additional Benchmark. (Thacker 2009)

The district abides by NCLB accountability rules through PSSA testing, locally has incorporated state standards into their assessments, and also tries to incorporate some performance-based measurements; the SDP uses school quality reviews as described in the performance-based accountability school of thought, in their lowest performing schools.

We have our monthly visits to these schools and those are to say "things are going terribly here" attendance is dropping rapidly, the principal is not doing anything and the teachers don't care about the kids [and then] we'll send an outside as one final group that's been able to look at schools around the country who are able to give us some degree of a report. And we did that to 14 schools this past year... in Philadelphia we use an organization called School Works and they go out and then write up a full report about the school. (Wiener)

In addition to AYP schools in Philadelphia are held accountable using two other measures, Annual Reports, which judge schools each year on a wide range of quantitative and qualitative criteria and are used solely to compare schools to themselves each year and measure
the growth that they make in a variety of criteria; and the School Performance Index, which uses
some of the data in the Annual Reports to compare schools to each other.

The Annual Report measures school quality within four categories: (1) academics; (2)
operations; (3) community satisfaction; and (4) school selected indicators. Academics focuses on
PSSA scores and graduation statistics; school operations covers criteria such as school safety,
students and teacher attendance, and budget spending. Weiner’s description of the third and
fourth category are helpful. Results for community satisfaction are found from an “extensive
surveying program... that we use and we kind of index the surveys so we have 10 questions all
matched up to parent engagement and we ask questions like, ‘is the school reaching out to you?’
‘Is the school communicating to you in your home language?’ And for all of these questions we
take their response and change them into parent engagement.” School selected indicators allow
schools to “select from a litany of different areas” that the district provides, so that each
individual school “could say okay this year we want to focus on, getting more African American
students into AP literature, or we want to focus on lessening our lateness we have a lot of kids
who are tardy and we want to lessen that.” 2 Once Annual Reports are completed, they are
published on the SDP website for parents and community members to see. Schools can also refer
to the Annual Reports to see how much growth they have made in specific areas, Weiner says.

The second way Philadelphia public schools are held accountable locally is though a
document called the School Performance Index. Whereas the Annual Reports are a measure by
which schools can track their own growth from year to year, the SPI is how schools within the
district are compared to each other. Weiner says, “we compare schools to the whole district and

2 See Figure 1 for the detailed criteria within each category.
then we compare them to a subset of similar schools. So schools get a ranking of 1-10 compared to similar schools and a 1-10 compared to the overall district, 1 is the best and 10 is the worst” (Weiner).

Figure 1
Science Leadership Academy

The Science Leadership Academy is a small, progressive, magnet school located in Center City Philadelphia, PA. The school opened its doors in the fall of 2006 after being planned by founding Principal Chris Lehmann along with The Franklin Institute Science Museum with the school has a partnership. "The overarching goal [of SLA] is to create young, civic-minded citizens with a solid foundation in science, research, and entrepreneurship" (Lehmann Principal Report 2010).

State Accountability

The Science Leadership Academy is no different than other high schools in the SDP in terms of how it is held accountable by the state or district. The juniors at the Science Leadership Academy all take the PSSAs in Math, Reading, Science, and Writing, though only Math and Reading are used in accountability data. SLA has to disaggregate test results for White and African American students. In order for a school to meet AYP in Pennsylvania between 2008 and 2010, 63% of its students must be proficient in reading and 56% proficient in math. Last year, SLA made AYP in all categories with 64.3% of its students proficient in math and 83.5% proficient in reading. Of the African American subgroup, 49/1% were proficient in math and 76.4% in reading.

Though they met AYP, the staff at SLA have feelings about the being held accountable for the Math and Reading test results that are similar to the feelings of teachers across the country that have been publicized for the last five to eight years in reaction to NCLB. The principal and staff of SLA that I interviewed felt that the PSSAs were a negative form of accountability because they were only one measure of student learning (and an inaccurate one), that they
limited real teaching and learning, and that despite the espoused connection between standards and testing; that the testing did not measure the state standards.

Principal Chris Lehman explained, “an over reliance on test scores as a major piece of the data is flawed because I don’t think test scores measure student achievement in the way we think they do, I think they’re a one shot snapshot of a specific kind of skill and I think that’s problematic. I also think the over reliance on reading and math at the expense of any other subject, especially say of a science high school is problematic.” The emphasis on testing is not problematic for SLA alone; Jahi (2008) writing for the Philadelphia Public School Notebook reported that Candace Carter, a 12th grade student at Sayre High School, a comprehensive public high school in Philadelphia described her curriculum as “boring.” She did not blame the boring curriculum on the teachers though. She believes that Sayre places an increased emphasis on math, English, and science because “they have to make AYP.” One of the ways in which Sayre had adapted to the high-stakes testing is through an extended second period, which lasts 90 minutes and is usually used for English or math. Recent studies conflict as to how much the No Child Left Behind Act (NCLB) has done to close the test score gap, but several reports suggest that the law is a factor making it harder for educators to deliver a full and rich education. (Jahi 2008).

Science teacher Rosalind Echols explained why, even though she saw the value of testing at times, she thought the PSSAs were “horrible”:

There also in the past have been questions on the test with no correct answer. I just feel that that is ridiculous. If there is no correct answer than there is no way for someone to get it right and they thought they had a correct answer so, I just find things like that frustrating because it’s careless and just shows a lack of thought and a lack of understanding about what is really valuable about science. And if you’re going to test something, okay, fine, I admit that maybe there is
some value in that but not if the test is faulty and also doesn’t measure the things that are important.  

Echols went on to explain that not only were the tests flawed, by accident, they are also poor in quality: “my frustration with the test is that despite that they claim that a large percentage of it is the nature of science there’s a lot of questions on there that test a very very specific set of discrete facts that are not actually good indicators of whether or not students understand science as a whole.” This comment relates to the common connection between standards and testing. English teacher, Matthew Kay saw a similar tension in English. He claims that “Actually, the English state standards are really not bad....[but] They’re not even testing what’s on the standards. There’s creative stuff all over the standards and that’s not even tested, all the writing skill stuff that’s not even tested.” Lehmann, Echols, and Kay all expressed how the tests do not promote or support meaningful learning, and beyond that only capture a snapshot of student success. The future relationship between SLA and the PSSAs is uncertain given the impending replacement of the PSSAs with the Keystone exams. In light of this, Lehmann rhetorically asked once in our interview and once in a Professional Development meeting if SLA would be allowed to work out an alternative system to the Keystone exam. Because SLA does not follow the SDP’s Core Curriculum and the Keystone exams will be linked to the Core Curriculum having the ability to present alternative achievement data for students at the end of certain classes will be essential.

District Accountability

Irrespective of how SLA is held, and will be held accountable by the state of Pennsylvania, it has to produce data for two other accountability measures more locally. SLA

---

3 Henriques and Steinberg (2001) and Ravitch (2010) document similar testing flaws.
participates in the Annual Reports however, had "insufficient information to calculate a 2009 SPI" (All Schools SPI). SLA will qualify for a SPI in the school year 2010-2011 after it has had two 11th grade classes take the PSSAs.

Lehmann reflected on the school's Annual Report (Figure 1) with two critiques. The first is that the reports prescribe what "good schooling" looks like and provides no room for alternative visions and the second is that it assumes continuous improvement in all categories at all times. He claimed "Obviously in certain aspects I don't [feel that the school report card accurately reflects the school quality], for example the idea that a school might choose not to give AP tests or offer AP classes, and there could be very good pedagogical reasons why they don't, I think is a problem." He continued:

I think the problem with the school report card also is that it assumes, the way it is currently conducted assumes improvement at all times in all categories and I think that's a major issue, for example last year we had 98.1% teacher attendance which is an astronomically high number by industry standards and any kind of standard and so this year we have to make 98.3 and we're not going to make it, you know we got hit with H1N1 this year and a few teachers who missed a week or two of school and we're not gonna make it. Does that mean our school is not quality? Does that mean my teachers do not work hard? No, we're still over 97.5%, which is still well over any kind of standard that you would expect. So the notion of continuous improvement in all categories, I think is a troubling one because if your focus is everything than your focus is nothing.

**Internal Accountability**

When we look at how SLA holds itself accountable the qualities that differentiate it from other Philadelphia public high schools become particularly important. Their internal system of accountability is grounded in the firm belief in and commitment to progressive pedagogy.

In addition to giving students grades for their courses, teachers at Science Leadership Academy write narrative report cards in the first and third quarter of the year. Narratives, as they
are called, are supposed to communicate to the student and parents justification for their numerical grade and make suggestions for improvement. Student's narratives and grades, in line with SLA's core values, overwhelmingly reflect performance on project-based assessments. Additionally, I have identified 10 of SLA's practices which in combination separate it from other schools and particularly influence the staff's outlook on accountability. They are below:

1. Re-imagining the Core-Curriculum  
2. Project-Based Learning  
3. Core Values  
4. Whole School Rubric  
5. Mandated Understanding by Design  
6. Advisory  
7. School Tool  
8. Daily Reports  
9. Observations  
10. EduCon  

In our interview, Lehmann explained:

we were founded with the mandate to re imagine the core curriculum. That was part of our mandate and we have taken our mandate very seriously. I don't think we ever said we were creating a core curriculum that could be lifted and blackboxed. What we said was given the standards of the core curriculum, which we really do teach, could you imagine it in a project based fashion and what we wanted to do was provide an example of not only what the outcome would look like but of what the process would look like and I think we've taken very seriously that piece of the puzzle.

This mandate has encouraged a philosophy of learning within SLA that is demonstrated in its core values and use of project-based learning and assessments. SLA's five core values are Inquiry, Research, Collaboration, Presentation, and Reflection. These values are the foundation of the alternative curriculum that Lehmann and the staff have created and part of the package which allowed the school to become exempt from the quarterly Benchmark Exams. Kay who was quoted above as criticizing the PSSAs for not testing the standards, commented on why
project-based assessments are so much more meaningful: “if the kid can pull something out of a reading and create a brilliant essay now that’s useful information - if they can pull out elements of a creative story and turn around and write one now you have useful information then a kid really intimately knows what’s going on.”

Because project-based assessments are used school wide, another one of SLAs distinctive features the school-wide rubric that is used to assess student work. The rubric asks teachers, regardless of discipline, to assess student project on the basis of 5 criteria:

*Design:* does the student plan and structure the project thoughtfully and purposefully?,  
*Knowledge:* does the student demonstrate the understanding of ideas through inquiry?  
*Application:* does the student use a variety of skills and strategies to apply knowledge to the problem or project?  
*Process:* does the student take the necessary steps to fully realized the project goals? and  
*Presentation:* does the student effectively communicate the central ideas of the project? (scienceleadership.org)

Understanding by Design (UbD) is a framework developed to assist teachers in “backwards planning” instead of starting to plan based on all of information a teacher wants to cover, UbD encourages teachers to layout the essential questions and understandings that students should get out of a unit and then structure their learning activities, or specific lessons, to work up to those goals (Wiggins and McTighe 1998). SLA has mandated that teachers complete UbD plans for each unit that include state or national standards (Sherif). This has caused Kay to feel more accountable since the Unit Plans are turned into the principal and posted on the schools website. It also has made him more accepting of the state standards, “if you’re not teaching kids what’s there [on the state standards] your not a good teacher,” he said.

SLA’s project-based curriculum, core values, rubric, and usage of UbD all refer to the
creation of curriculum, standards, and assessment. The remaining five practices deal more with how individuals interact with each other at the school.

Every teacher at SLA has an advisory, a group of no more than 15 students who they care for over their four years at SLA. Advisory is the place to process school happenings, the transition from middle school to high school or from high school to college. It is a nonacademic space where students can foster a sense of community. Advisors serve as allies for students and the liaison between students, parents, and the school. SchoolTool is an electronic program that allows all staff members to easily send email messages out to any student’s advisor, teachers, parents or the principle. This allows for an ease of reporting between all faculty and parents.

Daily Reports are a smaller system of SLA that affect between 40 - 50 students who are in danger of failing courses. These students are required to have each of their teachers sign-off on their participation and work in the following categories: came to class prepared, participated in class, outstanding assignments, pending assignments. The Daily Reports are checked off at the end of each day by a member of the support staff and sent home to parents. Advisory, SchoolTool, and Daily Reports work together to form a learning community where students are at the center and a team of people are responsible for their success.

Faculty observations and collaboration as well as EduCon are the two remaining hallmarks of SLA. Observations were brought to my attention by Kay who said, “Chris [Lehmann] is in my classroom so you can’t really not be on your game cause your principal is going to walk in and if he is giving you the freedom to craft your own curriculum you feel you need to earn that respect.” Lehmann conducts formal observations of his teachers twice a year but informally walks in and out of teacher’s classrooms daily. Faculty at SLA are also strongly encouraged to
visit each others classrooms and work together during shared planning time on Wednesday afternoons.

Lastly, SLA founded and has hosted EduCon an education and technology conference attended by 500 guests for the last three years. EduCon is a weekend long conference, hosted at SLA, in which educators, authors, and students can choose from over one hundred breakout workshops that address a variety of topics regarding the interplay between progressive schooling and instructional technology. EduCon is the most tangible example of a school opening its doors and allowing the community at large to traipse in and judge its philosophy, systems, structures, and student work. In addition to EduCon, SLA's interactive website where a majority of student work is published, Chris Lehmann's blog PracticalTheory.org, and the number of newspaper and journal articles published about SLA make it a very public public school.

The combination of these 10 features of SLA set it apart from other schools in the school district and earned them national recognition as an extremely successful and innovative school. The form of accountability used by the state and district does not allow for SLA to authentically publicize what they are doing well. For example, according to Kay, the SLA engineering teacher, Mr. “V.K. has them over there in science class building reactors and what not and that stuffs not tested. Our kids are scoring lower than some schools on the tests but they’re making stuff that’s getting patented. How does that not factor in?”

Kay’s point above, as well as the other arguments made by staff members document short comings with testings and standards—the PSSAs are a snapshot of student learning, that they do not test the standards, and that the Annual Reports, and SPIs, while comprehensive, have inflexible weighting systems that rely heavily on standardized tests. This, confounded with
SLA's belief in and commitment to performance-based assessments for the student and the school, highlight a void in how to authentically communicate SLA's successes and failures.

Standards Based Reporting

The remainder of this section documents the Science Leadership Academy's discussion on a new system of assessment and accountability called Standards Based Reporting (SBR). SBR would require disciplines to establish clear skill standards that each student at SLA should have mastery of by the end of each course and ask teachers to check in with students throughout the year on said skills.

Due to its partnership with The Franklin Institute, the uncommon publicity that SLA has received, and its commitment to continually reflecting and improving on the teaching and learning that occurs in the building, it is not surprising that the issue of data, reporting and "accountability" arose. The conversation about Standards Based Reporting came simultaneously from two different directions. The first was partially external, when Dennis Wint, President and CEO of the Franklin Institute asked Frederic Bertley, Vice President, Center for Innovation in Science Learning and direct liaison to Science Leadership Academy, what else SLA is doing to assess its students, if the leadership did not believe that PSSA data was a good measurement. He wanted to know what he could show to people to present the learning outcomes of SLA students (PD 3/3/10).

While Lehmann made it clear that SBR was not to be implemented as a system for external reporting, he noted that "outside audience is not inconsequential." Teachers inside of SLA, however, raised questions about measuring student success at nearly the same point Mr. Wint did at the Franklin Institute. Staff members both in Professional Development meetings and on their
online forums discerned four potential benefits of a new assessment system besides providing evidence for external purposes: (1) SBR could inform instruction, (2) SBR could reach and support students, (3) SBR could be used to connect with parents, and (4) SBR would drive a link between instruction and assessments.

What is Standards-Based Reporting? Lehmann explained: “What we intend to do...[is] say here is what it means to be a student at SLA, here is what it means to be a student of English and here is what it means to be a student of science and then actually use that. Where the goal is twice a year, in a really low tech kinda way cause it’s gonna be checkboxey. Here is where I see you based on the work that you’ve done...in these standards.”

In the planning process for each discipline’s pilot, a range of logistical steps for implementation were considered. The following is a description of what has been accepted.

Students will evaluate themselves based on a portfolio they create of their own work, and teachers will then review the student’s assessment. Echols explained her justification,

I’d like to see them evaluating themselves ... I had students evaluate themselves and then I evaluated them and for some students they had a good idea of where they stand and if this was their best work but for other students they are really not there and so I can see from that perspective how this could be a useful mechanism for really helping the student know how they are doing compared to how they could be doing or how they need to be doing.

Sherif and Kay suggested that the standards list be added alongside the school rubric. Kay offered, “so that every time I am grading a project I’m saying, alright a kid scores a 4 on this and a 3 on this so just to have that data so that when it comes time to do the standards based report card i’m not trying to test my memory on each kid’s paper when I’ve read 65 of them and I don’t really remember what they did six months ago” (Kay).

The reports that are produced from the standards-based meetings with students would not
be punitive. Lehmann notes that they will not prevent students from walking at graduation, or go on student’s GPA or permanent record (PD 3/3/10).

Echols saw its value to be quite straightforward: "I can see myself talking to a parent being like your child can communicate verbally but your child has a difficult time expressing their ideas with writing which is not a science skill necessarily or a science standard but is still something that I think is important and shows up a lot in my class.”

Resident Principal Mr. Cohen said, “I think you have an opportunity here to address another subset of students who are not truly hitting it out of the ballpark here.” One way in which it could allow teachers to better reach students is through the act of isolating specific skills in need of improvement and grading as Lehmann explained does not do this:

Grading measures gumption in most schools as much as it measures skill set. We all know both versions of those kids. The one who can be a solid B student not because their skills are not necessarily the best kid in the class but they turn everything in. The grinders. The bookers. We all know the kids that get by by just outworking everyone else. And we also know the other version of that the kids who are really really smart but don’t do a lick and end up as C and D students because they don’t achieve their potential despite the fact that their grade is the result from doing three quarters of the work really well as opposed to doing C work all the time.

Math teacher, Erin Garvey who often assists individual students in Math Lab, mentioned that “when we send students in for math lab, and I don’t know students, it would be helpful to know what areas they weren’t strong in.”

Logistically, this program has not yet been implemented at Science Leadership Academy. The conversations I have documented have been the staff’s discussion of it, they, by consensus voted to try out a pilot program in the fourth quarter of this 2010-2011 school year. Currently, the majority of the SLA staff is supportive of a pilot program because they, at least in theory,
recognize how SBR aligns with their values about measuring student work. Kay explained in
detail:

I think that when you’re looking at what they’re able to achieve, because they’re
creating things throughout the year you can compare what they are able to create
in September to what they are able to create in February to what they are able to
create in April and say okay this kid was struggling with this step of the process
and now they are doing really well so I think, because we look at the process for
every project we’re able to tell - we’re able to see where there is growth and in
much more authentic ways than a test gives and we’re able to have conversations
about it.

Kay’s justification for SBR recognizes value in student work that cannot be garnered from test
results. Noddings (2007) writes that “Under [a test-based] system, we can usually answer the
question, Had Johnny learned X? - but we cannot answer the far more interesting question, What
has Johnny learned?” (5). Both Kay and Lehmann, influenced by Nodding’s philosophy, act on
the theory that there is no “rich a data source as the work of student’s themselves” (Lehmann).

There are however, some strong concerns. Founding art teacher Marcie Hull does not think that
SBR will work at all. She said, “In reality – why this is a total no-go is that there are too many
children in your care, too many children per class, unless you divide you class, some this quarter
some others, unless you flipflop I don’t see this feasible with 30 kids in your classroom” (PD
3/3/10). Kay did not think that the information teachers collected through any kind of SBR
system should be made public to individuals outside of the community. He simply stated that
they were doing this for the benefit of students and teachers. Echols questions whether SBR data
should be reported to anyone other than students and parents, however because of reliability and
validity concerns. She said: “it’s our own internal measure and when it comes down to it its my
personal evaluation of my students and so its hard to tell how valid that is because it’s all based
on my own interpretation and someone might have their own interpretation of where they are.”
Based on these findings of the Science Leadership Academy and their intersections with a variety of accountability methods, the next section will discuss the value, possibilities, limitations and implications of the Science Leadership Academy and their development of Standards Based Reporting.
VI. Discussion

Based on eight weeks of empirical research at the Science Leadership Academy I have found compelling evidence as an answer to my question: given the outcomes of No Child Left Behind, how should schools be held accountable? In addition to discussing my findings from SLA, as independent verification that what SLA is attempting to accomplish can be implemented on a larger scale, I will also draw upon two case studies: The New York Performance Standards Consortium, a network of 28 public high schools that oppose high-stakes standardized tests and have implemented in their stead commencement-level performance-based tasks, and KERA, the Kentucky Education Reform Act of 1990 which included progressive reforms such as performance-based assessments to measure student’s achievement as well as high-stakes standardized testing to hold schools accountable (Whitford and Jones 2000, 11).

SLA, the Consortium, and KERA all incorporate(d) performance-based assessments into their alternative visions of accountability. The Consortium began in 1990 when 40 high schools joined together to introduce practices such as literary essays and creative writing, research papers and projects, oral presentations and debates, and mathematical problems and applications, all that reflected the adage, “learning is complex, assessments should be too.” The 28 schools today use four standardized performance-based assessment tasks that all students must complete as graduation requirements: an analytic literary essay, a social studies research paper, an original science experiment, and an application of higher level mathematics.

The Kentucky Education Reform Act had six main elements: outcome definition, student assessment, local control of curriculum, an accountability index, expected rate of improvement, and rewards and sanctions. When KERA was first phased in, performance-based tests were
developed to align with the instructional strategies that were being introduced into classrooms. The performance tests included: group-problem solving tasks, open-response questions requiring short, on-demand written portfolio, and a mathematics portfolio for example. All tasks were constructed with specific scoring criteria. “Both portfolios were to contain a student’s best work rather than show progress over time” (Whitford and Jones 2000, 12).

In all examples, the performance assessment practices strictly aligned to state standards. When Lehmann explained the motivation for standards-based reporting, he claimed: “by having standards - it reminds you of the skills you’re supposed to teach to. And that’s the whole point of standards based instruction anyway.” Cook and Tashlik (2005) leaders of the Consortium similarly explained “Who wouldn’t want kids to write well, read analytically, punctuate correctly, and solve geometry problems? Who wouldn’t expect kids to know the scientific method and that Seneca Galls convention was a turning point in the women’s rights movement?” They went on to confirm, “the Consortium’s curriculum not only met but exceeded State standards... Readings, assignments, discussions, debates, activities, and performance tasks required in academic courses, inter-disciplinary courses, and project-based work were scrutinized to determine alignment with the State’s learning standards” (Cook and Tashlik 2005). When NY State commissioned a Blue Ribbon Panel to assess the Consortium school’s assessment system, the panel found that each school’s system varied but the same single approach appeared in all of the schools. In support of the alternative assessment system, the report concluded, “we believe that some important aspects of the performance of individuals may be ascertained better by a contextualized evaluation than by standardized, external assessments...” (Everson 2001). Based on graduation rates and college-going standards SLA and Consortium schools are successful.
While SLA will graduate its first class in the Spring of 2010, as of March 1, 2010 90% of SLA's seniors applied to college and over 80% have been accepted to at least one college (Lehmann Principal Report 2010). Demographically, Consortium schools, when compared to the average NY high school have more students of color, more students who qualify for free lunch, more students receiving special education services, and more entering 9th and 10th grade students scoring below the state standard on reading and mathematics tests" (Foote, 2005). Despite this, 

Consortium graduates do very well in college. Most attend 4-year colleges, the vast majority of which are rated competitive or better by Barron's [Profiles of American Colleges: 25th Edition, 2003]. Consortium schools' graduates earn, on average, a 2.6 GPA, which is close to a B-, upon completion of up to three semesters of college. [Note: The ACT defines college readiness as the ability to earn at least a C, or a 2.0 GPA, in college level courses (3).] Consortium students remain in college as well, with 84% of those attending 4-year colleges and 59% of those attending 2-year institutions returning for a second year. In comparison, on a national level only 73% of students who enter 4-year colleges and 56% of those who enter 2-year institutions return for their second year. (Foote 2005).

These results were found from though a longitudinal study of college performance of Consortium graduates: The New York Performance Standards College Performance Study which asked the following questions “Do the graduates of the New York Performance Standards Consortium schools perform well in college? Do these schools prepare their students - mostly African American, Latino, and economically disadvantaged - for college-level work?” In total 666 transcripts were analyzed from 18 of the 28 schools over a period of two years. The student demographics of the transcripts returned approximated the demographics of the Consortium as a whole: 19.4% White, 27.6% Black, 43.5% Hispanic, 9.6% Asian and others, and 60.7% eligible for free lunch (Foote, 2005).
In Kentucky, while teachers used performance-based assessments in their classes and state tests were designed with performance tasks, the Kentucky Instructional Results Information System (KIRIS) the high-stakes testing component of KERA, required the use of test-results to determine 'school scores' which were used to determine appropriate rewards and sanctions (Whitford and Jones 1997) “The story [of KERA and KIRIS] is a powerful lesson about how such a high-stakes accountability system can distort and undermine the original visions for effective curriculum, instruction, and assessment practices” (Jones and Whitford 1997).

Echols was previously quoted raising one of the points often used to challenge performance-based assessments: reliability. One of the ways the Consortium proves reliability is through the moderation study which brings together hundreds of teachers from all of the consortium’s schools in order to cooperatively evaluate examples of student work in order to develop a common understanding of grades. Educators in Kentucky did not have the same opportunity, “Over the course of KIRIS testing, the increasing pressure for higher reliability and tighter alignment with a specified body of content, brought on by the high-stakes purpose of the state testing, has translated into a shift away from ‘open-endedness’ and performance in the assessments” (Witford and Jones 2000, 13). As time went on, multiple choice questions were added to KIRIS tests and the performance tasks and portfolios were dropped. This transition to more traditional high-stakes tests strained local decision making about curriculum—as it has been shown to do else where in response to NCLB (Whitford and Jones 2000).

Before this watering-down of performance based assessment tasks occurred, policy makers in Kentucky engaged in a debate over the outcome goals of KERA. Originally, the six
goals represented a combination of academic and non academic targets, similar to Rothstein’s proposal in the literature review. They were:

1. Use basic communication and mathematics skills for purposes and situations they will encounter throughout their lives;
2. Apply core concepts and principles from mathematics, the sciences, the arts, the humanities, social studies, practical living studies, and vocational studies to situations they will encounter throughout their lives;
3. Become self-sufficient individuals
4. Become responsible members of a family, work group, or community;
5. Think and solve problems in school situations and in a variety of situations they will encounter in life;
6. Connect and integrate experiences and new knowledge from all subject matter fields with what they have previously learned and build on past learning experiences to acquire new information through various media sources.

Out of the six goals, numbers two and three faced criticism from various public entities and relevant policy makers decided that those desired outcomes would not be assessed. Demonstrating conformity to the traditional academic purpose of school, the remaining four goals turned into 75 “valued outcomes” which were renamed as “academic expectations” (Whitford and Jones 2000, 11). Interestingly, goal number two, “apply core concepts,” if kept would presumably have required performance-based tasks as the assessment.

Beyond student achievement data, Jones (2008) and Rothstein (2008) presented school accreditations and school quality reviews as a way to measure non-academic variables within a school such as teacher quality, school safety, and facilities. In Philadelphia, we saw that school quality data was collected about school operations and community satisfaction through surveys that are filled out by administrators, teachers, students and parents and posted on Annual Reports and School Performance Indexes. As a result of my research on SLA, I found interest in Jones and Rothstein’s proposals. Interestingly, both staff members at SLA and David Weiner himself had alternative visions for more qualitative school reviews.
When asked how he would like SLA to be held accountable, Lehmann responded by saying, "I mean Middle States is amazing, you invite a team of people in here and they rip apart every process and program and system and structure you have, what a wonderful thing. I'm told it's a great big pain in the butt to get through and requires a lot of work but I don't mind that." Middle States is a school accreditation agency similar to NEASC, described in the literature review. When asked what, if any, new measure he would add the SDP's way of holding schools accountable, Weiner responded, "I would love to do reviews at all schools...as a principal as a district employee as a teacher I think they're extremely valuable they give you tons of great information so I would love to add that in."

On a similar note, Kay responded to the question with the response, "student interviews."

He clarified:

I think there's a lot of talk about honoring student voice but I don't think that anyone's actually about it... And if you ask the student, they're not dumb... A student will look you dead in the face, I know because it's happened to me, thankfully not that often, a student will say, "I didn't get anything... That month, that unit, I didn't get any thing... And they'll give you suggestions too. They'll say like there was this moment where I was starting to get you but you went and did something stupid and now I don't get you anymore. And students will have that discussion with you, struggling students will, students with disabilities will. They'll tell you about assessment about curriculum... I think that Lehmann, Weiner, and Kay are in agreement about the role of school reviews or accreditations because as Rothstein and Jacobsen (2009) wrote, "narrow test-based accountability plans can't possibly accomplish their stated intent - to tell the states and nation whether schools and related public institutions are performing satisfactorily and to indicate where improvements are required." Kay reflected on an experience he had where "Somebody came to my classroom last year a student doing something and he did a minute by minute thing
in my class said student A did this and student B did this when I read through it it was amazing.”

Whitford and Jones also raised the point about accreditations in their discussion of the outcomes of KIRIS.

We agree that schools should be held accountable, in fact, more accountable than they have been with KIRIS. Specifically, what schools are asked to account for should be broadened beyond student outcome measures to include professional practices and equity issues such as opportunities to learn. A school quality review process including school self-evaluations and periodic site visits is an appropriate means for developing and reporting such components of a schools precuts. (Witford and Jones 2000, 22)

This recognition of the importance of gathering data on schools via school reviews and accreditations is limited by one factor of which everyone is aware. Weiner explained that while he would live to have “School Works come in to all schools it’s just the cost is too excessive right now especially [because] we’re in a 70 million dollar budget deficit. [He added,] I know New York City spends about 40 million dollars a year we just don’t have that kind of money right now to be throwing at School Works.” Echols also noted that “it’s hard to think of a system that would be effective and valid without it costing a huge investment in time and money.” Kay, similarly but with more hope said, “you gotta pay people to do this but you gotta have people in people’s classrooms.”

Having “people in people’s classrooms” is one remaining idea that SLA can add to a conversation on alternative systems of accountability. Two of the ten practices that made SLA unique were observations and EduCon. Having a principal who will walk in and out of classes in a positive way and faculty that observes each others classrooms, as well as, in a broader sense a school that is made public annually when 500+ conference attendees walk the halls, observe classes on fridays and sit in workshops throughout the weekend, make SLA a public public
school. Echols weighs in on this idea with regards to their technology integration and the amount of student work posted on the internet: “One of the things I like about what we are doing, what we try and do here is that everything is public. If someone wants to see what the students are doing they can just go and oh- there it is...If everyone is going to see what my students turn in than I want to make sure everything my students turn in demonstrates that they actually learned something.” This idea of being public is immensely powerful. It takes the idea of publicizing data, most often test results, to another level. This is much bigger than SBR having potential as an alternative system. This is a new way of thinking about accountability, especially one that should be thought of carefully given the proliferation of educational technology.

The intersection of SLA, the Consortium, and KERA bring forth a number of implications for accountability policy. Namely, notwithstanding the benefits of performance-based assessments, they cannot coexist productively with test-based accountability measures. As the results of KERA illustrated, high-stakes testing can negate the usefulness of performance based assessments. In New York, performance-based assessments existed as alternatives to the state Regents exams however still have to frequently react to the demands of a high-stakes test-based accountability system. It is significant that Lehmann and Cook and Tashlik expressed dedication to using the state standards and that KERA was a state run program, as this implies the value of standards and performance based accountability coexisting.
VII. Conclusion

Educators, scholars, civil society organizations, business leaders, and policy makers have all contributed their thoughts to the national conversation on how schools should be held accountable. Many of their approaches are written in response to the current policy of No Child Left Behind. As was seen with the test-centered approach, many believe that No Child Left Behind with the strong role of the federal government, unequivocal goal of math and reading proficiency by 2014, and commitment to standardized testing, is the appropriate way schools should be held accountable even though they have suggestions for how to strengthen its implementation. Those who argue for standards-based accountability believe that schools should be held accountable for providing students with a content-rich standards based education. Performance-based accountability advocates promote the idea of a dynamic accountability system that re-envisions schooling.

It is with this performance-based discourse that I am most aligned. In spite of the positive data that comes out of SLA and the New York Standards Based Consortium, I recognize that performance based accountability has little chance of becoming new national education policy—as the blueprint for the reauthorization of ESEA has made clear. At the same time, the outcomes of NCLB have been unfavorable, so much so that I also accept that the remaining life-span of test-based accountability will be minimal. My research on SLA and their implementation of Standards Based Reporting suggests that a more achievable alternative accountability system must combine standards-based and performance-based accountability goals.
Future Research

At the end of a Professional Development Meeting in March, Lehman urged his staff to vote in favor of implementing a pilot standards-based reporting program. He exclaimed, “until we do it, we can’t know if it can’t be done.” It is in keeping with this theme that I encourage educational and political science researchers to more directly confront the issue of accountability in their proposals for the reauthorization of ESEA. Seeing as how the idea of accountability is dominating nearly all discussion of education policy, but so few empirical studies have been done on the success of alternatives, it seems that this should be a priority for future research. Additionally, as the Science Leadership Academy grows, students graduate, and the Standards Based Reporting system is implemented it will continue to provide rich data on a variety of educational questions.
IX. Works Cited


Chubb, John E. Learning from no child left behind: how and why the nation's most important but controversial education law should be renewed. Stanford, Calif.: Hoover Institution Press, Stanford University, 2009.


Goertz, Margaret E. *Standards-based Reform: Lessons for the Past, Directions for the Future.* Proceedings of Clio at the Table, Providence. 2007.


Hirsch, E. D. *The schools we need and why we don't have them.* New York: Doubleday, 1996.


Socolar, Paul, and Dave Mezzacappa. "Yes, you can transfer ... even in high school." Philadelphia Public School Notebook, Fall 2009.


