

**IDENTIFYING EFFECTIVE MEASURES FOR A  
STATEWIDE ACCOUNTABILITY SYSTEM**

Jeremy Plowman, B.A., M.A.

Todd Siebert, B.S., M.S.E.

Corey Willich, B.S., M.S.E.

An Abstract Presented to the Graduate Faculty of  
Saint Louis University in Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Education

2011

## ABSTRACT

This paper is a project report of a policy analysis of state accountability measures used to evaluate public schools. Current state accountability systems use a variety of accountability measures to evaluate their schools. The inconsistent measures can create misrepresentations of how well or how poorly schools are educating their students. Two identical schools in different states can receive entirely different rankings based on the measures applied from their respective state accountability systems. The purpose of this project was to provide recommendations to state policymakers and state department of education officials on what accountability measures should be included in their state accountability systems.

The project team conducted a review of literature to create criteria and indicators for effective accountability measures. An expert panel provided feedback to the project team to evaluate and strengthen the criteria and indicators. The four criteria are: (a) status versus growth models of measuring student achievement results; (b) setting content and proficiency standards on high-stakes tests; (c) components of high-stakes tests; and (d) other accountability measures. The team's project was a policy analysis comparing 17 selected states' current accountability policies to the team-created criteria and indicators. The team reported an overall analysis of its findings by state and by indicator.

The project revealed that there were vast differences among the states examined in how they determine the effectiveness of schools in their state accountability systems. The project team offered seven recommendations to state policymakers and department of education officials. The findings of the project are that some states have begun implementing many of the recommendations of the project team. However, it is

imperative that states use a growth measure for high-stakes testing results rather than using a status measure exclusively. States must also continue to create valid and reliable assessments that go beyond multiple-choice in math, reading, and science only. Measures other than high-stakes testing are just as important when evaluating schools. Current initiatives, such as Common Core State Standards (CCSS), Race to the Top, and SMARTER Balanced Assessment Consortium (SBAC), are providing incentives to states to adopt many of the accountability measures the team is recommending.

PREVIEW

PREVIEW

**IDENTIFYING EFFECTIVE MEASURES FOR A  
STATEWIDE ACCOUNTABILITY SYSTEM**

Jeremy Plowman, B.A., M.A.

Todd Siebert, B.S., M.S.E.

Corey Willich, B.S., M.S.E.

An Abstract Presented to the Graduate Faculty of  
Saint Louis University in Partial Fulfillment  
of the Requirements for the Degree of  
Doctor of Education

2011

UMI Number: 3475201

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent on the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3475201

Copyright 2011 by ProQuest LLC.

All rights reserved. This edition of the work is protected against unauthorized copying under Title 17, United States Code.



ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 - 1346

COMMITTEE IN CHARGE OF CANDIDACY:

Associate Professor Susan Toft Everson,  
Chairperson and Advisor

Associate Professor Gary Wright

## TABLE OF CONTENTS

List of Tables .....	v
List of Figures .....	vii
<b>SECTION 1: INTRODUCTION AND IDENTIFICATION OF THE PROBLEM</b>	
Introduction .....	1
Urgency of the Problem .....	2
<b>SECTION 2: PHASES OF THE PROJECT</b>	
Phase 1: Identify the Problem .....	9
Phase 2: Foundational Knowledge .....	10
Phase 3: Criteria Development .....	11
Phase 4: Preparation for the Policy Analysis .....	12
Phase 5: Policy Analysis Findings .....	12
Phase 6: Overall Analysis .....	13
Phase 7: Recommendations and Conclusions .....	13
<b>SECTION 3: BUILDING FOUNDATIONAL KNOWLEDGE</b>	
Foundational Knowledge .....	14
Criteria Overview .....	19
Criteria One .....	22
Status Model .....	22
Growth Models .....	29
Criteria Two .....	39
Standards-Based Accountability .....	40
Content Standards .....	41
Performance Standards .....	43
NAEP and Proficiency Standards .....	46
Criteria Three .....	49
High-Stakes Testing .....	49
Testing Frequency .....	50
Whose Results Count .....	53
Test Questions .....	57
Criteria Four .....	61
Other Measures .....	61
Attendance .....	62
Graduation Rate .....	64
Teacher Quality .....	68
<b>SECTION 4: CRITERIA DEVELOPMENT FOR POLICY ANALYSIS</b>	
Expert Panel Review .....	73
Panel Feedback .....	77
Criteria One .....	77
Criteria Two .....	79



Criteria Three .....	80
Criteria Four .....	81
<b>SECTION 5: PREPARATION FOR POLICY ANALYSIS</b>	
Selection of States .....	83
Arkansas .....	85
California .....	85
Colorado .....	86
Florida .....	86
Georgia .....	87
Louisiana .....	87
Massachusetts .....	88
Michigan .....	88
Mississippi .....	89
Missouri .....	89
New Hampshire .....	90
New Jersey .....	90
North Carolina .....	90
Oklahoma .....	91
Texas .....	91
Virginia .....	92
Wisconsin .....	92
Criteria and Indicator Analysis Process .....	93
Criteria One .....	94
Criteria Two .....	95
Criteria Three .....	96
Criteria Four .....	97
<b>SECTION 6: POLICY ANALYSIS FINDINGS</b>	
Analysis by Indicator .....	98
Criteria One – Indicator A .....	98
Criteria One – Indicator B .....	99
Criteria One – Indicator C .....	101
Criteria One – Indicator D .....	102
Criteria One – Indicator E .....	104
Criteria One – Indicator F .....	105
Criteria Two – Indicator A .....	106
Criteria Two – Indicator B .....	107
Criteria Two – Indicator C .....	108
Criteria Three – Indicator A .....	109
Criteria Three – Indicator B .....	110
Criteria Three – Indicator C .....	111
Criteria Four – Indicator A .....	112
Criteria Four – Indicator B .....	113
Criteria Four – Indicator C .....	115

State by State Analysis.....	117
Arkansas .....	117
California.....	118
Colorado.....	119
Florida .....	120
Georgia.....	121
Louisiana .....	122
Massachusetts.....	122
Michigan.....	123
Mississippi.....	124
Missouri.....	125
New Hampshire.....	126
New Jersey .....	127
North Carolina.....	128
Oklahoma .....	129
Texas .....	130
Virginia.....	131
Wisconsin.....	131
SECTION 7: OVERALL ANALYSIS	
Number of States that Match Each Indicator.....	133
Lessons Learned.....	136
SECTION 8: RECOMMENDATIONS AND CONCLUSIONS	
Recommendations.....	142
Recommendation 1 .....	143
Recommendation 2.....	143
Recommendation 3.....	144
Recommendation 4.....	145
Recommendation 5.....	146
Recommendation 6.....	146
Recommendation 7.....	147
Conclusion .....	148
Appendix A.....	151
Appendix B.....	153
Appendix C.....	160
Appendix D.....	172
References.....	173
Vitae Auctorum.....	189

## LIST OF TABLES

Table 1:	Status Measure in States .....	99
Table 2:	Growth Measure and Individual Student Growth in States .....	101
Table 3:	Validity and Reliability in Accountability Measures.....	102
Table 4:	Emphasis of Validity and Reliability in High-stakes Test Measures.....	103
Table 5:	School Effectiveness on Individual Growth .....	105
Table 6:	Ease of Understanding in States' High-stakes Testing Results .....	106
Table 7:	Strong Academic Standards in States .....	107
Table 8:	State Level Cut Scores Compared to NAEP Basic Level.....	108
Table 9:	States Not Adopting NAEP as their High-stakes Assessment.....	108
Table 10:	States Perform High-stakes Testing beyond NCLB Requirements .....	110
Table 11:	States Hold Schools Accountable for the Performance of 100% of their Students.....	111
Table 12:	Utilization of Multiple Question Formats on High-stakes Testing.....	112
Table 13:	States' Use of Attendance as an Accountability Measure .....	113
Table 14:	States Use the Adjusted Four-year Cohort Graduation Rate .....	114
Table 15:	States will Measure Teacher Quality in Schools .....	116
Table 16:	Arkansas.....	117
Table 17:	California .....	118
Table 18:	Colorado.....	119
Table 19:	Florida.....	120
Table 20:	Georgia.....	121
Table 21:	Louisiana.....	122
Table 22:	Massachusetts.....	122

Table 23:	Michigan .....	123
Table 24:	Mississippi .....	124
Table 25:	Missouri .....	125
Table 26:	New Hampshire.....	126
Table 27:	New Jersey .....	127
Table 28:	North Carolina.....	128
Table 29:	Oklahoma .....	129
Table 30:	Texas .....	130
Table 31:	Virginia .....	131
Table 32:	Wisconsin.....	131
Table 33:	Selected States' Measures of Accountability.....	137
Table 34:	Percentage of Strong Standards in Each State .....	172

PREVIEW

## LIST OF FIGURES

Figure 1:	Categories of Schools in Relation to Proficiency and Growth .....	34
Figure 2:	Summary of Criteria 1 by Indicators.....	133
Figure 3:	Summary of Criteria 2 by Indicators.....	134
Figure 4:	Summary of Criteria 3 by Indicators.....	135
Figure 5:	Summary of Criteria 4 by Indicators.....	136

PREVIEW

## SECTION 1

### INTRODUCTION AND IDENTIFICATION OF THE PROBLEM

#### Introduction

This project report is a policy analysis of state accountability measures used to evaluate public schools. The project team created research-based criteria and indicators for effective accountability measures and compared them to current accountability measures used in states. Based on the policy analysis, the project team made recommendations to state legislators and state education officials.

The paper is divided into eight sections. Section 1 articulates the rationale and framework for the policy analysis. In Section 1, the team also introduces the problem statement and the guiding questions. Section 2 describes the phases of the work the project team completed. In Section 3, the team conducts a review of literature to build foundational knowledge for the criteria. In Section 4, the team, with the help of expert panel feedback, develops criteria and indicators for the policy analysis of state accountability measures. Section 5 explains the selection of the 17 states used in the policy analysis. It also addresses the criteria and indicator analysis process. Section 6 contains the project team's analysis of current state accountability measures against the team-created criteria and indicators and presents the findings. Section 7 was an overall analysis of the project. In Section 8, the project team reports recommendations and conclusions.

## Urgency of the Problem

When No Child Left Behind (NCLB) was legislated in 2002, it was the culmination of two decades of the standards-based accountability movement in American public schools (No Child Left Behind Act of 2001, 2002). NCLB was going to, once and for all, hold school districts accountable for educating every student and closing the achievement gap between different subgroups. President George W. Bush, in his Nomination Acceptance Address at the 2004 Republican National Convention, stated when discussing NCLB:

We are transforming our schools by raising standards and focusing on results. We are insisting on accountability, empowering parents and teachers, and making sure that local people are in charge of their schools. By testing every child, we are identifying those who need help, and we're providing a record level of funding to get them that help. . . . challenging the soft bigotry of low expectations. And that is the spirit of our education reform and the commitment of our country: . . . We will leave no child behind. (Bush, 2004, p. 3)

However, since its inception, NCLB has been much maligned, as have the state accountability systems created as a result of the legislation. The problem with the current accountability systems is not the goals of the policies nor the focus on outputs. Rather, the problem is that the measures they are based upon, especially the high-stakes tests and how they are used, are flawed.

Policymakers need to have a thorough understanding of the measures which should be used in a statewide accountability system. Policymakers, researchers, and education officials use the measures to evaluate whether schools and districts are doing a satisfactory job of educating their students. An examination of measures could prompt policymakers to alter policies that have an impact on school districts and their students. It is vital that policymakers have knowledge of truly indicative accountability measures.

Poor measures that misidentify the success or failure of a district or school can lead to undeserved rewards and sanctions and failed accountability systems.

The American public and policymakers have always expected that the United States should be the premier country in the world. Beginning in 1983 with *A Nation at Risk*, numerous studies maintained that the United States' education system was falling behind those of other nations. According to William Evers and Herbert Walberg (2002):

Americans take great pride in the superior and ever increasing effectiveness and efficiency of most of our industries. Yet our schools fall behind those in other countries and have become less rather than more efficient, which is far from what we would want, given their central importance in the American economy and society. (p. 1)

For instance, in their executive summary of the 2007 Trends in International Mathematics and Science Study (TIMSS), the National Center for Education Statistics highlighted U.S. students' struggles in comparison to the achievements of their Asian and European counterparts in both math and science. The average U.S. fourth-grade mathematics score was lower than those of eight countries, while the average U.S. eighth-grade science scores were lower than those of nine countries all located in Asia and Europe (Gonzales, Williams, Jocelyn, Roey, Kastberg, Brenwald, 2009, p. iii). Although the findings of many of these studies, which show U.S. public schools are performing poorly, are dubious at best, the perception of the American public and policymakers is that U.S. schools are in crisis. In 1997, U.S. Secretary of Education Richard Riley commented in his state of education address, "we need to stop dumbing down our children, and reach up and set higher expectations" (Riley, 1997, ¶ 5). Although a majority of Americans still rate their local schools highly, the public view of the nation's schools as a whole are quite negative. The 2007 Phi Delta Kappan/Gallup poll found that



only 22 percent of Americans would give the nation's public schools an A or B (Bushaw & Gallup, 2008). Thus, whether U.S. schools are performing well or not, politicians and the public believe increased accountability in public schools is a necessity.

This fear that American public schools were not able to compete with the education systems of other countries or close the achievement gap between different groups led to the creation of outcome- and standards-based accountability systems. The shift towards standards-based accountability occurred first at the state level. In 2001, the No Child Left Behind Act was passed with great optimism as a federal accountability system. NCLB, and the state accountability systems created as a result, were widely lauded as a solution to identifying which public schools were doing a quality job educating students and which schools needed to change their ways. However, despite the initial support for the goals of NCLB, there has been opposition to how it has been implemented in the states. According to an American Research and Development Corporation (RAND) study in 2009:

After eight years of effort, many schools remain unable to meet NCLB's expectations for improvements in student performance. More importantly, fewer schools are improving, and the proportion slated for corrective action or restructuring is increasing. Existing interventions do not seem to be working at these schools, and more teachers and administrators report frustration with the law. (Stecher & Li, 2009, p. 2)

The Gallup Poll released results to a survey about No Child Left Behind. Frank Newport (2009), in a Gallup Poll Briefing, reported:

Of those familiar with the act, 21% say it has made the education received by public school students in the United States better, while almost half, 45%, say it has made no difference and 29% say it has made public school students' education worse. (¶ 1)

The negative opinion was more prevalent in those who consider themselves more familiar with NCLB. The negativity surrounding NCLB and state accountability systems has more to do with how schools and student achievement are measured and identified and less to do with the goals of the accountability systems.

The Thomas B. Fordham Institute illustrated the weakness of the measures used to evaluate schools under NCLB in “The Accountability Illusion” (2009). In it, the authors researched the various NCLB standards in 28 states. They measured 36 schools against the accountability measures in each state. The report found that what state a school is in greatly impacts whether that school is deemed acceptable under Adequate Yearly Progress (AYP). For instance, in Massachusetts, only one of 18 elementary schools studied would have made AYP, while in Wisconsin, 17 of those same schools would make AYP. As the 2009 report notes, “Same kids, same schools—different states, different rules” (Thomas B. Fordham, 2009, A few key findings, ¶ 2). Chester Finn, President of the Thomas B. Fordham Institute, wrote, “This report’s crucial finding is that—contrary to what the average American likely believes—there is no common nationwide accountability system for measuring student performance under NCLB. The AYP system is idiosyncratic, even random and opaque.” He added, “Without a common standard to help determine whether a given school is successful or not, its fate under NCLB is determined by a set of arcane rules created for each state” (Thomas B. Fordham, 2009, ¶ 3).

Richard Elmore (2003), an advocate for effective school accountability systems, stated, “Low-performing schools are part of a larger problem of educational accountability. Solving the problem will require accountability systems that can

distinguish between schools that are improving and those that are not” (p. 17). Current accountability measures are not able to draw the distinction mentioned by Elmore. Both Elmore and Finn point out the limitations in current accountability systems due to the measures utilized.

Leading experts on educational accountability systems believe that statewide accountability systems, not federal, are the best mechanism for ensuring improved student learning. Rothstein, Jacobsen, and Wilder (2008) stated the following:

The accountability system we suggest . . . should be a state, not federal, responsibility. Not only do we have a constitutional tradition of state control of education, but the failure of No Child Left Behind has made it apparent that in this large country, the U.S. Congress and Department of Education are too distant to micromanage school performance. Policy debates in education somehow become more ideologically polarized and less pragmatic when they move from state capitals to Washington. (p. 143)

Rothstein et al. (2008) added, “The federal government is too distant from the provision of educational services to be primarily responsible for holding schools . . . accountable. State governments can and should be the vehicles for doing so” (p. 5). However, whether it is a state or federal accountability system, the measures must be able to distinguish between schools that are effective in educating each and every student and those not living up to the intentions of the federal mandate and state accountability systems.

In response to this need of indicative measures, the project team developed the following three guiding questions:

1. What does research say are the most valid and reliable accountability measures in public school accountability?

2. How do the team-created criteria and indicators for accountability measures compare to current state accountability measures?

3. What recommendations should be made to state accountability officials and policymakers to improve the accountability measures used in their state systems?

The problem statement will address the role of statewide public school accountability systems, which has changed from a focus on resources and inputs to a focus on standards and student outcomes (Fuller, Wright, Gesicki & Kang, 2007; Isumi & Evers, 2002; Viadero, 2006). This focus on student outcomes has provided the proper direction for the accountability movement in public school education. The federal No Child Left Behind Act of 2001, with the goal of improving student achievement and closing the achievement gap, has dominated the public school accountability landscape throughout the past decade. Under NCLB, states have been compelled to adjust their state tests and accountability systems to comply with federal mandates (Perie, Park & Klau, 2007; Rothstein et al., 2008). Although the goals of current accountability systems are laudable, accountability systems are only as effective as their accountability measures. If a state's accountability measures in evaluating public schools are flawed, the entire system will be flawed (Ballou & Springer, 2009; Berliner & Nichols, 2007; Elmore, 2003; Hershberg, n.d.).

Through analysis of the literature and the use of the three guiding questions to focus its research, the project team developed the following problem statement:

State policymakers and state department of education officials need to formulate accountability measures which build trust and confidence among educators and the public. For accountability in U.S. public schools to be effective, accountability measures

and indicators must be valid and reliable. Under NCLB and current state accountability systems, schools are often misidentified or inaccurately described based on high-stakes testing and other measures used in a particular state (Cronin, Dahlin, Xiang & McCahon, 2009; Downey, von Hippel & Hughes, 2008; Finn & Petrilli, 2008; Hill & DePascale, 2003; Linn, 2000;). The project team analyzed current accountability measures against research-based criteria and made recommendations to the state policymakers and state education officials who implement the accountability systems.

The project team also developed a task list in order to provide the team structure and guidance for the research and analysis.

1. Synthesize the research on best practices for public school accountability measures in order to create a set of criteria and indicators necessary to ensure an effective statewide accountability system for public schools.

2. Identify and assemble an expert panel on accountability to review the project team's research-based best practice criteria of effective accountability measures. The team will consider the panel's feedback while revising the criteria and indicators.

3. Analyze measures in various state accountability models against the team's criteria and indicators.

4. Make recommendations to state policymakers and state departments of education to aid them in creating more effective and indicative statewide accountability measures and systems.

In the next section, the project team discusses the phases of the project beginning with the identification of the problem. Each step of the project is explained.

## SECTION 2

### PHASES OF THE PROJECT

The project team describes the eight phases of the project beginning with identifying the problem and finishing with recommendations and conclusions. The team writes about the tasks performed during each phase and provides the section where each phase was completed.

#### Phase 1: Identify the Problem

The first phase of the project was to identify the primary area of focus for the project. The team initially became interested in accountability while considering teacher tenure. Continued reading about teacher accountability in school performance became the most commonly discussed topic among the project team members. The readings highlighted the standards-based accountability movement in the United States, accelerated by No Child Left Behind. While standards-based accountability has been at the forefront of policy debates on public schools for over two decades, there is still discord over the effectiveness and methods of NCLB and state accountability systems.

Once public school accountability was identified as the primary area of focus, the project team tried to pinpoint why state and federal accountability systems were often viewed as failures by researchers, policymakers, and the public. As more research was uncovered, the measures used to evaluate public schools stood out as a detriment to effective accountability. Thus, the project team focused on identifying effective accountability measures for state education systems. A thorough description of this phase, including the problem statement, is found in Section 1.

## Phase 2: Foundational Knowledge

The project team set out to identify which accountability measures would provide a clearer picture of whether or not a school was doing an effective job educating its students. High-stakes testing became a primary focus of the team; not only how the tests were constructed, but also how their results were interpreted. Additional research improved the team's understanding of testing models and the impact the various models can have on illustrating the achievement of students and schools. Testing is not the only measure used in accountability, so other measures were examined.

Once the problem of effective measures for a statewide accountability system became the central focus, the team again went back to a unifying question, "What components must be in place for an effective accountability system?" The question helped the team realize there was not substantial research on the most effective measures to evaluate schools. Factors were considered that would apply to all levels of compulsory education, ranging from elementary through high school. The team concluded that a policy analysis comparing current state accountability measures to the research-based accountability measures created by the project team would be a worthwhile endeavor. In order to conduct the policy analysis, the project team set out to create criteria and indicators of effective accountability measures.

The team categorized its research on accountability measures into eight areas, which the team labeled "bins." Looking more closely at the research, the project team eliminated bins with limited supporting research, reducing the bins from eight to four. Four criteria and corresponding indicators were selected because they had applicability across all levels and they represented the most current research on measures in

educational accountability. The four criteria are: (a) status versus growth models of measuring student achievement results; (b) setting content and proficiency standards on high-stakes tests; (c) components of high-stakes tests (e.g., who tests, whose scores count, frequency of tests, makeup of tests); and (d) other accountability measures (e.g., attendance, graduation rate, teacher quality). The team collected research on all four topics. Next, criteria were written for each topic. Finally, the team developed indicators for each criterion. The foundational knowledge and criteria and indicators can be found in Section 3.

### Phase 3: Criteria Development

In order to validate the criteria and indicators, the team sought the feedback of an expert panel. The expert panel, having extensive knowledge of accountability in public school education, would provide guidance on the team-generated criteria and indicators. The project team was able to identify five researchers whose expertise was in accountability and sought their input after completing an Institutional Review Board (IRB) application through Saint Louis University. The team sent criteria and indicators to the expert panelists and asked for their evaluative comments for each of the indicators within the criteria.

Once comments and constructive notes were received from the expert panelists, the team re-evaluated the criteria and indicators. The project team used comments and additional research suggested by the panelists to revise criteria. The revised criteria and indicators, along with a description of the expert panelists and revision process, are found in Section 4.