

Implementing Federal Policy: Confronting State Capacity and Will

**Sara Dahill-Brown and Lesley Lavery
University of Wisconsin-Madison
October 24, 2009**

Abstract

For decades public policy scholars in the United States have investigated the process of policy implementation with two related goals in mind: to explain implementation failures, and, by identifying the causes of these failures, to improve the design of policy. “Top down” scholars have focused on design issues. “Bottom up” scholars have focused on the incentives and limitations confronting street-level policy administrators. A third generation of “hybrid” theorists have focused on the complications created by overlapping governmental authorities. But a coherent theory of policy implementation still eludes the discipline. Drawing insight from the attempts of “hybrid” theorists, we delineate a theory of policy implementation in the United States that accounts for the horizontal and vertical division of authority necessary in a highly fragmented federal system. We suggest that will and capacity – assessed at the state level where actors with primary responsibility must balance federal goals with local demands and structures – can explain implementation within a given state and the substantial variation in policy response and outcomes across states. We develop our theory around No Child Left Behind and test it using state-level data.

I. Introduction

In 2002, Congress reauthorized the Elementary and Secondary Education Act (ESEA) as No Child Left Behind (NCLB). News accounts described the act as “the broadest rewriting of federal education policy in decades,” (Milbank, 2002, p. A3) and President George W. Bush proudly proclaimed that it would “put American schools on a new path of reform and a new path of results” (Bumiller, 2002, p. A16). The bill’s new name and bipartisan support suggested a more serious commitment by the federal government to redressing educational inequality and to improving educational quality more generally. The body of the bill also suggested a new pathway; states were to impose clear mandates, goals, and accountability measures – specifically laid out in the bill – in exchange for federal funds. But eight years after NCLB’s passage, it is not apparent that achievement gaps are closing or that the quality of education is improving (Owens and Sunderman, 2006; Dillon and Rotherman, 2007). For most students in most states, the status quo persists.

This article addresses two important research questions: Why has a publicly appealing policy that garnered bipartisan support failed to produce greater change in student achievement? What can explain the variation in states’ response to NCLB? Existing theories of policy implementation that attribute failure to policy design, the goals of street-level bureaucrats, or a combination of competing top-down and bottom-up forces do not present a clear framework for analysts who seek to understand outcomes in a *federal* system. Therefore, in this paper we propose a modified theory of policy implementation. Our theory suggests that state compliance, or lack thereof, is a function of each state’s political will and capacity and variation in outcomes across states can be explained by some combination of these factors. It is imperative that scholars of implementation take seriously the need to better conceptualize the subfield; without such efforts, the different strands of work will continue to speak past one another and to explain less than they might otherwise.

We proceed in six sections. First, we describe top-down and bottom-up theories of policy implementation, highlight limits of these accounts, and propose our own “hybrid” theory of implementation. Next, we discuss the history of federal education policy and explain why education policy provides an ideal test for our theory. Then, we describe the individual tenets of NCLB in more detail. This leads directly to our hypotheses. We then present the data and

methods used to test our theory. In the paper's final section we conclude and discuss avenues for future research.

II. Implementation Theories: Top-Down, Bottom-Up and the Hybrid Approach

An often contentious implementation process begins after a bill becomes a law (Bardach, 1977). Compromises made to enact sweeping legislation fall apart as quieted interests gain power and voice in a policy's enactment (Matland, 1995). This appears to be especially true in the United States' federal system, which asks state level actors to implement policy developed in an institution often far-removed from these state actors' immediate concerns as well as those of contingencies on the ground (Shober et al., 2006). Several groups of policy scholars have attempted to explain the tumultuous implementation process in greater detail.

The first generation of implementation scholars typically began their analyses with a statute (or some other implementation decision) and examined the extent to which its legally-mandated objectives were achieved (Pressman and Wildavsky, 1973). They viewed implementation as a hierarchical execution of centrally-defined policy initiatives (Pulz and Trieb, 2006) and believed that if policy goals were clear and consistent (Van Meter and Van Horn, 1975; Mazmanian and Sabatier 1983), the number of actors involved could be minimized (Pressman and Wildavsky, 1973; Bardach, 1977), the extent of change or intrusion of the policy could be limited (Van Meter and Van Horn, 1975; Mazmanian and Sabatier, 1983), and the policy was placed responsibly with an agency sympathetic to the policy's goals (Van Meter and Van Horn, 1975; Sabatier 1986), success would follow. This brand of scholarship was termed "top-down" due to its emphasis on planning and design. Success was defined by such theorists in terms of specific outcomes tied directly to the statutes that created programs or initiatives (Pulz and Trieb, 2006).

Other scholars began to criticize the exclusive focus on administrative processes and the failure to consider political context. These "bottom-uppers" argued that a more realistic understanding of implementation could be gained by looking at a policy from the view of the target population and "street-level bureaucrats" who directly administered the policy (Lipsky 1971, 1980). These researchers looked at the behavior of public servants from teachers to social workers and police officers and the interactions between these public servants and citizens to explain local variation in policy outcomes (Weatherley and Lipsky, 1978; Hjern 1982). Some of

these theorists began to suggest that broader definitions of successful implementation (compromises that enabled strides toward a more distal goal) might also be warranted (Palumbo, Maynard-Moody, and Wright 1984).

A third generation of policy implementation theorists have attempted to bridge the gap between top-down and bottom-up approaches by incorporating insights from both, often using longitudinal and meta analyses (Goggin et al. 1990; Sabatier 1986). Goggin et al. argue that implementers are political actors in their own right and that policy outcomes depend on a complicated negotiation processes between implementers and central authorities (1990). Sabatier's more recent work suggests that implementation is a process that occurs in a policy subsystem. The important actors and advocacy coalitions within that subsystem are stable over time but major alterations in the policy core may occur as a result of external factors such as influence from other policy fields or economic development (1986). Elmore (1985) suggests that policy makers should start with a top-down consideration of the policy instruments available to them (forward mapping) and then use bottom-up perspective to identify the incentive structure of implementers and target groups (backward mapping). This work resembles Bardach's earlier work on scenario-writing and "fixing the game" (Bardach, 1977) but takes actors' incentives into account even more explicitly. Berman (1987) also attempts to bridge top-down and bottom-up thinking with his suggestion that policy implementation occurs at two levels; the macroimplementation level at which centrally located actors devise a government program, and the microimplementation level at which local organizations react to macrolevel plans. Finally, Matland (1995) suggests top-down and bottom-up approaches can be reconciled by concentrating on the theoretical significance of ambiguity and conflict for policy implementation. He develops a typology by which the degree of clarity (a top-down focus) and consensus (a bottom-up concern) in or for a policy can be used to predict both implementation success and policy content.

We do not dispute the findings of these authors. Indeed, we agree that the clarity of a specified policy goal and the incentive structure confronting a street-level bureaucrat, among a myriad other factors have significant effects on policy outcomes and substantial explanatory power when scholars look to explain variation. However, none of these theories provide scholars (or policy-makers) with a consistent structure for analyzing (or anticipating) the implementation process, and often they ignore the state-level institutions at work in the middle of the process.

We therefore join recent hybrid accounts of policy implementation in emphasizing the need to examine both intentions at the top and reception at the bottom, but improve upon past work by explicitly accounting for tensions present in a *federalist system* where semi-sovereign state actors serve as liaisons between federal policymakers and local-level service deliverymen and must negotiate working solutions to address the mismatch between federal prescriptions and local realities.

We suggest that state-level bureaucrats housed in an agency charged with the implementation of a particular policy take measure of both their state's institutional capacity and political support for the newly determined goals. Capacity to implement policy is dictated by resources available to address problems in a particular policy area (such as education). State and local tax dollars allocated toward a particular issue are an obvious constraint. Existing state infrastructure and human capital also restrict response. State-level political will cannot be measured directly and may be especially difficult to discern in response to new or very general policies, however, we believe that it can be estimated by measuring citizens' general views toward government, ideological and political leanings, or responses to previous policy in that issue area. Greater success should be expected for policies that align with a state's ideological or political ideals or build upon a popular, existing policy. When policies clash with states' strongly held beliefs and traditions or are seen as imposing by those operating under an existing policy, successful implementation is less probable. Before turning to specific hypotheses and operationalizing these concepts, we provide additional background on the history of federal education policy, a test case for our theory.

III. From ESEA to NCLB, Implementation and Federal Education Policy

Federal education policy in the United States, starting with the Elementary and Secondary Education Act (ESEA) of 1965, has been crafted to satisfy a multitude of social and political goals and has been implemented across widely varying local and state political contexts. Most scholars have described three phases of implementation challenges faced by ESEA: ambiguity between 1965 and 1980; a lack of national political interest between 1981 and 1989; and increasingly prominent political debates between 1990 and 2001 (Jennings 2000 & McDonnell 2005). This history of uneven compliance makes federal education policy a good case for testing our theory.

In 1965, ESEA's Title I was developed to provide categorical grants to districts with a high density of economically disadvantaged students. ESEA was a prominent piece of Johnson's great society legislation. While many conservatives opposed the federal government's involvement in setting educational standards and goals, they did not want to deprive states of federal assistance, and thus a debate took place over whether federal aid for education should be limited to high poverty districts or provided on a more general basis ensued. Those in favor of a watered down funding formula ultimately prevailed, ensuring that most districts received federal aid and could spend it as general aid rather than to provide supplemental instruction specifically to disadvantaged students (Graham 1984, cf in McDonnell 2005, 22). As a result of the conflict between the law's stated purpose and its funding formula, the Department of Health, Education and Welfare (DHEW) found that upwards of 15% of Title 1 funds had been blatantly misused. Scholars and civil rights activists voiced their concerns. In response, Congress proceeded to amend the law four times, tightening language until it became irrefutable that Title I funds were meant to supplement state and local funds, not to supplant them, and were specifically to be employed for the betterment of economically disadvantaged students (McDonnell 2005, 23-24). By 1980, most districts receiving Title I funds had developed clear cut, detailed means of accounting for federal dollars and had chosen to use the money to fund pull-out instructional services for struggling students (Jennings 2000, 519).

From the early 1980s forward, the law's goals were far less ambiguous, but political commitment on the national level began to wane. During the second phase of the ESEA's implementation, President Reagan criticized the quality of American schools and the fiscal accountability tied to federal education, leading the federal government to back away from its promise to provide educational aid to the poor (McDonnell 2005 & Jennings 2000). Title I continued to serve many students, but while Reagan sought to eliminate the newly formed Department of Education (ED)¹ and to minimize what he saw as a federal intrusion, funding was held at 1980 levels and thus the number of students serviced through the program declined.

The 1988 reauthorization of the ESEA marked the beginning of the third phase of ESEA's implementation – the first to explicitly call for academic standards and accountability. This began with the first President Bush funding the development of voluntary national academic

¹ The Department of Education was established under President Jimmy Carter from the Office of Education housed in the Department of Health, Education, and Welfare in 1979.

standards for multiple content areas (McDonnell 2005, 27). An emphasis on standards and accountability continued under the Clinton administration. In 1994, Congress passed Goals 2000, which provided grants to states who voluntarily developed content standards. Along with Goals 2000, came the 1994 reauthorization of ESEA which pushed states to ensure that Title 1 students were receiving high quality instruction that aligned with challenging state content standards; states were required to submit accountability plans to the department of education ED. At the same time, federal regulations on how Title 1 funds could be used provided for even more flexibility than had been created by the 1988 reauthorization (McDonnell 2005, 29-30).

Evaluations of ESEA have suggested little real policy change; most find that the status quo has been maintained in terms of quality and inequality (Borman and D'Agostino 1996; *ED* 1997). NCLB, the policy we seek to evaluate, was in part a response to these findings and what policy-makers perceived as prior implementation failures. In 2001 the law clearly specified its goals: increased student achievement and decreased gaps between different student populations. And, it specified a means of achieving them: rigorous standards would provide a framework for increased student achievement, and progress towards these standards could be measured by exams congruent with those standards. But recent research suggests that we continue to “muddle through” (Lindblom, 1959), making little substantive progress (DeBray 2006; McDonnell 2005).

NCLB requires states to create standards for each grade and subject area² and to test students on those standards yearly from grades three through eight and once in high school. The states are allowed to determine the standards themselves, how the standards will be tested, and what constitutes proficient, but they are required by the federal government to set Adequate Yearly Progress (AYP) goals for what percent of students should achieve the proficiency level each year in each subject area. In addition, states must keep raising expectations. For example, states may have only required 30% of students to be proficient in math or reading in 2003, but by the 2013-2014 school year 100% of students must reach this mark.

NCLB requires states to produce and report achievement at multiple levels. Schools, districts, and states, report the percentage of students who fall into each achievement category, focusing on those who are proficient or better.³ Districts and states then report the percentage of

² Standards were initially required for only reading and math but science and social studies standards more recently became requisite.

³ Achievement categories are generally defined as below basic, basic, proficient and advanced with some variance in specific terminology among the states.

schools making AYP, disaggregating by subgroups including special education students, limited English proficient students (LEPs), economically disadvantaged students (as defined by each state), and all racial minorities. A school either makes AYP or not, and is therefore judged to be a success or failure – if even one subgroup does not meet the state's AYP goal for the year, the school fails.⁴ In the next section of the paper we propose several hypotheses relating differences in state capacity and political will to variance in state responses to NCLB.

IV. Hypotheses

What factors explain the continued lack of progress towards federal education goals and increases in student achievement? What can account for the variation in response to NCLB? It is conceivable that states make less than ideal progress and respond differently to NCLB because they have different resources with which to respond – differences in **capacity** – or because they have different levels of political commitment to the law – differences in **will**. We highlight three hypotheses regarding state capacity and will and then identify ways to empirically measure these concepts and assess our hypotheses.

Different states have vastly different education-related capacity. Some states have well-staffed departments of education, stable and or/high state and local funding for schools and time-tested accountability standards and assessments. Others lack one or many of these education-related resources. We believe that high state education-related capacity can work in one of two ways. State education agencies with greater capacity in terms of education-related funding, professional personnel, or existing accountability programs might more easily implement federal standards because they have the resources and institutional strength necessary to do so. Alternatively, agencies with greater capacity may have resources deeply invested in existing programs, not necessarily at odds with the tenets of NCLB, but not readily and easily aligned with them. High-capacity states may in this case be more resistant to federal mandates and more likely to pay them simple lip service than those with low education-related capacity because a total change in state policy or framework (developing new standards, tests, and accountability

⁴ Under the law, schools that fail to meet AYP for more than two consecutive years (for the same subject and subgroup) are labeled as chronically failing and identified for improvement. Schools identified for improvement (SIFIs) are required to implement reforms. With each additional year of failure, these reforms require more drastic action on the part of the school, the district, and the state.

systems) may be quite costly and these high capacity states tend to shoulder more financial burden when it comes to education. We find the former explanation more likely. Therefore:

Hypothesis 1 (Capacity): States with greater capacity in terms of education-related funding, professional personnel, or existing accountability programs will establish higher standards and comply more readily with NCLB.

State political will depends upon the balance and perspective of state-level political interests. Liberal and Conservative states may support federal education policy to differing degrees. Scholars and researchers have long associated conservative citizens and elites with resentment towards federal intervention in state and local policy (Friedman, 1962; Niskanen, 1971) and the history of education policy also suggests that states with concentrated conservative interests might be expected to resist compliance with NCLB (Gill et al., 2001; Witte, 2000). Therefore, we might expect greater divergence from federal expectations in these states. The flip side of this argument implies that liberal citizens and elites may be more concerned with closing achievement gaps, less committed to local and state independence, and therefore less resistant to NCLB's prescriptions. If this is the case, we might expect greater alignment between national and state expectations from liberal states. Based on evidence from existing scholarship in this arena:

Hypothesis 2 (Political Will): States with more conservative and/or Republican populations will strive to preserve greater flexibility for the state and for local agencies.

Union strength is also an important component of state context, one that likely affects both a state's institutional capacity and its political will. Terry Moe has identified unions as the most powerful and active interest group in the country (2005) and has shown that the restrictiveness of union contracts can affect student performance and institutional flexibility (2009). Teachers unions opposed many of the accountability requirements in NCLB as the policy was being designed and have been largely critical of it throughout its implementation (DeBray 2006, Moe 2009). Therefore, we might expect states with strong unions to be more resistant to NCLB:

Hypothesis 3 (Political Will): States with strong unions will encourage state-level actors to resist tenets of the law that place restrictions on their members' freedom to teach or take away from

*their freedom to negotiate freely with local agencies. Therefore these states will establish less rigorous standards and assessments.*⁵

V. Data and Methods

In this section, we discuss our data and estimation procedures. Descriptive statistics of all measures appear in Table 1 at the end of the section.

Dependent Variable

To explore these hypotheses, we gathered two state-level measures of student achievement, one generated by state-designed assessments, the other by the National Assessment of Educational Progress (NAEP). Our dependent variable is a function of the difference between the state's estimate of its students' proficiency and the NAEP tests' estimate of its students' proficiency. Because this second measure is a national, arguably more objective assessment it can be thought of as an honest broker. When a state's test scores are subtracted from this measure the magnitude of the resulting variable is suggestive of the rigor of each state's tests and standards in comparison to all other states.

NCLB requires all states to administer math and reading assessments aligned with state standards in grades 3 through 8 and once in high school. Most states conduct these assessments in the spring of the school year, and must then determine whether or not each school and district has met the requirements of AYP based on the percentage of students who achieve proficiency or better.

NAEP assessments in reading and mathematics are administered every two years, during the spring, to a representative sample of fourth and eighth grade students within each state. Once the tests have been administered to a random sample of schools, the National Center for Education Statistics (NCES) publishes average scale scores as well as the percent of students falling into the below basic, basic, proficient, and advanced categories.⁶

The NAEP tests are the only objective, standardized assessment administered to K-12 students in all 50 states. They provide the only means of comparing student achievement among states. In order to measure how far from the national ideal states stand, we subtract the percent proficient and above on the state test from the percent proficient and above on the NAEP test.

⁵ We realize that states with strong unions often lean more Democratic in national elections. However, when it comes to policies like NCLB, unions' interests align with a more conservative ideology.

⁶ These data can be obtained at: <http://nces.ed.gov/nationsreportcard/naepdata/>

We calculated this gap for fourth and eighth grade math and reading scores using information reported by the states for the 2004-2005 and 2006-2007 school years in combination with NAEP 2005 and 2007 data.⁷

We obtained 161 observations for the 2004-2005 school year, when not all states were in full compliance with NCLB, and 200 observations for the 2006-2007 school year, when all states had their assessment and reporting systems in place. This provides us with approximately four measures of the gap between state assessments and the national assessment for each of two school years. A glance at the data suggest that the rigorous ideals of the NAEP tests stand in stark contrasts to a general lack of rigor across state-designed tests. Across all observations, on average, states' tests identified 35% more students as proficient than did the NAEP tests.

[Figure 1 About Here]

Independent Variables

Our hypotheses point to two broad factors – state **capacity** and **will** - that determine how a state implements NCLB and the degree to which it creates rigorous standards and assessments in accordance with federal mandates.

We first turn our attention to operationalizing state governmental capacity which may constrain states' ability to respond to NCLB. We identify three indicators. The first is the total education revenue per student, in thousands of dollars. Previous research has demonstrated that a state's ability to tax is indicative of governmental capacity (Levi 1989). However, education revenue is an incomplete measure, given the decentralized nature of education funding. We therefore also include the percent of the state's education revenues coming from federal and state governments. A high percentage of revenues coming from the federal government is indicative of a lower state capacity and an economically disadvantaged student population. A high percentage of revenues coming from the state government is indicative of a stronger state (in comparison to local education agencies). This variable matters, particularly when we consider state-district relations. When states finance a greater proportion of education, they have more

⁷ Student performance on state tests as well as demographic data for populations of students tested were gathered from Consolidated Performance Reports submitted to the federal government by each state. These reports provided us with information on the percentage of schools in each state that made or failed to make AYP during the 2004-2005 and 2006-2007 school years, as well as the percentage of students who achieved proficient or above on math and reading tests between grades three and eight. These reports can be accessed online at the US ED's website <http://www.ed.gov/admins/lead/account/consolidated/index.html#csp2005>

authority over local districts and therefore greater ability to strictly interpret and enforce implementation of NCLB.⁸

We employ several indicators of political will in an attempt to account for the ideological leanings and policy mood of each state's public and governmental institutions. To gauge public attitudes, we utilize Berry's measure of state-level citizen ideology. Berry's measure ranges from zero, conservative, to 100, liberal (Berry et al. 1998, updated in 2007). We then employ ideology scores from 2004 and 2006, matched to the appropriate school year. Finally, we capture the macropartisanship of each state using the Republican vote share from the 2004 election. While this second measure of public mood is clearly imperfect and partisan vote shares may be quite variable from election to election, we believe that it accurately captures the partisanship of one state *relative* to another. It also captures the attitudes of citizens in the state at a time when the state's response to NCLB's legislation was being negotiated.

To measure institutional will, we include an indicator for the state's governor's partisanship as well as a measure of the percent of a state's legislature that is Republican, averaged between the upper and lower houses. Both of these variables are matched to the beginning of a school year, 2004 and 2006 for the current analysis.

In addition to our measures of public and institutional will, we measure union strength in two ways. Our first measure, collected by Josh Cowen (2008), is a simple indicator based on state collective bargaining laws. States receive a zero if they have no law, a one if the law provides blanket coverage to public employees, and a 2 if the law explicitly mentions public school employees and/or teachers.⁹ Following Shober et al. (2006), we also chose to use the percent of the public sector that was unionized in the fall of the school years in the dataset. While the measure is imperfect - given that it includes additional, non-education related public employees - we believe that it does capture the relative strength of the teachers unions, and better data are simply unavailable. These statistics are a product of the current population survey and are compiled by Hirsch and MacPherson (2003).

⁸ From Common Core of Data. These data are publicly available at: <http://nces.ed.gov/ccd/>

⁹ We considered additional measures of union strength from Josh Cowen's dissertation (2008) including whether or not the state mandated collective bargaining, whether or not the state is a right-to-work state, how the state deals with strikes, and the overall scope (rated from 1-3) of the state's contract law. However, the measure of coverage seemed to over the best indicator of strength.

Test Factors and Control Variables

As a means of controlling for differences between grade levels and subjects, we include an indicator for whether or not the proficiency gap is based on a 4th grade assessment (as opposed to an 8th grade assessment) and for whether the assessment measures reading or math proficiency. We also include an indicator for the year in which the tests were taken. In addition, we include the number of students tested by the state, the percent of tested students who belong to a minority group, and whether or not the state is in the South. Finally, because the states of New Hampshire, Vermont, Rhode Island, and Maine entered into a consortium known as the New England Common Assessments Program (NECAP) and these states collaborate to administer the same test, we add an indicator for members. The gap between state tests and NAEP should be the same for these states. Table 1 provides the summary statistics for all variables used in the following analyses.

[Table 1 About Here]

Methods

In order to test our hypotheses concerning capacity and will, we estimate several OLS regression models. Due to multicollinearity among some measures of political will, we present four models including different combinations of these variables in Table 2 below. Because there are multiple observations of the gap between achievement on state tests and achievement on the NAEP tests for each state, we cluster standard errors by state. Model one includes Berry's measure of citizen ideology and the two measures of union strength. Model two includes these measures in addition to the partisanship of the state legislature and the governor. Models three and four replicate models two and one respectively but substitute Republican vote share in the 2004 presidential election for Berry's citizen ideology measure.

[Table 2 About Here]

VI. Results

Our hypotheses are supported to some degree in each of these models. All models are significant, and clearly we are explaining a substantial portion of the variance between the states

with our measures of capacity and will. However, the degree of multicollinearity among the variables in the model creates somewhat inconsistent results.

Only one of our measures for capacity achieved statistical significance in these models. The total amount of educational revenue per student in the state was significant at the .05 level in models one and two and achieved borderline statistical significance in models three and four. For each additional thousand dollars per student, the gap between the state and NAEP estimates of achievement should decrease by between 1.3 and 1.4 percent, supporting our hypothesis that states with higher levels of capacity would be more able to more fully comply with the expectations set forth in NCLB. Neither the percent of education revenues coming from the federal and state governments achieve statistical significant in any model.

Several Measures of a state's political will achieve statistical significance (or borderline significance) and lend support to our hypothesis that more conservative states are less likely to adopt rigorous assessments. Berry's measure of citizen ideology is borderline significant in our first model. More conservative states appear to adopt assessments less closely aligned with national expectations. The same effect is found in model three when we employ Republican vote share from 2004 instead of Berry's measure. States with a larger Republican vote share have larger gaps between the results their state assessments and their NAEP assessments. The effects of both of these variables are washed out when institutional measures of ideology are also included in models two and four, and no measures of political will achieve significance. We attribute this to multicollinearity.

Our third hypothesis, regarding unionization is not strongly supported. The coefficient for the percent of the public sector that is unionized is signed in the expected direction – states with a higher rate of public sector unionization exhibit larger gaps – but the coefficient does not achieve significance in any model. The coefficient for the state collective bargaining coverage is also positively signed, but only achieves borderline significance in models three and four.

Finally, there are some interesting results among our test factors that suggest states may be “gaming” the law. The gap between state and NAEP test scores is significantly and substantially larger for reading assessments than math assessments (8 points larger on average). We speculate, but cannot conclude, that assessments and expectations for reading are more easily watered down than assessments for math. The gap between state and national test scores is also significantly larger in the 2006-2007 school year than in the 2004-2005 school year. We see two

possible explanations for this increase. First, it is possible that students and teachers are learning how to better prepare for, take, teach and administer states tests. We think it unlikely that these improved test-taking abilities would not be reflected in the NAEP test, but acknowledge the that it is possible given legitimate curricular variation. Second, it is possible that states in 2006-2007, having considerably more experience with NCLB and test administration than they did in 2004-2005, adjusted either their assessments or their proficiency standards in order to reduce the number of schools identified as failing to meet AYP. Supporting this second explanation is the fact that numerous states submitted – and continue to submit – amendments to their NCLB accountability plans after the 2004-2005 school year.¹⁰ Data from the 2008-2009 school year and the 2009 NAEP assessments might adjudicate between these two explanations. If students and teachers are adapting to standardized tests, the increase should level off as improvements on tests due to increased test-taking skills do have an upper limit. If states are adapting their responses to NCLB in order to avoid the consequences of poor performance, then we might expect these gains to continue.

VII. Conclusions

We have tested our “hybrid” implementation theory on one major policy (NCLB) in one policy arena (education) and demonstrated that federal policy implementation is limited by states’ capacity and political will. With respect to NCLB, faithful implementation of the law’s policies may not be possible in some states where resources are limited and not palatable in others where political sentiments resist federal intrusion. A substantial portion of the variation in these implementation outcomes can be explained by state-level factors, a level of analysis often neglected by previous implementation studies. Education is an area over which state governments are constitutionally sovereign and thus any federal policy is potentially subject to distortion.

Future research might utilize state level capacity and will to structure investigations of education policy or the implementation processes in other policy arenas. Those who attend to different policy areas may find more or less variation at the state level than we find here, but this work suggests that this middle level can no longer be ignored by implementation scholars, and

¹⁰ These requests can be viewed in their entirety on the USED website at:
<http://www.ed.gov/admins/lead/account/letters/index.html>

that state capacity and will provide a useful means of organizing the many factors that can affect policy implementation.

Finally, though our focus has been on how federal policy is implemented at the state level, capacity and will may prove useful in examining variation in policy implementation at the local level as well. When one set of actors – at the national, state, or perhaps even district level - asks a removed set of actors at a lower level of government to implement a centrally-designed policy, the degree to which resources are available and ideals are shared may influence the success of this policy directive.

References

- Bardach, Eugene. 1977. The Implementation Game: What Happens After a Bill Becomes a Law. The MIT Press: Cambridge, Massachusetts, and London, England.
- Borman, Geoffrey and D'Agostino, Jerome. 1996. "Title 1 and Student Achievement: A Meta-Analysis of Federal Evaluation Results." *Educational Evaluation and Policy Analysis* Vol. 18, No. 4: 309-326.
- Berman, Paul. 1978. "The Study of Macro- and Micro- Implementation," *Public Policy* 26:2: 157-84.
- Berry, William D., Ringquist, Evan J., Fording, Richard C., and Hanson, Russell L. "Measuring Citizen and Government Ideology in the American States 1960-93." *American Journal of Political Science* Vol. 42, No. 1: 327-348.
- Bumiller, E. (2002, January 9). Focusing on home front, Bush signs education bill. *The New York Times*, p. A16.
- Cowen, Josh. 2008. *Bargain-Based Schooling: Teacher Unions and American Education*. Doctoral Thesis.
- Debray, Elizabeth. 2006. *Politics, Ideology, & Education: Federal Policy during the Clinton and Bush Administrations*. New York, NY: Teachers College Press.
- Dillon, Erin, and Rotherman, Andrew J. 2007. "States' Evidence: What it Means to Make 'Adequate Yearly Progress' Under NCLB," Education Sector.
- Elmore, Richard F. 1985. "Forward and Backward Mapping." In K. Hanf and T. Toonen, eds, *Policy Implementation in Federal and Unitary Systems*, Dordrecht: Martinus Nijhoff.
- Friedman, Milton. 1962. *Capitalism and Freedom*. Chicago, IL: University of Chicago Press.
- Gill, Brian P., P. Michael Timpane, Karen E. Ross, and Dominic J. Brewer. 2001. *Rhetoric versus Reality: What We Know and What We Need to Know About Vouchers and Charter Schools*. Santa Monica, CA: RAND.
- Goggin, Malcolm L., Bowman, Ann, Lester, James P., and O'Toole, Lawrence I. Jr., 1990. Implementation Theory and Practice: Toward a Third Generation. Glenview, Ill: Scott Foresman.
- Graham, Hugh Davis. 1984. *The Uncertain Triumph: Federal Education Policy in the Kennedy and Johnson Years*. Chapel Hill: University of North Carolina Press.
- Barry T. Hirsch and David A. Macpherson, "Union Membership and Coverage Database from the Current Population Survey: Note," *Industrial and Labor Relations Review*, Vol. 56,

- No. 2, January 2003, pp. 349-54. [Describes the database and web site www.unionstats.com.]
- Hjern, Benny. 1982. "Implementation Research-The Link Gone Missing." *Journal of Public Policy* 2:3:301-08.
- Jennings, John (Jack) F. 2000. "Title 1: Its Legislative History and Its Promise," *Phi Delta Kappan* Vol. 81, No. 7, 516-522.
- Levi, Margaret. 1989. Of Rule and Revenue. Berkeley and Los Angeles, CA. University of California Press.
- Lipsky, M. 1980. Street Level Bureaucrats. New York: Russell Sage.
- Manna, Paul. 2006. "Control, Persuasion, and Educational Accountability: Implementing the No Child Left Behind Act," *Educational Policy* 20(3): 471-494
- Matland, Richard E., "Synthesizing the Implementation Literature: The Ambiguity-Conflict Model of Policy Implementation," *Journal of Public Administration Research and Theory: J-PART*, Vol. 5, No. 2: 145-174.
- May, Peter J. 1993. "Mandate Design and Implementation: Enhancing Implementation Efforts and Shaping Regulatory Styles" *Journal of Policy Analysis and Management*, Vol. 12, No. 4: 634- 663.
- Mazmanian, Daniel, and Sabatier, Paul A. 1983. Implementation and Public Policy. Glenview, Ill. Scott Foresman.
- McDonnell, Lorraine. 2005. "No Child Left Behind and the Federal Role in Education: Evolution or Revolution?" *Peabody Journal of Education* Vol. 80, No. 2: 19-38.
- McLaughlin, Milbrey Wallin. 1976. "Implementation of ESEA Title 1: A Problem of Compliance," *Teachers College Record* Vol. 77, No. 3: 397-415.
- McLaughlin. 1987. "Learning From Experience: Lessons from Policy Implementation." *Educational Evaluation and Policy Analysis* 9:2:17-78.
- Milbank, D. (2002, January 9). With fanfare, Bush signs education bill. *The Washington Post*, p. A3.
- Moe, Terry. 2005. "Teacher Unions and School Board Elections," in Howell ed. *Besieged: School Boards and the Future of Education Politics*, 254-287.
- Moe, Terry. 2009. "Collective Bargaining and the Performance of the Public Schools," *American Journal of Political Science* Vol. 53, No. 1: 156-174.

- Murphy, Jerome. 1973. "The Education Bureaucracies Implement Novel Policy: the Politics of Title 1 of ESEA," in Sindler ed. *Policy and Politics in America*, ed. Boston, Little, Brown.
- National Conference of State Legislatures, "No Child Left Behind Quick Facts 2005," April 2005 and www.nea.org/lawsuitstateres.html cf in Mantel, Barbara. 2005. "No Child Left Behind: Is the Law Improving Student Performance?" *CQ Researcher*: Vol. 15 No. 20: 471-91
- Niskanen, William A. 1971. *Bureaucracy and Representative Government*. Chicago, IL: Aldine Atherton.
- Owens, Ann and Sunderman, Gail, L. 2006. "School Accountability under NCLB: Aid or Obstacle for Measuring Racial Equity?" The Civil Rights Project at Harvard University.
- Palumbo, Dennis J., Maynard-Moody, Steven, and Wright, Paula. 1984. "Measuring Degrees of Successful Implementation." *Evaluation Review*, Vol. 8, No. 1: 45-74.
- Phi Delta Kappa and Gallup Poll. 2009. "Polls of the Public's Attitudes toward the Public Schools." Available at: <<http://www.pdkintl.org/kappan/kpollpdf.htm>>.
- Pressman, Jeffrey L. and Wildavsky, Aaron. 1973. *Implementation*. Berkeley, CA: University of California Press.
- Pulzl, Helga, and Treib, Oliver. 2006. "Policy Implementation" in Frank Fischer/Gerald J. Miller/Mara S. Sidney (eds.): *Handbook of Public Policy Analysis: Theory, Politics, and Methods*. Boca Raton, FL: CRC Press/ Taylor & Francis, 2006, 89-107.
- Puma, Michael with Nancy Karweit, Cristofer Price, Anne Ricciuti, William Thompson, and Michael Vaden-Klernan. 1997. *Prospects: Final Report on Student Outcomes*. Cambridge, MA: Abt Associates.
- Sabatier, Paul A. 1986. "Top-Down and Bottom-Up Approaches to Implementation Research: A Critical Analysis and Suggested Synthesis." *Journal of Public Policy* 6:1:21-48.
- Schneider, Anne and Ingram, Helen. 1997. *Policy Design for Democracy*. Lawrence, KS: University of Kansas Press.
- Scholz, John T. and Wang, Cheng-Lung. 2006. "Cooptation or Transformation? Local Policy Networks and Federal Regulatory Enforcement," *American Journal of Political Science*, Vol. 50, No. 1: 81-97.
- Shober, Arnold F., Manna, Paul, and Witte, John F. "Flexibility Meets Accountability: State Charter School Laws and Their Influence on the Formation of Charter Schools in the United States." *Policy Studies Journal* 34(4): 563-587.

- Squire, Peverill. 2007. "Measuring State Legislative Professionalism: The Squire Index Revisited," *State Politics and Policy Quarterly*, Vol. 7, No. 2: 211-227.
- U.S. Department of Education. 2009. "10 Facts about K-12 Education Spending." Available at: <<http://www.ed.gov/about/overview/fed/10facts/index.html>>.
- Van Meter, Donald S., and Van Hom, Carl E. 1975. "The Policy Implementation Process: A Conceptual Framework." *Administration and Society* 6:4:445-88.
- Weatherley, Richard, and Lipsky, Michael. 1977. "Street-Level Bureaucrats and Institutional Innovation: Implementing Special Education Reform." *Harvard Educational Review* 47: 170-96.
- Witte, John F. 2000. *The Market Approach to Education: An Analysis of America's First Voucher Program*. Princeton, NJ: Princeton University Press.

Figure 1. Percent of Students Deemed Proficient on the State Test versus the NAEP Test

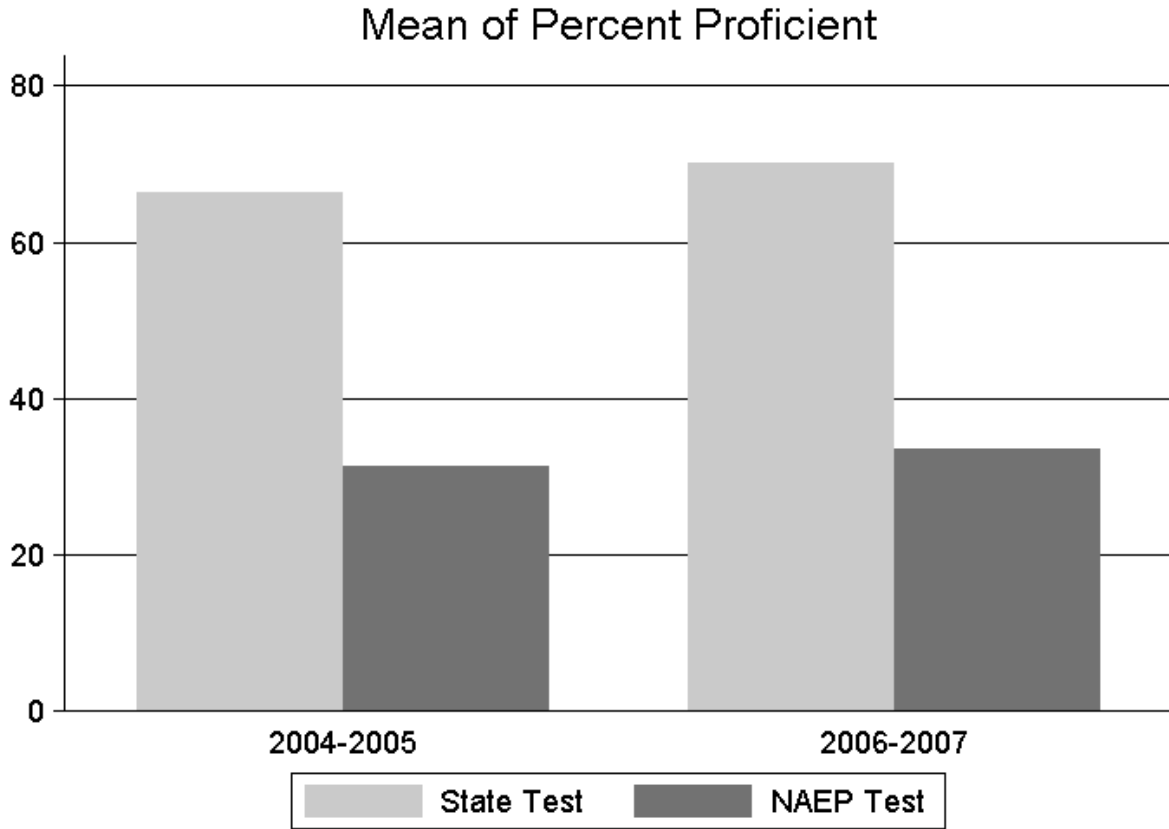


Table 1. Summary Statistics	361 Observations, 50 States			
	Mean	SD	Min	Max
Proficient or Above Gap % on State Test - % on NAEP	35.994	15.936	-32.675	71.405
Capacity				
Revenue Per Student (1000s)	10.693	2.526	6.408	17.707
% Revenue Federal Government	9.850	3.214	4.363	18.073
% Revenue State Government	49.431	12.199	26.884	89.698
Political Will				
Citizen Ideology (Liberal=100)	51.496	15.775	19.232	93.946
Republican Governor	0.557	0.497	0	1
% Republican in State Legislature*	50.653	14.511	14.688	78.571
Republican Vote Share 2004	53.906	8.365	37	72
% Public Sector Unionized	31.884	17.465	6.8	68.9
Collective Bargaining Law	0.803	0.701	0	2
Test Factors				
Grade 4 (1) versus Grade 8 (0)	0.493	0.501	0	1
Subject: Reading (1) & Math (0)	0.496	0.501	0	1
School Year: 06-07 (1) & 04-05 (0)	0.554	0.498	0	1
Controls				
Percent Minority	33.714	18.234	4.617	85.347
NECAP	0.555	0.229	0	1
South	0.316	0.465	0	1
N Tested (1000s)	493.622	640.471	18.933	3872.290

* Nebraska is excluded from the summary data for % Republican in State Legislature. There are 353 observations for that variable (within 49 states) Summary data presented are not adjusted for state clusters.

Table 2. Models of the Gap Between State and National Measures of Student Achievement

	Model 1	Model 2	Model 3	Model 4
Capacity				
Revenue Per Student (1000s)	-1.153† (0.588)	-1.207* (0.600)	-1.431* (0.617)	-1.402* (0.674)
% Revenue Federal Government	-0.144 (0.450)	-0.035 (0.503)	-0.422 (0.450)	-0.367 (0.492)
% Revenue State Government	-0.116 (0.118)	-0.086 (0.116)	-0.148 (0.119)	-0.137 (0.111)
Political Will				
Citizen Ideology (Liberal=100)	-0.243† (0.127)	-0.166 (0.123)	-	-
Republican Governor	-	-3.335 (3.040)	-	-4.667 (3.178)
% Republican in State Legislature	-	0.041 (0.189)	-	-0.050 (0.179)
Republican Vote Share 2004	-	-	0.746* (0.284)	0.768* (0.309)
% Public Sector Unionized	0.085 (0.109)	0.081 (0.113)	0.218 (0.133)	0.202 (0.140)
Collective Bargaining Coverage	4.448† (2.257)	4.333† (2.261)	3.866† (2.160)	4.085† (2.268)
Test Factors				
Grade 4 (1) versus Grade 8 (0)	2.526* (1.134)	2.639* (1.121)	2.532* (1.128)	2.627* (1.125)
Subject: Reading (1) & Math (0)	8.434** (1.066)	8.571** (1.077)	8.405** (1.067)	8.550** (1.079)
School Year: 06-07 (1) & 04-05 (0)	4.196* (1.672)	4.373* (1.696)	4.266* (1.739)	4.376* (1.827)
Controls				
Percent Minority	-0.002 (0.117)	0.011 (0.127)	0.074 (0.125)	0.079 (0.121)
NECAP	-10.678* (5.230)	-11.186* (5.456)	-7.981† (5.111)	-8.042† (5.224)
South	9.287† (4.624)	10.096† (5.801)	10.178* (4.463)	9.510† (5.266)
N Tested (1000s)	-0.002 (0.002)	-0.002 (0.003)	-0.002 (0.002)	-0.001 (0.003)
Constant	52.712** (13.661)	45.252** (21.974)	0.603 (24.199)	2.578 (25.680)
N	361	353	361	353
Clusters	50	49	50	49
R²	.321	.326	.344	.354

†p<.10, *p<.05, **p<.01

Nebraska is excluded from models that include the % Republican in the State Legislature.

