

Predicting the Results of School Finance Adequacy Lawsuits

Zachary Lynn

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy
under the Executive Committee of the Graduate School of
Arts and Sciences

COLUMBIA UNIVERSITY
2011

© 2011
Zachary Lynn
All Rights Reserved

ABSTRACT

Predicting the Results of School Finance Adequacy Lawsuits

Zachary Lynn

All but one state constitution contains an adequacy clause, which holds the state to some minimum standard of education that must be provided. Typically, when citizens allege that a state is violating its constitutional duty to provide an adequate education, the dispute winds up in the courts. This study attempted to identify what, if any, political or institutional factors led to plaintiffs' victory in an adequacy lawsuit, and also examined whether there was any strategy behind the order in which states experienced an adequacy lawsuit. Qualitative Comparative Analysis, a method designed to allow for structured comparison in small- and medium-n studies by identifying necessary conditions, was employed to test political and institutional hypotheses. Analysis revealed that there were no conditions (or combinations of conditions) that were necessary and sufficient for plaintiffs' victory across cases. The implications for this finding are explored, with attention paid to the possibility that either methodological or theoretical issues account for the finding.

TABLE OF CONTENTS

CHAPTER 1 - INTRODUCTION	1
Historical, Political and Educational Background	3
Theories of equity and adequacy	8
Methods of Determining Adequacy.....	11
Competing Explanations of Judicial Decisions in Adequacy Lawsuits	17
Institutions.....	18
Political and Institutional Theories Concerning Judicial Behavior	23
Judicial Decision-Making Primarily as the Product of Law.....	26
Other Possible Rival Hypotheses.....	29
Significance of the Present Study	32
CHAPTER 2 – LITERATURE REVIEW	34
Adequacy Lawsuits and ESEA	34
Studying Judicial Decision-Making.....	40
Courts and Social Change.....	40
Differences Between State Supreme Courts and the American Supreme Court	45
The Influence of Other High Courts on State Supreme Court Decision-Making.....	49
State Supreme Courts and the Separation of Powers.....	55
The “Electoral Connection” and State Supreme Courts	59
Rankings of State Supreme Courts	65
Studying Adequacy.....	66
Attacks on the Adequacy Concept.....	66
Previous Attempts to Study School Finance Litigation.....	70

Can Judges Effectively Evaluate Cost Studies?.....	78
Interpreting Possible Results.....	81
CHAPTER 3 – METHODS	85
Qualitative Comparative Analysis as a Research Method.....	85
Temporal Relationships and QCA.....	95
Why Other Possible Methods Were Not Employed	97
Procedures For Using QCA	98
Variables Employed In This Study.....	111
The Dependent Variable	111
A Brief Review of Theory	112
Selecting Independent Variables and the Problem of Omitted Variables.....	115
Conceptualizing and Coding the Independent Variables.....	117
Data Analysis Program Employed.....	123
Possible Findings and Interpretations	123
CHAPTER 4 – RESULTS	127
Brief Summary of Results.....	127
Terminology For Interpreting QCA Results	128
Detailed Results of Specific Models.....	132
Model 1	132
Model 1A	133
Model 2	134
Are the Best Cases Filed First?.....	135
Summary.....	136

CHAPTER 5 – DISCUSSION.....	139
POSSIBLE EXPLANATIONS AND IMPLICATIONS.....	139
There Are No Necessary and/or Sufficient Conditions	139
The Underlying Theory Needs to be Revised.....	140
Possible Methodological Issues	142
IMPLICATIONS FOR THE STUDY OF JUDICIAL POLITICS.....	144
GUIDANCE FOR ADEQUACY ADVOCATES	145
ADDITIONAL SPECULATION ON ALTERNATIVE THEORIES AND	
EXPLANATIONS	146
Are Conceptions of Adequacy State-Specific?.....	146
The Role of the Media in Framing Adequacy Issues.....	147
Do Some Institutions Benefit From Inadequacy?.....	148
Do Lawyers Play a More Important Role Than Political Scientists Give Them Credit	
For?	149
DIRECTIONS FOR FUTURE RESEARCH.....	150
REFERENCES	152
APPENDIX 1	161
APPENDIX 2.....	163
APPENDIX 3.....	165

Acknowledgements

The writing of a dissertation is a long and sometimes challenging process, and I'm grateful to a number of people who helped me along the way.

My committee deserves a hearty thank you. This dissertation would not have been possible without the fine advice of my sponsor, Jeff Henig, who kept driving me to sharpen my thinking. Thanks as well to Michael Rebell and Justin Phillips, both excellent teachers who provided valuable comments as early as the proposal stage. Thanks as well to Carolyn Riehl, who provided valuable comments, and to Jay Heubert, who chaired my committee and gave me some new perspectives to think about.

Several students in the Politics and Education Program provided indispensable feedback when this work was in its infancy. Thanks especially to Jonah Liebert and Stephen Q. Cornman, whose comments on earlier drafts are incorporated here.

Thanks as well to Valerie Reidy, Rosemarie Jahoda, David Colchamiro, and faculty and students at The Bronx High School of Science, who gave me the privilege of teaching New York City's finest scholars. The experience has been rewarding in its own right, and has doubtlessly informed my more academic work.

Thanks as well to the various friends and colleagues, who provided encouragement, and were willing to listen to my anxieties and venting about the possibility of not finishing. Particular gluttons for punishment were Michael Blumer, Dan Galimidi, Norm Rubin, Sophia Sapozhnikov, Shawn Tolidano, and Jon Waisbrot.

Lastly, I need to thank my brother, Josh, for his unwavering support, and my mother, Adele, who has always been my biggest fan. For five years, she made sure my tuition bills were paid and my Starbucks card was loaded, without letting me worry about

where the money was coming from. She gave me the freedom to pursue a doctorate, and I hope this dissertation makes her proud.

All these people have only made this dissertation better. Any remaining imperfections are, of course, solely my responsibility.

CHAPTER 1 - INTRODUCTION

A school finance adequacy lawsuit alleges, generally speaking, that a state is not fulfilling its constitutional mandate to provide an adequate education to the student within its border. The source of this mandate, according to the plaintiffs, is a clause in the state constitution¹ that specifies a state's obligation to provide an adequate education. The specific constitutional language, though, is fairly vague, relying on phrases like "thorough and efficient" or "sound basic education."

The current study attempts to identify institutional and political factors that are necessary and/or sufficient for plaintiffs to win an adequacy lawsuit. Such a study is important for a number of reasons. First, the adequacy clause is sufficiently ambiguous that "politics" must necessarily fill in the details and turn that constitutional mandate into an operational policy. Second, different courts (and indeed, legislatures) have come to different conclusions about the substantive meaning of adequacy, despite similar language. Additionally, plaintiffs have lost in some states where the objective evidence and rulings in similar states suggest they ought to have won. The reverse is true as well.

Perhaps most importantly, advocates could benefit from the guidance that would result from an empirical examination of the results of adequacy lawsuits. If there are political or institutional factors that make victory more likely, advocates would do well to ensure that they have maximized their chance of victory, to the extent that political and institutional conditions are open to change. And given the multiple forums in which advocates pursue adequacy reforms (lobbying legislators or attempting to influence public opinion, for example), advocates could use information on how those strategies interact, which the current study might also provide. While previous studies have

¹ All states except Mississippi have an adequacy clause.

examined the effect of one or a few legal, political or institutional conditions on the results of adequacy lawsuits, this study represents a first attempt to look at a number of conditions across all states, representing a middle ground between standard quantitative methods and case-study research which can, by definition, examine only a few cases in depth.

The current study attempts to provide a guide to the institutional and political conditions that might be necessary and/or sufficient for plaintiffs' victory in an adequacy lawsuit, although, as later chapters make clear, the present focus on particular conditions that may correlate with outcomes in adequacy cases does not mean that evidence and legal reasoning play no role. This study does explore two rival hypotheses. First, there is the possibility that judges decide based on the unique interaction of numerous factors in each case, and that, as a result, it is not possible to construct a model of necessary and sufficient institutional and political conditions for plaintiffs' victory. This rival hypothesis would be difficult to test directly, but would be partially supported by evidence that no institutional factor was necessary and/or sufficient. Second, there is the possibility that political or institutional conditions will have no effect on the decision because of strategic *filing* of cases. It is possible that the most winnable cases, representing the most egregious inadequacies, were filed first, and that later cases are factually weaker, which would lead to no consistent findings in the current study. This rival hypothesis, though, can be tested empirically.

While a primary goal of this study is the examination of adequacy lawsuits on their merits, it is possible that the results might also make a contribution to the broader literature on state supreme courts as institutions. Clearly, if theory-building were the

only goal, it would be desirable to examine other types of cases in addition to adequacy lawsuits. However, adequacy lawsuits might represent a fertile initial testing ground for new theories about state supreme courts for two reasons. First, the federal courts have removed themselves from school finance litigation, eliminating issues of precedent. Adequacy, then, offers a chance to see the full range of influences on a relatively unconstrained court.

Second, most studies of judicial decision-making (particularly Supreme Court decision-making) focus on “hot-button” issues, such as gay marriage, gun control, or abortion, for example. These issues, while relevant, timely, and worth studying in their own right, are at the heart of today’s Culture Wars, and are likely to arouse strong passions, and even religious sentiment, among judges, lawyers, and the public. People can be passionately “for” or “against” gay marriage in a way that it is nearly impossible for one to feel about adequate education for poor children. It is precisely for this reason that adequacy might make for a *better* test of institutional theories of court decision-making than those other issues. If the point of theories is to test propositions that can be generalized to a wide variety of specific instances, then it seems a more logical approach to stay away from Culture Wars issues when attempting to theorize about courts in general. Most of the decisions that courts make, after all, look more like their work in adequacy decisions, and less like the exceptionally polarizing culture wars issues.

Historical, Political and Educational Background

One might question the necessity of a study dedicated to adequacy lawsuits. In listening to the press reports about education, one has heard far more about accountability, testing, and No Child Left Behind (NCLB) in recent years than one has

about adequacy. However, adequacy lawsuits play a central—if unnoticed—role in current educational policy debates. NCLB mandates that schools make adequate yearly progress, both overall and for a number of subgroups. Schools, districts, and states that do not make adequate yearly progress are subject to sanctions. Out of deference to basic principles of federalism, however, the federal government mandates *the existence of* standards, accountability, and sanctions, but leaves it up to individual states to develop their own standards and accountability mechanisms for measuring progress towards those standards. Under NCLB, then, states have an incentive to set their standards very low, so that their schools face no sanctions and, incidentally, they look good in the inevitable comparisons with other states. In the context of NCLB-mandated federal sanctions, adequacy lawsuits function as a type of “fire-alarm” regulation as McCubbins and Schwartz (1984) define the term, in which the threat of adequacy lawsuits will prevent legislators from setting their standards too low. States face other incentives with regards to standards, however. States that invest in providing an adequate education are likely to see some return on investment in the form of economic growth. And, when sanctions are not the primary driver of federal education policy (as in President Obama’s “Race to the Top” grant program), one could imagine a scenario where the incentives provided by federal education policy and the goals of adequacy lawsuits are aligned.

Additionally, adequacy lawsuits link closely with another current educational policy objective as well, the narrowing the black-white achievement gap. The evidence presented in most lawsuits tends to focus on the education received by poor children and nonwhite children in urban areas. While rulings, by definition, find the funding formula unconstitutional for the entire state, as all districts are governed by the formula, the

evidence of inadequacy focuses generally on minority children. Notably, there are no lawsuits based mainly on the claim that wealthy, white or gifted students were receiving an inadequate education. Indeed, providing adequacy usually entails some redistribution of resources away from more affluent school districts. In this way, adequacy lawsuits can be a tool to narrow the black-white achievement gap, to the extent that they force states to redistribute money in ways that remedy the inequities of America's property tax-based school funding scheme that contributes to the achievement gap in the first place.

The premise that adequacy lawsuits could narrow the black-white achievement gap, or have any effect on achievement, comes from the assumption that money matters to school performance. While many policymakers implicitly assume a straightforward relationship between money and student achievement, the relationship has in fact been a source of considerable debate. Hanushek (1989) is one proponent of the viewpoint that spending more money on education will not generally improve student outcomes. He conducted a meta-analysis of the results of sixty-five studies that measured the effect of per-pupil expenditure on student outcomes, generally focusing on one type of student or one type of outcome, and concluded that there was no discernable pattern between per pupil expenditures and student outcomes. Specifically, only 16 of the 65 studies he used produced significant results, and three of those 16 results were significant in the wrong direction, indicating an inverse relationship between per pupil expenditure and student outcomes.

Hanushek's (1989) method was uncomplicated, as he merely counted and tabulated the results of previous studies and did not attempt an actual re-analysis of any data. Furthermore, he dismisses any worries about data quality or measurement

inconsistencies having contributed to the results he found by pointing out that he observed a pattern across several studies, so that while there might have been measurement or analysis problems in one study, it would be highly unlikely that several studies repeated this. Additionally, Hanushek offers an explanation for not finding an effect of per pupil expenditure on student achievement that primarily has to do with the way teachers are paid. Teacher quality is of paramount importance for Hanushek (although it was not directly measured in the same study), and he points out that effective teachers are not systematically paid any more than ineffective ones, as years of service is the general basis for teacher pay. Therefore, there is no reason to expect a systematic relationship between per-pupil expenditure and student performance, since the majority of expenditures are for teacher salaries. Hanushek's argument, then, is plausible if one accepts his methodology. However, Hedges et al (1994) found his methods problematic.

Hedges et al (1994) found two major flaws with Hanushek's (1994) methods. First, by merely tabulating the results of existing studies by sign of the coefficient and significant-or-not (commonly known as "vote counting"), Hanushek ignored the more technically sound methods of meta-analysis available. Vote counting, after all, cannot provide an estimate of the *magnitude* of an effect, only whether it is significant or not. Hedges et al. prefer to use "combined significance tests" and "effects magnitude analysis," techniques of meta-analysis that take into account the magnitude of an effect as well as its direction. Second, Hedges et al. attack Hanushek's formulation of his null hypothesis. If the null hypothesis is that per-pupil expenditure has no effect on student outcomes, there should be, Hedges et al. argue, an equal number of positive and negative effects, and only five percent of the studies should by chance, report significant results.

Since that was not the case in Hanushek's study, then there is evidence of a pattern, even if Hanushek's methods could not adequately describe it. Indeed, after criticizing Hanushek's methods, Hedges et al. reanalyze that data using their more sophisticated methods and determine that money does indeed matter to school performance. Hanushek (1994) and Hedges et al. (1994a) later converged on the opinion that money did matter *somehow*, although the relationship was complicated, and perhaps money mattered more in certain circumstances and when spent in certain ways. It seems that the debate about whether money matters now incorporates insights from economics: Hanushek (2003) now argues that what matters is not merely having resources or the budget, but also, having incentives for teachers and other school personnel to raise their performance.

The question of whether money matters is relevant to establish the connection between an adequacy lawsuit and better education. After all, advocates do not pursue adequate funding out of some abstract notion of fairness or distributive justice—or at least not *solely* because of such a notion. Advocates also believe that adequate funding is associated with better outcomes for the children on whose behalf adequacy lawsuits are filed. The discussion of whether money matters, while highlighting the enormous positive effect that winning an adequacy lawsuit can have, also brings into sharper focus some of the difficulties that judges face in deciding adequacy cases. If academics trained in education finance and research methodology, cannot come to similar estimates of the extent to which money contributes to an adequate education and the types of things on which money should be spent, then the decision-making process for judges becomes much harder and it becomes easier for judges to declare adequacy a non-justiciable question, although a number of judges have rejected Hanushek's claims on common

sense grounds. A brief survey of the history of school finance litigation will demonstrate the different ways that judges have dealt with school finance cases, and the different evidence that they have considered.

One of the most basic principles underlying the American constitution and federal-state relations is the reserved powers clause of the 10th Amendment. Simply put, it sets out the principle that the Constitution is essentially a list of things that the federal government can do, and that things that are not explicitly mentioned are “reserved” to the states or the people. The federal constitution says nothing about education; so therefore, this is reserved to the states, where it is a part of every state constitution. The federal courts, therefore, have been hesitant to enter the realm of education policy-making, unless there is some additional issue at stake that would constitute a federal issue. *Brown v. Board* is one example of the Supreme Court getting involved in education. The ruling, however, concerned mainly the need for states to ensure that schools were not segregated by race. What it did *not* do was delve into *how* education ought to be conducted, and how educational resources ought to be allocated. The substantive questions about how to educate were left to the states.

Theories of equity and adequacy

The issue of equity in school finance came to the Supreme Court in *Rodriguez v. San Antonio Independent School District* (1973). The plaintiffs alleged that the school funding system in Texas violated the equal protection clause of the fourteenth amendment by not providing equal per-pupil funding across all districts. Specifically, some school districts in the San Antonio metropolitan area could raise less money even

with higher tax rates than neighboring districts because they were relatively property-poor and funding for education came primarily from property taxes.

Justice Powell, writing for a 5-4 majority, ruled that Texas' system did not violate the equal protection clause and was, therefore, constitutional. Justice Powell gave three primary reasons for his finding. First, the poor were not a protected class for the purposes of equal protection analysis, and furthermore, even if the court were to consider designating the poor a protected class, there was not necessarily a correlation between *property-poor districts* and *poor children*, so there was no detectable relationship that "the poor" were being harmed. Although Justice Powell did not use the term, he was referring to what social scientists would call a levels-of-analysis problem in the plaintiffs' argument. Second, education is not a fundamental interest under the federal constitution, and the claim of inequitable educational resources should not then get the heightened judicial scrutiny that is granted to fundamental interests. Third, local control is a rational justification for the current system. Since the poor are not a protected class and education is not a fundamental interest, only a rational justification (as opposed to some higher bar, like a compelling interest) is required (*Rodriguez v. San Antonio Independent School District*, 1973).

The *Rodriguez* decision seemed to close the door on education spending cases in the federal courts, although the decision left state courts the opportunity to rule on equity or adequacy issues pursuant to their interpretation of equal protection. For example, state courts could find a fundamental right to education in their state equal protection clause, although the Supreme Court had failed to find such a right in the federal constitution.

Serrano v. Priest (1976), in California, was one of the first cases to bring an equity argument to a state supreme court. Although the plaintiffs won, the results were disastrous, and advocates achieved virtually none of their goals. The court ruled that per-pupil spending disparities were unconstitutional, and restricted such disparities to be no greater than \$100 between the highest- and lowest-spending districts statewide. Prior to *Serrano*, California had been one of the highest spending states on education, and advocates hoped that a victory would lead to rising spending statewide, so that everyone would catch up with the rich districts. However, just the opposite occurred. The case led to a leveling down, not a leveling up, of spending by school districts in California. The passage of Proposition 13 in 1978 severely limited the amount of money that could be raised through property taxes, and although the state increased funding for education, the total was still relatively low. Thus, in complying with the *Serrano* decision, lower per-pupil spending became the norm. Per-pupil spending was now more equal, but it was equally low. In terms of nationwide comparisons, California went from being one of the highest in per-pupil spending to one of the lowest in just a few years (Thro, 1989). Adding insult to injury, the richer districts managed to circumvent the spending limit anyway, by lobbying for special grants of categorical programs or supplementing the “official” district spending with money from nonprofit foundations, set up by residents in wealthy districts for the express purpose of helping fund programs in the schools (Zimmer & Jones, 2005).

In the aftermath of *Serrano*, advocates searching for ways to eliminate inequities in school funding developed the notion of adequacy as a different standard by which schools ought to be judged. Whereas equity focused on comparisons between school

districts, adequacy focuses on expenditures in absolute terms, whether the amount of money provided is enough to fulfill a state's constitutional mandate, regardless of how much money is provided in surrounding districts or states. The situation in Alabama provided some of the early theoretical basis for adequacy. In Alabama, spending is relatively equalized between districts, and yet educational results are dismal statewide (Eastman, 2007). Where such a situation exists, it is clear that equity alone will not boost educational outcomes.

The difference between equity and adequacy theories of school finance is similar to the difference between norm- and criterion- referenced tests. The first is primarily concerned with comparison and ones standing relative to others in the population. The second is concerned with the result in absolute terms. Beginning in the early 1980s, adequacy began to supplant equity as the favored legal rationale behind school finance litigation; however, there are two theories that emerged recently and used in concert, as in Missouri (CEE v. State, 2004).

If adequacy lawsuits allege that the state is not providing funding sufficient to fulfill its constitutional mandate to provide an adequate education, the implicit assumption is that there exists some way to calculate the cost of an adequate education. Adequacy, then, must be given an operational definition so that courts could order a remedy. There are four predominant models for costing out an adequate education (Duncombe and Lukemeyer, 2002; Odden, Goetz and Picus, 2007; Guthrie and Rothstein, 1999).

Methods of Determining Adequacy

While a full evaluation of the different methods to determine adequacy would require a full volume in itself, a brief overview is necessary for two reasons. First, it parallels the decisions that judges must face in evaluating expert testimony and deciding what constitutes an adequate education, and second, it allows one to see just how vague the constitutional language is that lies at the heart of adequacy.

Duncombe and Lukemeyer (2002) lay out three questions that any method costing out must consider. First, the “object and level of the standard must be defined.” In other words, some criteria must be selected that will demonstrate adequacy, whether that is an achievement test score or a certain amount of some resource or any other measurable item. Then, a spending level and the presence of adequacy must be linked in one “benchmark district” (or, conceivably, more than one). Finally, one must develop a way to adjust that benchmark figure as certain factors change from district to district (such as number of disadvantaged students or resource costs).

Duncombe and Lukemeyer (2002) then conclude by discussing different methods that have been developed to cost out adequacy. Guthrie and Rothstein (1999) discuss similar methods, although they use different terminology to describe them. The first method of costing out adequacy is what Duncombe and Lukemeyer call the “empirical identification approach,” which Guthrie and Rothstein call the successful schools model, which is the more descriptive name. The model is premised on the fact that the costs of adequate can best be computed by finding those districts (or states) that are already achieving adequacy and seeing how much they spend. The logic here is intuitive. If a district is meeting some standard of adequacy, by definition, that spending is enough to ensure an adequate education. Hanushek (2006) and others might point out, though, that

the successful schools model says nothing about efficiency. It follows, then, that depending on the district under study, this method likely will not yield the *minimum* required for an adequate education, because any inefficiencies in the successful school or schools under study would be institutionalized in the adequacy calculation. Indeed, Hanushek argues that all costing out methods suffer from the same pathologies, including the tendency to give the highest possible cost and to ignore issues of efficiency and trade-offs with other government services.

A second method for costing out is the resource cost model, which Guthrie and Rothstein refer to as the professional judgment model. Whereas the successful schools model is a top-down approach in which the starting point is a total expenditure that is then broken down, the professional judgment model represents a bottom-up approach. The model relies on panels of professional educators, asked to design an adequate school. Virtually every element is fair game for discussion and, at the end of the process, an itemized budget is developed. It is then up to the analyst to determine the costs of each element that the educators identified as part of an adequate education, and that sum becomes the cost of adequacy.

A third method for costing out is the “evidence-based approach” favored by Odden, Goetz and Picus (2007). The main idea of this approach is that the components of an adequate education should base on “evidence from research and best practices to frame its recommendations” (p. 4). It is different from both the successful schools model and the professional judgment model because, while both of those models end up with a dollar figure for adequacy, there is little attention paid to *how* that money should be spent to achieve adequacy. In the evidence-based approach, not only does one know the cost of

adequacy at the end of the process, but one has a guideline for how the money should be spent as well. This method is notable for taking issues of allocation and efficiency into account, even implicitly. It was also the model used by the Court in Kentucky in the nation's first successful adequacy lawsuit.

A final approach, favored by Duncombe and Lukemeyer (2002) is the cost function approach. This approach relies most heavily on statistical analysis and education production functions. The cost function approach “relates data on actual spending in a district to student performance, resources prices, student needs, and other relevant characteristics of districts” giving researchers the ability to construct “education cost indices, which measure how factors outside a district’s control affect the spending required to reach a given student performance level” (Duncombe and Lukemeyer, 2002, p. 12). While Duncombe and Lukenmeyer favor this approach and claim that it is the most accurate, others, such as Guthrie and Rothstein reject it because it is difficult for policy makers to understand, and the statistical model is dependent upon assumptions made by the analyst, which are often not “transparent” to those attempting to evaluate the model.

It would add some substance to the discussion to look at Kentucky, the state which had the first successful adequacy litigation. The court in Kentucky had to determine whether the legislature had upheld its duty of providing an “efficient system of common schools” throughout the state. The plaintiffs in *Rose* got a better ruling than they likely even hoped to get. Although much evidence was designed to show inadequacy in Kentucky’s poorer districts, the court further found that, “Kentucky’s system of common schools is under-funded and inadequate, is fraught with inequalities

and inequities throughout the 177 local school districts, is ranked nationally in the lower 20-25% in virtually every category that is used to evaluate educational performance, and is not uniform among the districts in educational opportunities.” The ruling, in other words, found that Kentucky’s *entire* system of public education was inadequate, even the education provided in the richer districts.

It is particularly interesting to see that the court used statewide assessment data and compared Kentucky with neighboring states to document objectively the claim of educational inadequacy. The court found that “Evidence relative to educational performance was introduced by appellees to make a comparison of Kentucky with its neighbors—Ohio, Indiana, Illinois, Missouri, Tennessee, Virginia, and West Virginia. It also ranked Kentucky, nationally in the same areas...Numerous well-qualified educators and school administrators testified before the trial court and all described Kentucky’s educational effort as being inadequate and well below the national effort” (*Rose v. Council for Better Schools*, 1989). .

The Court then proceeded to describe seven specific things for which an adequate school system would prepare every child. They are “(i) sufficient oral and written communication skills to enable students to function in a complex and rapidly changing civilization; (ii) sufficient knowledge of economic, social, and political systems to enable the student to make informed choices; (iii) sufficient understanding of governmental processes to enable the student to understand the issues that affect his or her community, state, and nation; (iv) sufficient self-knowledge and knowledge of his or her mental and physical wellness; (v) sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage; (vi) sufficient training or preparation for

advanced training in either academic or vocational fields so as to enable each child to choose and pursue life work intelligently; and (vii) sufficient levels of academic or vocational skills to enable public school students to compete favorably with their counterparts in surrounding states, in academics or in the job market.” (*Rose v. Council for Better Schools*, 1989).

Several more things are noteworthy about the *Rose* decision. First, the court gives a list (above) of what an adequate school system should be preparing students for that goes well beyond what many other people might consider an adequate education. The list is, indeed, quite extensive, and is the foundation of the court’s decision that no school in Kentucky is delivering an adequate education. Also interesting is that court made no effort to cost out its definition of adequacy, although it is clear that more money will be required than was being spent on education. That strategy was tenable in Kentucky, but political and institutional factors could make such a decision politically impossible in other states. Kentucky, after all, was unique not because it was the first adequacy case tied to standards but also because there was a large consensus among state officials and plaintiffs about what the verdict should be even as the case was first being filed. A final interesting point is that, in finding the Kentucky system inadequate, the court relies on interstate comparisons of educational outcomes. The meaning of adequacy in Kentucky, then, is both entirely outcome-based and relative.

The situation in Kentucky, and the possibility that institutions played a role, will be clearer if Kentucky is briefly compared with some other cases. In Kentucky, the court found the entire system of schools to be inadequate, and then proceeded to enumerate a list of characteristics of adequate schools. It took approximately one year from the time

the decision was made for the legislature to pass a bill that incorporated and funded those goals. Consider New York as a contrast, though. In New York, the court left it up to the legislature (and then the plaintiffs) to conduct a cost study, and specified that a “sound basic education” meant a meaningful high school education, but did not give the specific guidance that the Kentucky Court did. Additionally, the plaintiffs had to return to court to get the state to abide by the adequacy ruling. One plausible explanation for the greater judicial leadership in Kentucky than New York can be found in exploring the institutional relationships, although there are other rival hypotheses.

Competing Explanations of Judicial Decisions in Adequacy Lawsuits

This study investigates the institutional conditions that could prove necessary and/or sufficient for plaintiffs’ victory in an education adequacy lawsuit. Implicit in that formulation of the research is the assumption that institutions—and more generally, politics—have some important effect on judicial decision-making. In the forthcoming section on judicial decision-making, David Mayhew’s theory of congressional behavior is introduced below as a starting point for a theory that looks at the goals of legislators and how institutional arrangements embolden or constrain their pursuit of those goals.

There is, however, a rival possible explanation of judicial behavior that is much less concerned with politics. This more legal-oriented view of court decision-making emphasizes how judges apply the relevant law and precedent to the facts of each unique case. Feldman (2005) observes that political science as a discipline has generally been concerned with the political view, focusing on judges’ political preferences and the politics imposed by external actors. As a consequence, Feldman posits that the political science has not taken seriously the legal view, which is primarily concerned with the

application of legal rules. The legal perspective does not, for example, imply that judges and litigants robotically apply precedent and law, nor does it imply that judges are entirely unconstrained by the pressures of politics. However, it does imply that, if one wants to look for the driver of judicial decisions, one ought to look inside the court before one looks outside.

Legal theories of judicial decision-making are rather difficult to test directly. Only the written decision is observable, and while it might yield some clues as to how a decision was reached, much of the process remains obscured. And additionally, the institutional factors that some scholars believe affect judicial decision-making will sometimes be present whether or not they had any effect on a particular decision. In such a situation, those who want an institutional or political explanation will be able to find one (particularly if the sample of cases under study is large enough) and those who favor a legal explanation will be able to find one as well. If, however, no institutional or political explanation is apparent, and no political arrangement appears necessary or sufficient for a particular legal outcome, then there would be evidence for a legal explanation, even without any direct observation of judges' decision-making processes. The current study proceeds using this logic.

Institutions

According to Shepsle and Bonchek (1997) an institution is a formal structure endowed with resources and authority that emerges in response to recurring problems, so that the response to the problem becomes routinized—or institutionalized. Institutions are not physical things, but rather are the rules, procedures, and decisions of a particular group. Put another way, if the Supreme Court building in Washington, DC were to blow

up tomorrow, the Supreme Court would still exist, although presumably the justices would meet elsewhere.

If institutions dictate the rules of the game, then it makes sense that different institutions, designed with different purposes in mind, would employ different procedures that might lead to different results. Indeed, this is one reason why the design of new institutions is inherently contentious. The design of the institution inextricably affects the policies that come out of them. Partially because of the way different institutions are designed, they will each have different, although sometimes overlapping, strengths and weaknesses. Courts are the institution most explicitly responsible for disputes. In the course of making policy on the state level, courts interact with at least two other institutions of formal government (executives and legislatures) as well as countless other institutions that interact with these formal structures of government, such as lobbyists, litigants, and the bureaucracy. The point of emphasizing the institutional nature of these interactions is to point out that they occur repeatedly over time, across many different issues. When a lobbyist, for example, interacts with legislators or advises a client to file a lawsuit, he has certain expectations about what will happen based upon previous experience with those institutions and the strengths of his or her current case. Those expectations about institutional reactions must influence not only the course of action that people take in choosing to approach governmental institutions, but also the choices that one institution makes when interacting with another.

Ginsberg and Shefter (1990) analyze these interactions among different institutions, and conclude that interactions among institutions have become even more important than the typical partisan contests between voters in shaping American public

policy. Their analysis emphasizes that, since neither major party has been able to control all branches of government clearly and consistently the battles over policy are fought instead between Republican- and Democrat- controlled institutions, which serve to shape policy (Nelson, 2008). Ginsberg and Shefter call this phenomenon “institutional combat.” If their analysis is correct and there are indeed partisan- or ideologically-tilted institutions that are in combat with each other, then the specific institution that gets to decide a given issue matters a great deal. Potential adequacy litigants would be wise to consider the ideological tilt of not only the court that would hear the case, but also the court’s ideological relationship to other state-level institution and the likelihood that any order for additional funds would be acted upon by the legislature as opposed to actively fought, as an anti-reform legislature can, at the least, delay reform. Furthermore, since litigants and judges are likely aware of these institutional constraints, it should be expected that a preponderance of the cases would be either filed or decided for the plaintiffs under specific institutional conditions. Identifying and testing hypotheses about different institutional conditions will be the subject of later chapters.

Theories of the policy process (see Kingdon, 1995; Sabatier, 2007) mostly focus, to some degree, on the role of institutions. Of those theories, a subset takes an explicitly economic approach in explaining the policy process, which might be useful here. Komesar (1994) provides one such explanation, which he terms comparative institutional analysis. According to Komesar, policy problems should—and often do—wind up in the institution best suited to solving them. Komesar is concerned mainly about the courts, and his central point is that, too often, we examine the ability of courts to solve policy problems in isolation. We talk, for example, about the kinds of things that courts are

good at doing (interpreting laws) and perhaps not as good at doing (designing and delivering a complicated policy program). For Komesar, though, that analysis misses half the point. Instead, we should talk about the kinds of things that courts are better at doing *compared to legislatures or executives*. If different institutions are essentially different forums in which a policy problem can be dealt with, then rational and informed advocates will approach the institution that has the comparative advantage in dealing with a given policy issue, mainly because that institution will be in the best position to actually give advocates what they want.

If we accept Komesar's formulation, then it would be logical to make two assumptions. First, judges will issue adequacy decisions for the plaintiffs only in situations where they expect that their decisions would be carried out. Second, litigants would file adequacy lawsuits only in instances where the courts were the most efficient option for achieving adequacy (of course, sometimes a lawsuit is the most efficient option because it is the *only* option available at the time). Using school desegregation as an analogy could clarify these assumptions.

As a matter of institutional prestige, judges likely do not want to see their decisions ignored. Concerns about the judiciary's ability to enforce its decisions, and ways to ensure the legitimacy of the judiciary even while depending for enforcement on other branches of governments, are concerns that date to the beginnings of the republic. The Supreme Court declared separate but equal schools inherently unequal, and yet there was still quite a delay before that decision was realized, as other institutions resisted. It is well known that Chief Justice Warren wrote several drafts of his decision, in an attempt to ensure a unanimous decision. This is evidence of Warren's concern that, given the

controversial nature of the decision, unanimity could be an important way of legitimizing the court's decision in the face of resistance from other governmental actors, by adding a certain moral force to the court's legal opinion. Gerald Rosenberg (1981) uses similar logic in the *The Hollow Hope*. He argues convincingly that judges cannot bring about unconstrained social change, but rather that they put into law the social changes that society has already begun.

For the second assumption, that litigants would turn to the courts when they were the most efficient option, simply consider the success that blacks had had in desegregating school by going through other institutions. There had been only modest success, as neither legislatures nor governors were generally sympathetic, although some places, Baltimore, for example, were exceptions to that general rule (Baum, 2010). The point is, though, that litigation is very expensive and time consuming, and the upfront costs to begin a court case are very high compared to the costs of, say, lobbying or attempting to influence public opinion through writing a newspaper op-ed piece. Furthermore, litigation is even more expensive for "one-shotter" groups making opposed to those (like states) who are repeat players in the courts (Komesar, 1994; Galanter, 1974). Rational litigants would only turn to the courts, then, as their resort for being blocked from the more "political" levers of power.

In justifying the two assumptions above, however, institutional theories also create a paradox. On the one hand, litigants would turn to courts only when they have been unable to achieve adequacy through other branches of government. On the other hand, courts are only likely to find for plaintiffs when they are confident that other branches of government will respect and implement the decision. The very parties most

in need of judicial intervention would be least likely to prevail there. Clearly, reality does not support that conclusion, as there are certainly some adequacy cases where plaintiffs *do* win in court after getting no relief elsewhere, and that same pattern has been seen in other policy areas. Clearly something is missing from an institutional perspective on judicial decision-making. Adequacy advocates may be pursuing other strategies in concert with litigation, such as lobbying or public relations designed to pressure elected officials. Advocates may also believe that filing a case has benefits even if they lose. Perhaps the courts may have a significant power of moral suasion, and might be able to compel the legislature into action even while ruling against the plaintiffs, or perhaps the spectacle of a court case would be able to mobilize popular support for adequacy. It would be nearly impossible, however, to test hypotheses about whether plaintiffs filed court cases they expected to lose in order to reap some secondary agenda-setting benefit. No advocate would admit before a trial that they anticipated defeat, but would be similarly likely to take credit for some degree of strategy if indeed a losing case *did* lead to change.

The remainder of this chapter will explore ways to test the institutional theory, and will also explore how the rival legal hypothesis might be tested.

Political and Institutional Theories Concerning Judicial Behavior

The conception of judges as political and strategic actors might seem counterintuitive, but has a long history in the literature, dating from Walter Murphy's (1964) seminal work, *The Elements of Judicial Strategy*. If judges are political actors working to see their policy preferences enacted, any hypotheses about how judges might decide must consider the different goals that judges pursue. Epstein and Knight (1998)

point out that David Mayhew's examination of the goals of legislators provides a useful starting point. According to Mayhew, legislators have three goals. First, they must pursue reelection, because they must remain in Congress to accomplish any substantive goals they might have. Second, legislators seek power and prestige within the institution. As Mayhew was writing about the US Congress, where seniority matters a great deal (although perhaps slightly less in the Senate), he was referring to obtaining party leadership positions or committee chairmanships, as committee chairs are likely to exert a greater influence on policy than a committee member. Thirdly, legislators pursue the making of good policy; however, they define that for themselves.

With only minor adjustments, we can apply these goals to judges as well. First, judges seek to remain on the court, through partisan elections, non-partisan elections, retention elections, or reappointment. Second, they seek prestige, although, given the lower visibility that judges have and the restrictions on judicial campaigning, judges likely seek not only prestige for themselves, but also prestige for the institution. Third, judges seek to make "good" law; again, however, they properly define it.

The second goal, that judges seek prestige for the institution more than for themselves, requires some additional justification. Judges are, after all, rational people as well, and there is certainly something to be said for the goal of becoming a well-respected judge whose opinions are widely-cited. According to Mayhew, however, much of the credit-claiming that is behind the prestige-seeking behaviors of congressmen is due to the election incentive. This does not imply that legislators have no concern about the institution and its reputation—surely, some care deeply about the prestige of Congress or a state legislature. However, the incentive for personal prestige is stronger for legislators

than for judges, as legislators face a constant pressure to show what they themselves have accomplished for their district. Judges, even those who need to stand for reelection, face little of that sort of pressure to highlight their individual accomplishments. And, knowing a priori the differing roles of legislators and judges, it is likely that different types of people enter each field and are thereafter socialized to different professional norms. There has been, however, no work that I could find comparing the personality characteristics of legislators and judges.

Given these three goals, any factor that could affect them reasonably deserves empirical investigation as regards its effect on the outcome of an adequacy lawsuit. For instance, if we know that remaining on the court is generally a goal of judges and that the method of judicial selection differs among states, then, it would be prudent to test whether those different methods of judicial selection were related to differences in decisions. Similarly, if the hypothesis is that judges want to preserve the prestige of the court relative to other institutions and legislative or executive disregard of a court decision will damage the standing of the court, any factor that would make it less likely that an adequacy decision would be implemented should decrease the likelihood of such a decision.

The institutional perspective assumes, a priori, that politics matters. While political scientists are unlikely to question that premise, others might question the entire conception of a political model of judicial decision-making. Those skeptics might be more comfortable with a theory that focused entirely on formal legal rules. This study treats the understanding that politics matters as an empirical question, by setting up a more strictly legal conception of judicial decision-making as a rival hypothesis. Two

important pieces of evidence, however, point in favor of a political conception of adequacy lawsuits, and therefore deserve mention here. First, although examining the impact of a successful adequacy lawsuit on student achievement in a given state is difficult and controversial due to various confounding variables, (see, for example, Rothstein, 2004) no one has been able to conclusively link victory in an adequacy lawsuit to improved educational outcomes². Furthermore, among the states in which there has been a lawsuit, no literature exists that contends that plaintiffs in the more “inadequate” states win their lawsuits. Other political factors, then, must partially explain this variation.

Judicial Decision-Making Primarily as the Product of Law

The discussion thus far has been primarily concerned with how political and institutional variables might impact judicial decision-making. There is scholarly justification for that viewpoint, as well as for the use of Mayhew’s formulation to examine the goals of judges. However, the possibility exists that institutional or political factors might not play a role, or a *primary* role, in adequacy lawsuits. It could be that political and institutional factors only play a primary role in very politicized issues, and that adequacy lawsuits are less a politicized issue than, say, abortion. After all, judges could be either pro-life or pro-choice, in ways that they could never be pro-adequacy or anti-adequacy. Additionally, it is entirely possible that, as much of the research on institutions and judges comes from the US Supreme Court, that the theory simply doesn’t

² For example, Sol Stern (2006) points out that in New York, there were significant spending increases that took place while the case was still in litigation. From 1998 to 2002, NYC schools received an inflation-adjusted increase in spending of approximately 25%. Stern does not, however, pursue the question of what the cause of the spending increase was, and whether it was related to CFE. Joe Williams (2007) makes a similar point.

hold for state supreme courts—that they are either less visible or less concerned with issues of legitimacy.

There is a little research that directly addresses whether judges rely primarily on legal concerns at the expense of political or institutional concerns as an empirical question. The reason for this is likely twofold, because scholars investigating the effect of politics on the judiciary likely believe that politics matters, and that those who believe that judges rely primarily on the law (as most law journal articles implicitly assume) treat such a belief as nearly axiomatic, not necessitating investigation. Brisbin (1996), for example, points out that Segal and his colleagues, in testing the “attitudinal model” of judicial decision-making compare it to a “legal model” without also testing the legal model.

This legal model is, indeed, hard to test directly without any firsthand knowledge of how judges actually deliberate, because testing the legal model would require access to both the deliberation and the outcome, whereas testing a political model requires only the outcome. There is, however, some evidence for a legal perspective. Emmert (1992), for instance, conducted a logit regression of three thousand Supreme Court cases dealing with judicial review. He found several variables that exerted an independent effect on Supreme Court decision. Importantly, some were from the “political” camp (for example, the identity of the challenging party) and others were more “legal” in nature (for example, the specific constitutional arguments that the parties used).

Importantly, sitting judges have occasionally written about the judging process in general terms. This also provides support for the legal view, or at least a primarily legal view constrained at the margins by politics. Harry T. Edwards (1991), a federal judge on

the DC Circuit, writes about how he has tried to follow what he calls “principled decisionmaking.” He writes that he “still believe[s] in and subscribe to principled decisionmaking, but it is no longer entirely clear to me that partisan politics and ideological maneuvering have no meaningful influence on judicial decisionmaking.” He discusses three major challenges to maintaining principled decisionmaking, where “unprincipled” means a “focus on preferred results” at the expense of permissible legal reasoning. First, there is the expectation from litigants and observers that judges *will* behave politically, which leads to certain types of cases and arguments and creates a kind of self-fulfilling prophecy. Second, there are instances in which a judge consciously takes into account the expectations of others and how a decision will be perceived. Edwards discusses the “War on Drugs” and how judges might relax their standards on what constitutes a permissible search and seizure, particularly after the defendant has been caught red-handed, to avoid being labeled “soft.” Third, there are cases in which, in times of crisis, courts have essentially ignored conventional civil rights, due to the politics of the situation. Edwards discusses *Koretmasu*, which found the internment of Japanese-Americans during WWII to be constitutional, and also discusses the Cold War freedom-of-speech and freedom-of-association cases that limited civil liberties in the name of combating communism. Conceivably, Edwards would draw the same conclusions about many current day anti-terror laws.

What we have, then, is a legal perspective that puts the law front and center but leaves room for political and institutional constraints at the margins, where politics might play an especially large role only a small fraction of the time. Feldman (2005) describes the tension between what he calls the internal and external views of law, where the

internal is practiced primarily by lawyers and the external by political scientists. He also points to the recent attempts by political scientists to “harmonize” the two perspectives, specifically the “new institutionalist” approaches which recognize the importance of legal reasoning but also how it is shaped by political and social forces at specific points in history.

It would be hard to test this perspective empirically across many different states, especially because the theory specifies nothing *systematic* about the social, political, and legal relationships at any given point in time. It is, however, possible to offer evidence for such a perspective *if* a strictly political approach does not yield any results, as this legal perspective would be the next best alternative.

Other Possible Rival Hypotheses

There exist at least three other possible explanations for decisions in adequacy lawsuits, if the initial political hypotheses is not confirmed. They are less likely than the rival hypotheses discussed above, but nonetheless deserve mention for the sake of completeness. The first is the legal realism perspective. Alex Kozinski (1993) describes the legal realist perspective as believing that judges “glance at a case and decide who should win” and then, come up with reasoning to support their preferred conclusion because “judges can reach any result they wish” by using one line of reasoning as opposed to another. Kozinski wrote the article as a sitting judge, and discredits the legal realist perspective. What he offers instead sounds a lot like the new institutionalism discussed above. Judges have some choice, but that choice is constrained, foremost by legal factors and also by political concerns and also by the same errors of reasoning to

which all human beings are subject. It would appear, therefore, that a legal realist explanation would not explain adequacy decisions.

It is also worth considering whether there is in fact a pattern to be found, or whether judicial decisions in adequacy cases are essentially random. It is possible that some plaintiffs will win, and some will lose, for no discernible reason. Random outcomes, however, are unlikely in a court of law where the decisions have to be *justified* in writing. Some legal reasoning must be employed, although courts need not be equally rigorous. Importantly, though, this study does not attempt to evaluate the quality of the plaintiffs' case and their counsel's performance, although an effort is made to determine whether the objective level of inadequacy is related to the decision. The implicit assumption is that cases that reach this level are represented by equally able litigators who have a firm grasp of the applicable law and issues. While, strictly speaking, this might not be true, it would be difficult to objectively measure trial performance, and, in any event, beyond the scope of the current study. Trial performance would, however, be a kind of explanation of last resort if the empirical portion of this study yielded no significant results.

Importantly, though, the presence of no discernible pattern in *outcomes* does not necessarily imply randomness. It is entirely possible that there is systematic variation in some other part of the process—in this case, in the *filing* of adequacy lawsuits. If timing didn't matter at all, we might have expected to see fifty adequacy cases filed in 1989, in the days after the first case was filed in Kentucky. Clearly, that didn't happen. Instead, gradually, adequacy lawsuits were filed in more states. Filing a lawsuit is neither quick, nor cheap, nor easy, and so it is clear that there was *some* rational calculation involved in

why litigants in some states filed years earlier than others. And although the citizens of some states might be more or less satisfied with the quality of education in their state, the eventual proliferation of adequacy lawsuits to nearly every state suggests that litigants didn't wait to file a lawsuit because they thought education in their state was adequate.

The question then becomes why litigants in some states would file earlier than in other states. One possible explanation is that some states had particularly large variability between school districts, which would indicate inadequate funding of the districts on the lower end. While adequacy and equity are very different legal arguments, they remain inextricably linked in practice, because one way to demonstrate some inadequacy would be to demonstrate a substantial inequality in funding, unless one wanted to make the argument that states were over-funding some districts.

Given this argument, one could expect to see that states in which there was large variability between the high-per-pupil-expenditure districts and the low-per-pupil-expenditure districts were the first to experience an adequacy lawsuit. Additionally, the adequacy argument is relatively similar from state to state, which means that, all else being equal, it is possible that the first cases filed were actually the "better" cases, where plaintiffs stood a better chance of winning. In that case, one would expect to see later plaintiffs have less success not because of any political or institutional factor, but simply because they had worse objective cases to begin with, where, perhaps, the evidence of inadequacy was not as pronounced. In the event that no political or institutional factors are necessary and/or sufficient for plaintiffs to win, an alternative explanation (aside from the rather unsatisfying conclusion of randomness) might have to do with the declining levels of inadequacy as time passes and more cases are filed.

Significance of the Present Study

These adequacy lawsuits, and their results, present an interesting political science question for two reasons. First, there is unexplained interstate variation in the results of these lawsuits³. The strikingly different ways in which different courts have defined an “adequate” education, even among states with very similar constitutional language, would naturally lead one to question the factors that led to that decision. Clearly, *something* beyond “objective” observation of a state’s educational system is going on, because plaintiffs have lost in states where they would have been expected logically to win (for example, in Florida, with non-elected judges and a strong constitutional clause), and the reverse is true as well. An examination of the factors behind unexplained interstate variation could serve to shed some light on political or institutional factors that contribute to the decisions of state supreme courts⁴. Additionally, there are likely political and institutional factors that influence the decision to file an adequacy lawsuit in the first place. It is not at all clear that the worst performing states are necessarily the states in which lawsuits are filed.

The second reason that adequacy lawsuits present an interesting phenomenon for study is that, as the *Brown* decision acknowledged, education is one of the most important functions of state government. A ruling that a state is failing to provide an adequate education will have a significant impact on a state’s budget, necessitating either higher taxes, redistribution of wealth from richer to poorer districts, and/or cuts in some

³ There were thirty cases filed between 1989 and 2005. Plaintiffs lost in eight of them. As discussed in Chapter 3, there was also variation in the hypothesized necessary and/or sufficient conditions. Conditions with little or no variation were excluded because, by definition, they could be neither necessary nor sufficient.

⁴ I am aware that, in some states, the highest state court is referred to be a different name, and that New York uses the term “Supreme Court” to refer to its trial court. Nonetheless, I adopt the convention in judicial politics literature of using “state supreme court” as shorthand to refer to the state court of last resort.

other state-funded services. Political scientists interested in state legislative politics and/or tax policy ought to find the study of legislative response to an adequacy ruling, and the ensuing budgetary and taxation decisions, a promising prospect for study. The prospect that such potentially disruptive rulings could be made based on something other than an objective standard (or, more terrifying, that an objective standard is not possible) also has potential impacts on normative arguments about the proper role of the courts in policy debates and about the proper function of democracy in general, particularly in cases where the court is unelected or perceived to be unaccountable.

The next chapter provides a thorough review of the literature related to the study of adequacy lawsuits and judicial politics, and also expands on some themes presented in this chapter. Chapter 3 discusses Qualitative Comparative Analysis, the method by which the research question is investigated. Chapters 4 and 5 present results and discussion, respectively.

CHAPTER 2 – LITERATURE REVIEW

The purpose of the current study is to determine whether there are political and/or institutional factors that lead to a plaintiffs' victory in an adequacy lawsuit. In order to frame the study in a broader context so that its significance is understood, a broader review of the literature is required. This review will also examine possible variables that might be hypothesized to affect the outcome of an adequacy lawsuit, as well as variables that might not have such an effect. I proceed here from the general to the specific, attempting to first place adequacy lawsuits in a broader context before examining them—and the factors that affect them—in more detail.

Adequacy Lawsuits and ESEA

As discussed briefly in Chapter 1, adequacy lawsuits cannot be viewed in isolation. Rather, they are a policy tool, and the decision by litigants to file them, and by judges to decide them a certain way (or even to decide to decide them at all, and not dismiss them on grounds of non-justiciability), cannot be understood without also understanding the evolution of education in general. Specifically, as adequacy lawsuits are designed to hold schools accountable via the states in which they operate, and as adequacy lawsuits have come to imply the reallocation of resources on a vast scale, they need to be understood in concert with the evolution of accountability and resources in general, and ESEA in particular. Perhaps this relationship between broader accountability issues and adequacy lawsuits will emerge as a possible alternative explanation for decisions in adequacy lawsuits, especially if no political or institutional factor appears necessary for plaintiffs' victory across all cases.

No Child Left Behind (NCLB) is merely the latest incarnation of ESEA, and one can see obvious parallels between the evolution of education finance litigation and the evolution of ESEA. Cohen and Moffitt (2009), Thomas and Brady (2005) and McDonnell (2005) all give comprehensive overviews of the various incarnations of ESEA. ESEA began in 1965 as part of President Johnson's War on Poverty Program, and was designed to provide help for poor children. Unfortunately, four factors limited the ability of ESEA to fulfill its core mission in its early days. The first was that, in order to ensure political support for an expanded federal role in education, the benefits of ESEA had to be rather diffuse. Thomas and Brady (2005) point out that that, in the years following its inception, approximately 94% of all school districts received Title I aid. Even in a time of increased federal spending, the widespread distribution of Title I funds worked against the initial mission to target children in poverty. Poor targeting of funds was also behind the second factor that limited ESEA's effectiveness. While the funds were meant for poor *students*, they were in fact distributed to *schools* with *high concentrations* of poor students. Any student who was struggling academically and attended a Title I school was entitled to Title I services, whether or not that child was actually poor. Additionally, Title I money was being used to finance regular school budgets and allow localities to save money. Rather than supplementing local funds so that special programs could be offered, Title I became just another source of general revenue in many districts. Lastly, Cohen and Moffitt (2009) focus part of their examination on problems due to lack of infrastructure and capacity. ESEA purported to use federal power to help poor students, but decentralized school districts had no common infrastructure, standards, or methods of teacher education that would give

individual schools the capacity to systematically use Title I money. These problems were all exacerbated by the low status of the United States Office of Education, which was not able to fix these problems.

These problems with Title I implementation first received widespread attention after the publication in 1969 of a report entitled *Title I of ESEA: Is It Helping Poor Children?*. The report, written by Ruby Martin (Washington Research Project) and Phyllis McClure (NAACP Legal Defense and Educational Fund), highlighted abuses in Title I administration and led to several regulatory changes, including the explicit requirement that Title I money should be solely a *supplement* to ordinary school funding.

Most significant to the current analysis, though, is Phyllis McClure's status as an author of the report and a member of the NAACP legal defense fund. It demonstrates the multi-faceted nature of educational policy-making and the various venues in which strategic action can be taken by those hoping to affect change. The broad mission and goals of the NAACP are fairly clear, as is the link between the likelihood that a child is in poverty and the likelihood that the child is African-American. Therefore, one can see evidence of a coordinated strategy here on the part of the NAACP. At the same time that the Martin McClure report questioned whether Title I was serving its intended purpose, the NAACP Legal Defense and Education fund was also working in the courts to achieve related educational equity goals. In 1968, the NAACP won a favorable decision in *Green v. County School Board*, in which the so-called "freedom of choice" desegregation plan in New Kent County, Virginia was deemed unconstitutional. And in 1971, the NAACP would win an even bigger victory in *Swann v. Charlotte-Mecklenberg Public Schools*, which legitimized busing as a desegregation tool.

It is evident, then, that educational policy change can be pursued in multiple venues, and that different issues get decided by different audiences, in a manner which certainly implies a fair amount of strategy. The discussion of Komesar's Comparative Institutional Analysis in Chapter 1 should further illuminate this phenomenon, by pointing out the costs and benefits of litigation as opposed to other venues. Indeed, it will prove instructive in developing the theory of the current study to consider why the NAACP, working towards a set of common goals, pursued some items through the legislature (and indirectly, through the public) and others through the courts. After all, the NAACP was victorious in this case, so whatever calculation they made about venue selection and timing was successful (although, surely, the facts of the case played some part as well).

After the reforms prompted by the Martin-McClure report, ESEA continued relatively unchanged until the 1980s, when the overarching political climate began to change. McDonnell (2005) identifies the Reagan and George H.W. Bush presidencies as a time which Title I (renamed Chapter 1 in 1981) entered "a second phase." Given the Reagan Administration's ideological stance with regards to the size of the federal government, Title I survived mainly because it already had an entrenched constituency. It changed, though, in that the requirements were somewhat loosened and funding was cut. Additionally, Title I became part of a broader discussion that began with the publication of *A Nation At Risk* in 1983. The national educational debate shifted from equity to excellence,

President George H.W. Bush continued the Reagan administration's focus on excellence. In 1989, Bush hosted a "summit" with the nation's governors, and they

agreed in principle on broad goals for the American school system, which became known as America 2000. Bush's term did not include a reauthorization of ESEA, which would have given him the chance to tie the America 2000 plan with significant funding. And Congress proved unwilling to pass the bill as a stand-alone measure, as the conference version died in the Senate. Note that 1989 was the year of both the Bush education summit and the first successful adequacy lawsuit in Kentucky. In other words, in the same year that state governors and the president sent a powerful signal that the national educational dialogue was changing, the Kentucky Supreme Court issued a landmark decision, on a case that had been in the courts for three years.

While not identical, the America 2000 goals and the Kentucky court's definition of adequacy share some common concepts. Most notably, both sets of goals are ambitiously wide in scope, are concerned with results (in terms of student learning and later productivity in the workforce) and make comparisons of results. America 2000 talks about international comparisons, while the Kentucky decision references neighboring states. These two events represented the culmination of an important shift in educational priorities, from equity of inputs to responsibility for outputs.

There are two major lessons to be learned here. First, it would seem highly coincidental to think that the *Rose* decision and America 2000 were completely independent. Rather, this would seem to be an instance of judicial strategizing. The Kentucky Supreme Court could make its ruling, secure in the knowledge that other political bodies were moving in the same direction, and confident that some enforcement of its judgment might occur (and anecdotal evidence suggest that this calculation actually took place). Kentucky is, after all, known as a leader in standards-based reform, and was

one of the first states to embrace the principles of America 2000 when they became law as Goals 2000 under President Clinton (Superfine, 2005; Schwartz and Robinson, 2000).

Second, a kind of institutional sorting seems to have occurred, in terms of what was left to the more political branches of government and what wound up in the courts.

Conceivably, at least, America 2000, as it was developed in consultation with state governors, could have linked itself to the constitutional responsibility of those states to provide an adequate education. Instead, America 2000 (which, interestingly, became Goals 2000—nearly unchanged—under the Clinton administration) dealt with creating standards, tests, international competition, and ensuring that all students start school ready to learn. America 2000 left out any mention of how these goals would be financed. The question of financing wound up in the courts, and one can see a leveraging of America 2000 in the *Rose* decision, similar to how Rudalevige (2005) identifies more current adequacy lawsuits as leveraging the requirements of No Child Left Behind. Both the *Rose* decision and the America 2000 goals emphasize, for instance, outcomes in the job market and economic competitiveness, as well as sufficient schooling to exercise the duties of citizenship.

Here, then, we see another example of Komesar's comparative institutional analysis. In a kind of institutional sorting, policies that would require redistribution of resources wind up in the courts, as courts are the political organ that can most easily mandate potentially politically unpopular redistribution. However, courts do not act in isolation, and seem to take signals from other branches of government (in *Brown v. Board* as well, as I discuss below).

If my interpretation of the events at the Charlottesville education summit is correct, and my interpretation of the history of ESEA is correct, then there is some additional justification for the current study. Education reform can be pursued simultaneously in many different forums, a lesson that adequacy advocates have no doubt learned. In the case of Charlottesville and the *Rose* ruling, and also in the case of ESEA and desegregation, the politics in each venue seem to have reinforced each other, as when the goals of the Charlottesville summit appear to have given political cover to the court in *Rose* (although, since we are not privy to the internal discussions of the Kentucky Supreme Court, we can never be sure about the relative influence of different factors in their deliberations). Assuming that the prevailing politics played a part in the *Rose* decision (alongside the applicable law, of course), the current study concerns itself with identifying any political or institutional conditions that make it more likely a court will find for the plaintiffs in an adequacy lawsuit. This is viable because, given the similarities in constitutional language, the “law” is relatively consistent across cases. Importantly, however, the different multi-venue strategies pursued in each state introduce some political variability into relatively consistent legal and institutional factors.

Studying Judicial Decision-Making

Courts and Social Change

The study of judicial decision-making in public policy-oriented cases rests implicitly on the assumption that court decisions matter. Clearly, no one doubts that they matter for the individual litigants; the plaintiff either wins or loses (although some decisions give something to both sides, on the central issue there can generally be only one winner), and that decision carries the force of law. However, efforts that focus upon

analyzing and predicting judicial decision-making would seem to be making a more expansive argument, that judicial decisions matter *beyond* the named litigants, or, in others words, that judicial decisions matter because they set—or at least constrain—public policy.

Epstein and Knight (1998) and Rosenberg (1991) examine this question of whether judges can influence public policy, and they seem to come to opposite conclusions, with Epstein and Knight arguing that judicial decisions help shape public policy, and Rosenberg arguing that judicial decisions follow, rather than shape, public policy. While these works are usually seen as making opposite arguments, it is far from clear that that is actually the case, especially when one considers the different samples of cases that each book examined.

Epstein and Knight and Rosenberg used very different samples in their research. Epstein and Knight based most of their analysis on a single term (1983), and also used a second sample of landmark cases from the Burger Court (1969-85). Their single term consisted mostly of routine cases. Rosenberg, on the other hand, concentrates on substantive policy areas that nearly everyone would consider important, and follows those areas through multiple court terms. An entire section of the book is devoted to civil rights and another to abortion and women's rights, with extensive attention paid to *Brown* and *Roe*, respectively.

Therefore, while acknowledging that Epstein and Knight paint a much more optimistic picture of the judiciary's ability to lead social change than does Rosenberg, those differing conclusions need to be viewed in light of the different samples employed. Epstein and Knight intend for their book to serve as a wake-up call for judicial politics

scholars to pay attention to strategy, and to take seriously Walter Murphy's (1964) seminal work about strategic judicial behavior. They focus, therefore, on demonstrating that strategic behavior doesn't only take place in landmark cases. Rosenberg, on the other hand, focuses explicitly on landmark cases. As a matter of fact, he cites a study of educational litigation (Rebell and Block, 1982) as providing support for a "dynamic courts" hypothesis, but seems to consider education litigation after *Brown* unimportant. Rebell and Block, however, were writing before adequacy litigations began, and didn't restrict themselves to school finance litigation.

While both Epstein and Knight and Rosenberg deserve to be examined in detail, it is important to focus on where adequacy lawsuits would fit on the continuum of importance, whether Epstein and Knight or Rosenberg would select adequacy lawsuits to prove their point. Given the importance of education and adequacy decisions to state politics and finances, as discussed above, it would seem that adequacy lawsuits qualify as "major" or "important" litigation.

Epstein and Knight begin, as I do, with David Mayhew's examination of congressmen. As they are writing about the Supreme Court, Epstein and Knight emphasize the intergovernmental relations in which the court is enmeshed, as well as the court's need to maintain legitimacy. Reelection is obviously not a concern. Legitimacy, though, is two-pronged, as the Supreme Court needs to maintain legitimacy with both the other branches of government and the American people.

In order to demonstrate the existence of strategic behavior on the Supreme Court, Epstein and Knight examine changes over time in Supreme Court decisions regarding the Civil Rights Act of 1871. They demonstrate that, initially, the court did not take their

most preferred liberal position, because such a position would have put the court too far to the left of Congress and the American people.

Epstein and Knight do, however, leave two questions unanswered. The first is the extent to which their conclusions would apply to other courts. Admittedly, this was not their purpose, but, as I discuss below, it is by no means certain that insights about the US Supreme Court translate to state supreme courts, especially given the special reverence for the Supreme Court and the American Constitution. Additionally, Epstein and Knight set out to debunk Segal's attitudinal model of judicial decision-making, but, in the end, they leave the role of the ideology of individual judges only partially examined. The attitudinal model could best be described as positing that a liberal justice takes a certain position because he or she is liberal. Although Epstein and Knight demonstrate that there is strategy in judicial decision-making, they don't really debunk the attitudinal model. A liberal justice, after all, would adopt a certain strategy to get the most tenable liberal position, because he or she is a liberal. While Epstein and Knight explain how judges develop a voting strategy, at its core their argument serves to add nuance to the attitudinal model, not to refute it.

Rosenberg (1991) takes the view that courts almost never bring about social change. In studying substantively important policy areas, Rosenberg argues that courts generally cannot *bring about* social change as much as they set the stage for others to do so, although they are comparatively more effective in cases where the more political branches of government have failed to act. In articulating this viewpoint, Rosenberg emphasizes the two pathways through which courts can have an effect—the judicial route

(e.g. a ruling) and the extra-judicial route (for example, in legitimizing certain positions through opinion writing, or through granting cert to some cases and not to others).

The “constrained court” view that Rosenberg ends up espousing posits four reasons why courts will have trouble bringing about social change: “the limited nature of constitutional rights, the lack of judicial independence, and the judiciary’s inability to develop appropriate policies and its lack of powers of implementation” (p. 10). And therefore, in the rare case where it seems that courts have brought about social change, it is only because one of the constraints was uncommonly weak in that specific case, which itself would imply (given the nature of the constraints to be overcome) that the court was acting more as a legitimating authority of some other branch’s action than as a path-breaking advocate for social change.

In the case of adequacy and school finance, however, one of the constraints—the limited nature of rights—might not apply. Indeed, it is entirely likely that theories of judicial decision-making that derive from the American Supreme Court are influenced by the peculiarities of that institution. Douglas Reed (2001), for example, writes that the “direct transfer of interpretative frameworks from the national to the state constitutions is not tenable,” (p. 56).

This literature ought to be explicitly connected to the research question at hand. Specifically, the framework behind the current study, and its examination of whether any political or institutional factors affect adequacy decisions, assumes that judges who prefer adequacy will be more inclined to find for plaintiffs where they perceive a supportive political environment that will facilitate implementation. We cannot directly measure the support that other actors feel for an adequacy decision, but it *is* possible to identify some

institutional conditions that might make for either supportive politics of implementation (for example, adequacy decisions for plaintiffs' in similar states) or non-supportive politics of implementation (for instance, divided government). A review of Epstein and Knight and Rosenberg provides some justification the first part of this premise—that judges need a supportive political environment to bring about social change—so that a review of political and institutional factors that might provide such support is relevant.

Differences Between State Supreme Courts and the American Supreme Court

The frameworks that illuminate the American Supreme Court and its interpretation of the Constitution might not be directly transferable to the state context because of differences between the Constitution and the various state constitutions in form and substance, as well as historical evolution concerning the proper way to interpret and assign meaning to constitutional language.

G. Alan Tarr (1998; and Porter 1988), a noted state scholar of state constitutional law, examines the structural differences between the national Constitution and points out the practical implications of those differences for state high courts and constitutional interpretation. Of primary importance is that state constitutions are grants of plenary legislative powers. Whereas the national constitution is a listing of specific powers that the federal government possesses with the general understanding that other powers are reserved to the states, state constitutions operate in reverse. Plenary power refers to the fact that state constitutions grant states all the powers that are not explicitly prohibited, explicitly ceded to the federal government, or expressly prohibited by the federal constitution.

Tarr discusses other differences between state constitutions and the national Constitution as well. First, state constitutions are much longer and much more detailed than the federal constitution. Indeed, state constitutions tend to read more like statutes than statements of principles or broad outlines for the functioning of government, which presents a quandary for state supreme courts looking for an appropriate framework for constitutional analysis (Gardner, 2005).

Secondly, while both the federal and state constitutions recognize the existence of separation of powers, the federal Constitution separates powers by implication (it gives different powers to different branches of government), but does not explicitly state and define what a separation of powers entails. Many state constitutions, on the other hand, include a separation of powers clause. Tarr uses Wyoming as a representative case⁵. Clearly, the federal Constitution has no such clause, which is one reason that Chief Justice Marshall had the freedom to institutionalize judicial review. It is important, then, to consider the meaning of an explicit separation of powers clause. The explicit mention of separation of powers in state constitutions implies, at least in some states, that courts will give separation of powers concerns added deference, when compared to an analogous case in front of the United States Supreme Court. To put this in the terms that Epstein and Knight used, state supreme court judges might have comparatively less freedom to act as judicial policy-seekers, as they have to more explicitly avoid (or, at least, take pains to justify) decisions which might be seen as veering off into “legislating.” In terms of Mayhew’s model and its modifications, discussed in the

⁵ The relevant clause of the Wyoming State Constitution reads “The powers of the government of this state are divided into three distinct departments: The legislative, executive, and judicial, and no person or collection of persons charged with the exercise of powers properly belonging to one of these departments shall exercise any power belonging to either of the others, except as in this constitution expressly directly or permitted”.

previous chapter, Tarr's insights regarding separation of powers could conceivably imply that state supreme courts would pay very close attention to factors in the institutional prestige portion of the model.

A final difference that Tarr identifies between the federal Constitution and state constitutions concerns the attention paid to public policy issues. The federal Constitution says little about public policy—either in terms of means or ends. The broad outlines of the federal government and its power are prescribed by the Constitution, which partially addresses the means that the federal government can use in pursuit of public policy, but the Constitution does not commit the federal government to pursue any particular policy. This is in sharp contrast to state constitutions, which often lay out specific policies that the state must pursue. Tarr points to California, which incorporated an entire statute (The Marine Resources Protection Act of 1990) into its constitution, thus transforming what had been an ordinary statute to the level of constitutional law.

Recall that Mayhew's model, modified for use with courts, predicted that judges would respond to three sets of incentives—retention, prestige, and “good” law as they saw it, which involved seeing their preferred policy positions made into law. It seems clear, though, that state supreme court judges have a lesser degree of leeway than their federal counterparts, even in areas that require constitutional, as opposed to statutory, interpretation. If a state constitution either directs or prohibits a certain policy, judges lose the ability to use their preferred policy position as a starting point, and have to work within the boundaries that the constitution sets out. Even if those boundaries are rather vague, like “adequate” education, they still exist to a greater degree at the state level. The implication, then, is that the second plank of Mayhew's model would seem to take

precedence over the third, because the institutional context becomes more salient than judges' policy preferences.

Also of note is the ease with which state constitutions are changed compared to the federal Constitution. Whereas the meaning of the federal Constitution is clarified—and some would say constructed—through years of precedent that builds upon itself while the text stays relatively constant, the opposite process seems to govern state constitution. Tarr writes of the “states’ reliance on the formal mechanisms of revision....and amendment to promote constitutional change,” (p. 33). This finding would seem to have two implications for state supreme courts in adequacy decisions. First, state supreme court decisions that rely on constitutional rulings are likely to be considered less final and immutable than comparable decisions by the United States Supreme Court. After all, part of the respect accorded US Supreme Court decisions derives from the sense of legitimacy that the Constitution enjoys. Were the Constitution not as well respected, legal decisions based upon it would be less respected as well. Second, in terms of Mayhew’s model, the relative impermanence of state constitutions would impact both the prestige of the court, and its ability to see its preferred policy position made into law. With regards to prestige, the knowledge that the constitution could be amended in response to a ruling with which the legislature (or the citizens themselves) disagreed, a state supreme court is on unequal footing when engaging in institutional combat, and must consider the threats to its prestige and legitimacy that would result from its being consistently overruled via constitutional amendment. Thus, the politically acceptable range of decisions for a state supreme court would be limited, and could conceivably not include the court’s most preferred policy position. The only

way that the court could make a politically unacceptable ruling without losing prestige and legitimacy would be if the judges were sure that the public supported court action, and would vote out legislators who opposed it or otherwise act to secure the court's ruling. That calculation would be difficult for a branch of government that is not in regular contact with its constituents in the same way that legislators or governors are.

The Influence of Other High Courts on State Supreme Court Decision-Making

In the process of attempting to develop a new theory of state constitutional interpretation, James A. Gardner (2005) makes an observation that gets at the heart of what he considers the problem with state high courts. He contends that, too often, they fail to adequately explain their constitutional reasoning, or even to reason deductively from *their own* constitution. Gardner identifies *McDaniel v. Thomas* (1981), a case in Georgia where plaintiffs were alleging that education is a fundamental right under the Georgia Constitution, as an example of this phenomenon. As discussed in Chapter 1, the US Supreme Court in *Rodriguez* held that education was not a fundamental right under the United States Constitution, but left the door open for state courts to interpret their own constitutions. According to Gardner (p.9), the appropriate outcome in Georgia is clear, as the Georgia Constitution provides that "The provision of an adequate public education for the citizens shall be a primary obligation of the State of Georgia."

The Georgia Supreme Court, however, does not acknowledge that text in its decision. Instead, they decide that "The fact that education is not a 'fundamental right' under the U.S. Constitution provides some guidance to the state. Consistency in constitutional adjudication, though not demanded, is preferred." The Georgia Supreme Court, in other words, looked not to its own constitution, but to the US Supreme Court's

earlier decision in *Rodriguez*, even though the *Rodriguez* decision was based on a different constitution. Gardner, writing a normative theory of state constitutional interpretation, sees this as an example of federalism subverted. The decision is interesting here, though, for a different reason. It demonstrates that there is a human, and inherently political aspect, to state supreme court decisions. We can assume that the judges knew the language of their own constitution. The puzzle, though, is that they cited the decision of a different court, instead. Assuming that this diffusion of judicial decision-making is purposeful, judges would only engage in it if they thought that they would benefit. In this case, there is conceivably a legitimacy benefit to the Georgia Supreme Court, as they were able to make a decision that borrowed the accumulated capital and legitimacy of the US Supreme Court. Another possible explanation is that, although the legitimacy provided by the US Supreme Court didn't hurt, Georgia Supreme Court judges had a policy preference and were engaging in their version of forum-shopping. Advocates can choose which forum to use, just as courts can choose which other courts to look to for precedent.

Tarr and Porter (1988) refer to the process of state supreme courts citing each other as “horizontal federalism” (as opposed to the “vertical federalism” that exists between state supreme courts and the US Supreme Court), although it might be more easily understood as policy diffusion among state supreme courts. As constitutional law in other states is not legally binding on a state supreme court, the decision of when to cite another state supreme court, and which court to cite, reflects a choice. The circumstances under which rulings diffuse from court to court will be modeled in the current study.

Peter Harris (1985) was among the first to study this phenomenon among state supreme courts by studying patterns in intercourt citation among state supreme courts. Much like the Walker (1968) model of policy diffusion, Harris found that courts were more likely to cite decisions of geographically proximate state supreme courts. While his conclusions seem well-suited to the time in which he was writing, the explanation he gives raises the possibility that his conclusion might no longer be valid. After all, he points to the publication of adjacent states' court decisions in the same volume or set of volumes, which the citing judge was more likely to have read through. The advent of computer databases since Harris' initial foray into the area probably renders the conclusion anachronistic, especially since there is a limited theoretical basis that explains that finding. This is not to say that state supreme courts might not pay special attention to their neighboring states, especially if they consider those neighboring states to be similar in some way. However, Harris' theoretical basis for this conclusion might be false, even if the conclusion itself still holds.

Canon and Baum (1981) pioneered a second approach to studying patterns of diffusion among state supreme courts. They study the adoption of actual reforms in judicial doctrine, specifically tort reform. In a sense, Canon and Baum have the more reliable tool, because courts will adopt only the doctrine with which they agree, but might merely cite decisions of other state supreme courts either because they agree or because they are building a straw man to knock down. Canon and Baum failed to find a pattern. They found that region did not play an important role, but, on the whole, some courts were more innovative than others.

Although Harris and Canon and Baum were the pioneers of this research, others have continued to explore the topic more recently. Dear and Jessen (2008) and Kritzer and Beckstrom (2007) observe that intercourt citation analysis is a rather blunt instrument, because intercourt citation studies did not also examine whether the cited decision was actually implemented or heavily relied upon by the citing court. It would be more helpful, they argue, to model actual implementation of a decision by other state supreme courts, thus providing a fuller picture of diffusion. Dear and Jessen (2008), examining 24,000 state high court decisions that had been “followed” at least once as identified by Sheperd’s Citation Service and Lexis-Nexis, find that the Walker model of diffusion to geographically adjacent states is no longer valid, and they speculate that the reason has to do with improvements in communication and electronic legal reporting since the time of Walker and his contemporaries—improvements that, among other things, make their study possible. Instead, they find a pattern of diffusion of state supreme court decisions among states in the same federal circuit. This finding would seem to provide more rigorous empirical support for Gardner’s (2005) position discussed above, and would also suggest a variable to be modeled. If diffusion occurs through federal circuits, then state high courts might be more likely to find in favor of plaintiffs in an adequacy lawsuit if other states in their circuit have already done so.

Bird and Smythe (2006) reached a similar conclusion in their study of the diffusion of court’s adoption of three exceptions to the employment-at-will rule in American employment law. They find diffusion among state supreme courts within the same federal circuit, and they also find, contrary to their hypotheses, that political and economic factors did not play a significant role. In interpreting their results, however, it

is important to remember that it is a case study of the adoption of one type of legal standard related to one narrow type of law, and it, like any case study, might be idiosyncratic. Indeed, Bird and Smythe make no real claim of generalizability, although they do claim their study is important in its own right. From a legal perspective, however, state supreme courts within the same federal circuit might be inclined to follow one another because the federal circuit court plays a *potential* role. Employment law cases can eventually wind up in federal court, and would conceivably, if it were a landmark case, wind up in the federal circuit court. Put another way, state supreme courts might find a type of safety in numbers here, whereby it becomes easier to adopt a ruling once other states have done so without finding that they have been overruled.

While the federal circuit courts have potential jurisdiction in employment law cases, they have no such potential jurisdiction in adequacy cases, as the *Rodriguez* decision foreclosed the possibility of adequacy cases winding up in federal court. Therefore, the incentive for state supreme courts to look first to their federal circuit might not apply in adequacy cases. Interestingly, though, both Dear and Jessen and Bird and Smythe discuss, in slightly different terms, the issue of shared legal history among states in the same circuit as a more solid theoretical reason for diffusion or “borrowing,” as Dear and Jessen refer to it. They imply that diffusion within federal circuit regions is due to a kind of similar trajectory in legal development and constitutional interpretation that is common to that circuit. If that is the case, then it stands to reason that diffusion—due to similarities in habits and styles of interpretation—would persist within the circuit even in cases where there is no risk of a federal court ruling. This would suggest that federal

circuit be included as an independent variable in the empirical approach I'll be describing in the next chapter.

If the theoretical underpinning of the diffusion argument rests not on physical proximity but rather on perceived similarity or shared historical experience, as prior research seems to imply, then it is useful to consider what other groupings might serve as possible conduits for judicial diffusion. Political culture is one such variable, as it serves as a proxy for shared experience and norms that could conceivably lead to shared values regarding the role of the judiciary and the proper distribution of resources, or forum for deciding that distribution. In a study that is similar in many ways to the current one, Roch and Howard (2008) employ a political culture variable, based on Elazar's three categories. I will review the full range of similar studies below, but at current, it suffices to say that Roch and Howard were concerned with *any* court intervention at all regardless of the legal theory involved, and that they used a political culture variable as a proxy for many social and political characteristics that could usefully be subsumed under "culture." They define political culture as "what people believe and think about their government," (p. 334). One of the advantages of Elazar's typology is that it is widely used by other researchers to successfully link political culture and policy preferences (see Roch and Howard, p. 334 for a description of those studies). By using Elazar's typology, the current study could fit in with other studies on political culture and policy.

Another conduit for diffusion, especially in the judicial context, conceivably is the constitution's education clause itself. If the original hypothesis behind diffusion was that state supreme courts would look to the decisions of courts that were similar in some way in order to provide additional legitimacy for a decision, then, in a legal context, a similar

legal starting point ought to play at least some role. Thro (1989) developed a four-category typology of educational clauses in state constitutions, based upon the strength of the obligation they impose. If, as Thro contends, the education clauses in each of his four “categories” are similar, then we can reasonably expect legal diffusion to take place through that channel. If the logic of diffusion through a federal circuit holds because of perceived similarities in the legal context, it should also hold that judges, before deciding an adequacy case, will look to decisions in states that share the same constitutional language. If a court can cite a prior decision in another state with similar constitutional language, it serves to legitimize the decision, presumably both to the legislature and to the population at large.

State Supreme Courts and the Separation of Powers

State constitutions provide for a separation of powers, whereby the legislative, executive, and judicial branches have their own specific functions. Indeed, as discussed above, many state constitutions are even more explicit on that point than the federal Constitution; where the federal Constitution implies a separation of powers by its design of institutions, many state constitutions explicitly provide for a separation of powers.

Separation of powers could seem, at first glance, to present a problem for courts that rule in favor of plaintiffs in adequacy lawsuits. Indeed, every court that has decided to decide an adequacy case based on the substantive question at hand has found for the plaintiffs, other courts have declined to decide on the basis of separation of powers and justiciability concerns. The separation of powers concern is particularly cogent because some adequacy rulings (Kentucky, for example) read like something that might have come from a legislature. Kentucky’s definition of adequacy (quoted in Chapter 1) lays

out the court's definition of what an adequate education should entail, and the characteristics are expansive. This seems to stretch the court's traditional role of deciding who wins the case at hand, an issue on which the court could have ruled without also explicating the concept of an adequate education. Aspects of the Kentucky Supreme Court's definition, for instance the fifth component that requires "sufficient grounding in the arts to enable each student to appreciate his or her cultural and historical heritage," seem on their face to be questions of policy priorities that the legislature could, if it so chose, establish via the legislative process.

Certain courts are more willing to make those types of rulings than other courts. While the term "judicial activism" is normative—and pejorative to some—it does illustrate a certain ideological position on the role of the courts, and judges holding that position are more likely to have an expansive view of the court's proper role relative to other branches of government. While not all activist judges are likely to be in favor of adequacy, it seems unlikely that any court could find for an expansive definition of adequacy without facing accusations of judicial activism.

Equally interesting, as discussed above, is that, in every case, when courts have allowed adequacy cases to proceed to a finding on the *merits* (as opposed to dismissal on procedural grounds), courts have found in favor of the plaintiffs. Courts have never found in favor of the defendants solely on adequacy grounds, but rather where courts have found in favor of the defendants, they have invoked separation of powers, political question, and related justiciability arguments to do so (see also Chia and Seo, 2007 for a study that considers equity and adequacy lawsuits together). Simon-Kerr and Strum

(2009)⁶ find, somewhat differently, that courts in Massachusetts and Texas have both reversed lower court rulings in favor of plaintiffs, and ruled that the education provided was constitutionally adequate. Importantly, though, the decisions in those cases, while finding that the education was constitutionally adequate, referenced the court's concerns with separation of powers and the court's proper role. Issues about the proper role of the courts were an even more significant part of follow-up compliance cases, where the court was being asked to take on a larger enforcement role.

If courts that decide only on the merits find for the plaintiffs, and other courts frequently defer to separation of powers, political question, or justiciability concerns, clearly courts are thinking about separation of powers arguments, and whether they want to take on the legislature in a form of institutional combat, as Ginsberg and Shefter (1990) use the term. Had those courts really those that education was adequate, the path of least resistance would have been to examine the evidence and then simply say so.

Furthermore, Douglas Reed (2001) points out two particularly illuminating incidents demonstrating that courts take separation of powers concerns into account both when finding for the defendants and when finding for the plaintiffs. In ruling for the defendants, Justice McWade of the Idaho Supreme Court wrote that ruling for the plaintiffs "would be an unwise and unwarranted entry into the controversial area of public school financing, whereby this court would convene as a 'super-legislature'." More interestingly, in New Hampshire, even though the court found the existing school finance system inadequate, the court did not remand for a remedy, but rather stayed all further proceedings until the end of the legislative session to allow the legislature to act

⁶ Simon-Kerr and Strum concentrate specifically on the period 2005-2008. Data collection for the current study, however, ends in 2005 due to the availability of PAJID scores.

on the ruling of inadequacy. Even when finding for the plaintiffs, the New Hampshire supreme court showed deference to the legislature and let the political process work, by making a finding but letting the legislature come up with a remedy. And finally, even the Kentucky Supreme Court, after making an extensive list of what constituted adequacy, maintained that they were not mandating any specific piece of legislation or tax increase.

In this separation of powers context, neither the legislature nor judiciary wants to lose prestige. Legislatures have, in the course of their debates, explicitly considered what the court might rule, and courts have explicitly deferred to the legislature. Short of mind-reading, we can never be certain of the judicial motivations, but it is possible to speculate intelligently about the types of situations in which the state supreme court is likely to rule for the plaintiffs or defendants, respectively.

One obvious factor to consider is divided government. A state is under *unified* government when the governorship and both houses of the legislature are of the same party, and under divided government in all other circumstances⁷. Mayhew (2005, 2nd Ed.) has made the point that, on the national level, divided government can be overcome where the issues involved are important enough, but, on the whole, divided government is associated with gridlock. In light of the redistribution that adequacy implies, it is likely that divided government will preclude legislative action to provide adequacy. This would leave the court free to act with some added confidence that the legislature would not take any bold action, either to undermine the court's ruling, or to implement it if

⁷ I understand that it might be interesting to code the different possible subtypes of divided government (a governor of one party, and a unified legislature of the other party, for example, or a governor and a lower house of one party and an upper house of the other party). However, the limited possible sample size (N=49 states with adequacy provisions), the limited diversity of institutional and party arrangements within those states, and the ideal number of variables in QCA for this N before the procedure breaks down (about 5-7, see chapter 3), preclude a separate examination of all the different subtypes of divided government.

implementation necessitates any kind of wholesale changes. If we accept the premise that a ruling for the defendants often reflects separation of powers concerns, then divided government should lead to a decision for the plaintiffs, a proposition which this study is in a position to test. Implicit in this hypothesis is the classification of adequacy as an issue that is not important enough to overcome the barriers to action that divided government presents.

The ideology of the court itself should also have an effect on the decision, because ideology would likely inform both judges' views about the just distribution of resources and the propriety of forced redistribution that adequacy would require, as well as judges' views about the proper role of the judiciary in a federal system. In recent history, both judicial activism and greater equality have been considered "liberal" attitudes. While there have been periods of conservative judicial activism, namely in the Lochner Era prior to the New Deal and in the current Roberts Court, the time period of the current study (1989-2005) supports a hypothesis that links judicial activism with a liberal ideology, because the two were closely linked during that time period. Given these conditions, it stands to reason that liberal courts will be more likely to find for plaintiffs (a proposition that the current study will be in a position to test) because, among other things, liberals tend to view an adequacy ruling as being within the proper role of the courts and are also likely to favor the redistribution involved. Brace and Hall developed a system of measured the ideology of a court called the PAJID score (for PArtisan Judicial IDentification), which this study will employ (see, for example, Brace, Langer and Hall, 2000).

The "Electoral Connection" and State Supreme Courts

By design, the subheading should remind the careful reader of David Mayhew's work on Congress. However, I use it here mainly to illustrate the differences. Whereas life in the House of Representatives was centered on getting reelected every two years, state supreme courts are somewhat different, although no less political. If the judicial selection has an effect, the effect could manifest itself in one of three ways. The selection method could determine the types of people who become state supreme court judges (more or less minorities, more or less women, etc.). The selection method could affect the types of decisions that judges make once on the bench. And lastly, the selection method could affect the perceptions that other actors hold of judges and their actions. For example, if the perception is that judges selected through a partisan election will behave differently, and that perception colors the expectations and actions of other actors, then that perception becomes important regardless of whether or not it is eventually borne out by the facts. The perceived effect of selection method could become its actual effect.

There are four general methods of judicial selection used for state supreme courts: partisan election, non-partisan election, merit selection, and appointment (Choi, Gulati & Posner, 2008; see Berkson, 1980 for a detailed description of the historical context which gave rise to these systems). In appointment systems, the governor or state appoints judges, sometimes in consultation with the state legislature. In merit selection systems, a nonpartisan body or commission appoints judges, who then face an uncontested up-or-down retention election, usually after a year in office. In other systems, judges are elected. These elections can be further broken down into partisan and non-partisan elections. While it might be useful to include separate variables testing the effect of each different selection system, the concerns about sample size, number of variables, and

limited diversity that I discuss in Footnote 6 above (and in Chapter 3) still apply.

Therefore, it is imperative that a specific hypothesis is tested as to the effect of judicial selection, and not merely that “it matters.”

The type of selection system employed could conceivably affect the types of judges that are selected, which is important because some theorists hold that personality or background will influence judicial decisions. If getting considered in a merit selection system meant having access to the elites who make up a committee, we might expect judges to be disproportionately wealthy and white. However, current research (Glick and Emmert, 1987; Cann, 2007) does not show that to be the case. Additionally, no selection system is better at picking out a qualified and competent judiciary than any of the others. There does appear to be limited diversity on state supreme courts (Glick and Emmert, 1987), but this appears to be due to factors other than the method of judicial selection, and might have more to do with the need to recruit more minorities into law schools and the practice of law in the first place.

There is evidence that suggests that the selection system affects how judges rule. Pinello’s (1995) book on the subject was one of the first scholarly attempts to study the effects of judicial selection method. While his comparative case study approach demonstrated that, at least in his cases, judicial selection method mattered for judicial decision-making, other more recent work has attempted to build a more generalizable model.

In a review of the relevant literature, Roch and Howard (2005) write that “Scholarship shows that the selection and retention processes also are determinants of outcomes...and that elected judges often act as representatives of the electorate,” (p.

336). Additionally, they discuss recall voting in retention elections as evidence that the public is informed about judges and their rulings, and that rational judges would behave accordingly. One can infer, then, that the selection system *does* affect how judges rule, and that judges will take the policy preferences of voters into account, because it is still relatively rare that judges lose retention elections, in spite of the fact that voters do seem to be aware of the rulings that these judges make.

There is evidence to suggest that partisan judicial elections are particularly competitive. Hall (2001) even concludes that partisan judicial elections are just as competitive as those in the House of Representatives (which, although surprising, needs to be viewed in light of the typical bias towards the status quo—the incumbency advantage—in House elections). Perhaps more significantly, Schotland (2007) writes at length about the deteriorating tone of judicial elections, and the increasing reliance on attack ads and other trappings of normal political campaigns that had previously been viewed as unbecoming a judicial candidate. He advocates lengthening judicial terms as the solution, so that the public perceives judges as relatively free of political pressure.

The perceived effect of judicial selection method might be nearly as important as any actual effect. Since courts can only rule on cases brought to them, it is conceivable that perceptions of probable judicial behavior might determine whether a case is filed, and therefore, the universe of cases under study. Additionally, these perceived effects of selection method could shape the socialization and behavior of newly selected judges. Cann (2007) conducted a study of the perceived quality of state supreme courts by surveying judges themselves, and comparing their perceptions of their state's courts with the structural characteristics of those courts. Cann hypothesized that judges from merit

selection and appointment states would rate the quality of courts in their state higher. His hypothesis is confirmed, as he finds that “judges in states where most judges are elected by merit selection or appointment rate their state court system significantly higher than judges in states where most judges are elected in partisan elections,” (p. 230).

Furthermore, he finds that partisan and non-partisan election states are not significantly different from each other.

There is also some evidence that judicial selection method does not have an effect, at least where merit selection is compared to elections. Merit selection was a progressive reform, intended to shield judges from politics. In a merit selection system, judges run unopposed in a retention election after their first year on the bench. It might be expected, then, that merit selection performs differently than regular judicial elections, or so the progressives hoped. Cann (2007), however, writes that “elections may be no worse than merit selection in terms of accountability and independence,” (p, 228). In other words, whatever incentives judges are subject to by having to face the voters, according to Cann’s interpretation of the literature, are equal in an election or merit selection system. Cann does not, however, distinguish between partisan and non-partisan elections. Additionally, Cann looks at accountability and independence, but not directly at the types of policies that result from each selection system.

The appropriate hypothesis for the current study, then, would seem to be that elected courts (whether partisan or not) will behave differently than courts selected either by merit selection or by appointment. Such a hypothesis would be further supported by evidence that there was some opposition to adequacy lawsuits on the part of the public—or, at least, the part of the public more likely to vote in judicial elections and/or contribute

to judicial campaigns. A useful starting point is Douglas Reed's (2001) observation that citizens tend to resent judicial interference in their school affairs, as it impinges on local control. He implies that, even in the case where a court is doing something good, it can still breed resentment because it comes in the form of a court order. And Molly Hunter (1999), in a case study of the *Rose* decision in Kentucky, lauds the plaintiffs' public relations work in gaining business community's support for adequacy, but acknowledges in the footnotes that, according to a Roper Center survey, one-third of Kentucky residents were *not* prepared to pay higher taxes in exchange for better schools, even *after* the outreach effort (compared to 53% before). Furthermore, Hunter does not examine whether those who would support higher taxes in exchange for adequacy are the likely voters or contributors in judicial elections, so that an examination of the incentives facing judges in Kentucky is not possible. In any event, Kentucky, as the site of the first successful adequacy lawsuit, may anyway be something of an anomaly, in that public opinion on adequacy litigation may have hardened in the aftermath of the *Rose* decision.

Swenson (2000) takes a different view, arguing that "public opinion poll data sometimes indicates that the majority of the public favors school finance reform; *aside from the minority of students residing in the richest school districts*" (emphasis mine). Swenson interprets this to mean that the majority of the population supports school finance reform. But, neither Swenson, nor anyone else in the literature that I could find, makes the connection between this public opinion and the likelihood of involvement in judicial elections. In the absence of any evidence to the contrary, then, it is safe to assume that judicial elections follow the pattern of elections generally, in that the wealthier are more inclined to vote and to be otherwise engaged in the political process.

By Swenson's own review of the evidence, those wealthy people, living in the richest school districts, oppose school finance reform. That is likely a force that elected judges will take into account, as the rich will be the ones to contribute to a judicial campaigns.

Rankings of State Supreme Courts

I have alluded above to the limited number of possible variables (or, as QCA defines them, necessary and sufficient conditions) that are feasible in the current study, due to the small sample size and limited number of causal combinations. There are, however, a large number of possible variables that could influence court rulings—anything from docket size, to professionalism and compensation. In short, there are a number of ways of evaluating the performance of state supreme courts. The next logical question, then, is whether better-performing courts are likely to decide differently than other courts—either because they have more political capital to spend on an adequacy ruling or because they are actually better at what they do and can therefore more easily grasp the complexity involved in adequacy litigation and costing out.

The National Chamber of Commerce ranks state courts yearly on a number of measures based on a nationwide representative survey of lawyers. For our purposes, though, those rankings are useless because they consider the performance of the state court system as whole and not merely the state supreme court. One could also question the technical expertise and information that the National Chamber of Commerce, a trade group, has in regards to state judiciaries. Other rankings consider mainly the rate at which state supreme court opinions are cited by other courts as the sole measure of judicial quality (for example, Caldeira, 1983).

Choi, Gulati and Posner (2008), however, create a composite measure of the performance of state supreme courts. They measure productivity, influence or opinion-quality, and independence, and also take account of state-specific factors that might affect the types of cases a court sees and consequently the likelihood that it would have the chance to make influential and widely-cited decisions. These factors include urbanization and the amount of commerce that takes place in a given state. By productivity, they mean the number of opinions written each year, per judge. By influence or opinion-quality, they are referring to the extent that a court's decision is cited in the majority opinion of another court. Choi, Gulati and Posner use "influence" and "opinion quality" simultaneously, on the theory that courts will only cite opinions which they consider to be of the highest legal quality, and therefore influential opinions are high-quality opinions. To measure judicial independence, Choi, Gulati, and Posner take into account judges' partisan affiliations (even in courts selected in a non-partisan manner), and give judges a high score if they are more likely to vote with a member of the opposing party, and low scores if they are not. The theory behind their measure is that "a judge exhibits independence when she writes an opposing opinion against a co-partisan," (p. 11).

Including the composite measure will allow for a more comprehensive study while still keeping the number of variables manageable for a QCA analysis. In the following chapter, it will become clear how this variable will be transformed into the binary format required for QCA's analysis of necessary and sufficient conditions.

Studying Adequacy

Attacks on the Adequacy Concept

Thus far, although I have examined judicial behavior in an effort to understand rulings in an adequacy case, I have spent relatively little time on the concept of adequacy itself. Part of the reason is because the adequacy clauses themselves are ambiguous, as are attempts to quantify and measure adequacy, as I discussed. Additionally, adequacy clauses are not *legally* defined until a court interprets the relevant clause of a state constitution and says what the constitution means. Nonetheless, as I discuss in Chapter 1, there is some broad agreement about what an adequate school system might accomplish, generally built off the *Rose* decision in Kentucky. It is important, however, to pay some attention to those who attack “adequacy” as a useful concept, or adequacy lawsuits as a way to achieve equity and excellence for all students. Understanding these objections to adequacy lawsuits might, at the very least, allow future litigants to construct better arguments.

Fundamentally, adequacy lawsuits are about money. Plaintiffs’ arguments might be illustrated with examples of concrete differences between well-funded and under-funded districts (old schools and crumbling infrastructure in the Bronx, and new technologically-sophisticated buildings in Scarsdale, for example), but the proposed remedy nearly always—directly or indirectly—involves money, either by mandating more of it directly or by mandating a definition of adequacy that will necessitate the spending of money. The central role of money partially explains the heated controversy over the different methods of costing out discussed in Chapter 1. Part of the reason for an emphasis on money is surely an effort by both courts and litigants to keep the courts out of the complexity of school and district budgeting—keeping in mind that the current study focuses on *initial adequacy* decisions (as opposed to follow-up cases related to

compliance) one could hardly imagine, for instance, a court initially ordering a state to provide x number of new textbooks, and hire y number of teachers, although this has happened in follow-up cases where plaintiffs allege that the state has not complied with previous adequacy rulings. There is, though, an implicit premise in plaintiffs' arguments that adequacy is equated with either more money, or concrete improvements that cost money. As Michael Heise (2005) points out, "the legal remedies sought by districts pushing adequacy lawsuits invariably involved resource increases." And although the previous chapter established that money matters, nearly all researchers would argue that other factors mediate the effects of money. Some representative objections are discussed below.

Roza and Hill (2006), for instance, would agree that money matters. They fault the logic behind adequacy lawsuits on two counts. First, they argue that current district spending and accounting practices make it difficult to get an accurate figure of what is being spent on the average student, since so much money goes either to exceptional students or to central office employees, who do not add equal value to all schools in a district. A failure to get an accurate idea of spending beyond the crude per-student average makes it hard to engage in the type of marginal analysis that is the backbone of good economics. More importantly, though, the focus on district-level spending in adequacy lawsuits ignores significant intra-district variation. According to Roza and Hill, complaints about the under-funding of inner-city school districts were at the core of the plaintiffs' cases in New York and Texas. Yet, in each state, they found that the level of *intra*-district variation in those districts was greater than the level of *inter*-district variation about which they were complaining. Roza and Hill, then, see devolution of

money to the school level, and an end to intra-district inequality, as a necessary precondition to adequacy.

Other researchers question the premise behind adequacy, because they point to schools or districts that are doing well with the same funding that others consider inadequate. Herbert J. Walberg (2006) identifies a number of schools that seem to “beat the poverty odds,” and presents a regression of poverty and proficiency among South Carolina districts to illustrate the relationship that he discusses. He does seem, however, to pay short shrift to the statistics underlying his analysis. Indeed, in any properly computed linear regression line, there will be cases above (over-achieving schools in his analysis) and below the regression line. The fact that some schools beat the odds does not, by itself, disprove the theory behind the computation. Walberg also discusses “African-American private schools” and Catholic schools, and how they overcome their poverty demographics by being more cost-effective. However, his description of teachers’ duties and of the ways in which money is saved ignores the very real fact that many of those reforms are not possible in the public schools, in part due to the power of unions. Walberg seems to be making an implicit argument, then, for vouchers or privatization in inner city schools, which could bring his entire analysis of adequacy litigation into question. After all, he is not only attacking the effectiveness of adequacy lawsuits, but also providing support for another controversial policy at the same time, and one is left to wonder which part of the argument came first.

Evers and Clopton (2006) make the converse of Walberg’s argument, in that they identify six high-spending and low-performing districts, and conclude that since these high-spending districts are low-performing, adequacy lawsuits will not work, because

school districts will not spend the money wisely. There are clear concerns about sampling—concerns that are present in any comparative case study—but a more fundamental flaw with the study concerns its mode of inference. Because corruption occurred to some extent in some of these districts, for example, does not mean corruption will occur in all districts that are high-spending. And because some districts spend money unwisely, it does not follow that all districts will. Most concerning, though, is that Evers and Clopton fail to give adequate credit to those who support adequacy litigation, writing in their first paragraph that “A premise of the adequacy campaigns is that it is *easy* to quantitatively measure and objectively determine the cost of providing an adequate public education...” (p. 103, emphasis mine). I can find no record of any education researcher writing that it is easy—even those ardently in favor of adequacy recognize the difficulties involved in costing out and the different mechanisms that can be employed to do so. Evers and Clopton, then, offer ideological, but not theoretical, objections to adequacy.

Previous Attempts to Study School Finance Litigation

Having discussed the theories behind adequacy lawsuits, and some representative objections to adequacy lawsuits, it would be prudent to review some of the previous research that attempts to identify the predictors of successful school finance litigation and the effects of such litigation.

Baker and Green (2008) briefly survey the literature that attempts to look at the effects of school finance litigation. Typically, the studies of school finance litigation have attempted to look at what might be considered an intermediate effect—the effect on the distribution of money, as opposed to the ultimately desired effect on student

achievement. Other studies have attempted to look at the predictors of plaintiffs' victory in school finance litigation, but have tended, with the exception of one recent study that I review below, to group adequacy and equity cases together into a single category of court-ordered school finance reform. And, as Baker and Green (2008) make clear, studies that attempt to do things other than case study research (e.g. statistical analysis or systematic comparative analysis) face an inherent trade off. Specifically, "studies of the influence of court ordered reform on spending level and distribution have not typically included detailed indicators of political context, such as party balance of legislatures. Others have attempted to measure the influence of political variables...but with little attention to legal constraints," (p. 313). It is worth noting that the current study faces that same trade-off between depth and breadth. Even though QCA is much more a qualitative rather than a statistical tool, it is subject to the same limitations of sample sizes and numbers of different variables, so that, in every study on the topic that is something other than a case study, a possibly important variable might be left unexplored.

Swinford (1991) was one of the earliest researchers to attempt a multivariate model of judicial decisions in school finance reform cases, although it is important to note that he combined equity cases with the first adequacy case, so that his universe of cases began with *Serrano* and ends in 1989. He broke down his model into court characteristics, strength of the existing law, and "legal history," which was essentially a crude diffusion model that looked at the number of previous cases in other states that had upheld the financing system, and subtracted those that had struck down the school financing system, on the rationale that courts will be likely to borrow from "sister courts." Note that Swinford included the standard deviation of the distribution of district

spending within a state as a measure of inequality, which, is subject to the pitfalls of district level figures that I discuss above.

He attempts to analyze results using probit regression. Unfortunately, he ran into many of the same problems that I once encountered trying to run data for the current study in a regression. Swinford had twenty-eight cases, and while he was able to get the regression equation to converge, he then had had to explain why he considered his study valid and important although nothing was statistically significant. He writes that “traditionally accepted indicators of relationship strength (t statistics) are less than ideal for this analysis. Such scores are somewhat ambiguous and problematic in terms of interpretation. For the purposes of this research, the most appropriate strategy is to pay attention to the direction of the relationship discovered...” He proceeds to do that, and reports that all the variables in his model perform in the expected direction with the exception of the “legal history” measure described above. He concludes that he developed a model that shows that “judicial decisions are the product of a host of interactive and competing forces,” (p. 348). Those variables that had effects in the expected direction included court characteristics (liberalism, propensity towards judicial activism), state constitutional characteristics (the strength of the educational and due process clauses), legal facts (standard deviation of district per-pupil spending, state ranking on spending per pupil), and legal history.

Swinford, conclusion, however, is puzzling because he interprets a lack of statistical significance for any of his predictors as indicating that many predictors matter and that they interact in various ways. It might very well be that standard methods involving testing for statistical significance were unsuitable for the study at hand, but,

given that the sample size was known at the outset, Swinford might have been better advised to select a different research method to use either instead of or in addition to a probit regression model. His conclusion, that is quoted above, might be correct, but is not necessarily supported by his empirical evidence, and he fails to consider seriously any rival hypotheses.

In performing his analysis, Swinford has offered very little that is actually new. It was well known, for instance, that liberal judges would be more likely to find in favor of an equity claim. More important for the current study, though, is Swinford's understanding of the limits of significance tests for such small sample sizes. Indeed, by using regression but ignoring standard statistical tools like significance tests or effect sizes, Swinford demonstrates the unsuitability of regression analysis for the purpose at hand. Additionally, Swinford observes that, even if he were inclined to use the regression output, the interpretation of coefficients in this context is difficult. One of the very reasons that I rely on QCA—a method that find both necessary and sufficient conditions—is because it is more easily interpretable. If one of the purposes of the current study is to guide advocates of adequacy, phrasing statements in terms of necessity and sufficiency (or even near-necessity and near-sufficiency, as discussed in Chapter 3) is more likely to be useful than providing difficult to interpret regression coefficients. It is also true that the current study, using QCA, might not find any necessary or sufficient conditions. However, QCA is at least a better tool for such an analysis, as it is designed to handle small and medium sample sizes.

Murray, Evans and Schwab (1998) perform a more sophisticated econometric analysis to try and determine whether school finance litigation that results in court-

ordered reform actually reduces inter-district inequality within a state, using four different measures of inequality. Importantly, they consider both equity and adequacy cases together, and create a binary variable signifying whether or not a state is under court-ordered reform. They conclude that court-ordered reform served to decrease within-state inequality, and that, in addition, the funds to reduce inequality were raised in most cases through higher taxes, and not through cuts to other programs. It is worth noting, though, that Wood and Theobald (2003) have criticized this study on the basis of failing to control for other possible variables that could explain these changes, and for using data at five-year intervals after the court decision, although Wood and Theobald fail to suggest any variables for which Murray Evans and Schwab ought to have controlled.

Murray, Evans and Schwab encounter two methodological issues which are also present in the current study. First, they recognize that some states (New Jersey, in particular) have had many state supreme court decisions, mainly because plaintiffs pursue litigation with regard to compliance and to persisting inequities or inadequacies. Nonetheless, they count each state only once, regardless of the number of decisions there have been. In addition, they code states only as in court-ordered reform or not. In explaining that decision, they recognize the ambiguity in the constitutional language that governs school finance, as well as the difficulty of coding complex legal decisions or legal re-arguments in any theoretically meaningful way. I employ a similar logic in the current study, counting only the initial decision in an adequacy lawsuit, for much the same reason. Once the legal reasoning behind the need for any subsequent compliance litigation was coded, there would be too many uniquely coded cases (indeed, depending

upon the depth of legal analysis, probably every case would be unique). As Murray, Evans and Schwab realized, counting each state only once and not attempting to code complex legal reasoning were the only ways to ensure comparability between cases, and to strike a balance of depth and breadth required for meaningful comparison.

Wood and Theobald (2003) analyze per-pupil expenditures from 1992-1996 in school all districts, classifying districts as either being in a court-ordered reform state or not for each of the years in the dataset. While their results must be viewed with some caution due to the limited timeframe, they nonetheless leave the reader pessimistic. While Wood and Theobald find that court ordered reform (either equity or adequacy based) does lead to greater inter-district equity in a given state, they find that the amount is statistically significant but not clinically significant, as “on average the overall allocation to local districts in states under judicial mandate was larger by about \$87.18 per pupil,” (p. 733). Furthermore, they find that “judicial mandates are generally unsuccessful in producing more equal allocations but are more successful when accompanied by receptive citizens and institutions,” (p. 718). This would seem to be a serious blow to supporters of adequacy lawsuits. Recall from Chapter 1 that Komesar theorized that disputes ended up in the institution that could best handle them. However, if courts are the option of last resort for the disadvantaged who cannot get relief in the legislature, but courts can only be effective if citizens want them to be, then it stands to reason that courts would have trouble bringing about social change. Clearly, courts can and do succeed in bringing about some instances of social change, but the point of emphasizing this apparent paradox is to illustrate the political and institutional barriers to such change that courts must work around.

However, in carrying out their analysis Wood and Theobald computed citizen ideology using a method first pioneered by Berry to infer citizen ideology. This method used the election returns from the most recent House of Representatives election, weighting each candidate's ideology score by the proportion of the vote he or she received to compute a statewide ideology.

Judges, though, have ideologies as well. And in states where judges are elected, those ideologies are generally available and known to voters, as discussed above. Even allowing for the fact that voters take factors other than ideology into account, it seems likely that, in states where judges were elected, their preferences would be closer to the elected legislators. While I understand that the time horizon for House elections is generally shorter than for judicial elections and that House elections are probably the best reflection of the mood of the electorate, it still seems that Wood and Theobald would have been able to make a stronger causal argument about the role of ideology if they had included a variable about the judicial selection mechanism. Even if Wood and Theobald didn't care about judicial selection per se, the relationship between the court's decisions and the general public could be different depending on the presence or absence of competitive elections.

Springer, Liu and Guthrie (2009) take the analysis of court-ordered school finance reform one step further, and code equity and adequacy decisions separately. They find that, while court-ordered school finance reform does tend to lead to a decrease in interdistrict inequality (using multiple methods to compute inequality similar to Wood and Theobald), there is no difference in that reduction between equity and adequacy states. Specifically, they find evidence "suggesting – contrary to the expectations of

some theorists – that resource distribution patterns are not significantly different following court-mandated equity and adequacy reforms.” (p. 440).

This finding is noteworthy for the current study, particularly when considered in concert with their discussion on how their data were coded. Springer, Liu and Guthrie discussed the difficulties in gathering data for their study, mainly because there still exists substantial disagreement about what might constitute equity, adequacy, or even victory, outside of the ideal examples given in theoretical articles. They highlight, for example, one instance where different sources had the same case occurring in different years, depending on whether one used calendar year, or end of court term as the standard. Distinctions like this are important when you have collected yearly data from other sources or created lagged variables. Additionally, they describe the Rhode Island case where, apparently, the judicial decision gave something to both sides, because there was disagreement among existing lists of decisions about which side had even won that case.

Springer, Liu, and Guthrie, then, have essentially described all the reasons why I decide not to code the content of the legal decision in the current study, and why I use only the initial adequacy decision in every state. As a state’s litigation evolves, it embarks upon a sort of path dependence, whereby compliance decisions in a given state are dependent upon prior developments in that state, and cross-state comparisons would no longer be valid because of the unique histories involved. Additionally, if there are no significant differences in inequality reduction between equity and adequacy cases, it probably makes little sense to model per-pupil spending or test scores. If, judicial decisions and their results are context dependent, and spending figures taken out of a meaningful context are unlikely to be helpful for the current study.

Given the uniqueness of each case, and the research I cite above, one might question the wisdom of proceeding with the current study. There are two possible answers. The first is that even though each case is unique, it can still be useful and instructive to identify commonalities, particularly where the results have the potential to make a large impact on public policy. Indeed, many studies in comparative politics deal with countries that are fundamentally different, but that have enough in common, at least as regards the question under study, that meaningful results can be gained. Second, it would be incorrect to assume that, simply because adequacy lawsuits are complex, there are no necessary and/or sufficient conditions for plaintiffs' victory without any empirical support for that statement, in the same way that it is wrong for commentators to draw conclusions about "liberal judges" and "liberal courts" and adequacy without empirical support for those statements.

Can Judges Effectively Evaluate Cost Studies?

The discussions in previous sections of objections to the adequacy concept and to the nuances of "costing out" adequacy serve to highlight the highly technical nature of adequacy cases. While the use of social science in education litigation dates at least to the citation of Dr. Kenneth Clark's research in Footnote 11 of *Brown v. Board*, it is still unclear that judges have the capacity to weigh social science evidence appropriately. After all, even Dr. Clark's study, which was used to support a critically important claim in a case with far-reaching implications, didn't stand up to scientific scrutiny (Heise, 2002). Courts' inability to consistently evaluate social science evidence properly remains a problem. Writing about courts' "comparative abilities to deploy social science research," Michael Heise finds that "institutional costs from the courts' use of social

science and empirical evidence include the stresses incident to pushing courts and judges into relatively unfamiliar intellectual terrain...[and] also uncovers institutional and capacity limitations of the courts' ability to work with such information," (Heise, 2002, pp. 1311-1312)⁸.

Lindseth⁹ (2006) points out, for example, that the judges in New York did not require proof of *causation* that funding levels were responsible for student achievement, but were instead satisfied with *correlation*¹⁰. The implication was clearly that the confusion between correlation and causation is an elementary methodological mistake, and that the judges of New York's Court of Appeals were trying to interpret evidence that was over their heads. Although Lindseth (as defense counsel in New York) doubtlessly knows the evidence in question and that there *are* times when correlational research is the best available method, he seemed to imply that the judges in New York had not made a reasoned methodological judgment, but rather had simply confused correlation and causation. In Massachusetts, Pullin (1997) discusses how the Supreme Judicial Court, although they made some use of social science evidence, failed to appreciate the complexities involved, and failed to use test scores appropriately. Pullin further observes that the Court didn't have the most recent research to consider in forming its opinion, which could conceivably be a result of the judicial reliance on attorneys to find and present evidence, as it is in the litigants' best interests to present evidence that buttresses

⁸ There are conflicting points of view on this point. Koski and Levin (2000), for example, argue that judges do a good job of defining adequacy, and that the part of the problem is how legislatures implement judicial decisions.

⁹ I am aware that Lindseth was defense counsel in CFE, and certainly has a bias in discussing adequacy lawsuits. While I would not rely on him for an unbiased review of adequacy or CFE, he does bring up a useful point here that is worth discussing, and, in any event, has also been pointed out by others as well.

¹⁰ See Tropp, Smith and Crosby (2007) for a broader discussion of how judges consider social science evidence, and differences between the legal and scientific standards for proof and evidence, in the context of school desegregation.

their points. Conceivably, if newer research muddied the waters, perhaps by presenting a confusing or inconclusive result, then neither side would present it and the court could be left unaware of the latest research.

Perhaps the most convincing argument that judges have trouble interpreting social science evidence comes from an experimental study. Kovera and McAuliff (2000) conducted an experiment where judges were asked to analyze a hypothetical fact pattern, and also given some expert testimony and asked whether they would admit it as evidence. The study took place in Florida, where the *Frye* standard for the admission of scientific evidence is still used, and judges are to consider whether the scientific evidence has achieved acceptance in its field. In making their determination, judges can consider such things as whether the research was conducted in accordance with the scientific method and whether the research was well-designed. Kovera and McAuliff manipulated the scientific evidence, though. They found that judges relied too much on whether the research had been published in a peer-reviewed journal, and did not often pick up such elementary research errors as lack of a control group, and possible experimenter bias or lack of an experimental blind. There do exist some law review articles that come to different conclusions regarding judges and the interpretation of social science evidence, but, perhaps tellingly, I can find none that use a randomized controlled trial or even a statistical model in reaching their conclusions. The most compelling evidence, then, is that judges do indeed have trouble interpreting such evidence.

If judges ordinarily have a bit of trouble interpreting scientific evidence, and adequacy decisions can frequently turn on which scientific evidence is given credence, then it seems plausible that courts that can better evaluate such research might make

higher-quality decisions. While social science evidence might not often be the deciding factor, it nonetheless plays some role, especially given the statistical and research-based definitions of adequacy. Judges who have the resources and inclination to engage in their own research and interpretation, instead of trusting lawyers to provide an accurate summary of the research, are likely to have an easier time with the interpretation of scientific evidence. In other words, more professional courts will do a better job, simply because they have more resources at their disposal to assist with interpreting this evidence.

There are a variety of ways to operationalize the professionalism and resources of a court, and adding all the possible variables would overload the QCA model. Choi, Gulati and Posner (2008) develop an index to rank State Supreme Courts and separate the identify the highly performing courts. Their ranking system takes into account compensation and resources, the number of times that the court is cited by other state supreme courts, the number of opinions the court produces, and provides, according to the authors, a reliable ranking of state supreme courts. The advantage of this ranking for the current study is that it allows many different court characteristics to be examined using this one variable. Following Choi, Gulati and Posner, I will consider the top ten courts to be highly performing and, therefore, more professional.

I hypothesize that more professional courts will be part of a necessary and/or sufficient combination of causes, but not necessary and/or sufficient on its own. QCA allows for causes to be necessary in combination, so that this hypothesis can be tested.

Interpreting Possible Results

The next chapter will explain Qualitative Comparative Analysis (QCA) in extensive detail. For now, it is sufficient to explain that QCA is a technique that is designed to identify conditions that are logically necessary and/or sufficient for a given outcome—in this case, the ruling in an adequacy lawsuit. Also of note is that QCA requires variables in binary form. Necessary and/or sufficient conditions must be either present or absent¹¹. Therefore, the number of possible causal conditions under study is necessarily categorical and somewhat limited. For instance, although there are four distinct methods of judicial selection, only one variable will be coded—elected or not, based on the theoretical knowledge and reasoning behind the hypothesis. It is roughly analogous to the “degrees of freedom” problem—if four binary variables for judicial selection were included (and each state were absent on three and present on one), then, not only would many more cases wind up being “unique,” but there would also be less room in the model for other theoretically important variables.

One possible outcome of the analysis is that certain conditions (or combinations of conditions) prove necessary and/or sufficient (or nearly necessary or nearly sufficient, depending on the consistency threshold employed in Chapter 3) for plaintiffs’ victory. Another possibility is that there exists different causal configurations for different groups or clusters of cases—if, for instance, the first several adequacy cases needed x and y to be successful, but a second group of cases needed only z to be successful.

A third possibility is that virtually every case is unique, and there exist several possible configurations of necessary and sufficient conditions, each of which covers only

¹¹ There is a variant called “Fuzzy Set” analysis, where cases can be partial members of a set and data is numerical between 0 and 1, but, as I discuss in Chapter 3, there are inherent problem with the researcher assigning the fuzzy set values and then testing a hypothesis, as that would be similar to post-hoc theorizing or data-snooping. There exist no hard and fast rules for creating fuzzy set values, and they are supposed to be based on the researcher’s judgment and knowledge of the cases under study.

a few cases, so that no blanket statements about necessity or sufficiency can be made in any way that would be theoretically meaningful. Some people might interpret that outcome as a failure of the study. On the contrary, though, such a finding would give us meaningful information about judges and about adequacy litigation. If no pattern could be found, it would imply that judges decide adequacy cases focused mainly on the particular merits and circumstances of each case, and that adequacy lawsuits were free from the types of diffusion and social learning among courts that Reed (2001) identified as central to desegregation efforts.

Finally, a fourth possibility is that there is a pattern to be found based on *when* the cases are filed and decided. It is known that adequacy advocates communicate with each other, and, also, that the timing and choice of forum in lawsuit represents a conscious and usually strategic decision. Therefore, it is possible that the “best” cases, in states with the most inadequate systems, were filed first. Theories of policy diffusion would predict that early wins in a few states would embolden advocates in others, just as early losses might deter them. This theory is tested outside of the QCA method detailed in the next chapter, both because resists easy conversion to binary data and because it arose as a reasonable alternative only after QCA was selected for the current study.

The median district per-pupil expenditure is calculated for each state in the year before the decision occurred, as is the standard deviation¹². The year before the decision is used, as opposed to the year of filing, because it is difficult to know *when* plaintiffs first signaled their intent to file, which can be different from the actual filing year, and

¹² Stephen Q. Cornman, Esq., of the federal Institute for Education Science, graciously supplied the data. Unfortunately, neither he nor his colleagues were able to provide data prior to 1994, when NCES switched to the computerized system they currently use. This truncates this analysis somewhat, but still should allow for the detection of a trend if one exists.

furthermore because, except for the states that settled out of court, one would not expect any school finance reforms while a trial was ongoing. If this theory is correct, one would expect to see high standard deviations, low medians, and plaintiffs' victories in the early years, changing gradually to low standard deviations and decisions for the defendant states as time went on.

CHAPTER 3 – METHODS

Qualitative Comparative Analysis as a Research Method

Walker (1968) referred to the American states as a “laboratory,” in which policies could be tried in different states, and the results compared. Inherent in that position is the notion that the American states can be viewed in a comparative context, utilizing some of the standard tools of comparative politics. The current study uses that framework of comparative politics to compare the institutional arrangements and outputs in the American states as they relate to filing and decision of an adequacy lawsuit. Most importantly, the method employed is qualitative in nature, although the procedures are systematic so that hypotheses can be generated and tested analogous to the procedures used in quantitative methodology.

The reason for a qualitative methodology stems from the fact that there are only fifty states. Given the number of independent variables that I want to incorporate, it is impossible to use quantitative tools such as logistic regression (or variants of it, such as event history analysis), because the regression equations will not converge. The normal procedure in such situations is to increase the N (see King, Keohane and Verba, 1994), but that is impossible in this case, as one cannot go out and find more states to sample once all fifty have been accounted for. These limitations do not stop people from using quantitative methods inappropriately. Lundberg (2000) and Swenson (2000), writing in the same issue of the Albany Law Review, both use logistic regression to analyze adequacy cases, and find that none of their predictors are significant. When I used data on adequacy cases in an attempt to replicate those analyses on SPSS (the program that

one author employed), I received an error message that the equations did not converge, rendering the estimates that the program generated meaningless.

The current study will use the method of Qualitative Comparative Analysis (QCA) developed by Charles Ragin (1987). This method allows for the systematic collection of data and testing of hypotheses by relying on the tools of symbolic logic and Boolean algebra. In other words, while the technique is qualitative in nature, as it doesn't rely on statistical inference but rather on inductive reasoning, its systematic nature allows for more than just "thick description." The development and use of a tool like QCA addresses many of the problems that King, Keohane and Verba (1994) discuss in their critique of qualitative methods. Indeed, although Ragin first wrote about QCA in 1987, the method gained wider acceptance after 1994, and with Ragin's introduction of "fuzzy sets" (2000, 2007). QCA also allows for explicit testing of hypotheses about necessity and sufficiency, which, as I discuss below, approximates the "cause-effect" reasoning present in statistical studies.

The study of American politics has tended towards quantification and the use of formal theory wherever possible (Cohn, 1999). As I will depart from that standard practice, it becomes important to examine at the outset what one can reasonably expect to conclude at the end of the current study. A brief examination of the role that qualitative techniques have played in political science would be instructive. King, Keohane and Verba (1994) state that the fundamental goal of all research is inference, and they imply that qualitative researchers fail at that goal except when they attempt to use the quantitative template. For them, good research involves establishing causality, increasing the N to ensure causal leverage, and using deductive reasoning and formal

theory. Clearly, then, the current study would be a failure under King, Keohane, and Verba's rather strict criteria. No matter how much data I gather, I will not be able to conclude that a particular institutional arrangement (method of judicial selection, for example) caused a court to decide in a particular manner, or caused litigants to file a case in one state as opposed to another. Like many studies, this one takes place without a laboratory control group or even a constructed control group to allow for a "quasi-experiment" and so I would never be able obtain enough cases to make a claim of causality as King, Keohane, and Verba define the term.

Fortunately, contributors to Brady and Collier's (2004) volume, *Rethinking Social Inquiry*, have addressed many of these concerns. Collier, Brady, and Seawright (2004) point out that increasing the N is not the only way to say important things about effects and causality. Also, Rueschemeyer (2003) makes a persuasive argument that a few cases—or even one—can make important theoretical contributions depending upon the theory which drives the study. A study that produced a counter-example to a generally widely-held belief would be important in its own right, for instance. He points to the work of noted political sociologist Theda Skocpol as an example of the type of small-N research that he believes yields theoretical gains. And prior to Rueschmeyer, Harry Eckstein made a valuable distinction between case studies which are description for their own sake and the more valuable "critical case study," which is valuable in the building and testing of theories. Furthermore, Eckstein observed that, even in medicine, where the classic experimental design can be employed fairly easily, clinicians still sometimes report on individual cases where such reports provide some analytical leverage.

Additionally, quantitative and qualitative researchers seem to mean very different things when they talk about causes. The current study will employ QCA to determine necessary and/or sufficient causes (or combinations of causes) for filing and for plaintiffs' victory. The logic of necessary and sufficient conditions is mainly confined to settings where large-scale statistical analysis is not appropriate, due to small sample size. According to Gary Goertz and Harvey Starr, *Designing Social Inquiry* is an "attempt by quantitative scholars to deal with qualitative methods" in which we "see correlation as the dominant causal concept and the complete absence of the necessary condition concept" (Goertz and Starr, 2003, p. 16).

Goertz and Starr's main point is an important one. Quantitative researchers tend to phrase their hypotheses in terms of magnitude and effects on average, whereas good qualitative researchers, while being no less rigorous, phrase their hypotheses in terms of the presence or absence of certain variables—e.g. whether a variable is necessary and/or sufficient. A researcher using statistical techniques, then, would be interested in whether a greater amount of ideological polarization on a court (however properly measured) is associated with a court making a smaller estimate of the cost of an adequate education, while a primarily qualitative researcher might be interested in whether ideological polarization on a court is necessary and/or sufficient for a decision for the defendant states.

Gary Goertz (2003) argues that this dichotomy in the way that different researchers define a "cause" has existed since the earliest days of western philosophy. According to Goertz, a cause can be either "the constant conjunction of cause with effect" or "the effect not occurring without the cause." The first definition is consistent

with a statistical and correlational view of causality, and can be found in the writings of Carl Hempel. The second is consistent with hypotheses concerning necessary and sufficient conditions, and can be found in the writings John Stuart Mill. Most interestingly, David Hume incorporates both formulations into his definition of “causality” and sees no contradiction.

The preceding discussion ought to have convinced the reader that political scientists can make valid inferences even when not using the statistical template, and that testing for necessary and sufficient conditions represents at least one possible way to do so. As QCA is a relatively new technique for testing hypotheses about necessity and sufficiency, an explanation of its benefits requires an understanding of prior methods for testing these hypotheses. Those methods were John Stuart Mill’s methods of agreement, and methods involving symbolic logic pioneered by Braumoeller and Goertz (2000).

Mill’s methods were perhaps the least “scientific” in the usual sense of the word, although the logic was sound. The method of agreement argued that “if two or more instances of a phenomenon under investigation have only one of several possible causal circumstances in common, then the circumstance in which all the instances agree is the cause of the phenomenon of interest,” (Ragin, 1987). Essentially then, the method of agreement is a thought experiment that proceeds by process of elimination, listing all the causes of each instance of a phenomenon until the investigator finds one in common. Note, though, that there is no inherent systematic way to go about doing this, and so a lot of judgment is left in the hands of the investigator. Mill’s method of difference also does not allow for causes to be considered in combination, or for different causal pathways to the same outcome. Perhaps most importantly, Mill recognized these limitations,

recommending that method of agreement not be used in the social sciences, and further recommending that experimental methods be used wherever possible. It is likely, then, that Mill meant the method of agreement to be used as a tool for thinking and philosophizing, and not necessarily for scientific investigation where better methods existed.

Braumoeller and Goertz (2000, 2003) were among the first group of scholars to try to formalize the testing for necessary and sufficient conditions. They started with the point that, since many hypotheses are expressly framed in terms of necessity and sufficiency, then an explicit methodology for testing those hypotheses is necessary. They provide a number of examples of specific hypotheses in the literature that are framed in terms of necessity and sufficiency, but are not treated as such. For instance, Braumoeller and Goertz cite Larry Diamond's article about democratization, in which Diamond writes "It is important to emphasize as well that democracy can occur at low levels of development if the crucial mediating variables are present. Economic development is not a prerequisite of democracy." The critique is that Diamond tests that hypothesis using regression models, where a separate test of the hypothesis of necessity would have been warranted.

Furthermore, they find the two most common existing methods for testing necessary conditions to be inadequate. Yule's Q, a method favored by Bueno de Mesquita and Lalman, is inappropriate because its value is "influenced by data that are irrelevant to the hypothesis of necessity" (Braumoeller and Goertz, 2000). Specifically, it varies between 0 and 1, and while it *will* give a value of 1 in the case necessary condition hypothesis is true, it will also give a value of 1 in other cases as well, particularly if there

are a shortage of cases where the hypothesized necessary condition occurred without the outcome occurring. As explained previously, cases where the outcome does not occur are irrelevant to hypotheses of necessity, and therefore should not influence a statistic that tests for necessity. Braumoeller and Goertz reject Yule's Q on those grounds, because it is influenced by data irrelevant to the hypothesis at hand.

The Proportional Reduction in Error Statistic ("del") is also inappropriate, because it "amounts to a measure of the *degree* of necessity of X for Y, not a *test* of the proposition that X is necessary for Y."

They recommend a test in which cases are placed in a 2x2 matrix, with the columns represent the absence or presence of X and the rows representing the absence or presence of Y. Logically, X is necessary for Y if there are no cases in the lower left cell (X absent/Y present) and cases in the lower right cell (X present/Y present). While the method they present for checking significance is not germane to the current study—there is no sampling as the entire population of states is studied—two of the insights that underlay their model are important. An adaptation of their matrix appears below.

	X = 0	X = 1
Y = 0	N/A	N/A
Y = 1	0	100

The first important lesson to draw from Braumoeller and Goertz's "matrix" approach is that they look only at the two cells of the table in which Y occurred. In their model, one can only learn anything about a hypothesis of necessity by looking at

instances of the outcome in question. Their argument relies on symbolic logic: the statement $P \rightarrow Q$ ("if P, then Q) is only false in the case when P is true but Q is false. The statement is true any time P is false, regardless of the value of Q. Seawright (2002) criticized strategy of testing for necessary conditions by looking only at the cases where the outcome of interest occurred, arguing that all cases should be examined. He reasoned that "Nature" assigned each case to one of the four boxes in Braumoeller and Goertz's matrix, and that the only thing that a necessary condition hypothesis implied was that no case would fall in the bottom left cell (cause absent/outcome present). Braumoeller and Goertz (2002) responded with a more detailed defense of their method and a discussion of conditional probability that supported their method, but the debate was instructive for two reasons. First, it illustrates the minefield of inference in necessary conditions hypotheses which I avoid by examining all fifty states. Second, it contrasts with standard statistical methods, which avoid selecting on the dependent variable.

Seawright also believed that a necessary condition hypothesis could be falsified by a single counterexample, implicitly using the "critical case study" logic discussed earlier, where the nonconforming case would be the critical one to study. Braumoeller and Goertz and Ragin (---) disagree with this claim, however, for reasons that are understandable when dealing with research that involves human behavior, although from the standpoint of strict symbolic logic, Seawright's logic is correct, as Braumoeller and Goertz acknowledge. They advocate a probabilistic testing of necessary conditions hypothesis for two reasons. First, in the social sciences, there is inherent error in the data, in both concepts and measurement. Second, and perhaps more importantly, necessary conditions hypotheses can be theoretically valuable even if there are a few

counterexamples. Depending on the specific concepts at hand, there could be theoretical gain from the discovery of factors that are nearly always or almost always necessary. Such factors could also provide some degree of predictive power if the question at hand is a pressing policy issue.

One might object at this point that once probabilistic reasoning comes into play, the need arises for conventional statistics and significance tests. That objection, however, could be refuted on three grounds. First, since necessary condition hypotheses are fundamentally qualitative in nature, “significance” only applies in its Standard English meaning, it is up to the researcher to demonstrate the relevance or significance of his or her findings. Put another way, in a multiple case-study approach, there is no numerical way to determine whether the outcome in question is due to chance, and the acceptability of the findings rests on the researcher’s ability to persuade others of his argument that one or another such factor was important or decisive (or significant). A second way to refute that objection is to point out that in the case of Qualitative Comparative Analysis, Boolean Algebra and the reduction algorithm it employs will eliminate conditions that are trivially necessary or sufficient, and has a procedure in place to deal with contradictory observations (discussed below), which together address the same theoretical concerns that are addressed by statistical significance tests. Third, in the current study, there is no sampling or inference taking place. The entire population of states is being sampled, and, therefore, even if standard statistical techniques were being employed in the current study, one could still question whether significance tests were mandatory. Charles Ragin did use Bayesian statistics to develop a way of approximating conventional significance when testing for necessary conditions, but his method has not been widely adopted, and

additionally would not be compatible with QCA, although they are compatible with some other methods of testing necessary conditions.

To use a hypothetical example, assume that a researcher hypothesizes that a small school (less than 500 students) would be necessary for higher achievement in urban schools. If that pattern is found in all but one case under study, Seawright would argue for rejecting a hypothesis of necessity as far as small schools are concerned. From an applied perspective, though, it would make more sense to take these results at face value. Small schools would be almost always necessary for increased academic achievement. A probabilistic conception of necessary conditions hypotheses holds the additional advantage of giving researchers an incentive to try and explain nonconforming cases, perhaps by better defining the concepts under study. Ragin (2006) and Dion (2003) both felt strongly enough about the utility of probabilistically necessary conditions that they developed methods to test them for significance at a level that is analogous to the 5% p-value typically used in statistical studies. In his initial work in the area, Ragin developed categories whereby things were "almost always" or "sometimes" necessary or sufficient, where the percentages required for those categorizations corresponded to $p=.05$ and $p=.10$, respectively. Dion employed a somewhat more sophisticated model that incorporated Bayes' rule to derive the number of conforming cases that needed to be observed before a cause could be deemed to be necessary at the $p=.05$ level. The main drawback to Dion's method is that it depends on knowing, a priori, the probabilities of the outcome either occurring or not. In some cases, the probabilities can be assumed to be symmetrical, as if one were flipping a coin, but such simplifying assumptions do not

necessarily work when dealing with real data, especially if the true probabilities might have been different in different cases.

Ragin's method of Qualitative Comparative Analysis (QCA) allows one to sidestep these issues of what to do with counterexamples and how to test for significance by allowing for what he calls "multiple conjunctural causation." In the methods for testing necessary conditions hypotheses discussed above, only one cause could be considered at a time. Braumoeller and Goertz, for example, discussed *a cause* and *an effect* in terms of X and Y. As I demonstrate below, QCA is a superior analytical technique because it accounts for as many possible necessary conditions as the researcher cares to analyze, and allows one to deduce whether each cause is necessary by itself but also which *combinations* of causes are necessary. In addition, it is possible—even likely—that any one of several causal combinations can be necessary or sufficient. QCA explicitly accounts for that as well. For instance, it is entirely conceivable that either divided state government *or* the election of judges in a non-partisan manner (or, logically, both) is necessary for plaintiffs to win an adequacy lawsuit. QCA would allow the researcher to uncover that pattern, whereas the methods proposed by Braumoeller and Goertz and by Dion would not. Instead, those methods would test each variable individually and find that *neither* of them was necessary, because, individually, either one of them would present too many counterexamples. Such a conclusion would retard the development of theory and would also have negative practical outcomes if the "effect" in question were a desired policy goal and policymakers were mistakenly led to believe that there was *no* necessary condition instead of *two possible necessary conditions*.

Temporal Relationships and QCA

For all the strengths QCA offers, there is one aspect of it that some might consider to be a weakness. It does not directly account for temporal relationships, either among the different hypothesized necessary conditions in a single case, or among their appearance in the various cases. Put another way, QCA fails to account for that fact that events in one case might have been influenced by events in an earlier case, and essentially treats all cases as independent. And even in a given case, QCA tends to ignore the order in which causes appear. The only exception to these limitations is if the researcher self-consciously includes some kind of measurement of other cases as a necessary or sufficient condition, but even then it is restricted to a binary indicator, as are all variables in QCA. Paul Pierson (2004), for example, is one of a number of scholars to stress the fundamental importance of temporal relationships and path dependence, and so it is worth considering how the choice of QCA as a research method—and its lack of a temporal component—will affect the analysis.

First, it is important to acknowledge that the adequacy arguments share some very similar characteristics that have not changed over time. The definition of adequacy, and the arguments that support it, are not dependent only on the most recent set of educational statistics, or upon a particular election cycle. Indeed, as state after state faced adequacy litigation, the arguments begin to sound similar, and even involve some of the same key players (Erik Hanushek, for instance). Additionally, to the extent that temporality is important because rulings “diffuse” from one state to another state, the current study incorporates two diffusion variables, which measures whether, on balance, the court in question faces a net force towards an adequacy ruling or towards a ruling that favors the defendant states, based on rulings in similar states. It is true that such a variable doesn’t

take account of the specific amount of time that passes between rulings (months or years) in the same way some other more quantitative models might, but there is no theoretical reason to assume that temporal relationships matter for decision in a way other than what the diffusion variables take into account.

Where time might play the biggest role is in *filing* the cases in the first place, and, when one merely glances at case outcomes over time, it becomes clear that plaintiffs are having less success now than they once did. This could be due, however, to the levels of inequality or inadequacy being different in these later states, as strategic litigants went first to states where they were most sure of winning. While QCA cannot test for this type of strategy, it will be the subject of a separate analysis.

Why Other Possible Methods Were Not Employed

Before reviewing the procedures for using QCA in detail, it is important to devote time to an explicit discussion of possible alternative methods, and the reasons that they were not used in the current study. There were both quantitative and qualitative alternatives to QCA, and therefore the questions about the merits of QCA relative to these other alternatives are likely to come from methodologists of all persuasions.

A quantitative study would have required a larger sample size, and would have looked more like the studies seen in leading political science journals. The primary way to increase sample size would have been to consider a wider universe of cases, relating to adequacy issues and many others, and to create an “adequacy case” dummy variable. This approach would have allowed an explicit comparison of adequacy cases to other types of cases. However, it would have been difficult to find many areas from which the federal courts have nearly completely removed themselves as they have in adequacy.

Additionally, such an approach would have required the reading of many decisions and coding of case histories, which would have required that the author receive legal training or take on a coauthor.

Case studies represent a qualitative alternative. However, there were two reasons why a case study was not employed. First, case studies are already common in educational research, both on adequacy and a multitude of other topics. Many scholars have examined adequacy litigation in one or a few states in some depth, and it was doubtful that another such study could make a new contribution to the knowledge base. And education research seems to be overrun with case studies, which was one reason that NCLB tried to shift the focus back to randomized controlled trials. Indeed, Koski and Levin (2000), in reviewing the literature on the effectiveness of court involvement in school finance post-*Rodriguez*, bemoan the very fact that most studies consider only one of a handful of states, making generalization difficult.

Second, most of the key players in adequacy are still practicing and writing. They can tell their own stories far better than I can. And while document review and other historical methods are perfectly acceptable for case studies in history, it makes little sense to employ them when at least one object of study can be found in an office down the hall.

Procedures For Using QCA

QCA relies on combinatorial logic, and therefore the first step in employing QCA is to construct a truth table. In a truth table, each observed causal combination is a row in the table, and data are converted to binary. For instance, if there are three causal variable (A,B,C) and an outcome variable Y, then, in a case where all four variables are present, that row of the truth table would read 1,1,1,1. Notation in QCA is such that a capital

letter indicates the presence of a variable and a lowercase letter indicates its absence, so that in the case where only A is present, and the outcome does not occur, the row Abc would read 1,0,0,0.

The conversion of variables to binary is straightforward in the case where there exist only two categories. If there are three categories (for example, in Elazar's classification of state political culture as traditionalistic, moralistic, or individualistic), then that variable is recoded as three binary variables—the presence or absence of the first category, the presence or absence of the second category, and the presence or absence of the third category.

Interestingly, the frequency with which a causal combination occurs is not incorporated directly in the analysis. QCA's focus on multiple conjunctural causation means that all causal pathways are analyzed to see whether they are necessary or sufficient, irrespective of how often they occur.

Once a truth table is developed, all the causal combinations that lead to the outcome are put in the same equation, and the equation is simplified. For example, assume that combinations Abc , ABc , and abC all lead to outcome Y . (Recall that an uppercase letter indicates the presence of a variable, and a lowercase letter represents its absence. That would be represented algebraically as $Y=Abc+ABc+abC$, where addition represents the logical operator OR and multiplication represents the logical operator AND. Looking at that equation, though, one should see immediately that it can be simplified. The first two terms, Abc and ABc , should be combined and simplified to Ac , because the equation states that the presence of A and absence of c will bring about Y *regardless of whether B is present or not*. The two terms, then, are redundant.

The simplified equation can then be examined for necessary and sufficient conditions. Ragin (1987) gives a precise definition of necessary and sufficient causes in QCA, and uses the term “cause” to mean “causal combinations” except where he specifies a singular cause: “A cause is both necessary and sufficient if it is the only cause that produces an outcome and is singular...A cause is sufficient but not necessary if it is capable of producing the outcome but is not the only cause with this capability. A cause is necessary but not sufficient if it is capable of producing an outcome in combination with other causes and appears in all such combinations.” After necessary and sufficient causes for the outcome have been obtained, one can apply DeMorgan’s rule, a logical operation that gives the inverse, to identify characteristics that are necessary and sufficient for the *absence* of the outcome. Note that just because a condition is not sufficient for the outcome, it does not logically follow that the condition is sufficient for *not* having the outcome. Rather, you have to run a separate test to determine conditions that are necessary and/or sufficient for the absence of an outcome.

Ragin emphasizes the need to use theory in developing and testing necessary conditions hypotheses. After all, the contention that a cause is necessary or sufficient is of little value unless it sheds light on a mechanism that is identified and supported by literature in the field. It generates little new knowledge if the mechanism by which necessary and/or sufficient operate is obscured in a “black box.” Ragin’s instruction to rely upon theory to substantiate the necessary and sufficient causes derived by QCA also serve to answer one of the main arguments of his critics.

One of the main criticisms of QCA is encapsulated in the abstract of Monroe and Gold’s (2004) paper. They pose the following scenario: a man drinks too much on four

consecutive nights: bourbon and soda the first, scotch and soda the second night, rye and soda the third night, and gin and soda the fourth night. Monroe and Gold make the point if the heavy drinker in question used QCA to try and explain his hangover, he would deduce that soda was necessary and sufficient for a hangover. He would then proceed, according to Monroe and Gold's reasoning, to down a bottle of scotch confident that he would be headache-free the next morning. Furthermore, even if the different alcoholic ingredients were recoded as "alcohol," the conclusion would be that alcohol and soda were both necessary and sufficient for a hangover, so that the man would still feel free to drink alcohol without threat of a hangover.

Monroe and Gold's example, however, far from being a criticism of QCA, is rather a criticism of the failure to meld a research tool with an appropriate theory. The careless researcher detailed in countless introductory statistics books could find a statistically significant correlation between ice cream consumption and the murder rate. The researcher's underlying theory would justifiably be questioned, but by itself that example would do nothing to discredit the potential usefulness of the Pearson correlation coefficient if used correctly. Put simply, any analytical tool must be paired with a good theory that specifies a hypothesized causal mechanism. As Monroe and Gold demonstrate, QCA is no exception to that general rule. A researcher should be able to at least hypothesize reasons why a given factor is necessary or sufficient to a process, even if that process cannot be directly observed, as will be the case in the current study. After all, if I were to demonstrate through QCA that unified government (executive and legislature controlled by the same party) is necessary for plaintiffs to prevail in an adequacy lawsuit, I should be able to offer a theory as to why that is the case, even

though I clearly will not be able to confirm that theory empirically by watching judicial deliberations.

There are two remaining important issues that one has to deal with when using QCA. First, there exists the possibility that the truth table can have what Ragin refers to as “contradictory rows.” In the example discussed above, a causal combination either led to the outcome in all the cases in which it occurred, or it led to the outcome in only some of the cases in which it occurred. It is certainly possible the data I employ will show that a causal combination leads to a decision for the plaintiffs in one case but not in another case. The construction of the truth table, then, is complicated, because only one outcome can be coded. Ragin offers two suggestions for what to do in a situation where contradictory rows exist. If the cases are nearly evenly split between outcome-present and outcome-absent, Ragin suggests it would be prudent to investigate other possible independent variables that should be included, because *something* is clearly different between cases that show and don’t show the outcome. It is conceivable that, in reading more about cases involved, the researcher will discover an additional independent variable that, when included, eliminates the contradictory rows. The second option, which could be useful if the contradictory row displays a clear preponderance for outcome-present or outcome-absent, is to code that row according to the outcome that most of the cases display, and then to explain the deviant case using more conventional case-study methods.

Ragin also discusses the phenomenon he calls “limited diversity.” Simply put, unless you are in a laboratory setting, it is likely that all possible causal combinations will not be observed. QCA depends on recording and analyzing extant characteristics, not

manipulation. Therefore, if no state displays a given causal configuration, no data will exist about it. It *might* be necessary, but we have no way of knowing. Ragin suggests erring on the side of caution and coding unobserved causal configurations as if the outcome did not occur, or leaving them out of the analysis entirely, as the end result is the same. Furthermore, it is possible that some of the causal configurations that did not occur are unlikely *ever* to occur, as certain characteristics in a political system (or any system) tend to occur together.

While the above discussion focused on the aspects of Ragin's methods that I will be using, it is important also to discuss one of his recent advancements that I have purposely decided *not* to use. Ragin (2000) expanded the concept of QCA to try and incorporate continuous variables through a concept that he termed "fuzzy sets." In this method, cases are given a number between 0 and 1 reflecting the extent to which they are either "in" or "out" of a set, with a score of 1 representing a case that is fully in the set. If the variable in question was "democracy," for example, the United States might get a 1, but Russia, which has some undemocratic features, would get some number other than 1 to reflect the fact that it didn't fully fit in the set of democracies. This all sounds good in theory, but the catch is that there exists no scientific or objective method for converting empirical data to fuzzy set notation. Ragin expressly leaves it up to the researcher to assign values based on their knowledge of the variable and the cases at hand. If there existed *a priori* fuzzy set values that another researcher had created for my variables, then I would employ them. However, since I have full knowledge of my hypotheses and the numbers that I would have to plug in to make it correct, it seems inappropriate that I should both assign the values and develop the hypothesis. It reminds me somewhat of the

prohibition against post-hoc theorizing or data snooping that one hears in an introductory statistics class.

Some examples of how QCA has been used to study a wide range of phenomenon might be useful. One interesting article is Amenta and Poulson (1996) which dealt with state spending on social programs at the conclusion of the New Deal, specifically old age assistance and the Works Progress Administration. Specifically, they test an institutional politics model, which dealt with nondemocratic state political systems (using a voter turnout measure), patronage party organizations, and state administrative strength. Most interestingly for the current study, however, is that they used multiple regression and then followed up with QCA, because, as they write “multiple regression is not always technically well suited to analyzing contextual arguments, such as those of institutional politics theory...we employ qualitative comparative analysis (QCA)—a method explicitly designed to assess contextual arguments.” In other words, Amenta and Poulson’s use of multiple regression was mainly to show how it would be ineffective and yield different results than QCA, a kind of straw man that they built up to knock down. They argue that QCA provides a viable method to assess contextual and institutional hypotheses in which the population is the American states, and that it is superior to multiple regression.

QCA has, in fact, been used to study a wide range of political and social phenomena, encompassing topics as diverse as the success of homeless social movements in US cities (Cress and Snow, 2000), an examination of the processes by which women won the right to sit on juries (McCammon et al., 2008), the reasons that left-libertarian parties are more successful in some Western European countries than others (Redding

and Viterna, 1999), labor politics, worker insurgencies, and strikes (for example, Dixon, Roscigno, and Hodson, 2004; Biggert, 1997) and a longitudinal study of the Florida legislature's consideration of healthcare reform bills that identified causal combinations that were prerequisite to a bill's passage (Harkreader and Imershein, 1999).

The Harkreader and Imershein study is of great use in informing, and legitimizing, the current study. Harkreader and Imershein examine the Florida legislature over time, treating each legislative consideration of healthcare reform from 1965 to 1993—operationalized as the introduction of a formal bill that the legislature considered—as a “case.” They coded each case for the presence of seven factors: 1) if there were federal financial incentives promoting the policy; 2) if there was a unified health-care provider policy in opposition to state policy; 3) if there was a state agency supporting the policy; 4) if legislative leadership in the house and senate supported the policy; 5) if the governor supported the policy *or* did not oppose it versus if the governor opposed the policy; 6) if the health-care provider position was undermined by fiscal matters; 7) if the state policy position was undermined by fiscal matters. They identified three necessary conditions for the legislature to pass a healthcare reform bill “the leadership of legislative health-care committees agreeing on state action, the support of the leadership of the state agency with jurisdiction, and a credible policy position in light of prevailing economic circumstances.” These corresponded to factors four, three, and seven in the list above.

Harkreader and Imershein demonstrate that QCA can credibly be used to examine the actions of a government institution. Additionally, they make two questionable decisions that illuminate important issues for the current study. First, their analysis

appears divorced from any underlying theory. While there is discussion in the literature review of the politics of healthcare in Florida, there is little discussion of an underlying theory or model of the policy process in general or even a health policy process in particular. This appears to ignore Ragin's exhortation to connect the search for necessary (or "prerequisite," as they call it) conditions, and leaves them vulnerable to the kind of nonsensical conclusions that Monroe and Gold warned against. Harkreader and Imershein's lack of a theoretical base is most evident in their fifth condition, whether the governor supported a proposal or did not oppose it. Note that they treat the two as equivalent, coding support and non-opposition as "1" and opposition as "0." Such a conception of the governor's role is questionable, especially without any theoretical support. It seems to ignore any bully pulpit or agenda setting role that a governor might have, which would either directly force the legislature into passing the bill or would encourage popular support for the bill which would in turn lead to legislative support. Concerns such as these are why the current study will test hypotheses that were deduced from a broader theoretical model, as briefly discussed in the previous chapter.

Harkreader and Imershein also chose to study the same institution over time without taking into account any of the contextual factors that might vary over time, which is itself interesting because QCA was *designed* with those types of variables in mind. They make no attempt, for instance, to code for *which* party is in power, although they account for whether the party in power supported the legislation, ignoring the fact that Republicans and Democrats will tend to support different types of plans irrespective of many other factors. From their data, there is no way to deduce any relationship between which party was in power at the time, and whether the bill passed, although one would

hope that they examined whether such a simple relationship fit the data before proceeding to a more complicated hypothesis. They also made no attempt to look at whether the governor and legislature were of the same or different parties (unified versus divided government). All else being equal, one would assume that a bill would be more likely to pass under unitary government. Attention to these and other variables that shaped the context and institutional relationships within which individual actors operated would have allowed for Harkreader and Imershein to create a more complete picture of the policy process and one that would have contributed to the policy and institutions literature at large.

If Harkreader and Imershein looked at the Florida legislature over time, though, then my decision to look only at the *initial* adequacy filing and decision in a given state deserves some explanation. By initial adequacy filing, I mean that I examine only the first time a state supreme court considers an adequacy argument. I don't look at the number or decisions in follow-up cases. In New Jersey, then, I would examine the initial Abbot decision, and not the multitude of compliance and enforcement decisions that came afterwards. The reason for this is rather simple. It is relatively easier to develop comparative hypotheses about how institutional and political factors will affect adequacy decisions when adequacy in the state is a "new" issue. But, in follow-up cases alleging such things as a reversion to inadequacy, or deficiencies in physical plant and construction, each state carries the baggage of its unique history and also unequal judicial enforcement powers. Put another way, comparing the first adequacy decisions made in two states is relatively easy, but comparing institutional and political factors and re-arguments about compliance in two states would leave out an important intervening

variable: what actually happened in those two states. In theory, I could actually code what happened in each state and analyze follow-up cases, but two issues make this very difficult in practice. First, in the process of coding, I would either wind up with an unwieldy number of categories or have to collapse results into a few categories of compliance and follow-up practices so that I would lose most of the variability that the process was trying to capture in the first place. Second, with each successive follow-up filing over compliance issues, the sample size of states that had experienced that sequence of events would get very small (possibly even $N=1$ after a few iterations) and the necessary and sufficient conditions that I could generate would be pedantic in nature (along the lines of given a plaintiff's victory followed by a plaintiff's victory in a follow-up about funding and a partial state victory in a second follow-up about facilities funding, the following conditions are necessary for plaintiffs to succeed in an action about compliance among inner-city districts...).

Lastly, it would be prudent to discuss some scholars who have used QCA in the context of studying court decisions. Perhaps the earliest occurrence in the literature is Musheno, Gregware, and Drass (1991), who use QCA to analyze a sample of AIDS-related cases during the 1980s. Their theoretical framework, with its emphasis on critical legal studies, power, and social class is clearly open to debate, and readers might therefore question the wisdom of the hypotheses that they choose to test. Nonetheless, from a technical standpoint, they correctly employ QCA in two phases of their analysis. They "trace how relational attributes, evident in the social and legal characteristics of contestants...combine in various ways to define wins for various parties...[and] how interpretational attributes, embedded in the text of judicial rulings (e.g., use of divisive

AIDS metaphors) combine in various ways to legitimate wins for each set of parties,” (pp. 755-6).

McClurg and Comparato (2004) carry out a study that appears to be the most similar in concept to the current study, albeit while examining a different theoretical question. McClurg and Comparato are concerned with the decision-making of state supreme courts and the extent to which state supreme courts respect precedent of the US Supreme Court. Specifically, the study dealt with the Supreme Court decision in *Gates v. Illinois*, a search and seizure case regarding specifically the use of anonymous tips. Some state supreme courts have refused to apply the standard for determining the legality of a search that the Supreme Court developed in *Gates*, and have criticized the decision in their opinions.

McClurg and Comparato specifically justify their use of QCA as superior to either case-study methods or quantitative modeling methods, which are “unable to capture the complexity of judicial impact as thoroughly as we would like.” QCA is, to them, a middle ground that “balance[s] the depth of qualitative research with the power gained by comparing multiple cases,” (p. 9). They ran analyses for two different dependent variables. First, the analysis concerned whether state supreme courts respected the *Gates* precedent. Second, they looked at whether the state supreme court decision made it easier for the police to use anonymous tips (as in *Gates*)—separately from what the court said about the *Gates* precedent. This represents an innovative way of analyzing the policy impact of a judicial decision in cases where researchers might care more about policy outcome than the nominal textual outcome.

McClurg and Comparato used four independent variables, and the theory from which they were deduced paid explicit attention to legal, political, and institutional factors. They were 1) whether *Gates* was useful to the decision of the state supreme court; 2) are state supreme court judges selected by election; 3) Is it a liberal state court; and 4) if the Supreme Court has sent signals that they are monitoring the state court. McClurg and Comparato operationalized the second and third variables in a particularly useful manner. The liberal leaning of State Supreme Courts was determined using the Partisan Adjusted Judicial Ideology (PAJID) scores measure developed by Brace and Hall and introduced in Chapter 2. The current study will also employ PAJID measures to determine the ideological leaning of state supreme courts. Additionally, they examine the second variable using only “states that retain their justices via popular election and those that do not,” making no distinction between partisan elections, non-partisan elections, and retention elections. This seems at first glance to be a rather simplistic way to code that variable, and that simplicity is reflected in the fact that, of the 25 cases examined, 80% were coded as having elected judges.

McClurg and Comparato were clearly excited about the use of QCA in analyzing the decisions of state supreme courts, remarking that “in order to move the theoretical debate forward, it is desirable to move the methodology forward as well.” The current study represents the next logical application of QCA to state supreme court decisions. Like McClurg and Comparato, I attempt to take account of the political, legal, and institutional factors that influence decisions. However, the area that I study is not bound by Supreme Court precedent, allowing, perhaps, for an arena where a greater variety of

political and institutional effects will manifest themselves as state supreme courts confront adequacy with few legal limitations.

Variables Employed In This Study

The Dependent Variable

Having settled the question of what method ought to be used in this study, it is still necessary to justify the how the specific variables in this study will be measured. Since I rely solely on secondary data analysis, the methods by which data are collected and coded take on added importance.

The dependent variable, plaintiff or defendant victory in a school finance adequacy lawsuit, is slightly more complicated than it first appears. As West and Peterson (2007, p. 345) write, “school finance cases do not fall neatly into equity and adequacy boxes.” Nonetheless, they develop and consistently apply a classification system for all school finance equity and adequacy cases from 1971-2005 in the appendix of their co-edited volume, which forms the basis for the coding of the current study. The table they present will serve as the source of data about case outcomes for the current study, with the database on the website of the National Access Network at Teachers College (www.schoolfunding.info) serving as a secondary check.. Note that the West and Peterson table has become kind of an industry standard, and was the source that Springer, Liu and Guthrie (2009) relied on for their coding of equity and adequacy decisions, as discussed in Chapter 2¹³. Only cases that reached the state supreme court are included.

¹³ The one controversial coding in the West and Peterson volume deals with three cases prior to 1989. They code the 1981 Georgia decision, the 1978 Washington decision, and the 1984 West Virginia decision as adequacy decisions that the plaintiffs won. Most scholars, including the National Access Network, though, begin their adequacy calculations in 1989, which reflects the first time that adequacy lawsuits were

A Brief Review of Theory

Chapter 1 discussed Mayhew's (1974, 2nd Ed. 2004) hierarchy of goals pursued by members of the House of Representatives. In Chapter 2, Mayhew's theory served as a kind of guide to considering the various factors involved in judicial decisions in adequacy lawsuits. There are numerous examples in the literature of judges as political decision-makers, weighing costs and benefits and taking the action that most furthers their interest, however that interest is defined. Mayhew's model provides a framework for considering the different goals that judges might pursue and the variables through which those goals might be expressed. A brief restatement of that theory, and its application to the current study, would be helpful here.

Mayhew identifies three goals that legislators pursue. A legislator's first goal is reelection. The pursuit of any goal that one has as a legislator is predicated upon remaining a legislator, and so reelection is, according to Mayhew, the first goal on which all others depend. In the House of Representatives, which is Mayhew's focus, the two-year term makes the pursuit of the reelection goal a near-constant concern. Conceivably, in legislatures with four- or six-year terms, the pursuit of the reelection goal could be muted somewhat in the middle years of a term.

The second of Mayhew's goals for legislators is attainment of prestige or power.

What is considered prestigious for a legislator is, of necessity, shaped by the structure and

connected to the standards movement, as has been the case in subsequent cases. In any event, the coding of those cases does not affect the results of the analysis presented in Chapter 4. I also adopt the National Access Network coding for Georgia, which uses the 2005 decision as an adequacy decision in favor of the defendants. I adopt the National Access Network coding for Washington as well, which lists no adequacy decision in the background text available on their website, as the parties agree to stay the lawsuit as the state implemented reforms. The website also reports no state supreme court involvement in West Virginia, where the trial court has required reform and then ended its jurisdiction, while relying primarily on equity reasoning. Given the results that I discuss in Chapter 4, and how far any causal combinations were from being necessary and/or sufficient, it is not plausible that these recodings had any effect on findings of necessity and/or sufficiency.

norms of the institution. In the context of the House of Representatives, appointment to powerful committees (or, even better, the chairmanship of said committee) is one of the most common ways that legislators might gain prestige or power. Appointment to a party leadership or fundraising post would be another way.

Mayhew's third goal is the pursuit of "good" policy, which can be pursued most easily once the first two goals have been reached. Presumably, it is comparatively easier for more powerful legislators to see their preferred policies enacted.

With slight modification, Mayhew's three goals for legislators can be applied to judges. Judges would seek 1) retention of their position on the bench; 2) maintain or enhancing prestige for themselves or, more broadly, for their institution; and 3) to make "good" law. There are two changes from Mayhew's goals to this formulation for judges. The first goal has been reworded, and the second expanded. The first goal was modified simply because, as discussed in Chapter 2, not all state supreme court judges are elected. The methods of judicial selection differ among the states, and some judges will face partisan elections, others retention elections, and yet others appointment.

The reasoning behind the modification of the second goal is slightly more complicated. Legislators, because of the need to frequently win reelection, will seek prestige for themselves. For state supreme court judges, the calculus is slightly different, partially because judicial elections are, in most cases, less competitive and different in tone than congressional elections. It is possible that judges seek prestige for themselves, for example, by writing an opinion that will frequently be cited, but it is also the case that judges seek to preserve and enhance the prestige of the court as an institution relative to other institutions of state government, and that such concerns about institutional prestige

weigh more heavily on judges than on legislators, as the judiciary is nearly always dependent on other branches of government for enforcement.

In *Marbury v. Madison*, for instance, Chief Justice Marshall was cognizant of the court's limited institutional power, and its reliance on other branches of government for enforcement of judicial decisions. Indeed, some accounts depict the very creation of judicial review as a way for Marshall to avoid the damage to the court's institutional prestige had President Jefferson refused to abide by its ruling. And in *Brown v. Board*, Chief Justice Warren made a substantial effort to craft a ruling that would result in a unanimous decision, as he knew that the decision would be unpopular in the south and wanted to use unanimity as a tool to strengthen the prestige and legitimacy of the court's decision over what would be significant resistance in implementation.

Courts, then, have to take the issue of institutional power and prestige much more seriously than other branches of government, as they are wholly without enforcement power. Indeed, it has been suggested that courts that rule in adequacy lawsuits are creating a significant separation of powers issue, by circumscribing the power of the legislature to pass a budget and appropriate funds (Dunn and Derthick, 2007). Courts depend wholly upon other branches of government for enforcement of their adequacy rulings, and, in light of the historical record, ought to be less likely to rule for the plaintiffs in adequacy cases where there is a likelihood that their decision will be either ignored or not implemented. After all, a court whose decisions are not implemented in one case will lose the institutional prestige and ability to make important policy decisions in future cases. Variables that are included in the current study in an effort to capture this goal, then, will measure those things which are likely to lead to non-

implementation of a court decision and which would therefore damage the prestige both of the court as an institution and of the individual judges involved.

Mayhew's third goal of making "good policy" translates nicely to judges' goals of making "good law." I begin with the assumption that, in any complex case, there are multiple possible interpretations that would be acceptable as a matter of law. Judges, however, like other politicians, have their preferred solutions, or policy positions, which they would advocate for within their institutional constraints. Some judges might consider good solutions to be ones that guarantee individual freedoms and keep government small, for instance, while other judges might favor solutions that call for a more expansive government role.

Nothing in the above paragraph is meant to offend any judge. Clearly, the facts—as judges see them—and the applicable precedents play a defining role, and it seems unlikely that a judge would willfully come to a decision unsupported by facts, precedent, and legal reasoning. What it does mean, though, is that judges are products of their experience, ideology, and training, and that all of that is likely to influence judicial decision-making. Clearly, were judging entirely fact-based, it would also be entirely predictable with nearly 100% accuracy. In areas of the law where there is no precedent and where the intent of the law is vague, like adequacy, judges have a blank slate. How judges fill that blank slate clearly will depend upon factors like ideology, experience, training, and philosophy, and not merely upon the facts of the case.

Selecting Independent Variables and the Problem of Omitted Variables

The independent variables for the current study are listed in Table 1 (below). All of the independent variables stem from the literature reviewed in Chapter 2, and the

theory which guided that review, discussed above to refresh the reader's memory. Each independent variable corresponds to one of Mayhew's hypothesized motivations.

While the independent variables selected for the study are theory-driven, there is still the question of possible omitted variables. While it is impossible to justify every possible variable that wasn't included (or even to list them, since the possibilities are theoretically endless), I did discuss the concepts of theory and trivially necessary conditions in Chapter 2. The variables selected for this study are credible proxies for the underlying theoretical construct. And, many other conditions that might be necessary either don't vary (for instance, the presence of a state court system) or are necessary but trivially so (the existence of intrastate disparities in achievement). Another limitation to keep in mind is that variables selected for QCA must be able to be turned into binary in some systematic and useful manner. As the coding of the judicial selection variable indicates, a variable with too many attributes, each of which is examined separately for necessity, will overload the model.

The current study does not have a variable that takes the facts of the case into account. While this is a conscious decision, the fact that many other studies of judicial decision-making *do* have such a variable necessitates some discussion as to how its omission here does not compromise the study. Put simply, the facts "on the ground" simply aren't very important. Adequacy cases rely on essentially the same argument from state-to-state, and, in every state, it is possible to find some evidence of inadequacy or inequity, particularly when inner city schools are compared to richer areas. The logic is analogous to explanations of the Supreme Court's decision in *Brown* or in *Roe v. Wade*. In neither case did the fact pattern play a central role, rather there was a broader

constitutional question at issue. While both *Brown* and *Roe* have been analyzed by generations of scholars, no one that I can find has suggested that the facts of the case—exactly how far Brown lived from the White and Black schools, or the psychology that caused Roe to want an abortion—played a decisive role in the decision. The same logic applies in adequacy cases. Courts have the freedom to reasonably find for either party, and the fact pattern under consideration is not nearly as important as the bigger constitutional question at issue.

Conceptualizing and Coding the Independent Variables

The independent variables are summarized in Table 1 and are based on the literature reviewed in Chapter 2. What follows is a discussion of how they are gathered and coded. This is particularly important in QCA, because, as I discuss in Chapter 2 and again at the beginning of Chapter 3, variables must be binary, and the number of variables that the model can handle is finite. Therefore, it is impossible to test separately for the effect of *every* judicial selection system, or *every* political culture, and so only the relevant differences, as elucidated in Chapter 2, are coded for.

TABLE 1

THEORETICAL JUDICIAL MOTIVATION	INDEPENDENT VARIABLES	HOW VARIABLE IS MEASURED
Retention	- Method of Judicial Selection	Binary – Elected or Not
Prestige <i>Contextual Factors</i> <i>Institutional Relationships</i>	- Diffusion Through Federal Circuit - Diffusion Through Strength of Constitutional Clause - Diffusion Through Political Cultures - Divided Government - High-Performing Court (professionalism)	Diffusion calculations described in text. Coded 1 for a positive number (indicating a plurality of decisions for plaintiffs), 0 otherwise. Coded 1 if governorship and both houses of legislature controlled by the same party, 0 otherwise. Coded 1 if “Top 10”, 0 if otherwise. Source and description in text.
Good Law	- Median Court Ideology	PAJID Scores. 1 for courts that are more liberal than the median state supreme court in that year, 0 for courts that are more conservative. No court in the study fell exactly at the median.

The first independent variable is the method of judicial selection. As discussed in Chapter 2, the relevant difference here is whether judges are subject to elections, and therefore subject to the incentive that Mayhew discusses, or not. Partisan and non-partisan elections are treated the same (as “elections”) and are coded 1, all other systems are coded 0. Recall that the hypothesis is that elected judges will be *unlikely* to decide in favor of plaintiffs, because of the redistribution involved and the fact that the affluent, who lose under a redistributive plan, are more likely to vote and to contribute to campaigns.

Recall that policy diffusion was discussed and applied to the judiciary in Chapter 2. According to the theory, rulings and the logic behind them tend to diffuse from one state supreme court to another (through mechanisms like inter-citation), in much the same way that Walker demonstrated diffusion of legislative innovations. Three possible channels of diffusion were discussed, and, in each case, the possible channels of diffusion are based upon different ways that states can be similar. The three ways were: 1) diffusion through the same federal circuit 2) diffusion through the strength of the constitutional adequacy clause; or 3) diffusion through political culture.

While these data are easily available, a method of coding them as binary variables for QCA requires some justification. Walker, for instance, used a proportion of states that had implemented a policy as a measure of diffusion, but such a measure is not usable here. Recall that the underlying theory is that a court will be more likely to find for plaintiffs once another court in a state that is similar in some way has made such a finding, as this will serve to provide legitimacy to a court's decision in the face of what is likely considerable opposition to adequacy among the political elites. If there were no opposition, it is likely that the legislature would provide adequate funding without the need for court intervention. Therefore, in coding diffusion variables, I examine all the similar states that have had an adequacy decision before the state in question, and create a scale whereby I add one point for every decision in favor of adequacy, and subtract one point for every decision where a court failed to find for the plaintiffs. If a state winds up with a positive number, that gets coded as a 1 in the QCA analysis, because there exists the added legitimacy that diffusion can provide. If a state winds up with a 0 or a negative number, that state will get coded as a 0.

Although this measure is the best possible way to code for diffusion in the QCA analysis, there are two possible objections that one might raise. The first is that coding diffusion this way ignores the magnitude of the effect, in that three prior decisions for the plaintiffs will be coded the same as one prior decision. The second is that the coding method ignores mixed signals, in that it treats two positive decisions and one negative the same way that it would treat a single positive decision.

The response to both these objections is that, the way I have presented diffusion, it is designed to give added precedential authority to a court that wants to make a certain decision by being able to point to other state supreme courts that have made a similar decision (perhaps through citing their decisions). To do so requires only a general trend in that direction, which, in the current scheme would be illustrated by a positive number. The existence of a single countervailing case would not undo the mechanism of diffusion, especially since courts will cite strategically. Additionally, I would argue that the number of cases isn't as important as the prevailing trend, which is why I feel comfortable with a method that does not code the number of positive decisions. Because diffusion of a policy lends legitimacy to the policy, it is more important to look at what has happened as a whole in the relevant comparison states than the exact number of decisions. Put simply, I am measuring whether there is previous support for adequacy that courts could draw on if they wished or not.

The computation of the diffusion variables is rather straightforward. Thro's analysis of the strength of the constitutional adequacy clause will be used to create diffusion groups, as will Elazar's classification of political culture. Federal circuits are a matter of public record. Thro developed a typology of education clauses in state

constitutions that separates them into four “Categories” with higher numbers denoting a higher obligation on the part of the state, which he based largely on earlier work by Gershon M. Ratner. Of course, the categories are categorical variables, and, as such, the difference in the level of legal obligation between a Category I and Category II clause will be different than, for example, the difference between a Category III and Category IV clause. Thro identifies Category I education clauses as those that “impose the minimum educational obligation on a state.” Category II clauses “mandate that the system of public schools meet a certain minimum standard of quality such as ‘thorough and efficient.’” Category III clauses are different than category I and II clauses because Category III clauses have a “stronger and more specific education mandate [and] purposive preambles.” Category IV clauses, the strongest, “impose the greatest obligation” and “typically, they provide that education is ‘fundamental’, ‘primary’, or ‘paramount.’” As discussed, QCA requires that the variables be converted to binary. Therefore, Thro’s four-category typology was converted to two categories, where the state constitution either imposed a minimum educational obligation or did not, which is to say either a Category II or higher clause (coded as 1) or a Category I clause.

Divided government is a matter of public record, and the particular data were obtained from a spreadsheet prepared by the National Conference of State Legislatures. Governments were coded as a 0 if unified (both houses and the governor of the same party) and a 1 if otherwise.

Brace and Hall developed PAJID scores as a measure of judicial ideology based on a judge’s writings and the decisions that they author and/or sign. The measure has been validated and is used extensively in the field. Higher PAJID scores indicate

increasingly liberal ideology. Since we are concerned in the current study with the relation of one state court to another, I compute the median¹⁴ PAJID score for each year under study, and code 1 for a court that is more liberal than the median, and a 0 for a court that is more conservative. In no case does a court under study exactly equal the median. Brace, Hall and Langer typically publish new PAJID data every few years. If in any case the data cannot be located because it has not yet been computed, then the last available value will be used, provided that the party balance on the court has not changed. As a last resort, that data point will be coded as missing. The PAJID data used in this study was collected by Stefanie Lundquist at Vanderbilt University as part of the State Politics and the Judiciary Project, who graciously made her data available for this study¹⁵.

The data on high performing courts was computed by Choi, Gulati and Posner (2008), as discussed in Chapter 2. Inclusion of this composite variable will allow for the measurement of a variety of characteristics—court independence, productivity, and influence—while taking account of state-specific factors and allowing for the limited number of variables appropriate for the QCA analysis. “High-performing” courts, which they consider as those in the Top 10, will be coded as 1, others will be coded as 0.

Although Choi, Gulati and Posner take their measurement at one point in time, they

¹⁴ In many cases, the current study converts continuous data to binary by coding a data-point as either above or below the median. This practice follows the advice that Ragin offers in the documentation for the fs/QCA software, where he advises using the median when no other theoretically-driven cut point is evident. As an exercise, though, I did try using running the model with different cut points (e.g. top quartile or not, bottom quartile or not) with no change in the substantive results.

¹⁵ I am aware that the State Politics and the Judiciary Project includes some state-level education variables (for instance, per pupil spending, an equity index, and education spending as a percent of Gross State Product) which at first seemed ideal for this project. However, Lundquist computes these variables only at certain points in time, or, in a few cases at one point in time. Since her dataset offers no way of knowing which variables might have changed between the year of collection and the year of an adequacy decision, it would be questionable to use that data.

demonstrate, through a review of prior literature, that the factors on which they are relying are unlikely to change much over time. For instance, the California Supreme Court has long been recognized as an influential court, and has long led the nation in inter-citation.

Data Analysis Program Employed

The current study employs the fs/QCA program developed by Charles Ragin and available for download on his website at www.fsqca.com. The crisp sets procedure is employed, following the procedures set forth in the documentation, available for download with the software. The recommended thresholds for frequency and consistency are applied¹⁶.

Possible Findings and Interpretations

This study uses a number of independent variables, summarized in Table 1. When a number of independent variables are used, it is helpful to spend some time thinking about possible outcomes, and the conclusions that those outcomes might support. This will make it easier both to interpret the findings and to avoid, to some extent, charges of post-hoc theorizing.

As discussed earlier, QCA put the independent variables in binary form and allows for test of necessity and sufficiency. Furthermore, QCA also allows for multiple conjunctural causation, which means that there can be multiple combinations of variables, any one of which is necessary or sufficient for the outcome in question. I expect to find

¹⁶ As discussed above, consistency refers to the extent to which the set/subset relation is maintained, which is to say the proportion of cases that display the necessary condition and the same outcome. Ragin recommends a minimum threshold of .75 for QCA to be useful, and compares this convention to the convention of using the .05 p-value.

multiple conjectural causation in this case, which means that several possible combinations of variables would lead to a plaintiffs' victory in an adequacy lawsuit.

The three rows of Table 1 list the three possible motivations of judges, as represented by Mayhew's typology, in the left-hand column and the variables that represent those motivations in the current study in the right-hand column. For instance, whether or not judges are elected is an operationalization of the retention incentive. And PAJID scores, a measure of judicial ideology, is a measure of a judge's desire to make good law, where "good" is used in the normative political sense of the word, and judges will consider good law to have been made when it matches their policy preferences, which are heavily influenced by ideology.

If variables from only one motivation are necessary, I could conclude that the motivations of judges are very different than the motivations of legislators—perhaps judicial motivations are not dependent upon whether or not they are subject to elections. I could also find different "clusters" of states that have different combinations of necessary conditions. For example, perhaps some state supreme courts look to decisions in states with similar political cultures, and for only those courts, the divided government plays no role at all.

Additionally, a failure to find *any* necessary conditions could indicate that a different operational definition of one (or more) of Mayhew's categories is warranted. The null hypothesis here is that the absence of a relationship between these variables and case outcomes indicates that decisions are based on idiosyncratic state-specific and case-specific factors. It could, however, indicate the need for a different, and less ideological

conception of good law, where good means respecting the facts and context of each individual case, instead of putting your policy preferences into law.

Lastly, I expect to find that every combination of necessary conditions will include the PAJID score variable. While judicial decisions might not be exclusively (or even primarily) driven by ideology, it would be surprising to see a combination of necessary conditions that did not involve ideology at all. Perhaps most importantly, examining the different combinations of necessary conditions might give us insight into complex institutional relations about which we could not speculate beforehand, perhaps suggesting other relationship or variables that future researchers might want to look at more closely. Perhaps, even, the relationships discovered in this study might give us some degree of insight into judicial decisions about other social welfare issues.

There is also the possibility, however, that the results will not be as illuminating. QCA allows for multiple conjectural causation, but that also means that there could be so many paths to plaintiffs' victory, each covering one or two or three states, that few interesting general conclusions could be drawn. In that case, one could conclude that, at least in adequacy cases, judges make decisions based primarily on the merits, or on state-specific factors so that no generally applicable model is possible. It might also be valuable to recode some of the diffusion variables for the presence or absence of specific traits (e.g. a moralistic political culture, or a strong education clause) if the diffusion construct turns out not to be applicable.

The next chapter will conduct the data analysis to answer these, and other, questions.

CHAPTER 4 – RESULTS

The data for this study were analyzed using Qualitative Comparative Analysis (as described in Chapter 3). Specifically, the data were analyzed using the Windows Version fs/QCA software written by Charles Ragin and his colleagues, specifically for QCA (available for download online at www.fsqca.com). The accompanying manual, also available for download, describes the procedures for running QCA using the crisp-set procedure. The current study followed those procedures.

Brief Summary of Results

When QCA was run on data collected for the current study, using three different models discussed below, one could conclude that there are no conditions that are non-trivially necessary and/or sufficient for plaintiffs' victory in an adequacy litigation. This is due to the high number of rows in the truth table that fail to display a consistency score of at least 0.75. In other words, for the hypothesized combinations of necessary conditions tested in the current study, the set-subset relationship was not maintained frequently enough for any of these variables, either alone or in combination, to be necessary. According to Ragin (personal communication, 2010), when a large proportion of the cases fail to display sufficiently high consistency, it is correct to interpret that as evidence of the absence of necessary and/or sufficient conditions, provided there are not important omitted variables.

I discuss the implication of this finding below, in terms of the null hypothesis discussed earlier. Recall that the null hypothesis for the Good Law category could be understood as a *different conception* of good law, which largely excludes policy

preferences and instead focuses on the evaluation of evidence and constitutional interpretation to produce a legally sound finding.

Terminology For Interpreting QCA Results

The theoretical basis behind QCA and its interpretation was discussed in Chapter 3. However, it is useful to review the nuts-and-bolts issues of terminology and notation, to allow for easier interpretation of results

When using statistical analysis, one or more independent variables are said to lead to one dependent variable. QCA, however, is not a statistical method, but rather one that is concerned with finding necessary and sufficient conditions. To reinforce the conceptual difference, Ragin advises against calling the results of QCA analyses independent and dependent variables. Instead, he prefers the terms “conditions” and “outcome.” In other words, QCA is not investigating whether unified government leads to plaintiffs’ victory in an adequacy lawsuit, but rather whether unified government is a necessary and/or sufficient condition for plaintiffs’ victory in an adequacy lawsuit.

The evaluation of hypotheses about necessary and sufficient conditions requires some numerical tools, in much the same way that statistical methods rely on the concept of statistical significance. Chapter 3 detailed the concepts of consistency and coverage and Ragin (2006) gives a detailed description of the mathematics behind them. The definitions, however, deserve repeating here.

Consistency is the degree to which the set-subset relationship is maintained. For instance, if one hypothesizes that A is necessary for X, and all instances of outcome X display condition A, then the consistency of that relationship is 1.0. While, of course, it would be easier if set relations were always perfectly consistent, Ragin and others made

the point (detailed in Chapter 3) that necessary and/or sufficient condition hypotheses could be valuable even without perfect consistency. He proposed a threshold of 0.75 for consistency, which has become the standard in the field. With a consistency level of less than 0.75, it becomes nearly useless to speak of necessary and/or sufficient conditions because the set-subset relationship does not hold.

Coverage refers to the number or proportion of cases displaying the outcome that are accounted for by a given necessary condition or combination of conditions. While Ragin offers no hard and fast guidelines here, because the relevance depends more on the theory and the nature of the explanation than on a specific number, he does warn against interpreting necessary conditions that have low coverage in the context of a particular study.

Consider, for example, a study that seeks to examine ways in which low-income residents of a certain city avoid falling into poverty. One of those residents wins the Powerball lottery, and consequently avoids falling into poverty. Also, assume that this resident was destined for poverty had he not won the lottery, as he showed none of the other necessary conditions for poverty avoidance: he did not hold a high school credential, attend job training, etc. For this one resident winning the lottery would be a necessary condition for staying out of poverty, and, as he was the only lottery winner, the consistency of the relationship would be 1.0. The coverage of this necessary condition would be very low, however (only *one person* out of a whole city), and we would clearly be advised to treat this as irrelevant for the purposes of theory development or practical application.

It would be prudent to briefly review the logic behind each of the variables tested in this study. It was hypothesized that method of judicial selection—specifically whether judges are elected or not—would impact decision because elections should serve to link judges more closely to the public. The people most likely to turn out for judicial elections are also those most likely to oppose the redistribution that adequacy implies, meaning that elected judges should be less likely to decide in favor of plaintiffs.

It was hypothesized that judges would be more likely to find for plaintiffs if the prevailing trend in states that were somehow similar was to find for the plaintiffs. In other words, rulings would diffuse across similar states. There were three hypothesized channels of diffusion, reflecting three different ways in which states could be similar: 1) federal circuit; 2) strength of constitutional clause; 3) political culture.

Two types of institutional relationships were hypothesized to affect decisions. Divided government should affect judges' decisions because legislative gridlock leaves the courts somewhat freer to make a broad ruling, secure in the knowledge that the legislature (or the governor) is not in a strong position to successfully undermine a court decision.

It was also hypothesized that a high-performing court, which had, by definition the institutional support necessary to evaluate the complex social science evidence in adequacy cases would be more likely to decide in favor of the plaintiffs if it were already inclined to do so. In other words, a high-performing court does not by itself imply a victory for the plaintiffs, but rather could be part of a combination of necessary and/or sufficient causes for plaintiffs' victory.

Judicial ideology was hypothesized to affect decision because more liberal courts would be more likely to find for the plaintiffs, as liberal judges would favor the redistribution that adequacy lawsuits represented. Judicial ideology was measured using the median ideology score for the state supreme court for the year of decision, and coding whether it was more or less liberal than the median state supreme court for that year.

In this study, as in all well-constructed studies, it is possible that the hypotheses might not be confirmed. In the current study, two types of non-confirmation are possible. In the first, some variables might be necessary and/or sufficient, while others might not. In that case, there would be a failure to confirm some part of the theory, whereby one hypothesized judicial motivation either was improperly operationalized in the translation from legislative to judicial motivations, or perhaps does not apply to courts at all. The second type of non-confirmation involves finding no necessary and/or sufficient conditions that either cover many states or maintain a high enough degree of consistency to be theoretically relevant. If the problem is primarily one of coverage, then the interpretation would be that some substantively important variable has been omitted, which would, when included, tie together two or more subsets. If both consistency and coverage are problematic, then the appropriate conclusion would be that there are no necessary and/or sufficient conditions that derive from the current conceptual argument. For example, the failure of the current study to identify necessary and/or sufficient conditions due to both low coverage and low consistency could imply either that no such conditions exist, or that a different conception of “good law” or “prestige” is necessary, perhaps because the incentives facing state supreme courts are different than originally hypothesized.

One might be concerned that necessity and sufficiency (or even near-necessity and near-sufficiency) are very high standards, when compared to the significance tests by which multivariate analyses are judged. In other words, the concern is that, too often, QCA analysis can never be successful, even in places where regression might have been. There are two possible answers to this objection. First, regression and QCA are designed for different purposes. QCA is designed for small and medium-sized datasets, where the sample size might not be big enough for regression analysis. Second, QCA yields logical results about necessity and/or sufficiency, while regression yields standardized coefficients and effect sizes. These results provide different types of interpretations suitable to different research questions. In a large dataset, one might be interested in interpreting the standardized coefficients of each independent variable, but it is difficult to see how they would be very useful here in attempting to construct a guide for advocates, who could more likely use results about necessary and sufficient conditions. In short, regression and QCA are two different tools suitable for different types of data and research questions.

Detailed Results of Specific Models

Model 1

The first model tested involved all the variables discussed in Chapter 2 as hypothesized necessary and/or sufficient conditions. The variables in Model 1 include: 1) Judicial Election; 2) Divided Government; 3) a high-performing court; 4) a liberal court; and 5) the three diffusion variables, based on federal circuit, strength of constitutional clause, and political culture. The truth table for this model, and for all subsequent models, is provided in Appendix 2.

Using the standard consistency value of 0.75, it is clear from inspecting the truth table that only 18 of the 22 plaintiffs' victories meet the consistency threshold.

When the model is run, and the parsimonious solution¹⁷ interpreted, we see that each combination of conditions contributes very little to overall coverage. The two largest contributors to coverage are (the absence of diffusion through political culture and unified government) and (an unelected judiciary and the absence of diffusion through federal circuit and a liberal ideology)¹⁸, but each solution can account for only four cases, and there is overlap in states accounted for. Since no combination of conditions that accounts for only four cases can conceivably be called necessary for the outcome in any meaningful sense, a truncated model was constructed in an attempt to increase the coverage of a single solution.

Model 1A

Three of the variables in Model 1 relied upon the concept of diffusion of judicial rulings from state to state. However, there exists the possibility that hypotheses regarding the diffusion concept might be incorrect, as prior research has identified diffusion in legislatures and courts, but not specifically in adequacy litigation. Given the failure of Model 1 to produce necessary conditions, Model 1A was a re-running of Model 1 excluding the three diffusion variables.

In this case, the results are even worse, with only 11 states covered by solutions that reach the required consistency threshold (0.75). There is one solution (the presence

¹⁷ The parsimonious solution refers to the solution after all logical canceling has been completed. For instance, if one necessary condition is $A*B*C$ and a second is $A*B*\sim c$, and both lead to the outcome, then it is clear that C is irrelevant to the outcome. Since the outcome occurs either in the presence or absence of C , the parsimonious solution will ignore C entirely and report $A*B$ as necessary conditions.

¹⁸ The solution can be represented symbolically as follows: $\sim pol_cult_diff*\sim divgovt$ and $\sim judicial_electi*\sim circuit_diff*\sim lib_court$

of judicial elections and a conservative court) that covers five states (Idaho, Ohio, Arkansas, Texas, Montana)¹⁹. On the whole, however, the model is not useful because the model, and the three causal combinations that it generates, accounts for the plaintiffs' victory in only eleven states.

Model 2

A second model was run, to account for the possibility that the hypothesis regarding the presence of the diffusion mechanism was incorrect, but to take account of political culture and strength of the constitutional clause. These two variables could, conceivably have direct effects on decisions in adequacy litigation outside of the diffusion mechanism. The political culture variable was recoded to take account of the fact that moralistic political cultures would likely tolerate the redistribution involved in a successful adequacy lawsuit. Therefore, the presence of a moralistic political culture was coded as 1, and either of the other two political cultures was coded as a 0. Additionally, the constitutional clause variable was recoded to take account of whether the constitution mandates at least a "minimum standard of quality" (Thro, 1989, p. 1663). According to Thro's analysis, that requirement is characteristic of clauses that he classifies as Category II or higher, but not characteristic of Category I clauses. Therefore, the constitutional clause variable was recoded as either requiring least a minimum standard of quality (Category II or higher) or not requiring that minimum standard of quality.

Additionally, the variable that measured whether a court was more liberal than the median state supreme court was removed, and replaced with a different variable, computed from the same State Politics and the Judiciary dataset, that measured the difference between elite ideology and citizen ideology. After all, with the diffusion

¹⁹ The solution can be represented symbolically as follows: `judicial_electi*~lib_court`

variables no longer in the model, it no longer made sense to compare a court to court's in other states. Instead that court ought to be compared to the citizens, who need to consider it legitimate. A 1 on this variable meant that the elite ideology was more liberal than the average citizen ideology, and a 0 meant the opposite.

In this model, only 14 of 22 positive cases met the consistency threshold to be included in the model. When the model was run, no combination of necessary conditions accounted for nearly all cases. Interestingly, eight cases displayed a combination of necessary conditions that is strangely counterintuitive, (divided government and a weak constitutional clause)²⁰. There is no theory that would support those two variables operating in concert to produce a victory for the plaintiffs, and Ragin (2006) specifically advises that there can be no necessary conditions independent of a theory that specifies a causal relationship. Indeed, that result is so theoretically unexpected, that one might conclude that there is an omitted variable that accounts for the appearance of necessity in that case, and that the current study has uncovered a logical equivalent of a spurious correlation.

Are the Best Cases Filed First?

Although QCA analysis is a useful tool, it is not suitable for examining temporal relationships, as discussed in previous chapters. Therefore, an analysis was also performed of the level of inadequacy in cases over time. If the "best" cases were filed first, one would expect to see that plaintiffs with lose with greater frequency as time went on, To test this hypothesis, the per-pupil expenditure of the median district in each state was calculated, as well as the standard deviation of the district per-pupil expenditures. The standard deviation is likely the better measure of inadequacy (or inequity, for that

²⁰ Represented symbolically as: $divgovt^* \sim strong_const_clause$

matter), because comparing medians across states can mask legitimate between-state variation in the cost of inputs, and could lead to false claims of inadequacy.

No pattern is evident when examining the data, which can be found in Appendix 3. When the cases are considered in temporal sequence, there is no clear trend to indicate that later cases showed less inadequacy or inequity. Furthermore, there is also no evidence that low levels of inadequacy make it any more likely that plaintiffs will lose. For instance, plaintiffs lost in the state with the lowest standard deviation (Florida) but won in the second-, third-, and fourth-lowest (North Carolina, Maryland, South Carolina).

Summary

If there had been necessary conditions discovered for plaintiffs' victory, it would then be appropriate to look for necessary conditions for defendants' victory. However, as stated above, it appears that the best interpretation of these results is that there is no condition or combination of conditions that is necessary and/or sufficient to for plaintiffs' victory among the variables explored in the current study²¹. While each model provided necessary conditions for some of the cases, the counterintuitive nature of the findings in Model 1A and Model 2 would lead one to use extreme caution in interpreting them. There are three possible interpretations of these results, the implications of which I explore in the next chapter.

First, it might be the case that judges look at the facts and law in each case, and reach a decision based on legal process factors, and that, in adequacy cases, political and institutional factors do not play much of a role. This would be a different conception of

²¹ Nonetheless, all three models were run for defendants' victory as well. The results were similarly unsuccessful.

the “good law” motivation of judges. Perhaps judges, at least where school finance is concerned, are not concerned with seeing their policy preferences enacted, but rather are concerned with constructing a sound legal argument. “Good” law could, for many judges in most cases, be good because it accords with the principles of legal reasoning and follows precedent where one exists. While there are those Culture Wars issues in which policy preferences do play a role, perhaps judges are content, most of the time, to simply construct the best legal argument.

Second, it might be the case that there are important omitted variables that would tie many cases together, or explain the counter-intuitive results of Model 2. These omitted variables, which perhaps represent different conceptions of variables already included, could have resulted in a problem that is analogous to spurious correlations. Perhaps there is an unobserved necessary condition which, by either its presence or absence, is driving some of the relationships that have been observed, or which would be present in a larger number of cases.

Third, it might very well be that there are no necessary and/or sufficient conditions for plaintiffs’ victory in an adequacy lawsuit, either because they don’t exist or because, in practice, the court decision cannot be disentangled from developments involving other institutions and other forums to pursue adequacy. While this might be a good thing from the standpoint of normative legal theory, it does represent a serious blow to the hypothesis under study. The discussion in Chapter 2 of ESEA and advocates’ pursuit of equal educational opportunity through multiple venues could be relevant in interpreting these results. If advocates pursue adequacy simultaneously in multiple venues, and results interact across venues, then it is conceivable that that introduces some

idiosyncrasy into the process. There appear to be no necessary and/or sufficient conditions because either there aren't any or because, due to action in other venues that varies between-states, the hypothesized necessary conditions have different effects in different states. Additionally, it appears that the objective facts of the case, as measured by the level of inadequacy, are not a predictor of case outcome, and that temporal sequence is also not a predictor of case outcome.

The next chapter will explore the implications of these results in greater detail, and lay out an agenda for follow-up research.

CHAPTER 5 – DISCUSSION

As discussed in the previous chapter, this study discovered no necessary and/or sufficient conditions for plaintiffs' victory in an adequacy lawsuit. The mere fact that the evidence did not support the primary hypothesis does not, however, mean that the study itself failed in generating new and useful knowledge. Indeed, because this study discussed credible rival hypotheses, the non-findings are themselves valuable.

Explanations for the results of this study fall into three broad categories. The results could be explained by 1) concluding that there are no necessary and/or sufficient conditions; 2) concluding that there were theoretical problems, and that the results support a revision of the theory presented in Chapter 2; or 3) concluding that there were problems with the methods employed to test the research question. I consider each of these interpretations, and the implications that follow from them, below.

POSSIBLE EXPLANATIONS AND IMPLICATIONS

There Are No Necessary and/or Sufficient Conditions

One possible interpretation of these findings is that there are no conditions that are necessary and/or sufficient for plaintiffs' victory in an adequacy lawsuit. This would explain both the low consistency values observed in the truth table, as well as the fact that no condition or combination of conditions was necessary in a majority of cases.

This interpretation would lead one to conclude that there is, at best, a probabilistic relationship between the hypothesized necessary conditions and outcome in an adequacy lawsuit. Put another way, there is enough uniqueness among states that there can be no necessary and sufficient conditions that apply to a large number of them. Each case could also become more unique as it winds its way through the legal and political

processes unique to each state. In such a situation, there is not enough similarity between cases to make meaningful comparisons.

The discussions of desegregation and ESEA in Chapter 2 would help to justify such a finding, particularly when one considers the simultaneous actions undertaken by advocates in different venues. Just as desegregation and ESEA resulted from complex interactions across multiple forums (for example, legislative action, court action, government committees and reports), adequacy is also pursued in multiple forums in each state, and the choice of forum and subsequent result helps to frame the issue. It is conceivable that the state-specific interactions that lead to a case being filed in the first place set each state towards its own path-dependent result, and that the *way* a case gets to court matters a great deal, which this study did not take into account. This pursuit of adequacy in multiple forums, and the politics that accompanies the filing of a case in the first place could matter a great deal, and also represents a possible omitted variable in the current study.

It could also be the true that states are comparable but that no political conditions are either necessary and/or sufficient if judges employ a strict legal standard. Perhaps state supreme court judges try to look only at the facts and applicable law, so that the legal process is more significant than any political considerations. This would imply that a different underlying theory is appropriate, which I consider in the next section.

The Underlying Theory Needs to be Revised

Recall that the current study was premised on Mayhew's three goals for legislators, adapted for judges. The way that those goals were defined and then measured surely must play a role in the conclusions. Specifically, the current study operated under

two assumptions, which, although justified in Chapter 2 on the basis of existing literature, could still be incorrect. Specifically, this study operationalized the “prestige” goal to mean prestige for the institution, and “good law” for judges to mean seeing their preferred policy position enacted, as is the case for legislators.

Both of those conceptions could be incorrect. It is possible that judges are much more like other politicians than I gave them credit for, in which case judges are self-interested like other politicians, and there is the same variation between judges as there is between legislators. Just as there are some senators who have a great deal of respect for the institution, and others who seem to have less respect for those institutional norms, judges’ attitudes could run the gamut from valuing institutional prestige over all else to personal prestige over all else. While some judges surely care about maintaining the prestige of the court and its independence, there could very likely be others concerned primarily with ensuring their own place in the history books by writing a famous and widely-cited opinion. The implication is that self-interested judges will find themselves constrained—in the same way that other politicians are also constrained—by different political conditions and opportunities that can lead to different behavior. In that case, the variables I employed to measure this goal would not be necessary and/or sufficient to plaintiffs’ victory because they are simply the wrong variables, as my perception of what judges value was incorrect.

The conception of good law employed in this study could also be questioned by raising similar concerns validity concerns. The desire of judges to make good law was equated with seeing their preferred policy outcomes become law, and preferred policy outcomes were inferred from PAJID scores. It was assumed that liberals would favor the

redistribution inherent in a successful adequacy lawsuit. As discussed in Chapter 2, however, it is possible that judges' conceptions of good law are divorced from ideological concerns, at least at the level of state supreme courts. Judges could be more concerned with using sound legal reasoning, and with properly weighing the evidence and reaching a conclusion that stands up to rigorous logical examination. In that case, a judge's conception of good law and legal reasoning might have more to do with the legal training he or she received and his or her adherence to the legal process theories that the school teaches. One could imagine, for instance, more law and economics reasoning coming from judges educated at Chicago.

It is also possible that although ideology might play a role in some cases, adequacy is not an issue that is likely to break on ideological grounds. Perhaps, for state supreme court judges at least, ideology plays a role only in a very small subset of cases (abortion, for instance) and does not play much of a role in others. Clearly, education finance does not tap the "culture wars" in the same way that abortion does, or that gay marriage does, and maybe only those types of issues activate the ideological preferences of state supreme court judges. Additionally, the role of political, institutional, or ideological factors might be marginal, and secondary to the legal reasoning at play.

Possible Methodological Issues

I attempted to use sound methodology and to take account of all theoretically compelling variables, and I gave a very detailed explanation of QCA in Chapter 3. Nonetheless, it remains possible that the current study suffered from some methodological problems. First and foremost, it is possible that QCA was simply a poor tool for this purpose. Despite its promise, and the ease of interpretation that it offers

insofar as “necessary” and “sufficient” are terms that advocates can use and understand, it remains at least possible that QCA was not the best choice. Of course, as standard statistical models were unsuitable due to the small sample size, that leaves qualitative comparative case-study designs as the other possible alternative. I was attempting, however, to look at *all* adequacy cases, and a comparative case study design not only seemed daunting, but also seemed to have been done already. It is conceivable that in my attempt to break new ground, I allowed the pendulum to swing too far towards breadth at the expense of depth.

And even if QCA was the proper choice, it remains possible that an important independent variable was omitted. I discuss above the possible theoretical problems in light of the findings, but, even beyond that, choices must be made in the definition and measurement of variables. It is possible that, in that process, some important (and possibly necessary and/or sufficient) variable was unaccounted for.

In reviewing Ragin’s work on QCA and the QCA manual, however, it still does seem as if adequacy lawsuits meet any criteria given for areas in which QCA is likely to be useful. Most of his examples are of different countries, but there is no a priori reason to expect that states cannot be compared using comparative methods commonly applied to countries. The findings of the current study raise the question, then, of when exactly QCA might be useful. One obvious answer is that QCA is useful when there is a strong reason to believe that a cause is nearly always necessary or sufficient. However, in such a case, the necessity or sufficiency of the cause is likely not to be in dispute, perhaps because it has been demonstrated through case study research, and this leads one to wonder about the usefulness of QCA as a method for generating new knowledge as

opposed to confirming old knowledge. Perhaps, then, QCA is most useful as a systematic way of combining findings from a large number of case studies that look at similar variables but different cases. In other words, QCA might be best used as a kind of meta-analysis for qualitative research, to demonstrate in a rigorous way that, once case study data is aggregated, certain factors might be necessary or sufficient.

There is, though, the more fundamental issue that certain variables simply cannot be used in a QCA analysis. As discussed above, temporal relationships are excluded. Additionally, variables need to be coded into binary. While useful, either as a simplifying assumption or in cases where the variables really *are* binary (workers strike or they don't, for example), it does limit the types of variables that can be chosen, and therefore structures the research question in very artificial ways. Reliance upon values being either greater or less than the median (or the top quarter or any other such cutoff) means that all measurements are relative, and that very few absolutes are being measured. In retrospect, it is not ideal that a court could be liberal one year and conservative the next if its composition hasn't changed, but the composition of other courts has changed.

IMPLICATIONS FOR THE STUDY OF JUDICIAL POLITICS

There are two main lessons to be drawn from this study in regards to the literature on judicial politics and judicial decision-making. First, judges are people, just like other politicians, and their decision-making is influenced by a wide variety of factors, each of varying importance depending on the particular circumstances of a given situation. Political scientists who write journal articles focusing on politics to the exclusion of the actual written law are likely to overlook some important things, as are academically-oriented lawyers, who try to explain legal decisions entirely in terms of legal reasoning

and precedent. A useful example might be Allison and Zelikow's (1999) examination of presidential decision-making in the Cuban Missile Crisis. They presented three different models, each of which lent something valuable to explaining the event as a whole, and they spent more time exploring the models and their implications than trying to find the one correct model. In terms of adequacy lawsuits, a political perspective and a legal perspective might each offer a useful analytical lens in concert with each other, even if we cannot determine a predictive model.

Second, the current study has illuminated that the context of the specific court matters a great deal. Epstein and Knight, for instance, worked exclusively on the US Supreme Court, and it now seems that their findings, and those of other Supreme Court scholars, reflect less a theory of judicial behavior and decision-making than a theory of judicial behavior *of US Supreme Court justices*. There is absolutely nothing wrong with such a theory, and, indeed, the US Supreme Court is so important that it does deserve study in its own right. However, theories developed at the Supreme Court level might or might not apply to state supreme courts, or to courts in general.

GUIDANCE FOR ADEQUACY ADVOCATES

I would be remiss if I did not, in concluding this study, address the purpose which motivated the entire project. Specifically, this study set out to find the necessary and/or sufficient causes of plaintiffs' victory in order to better guide advocates.

Given the results of the study, the only advice I can offer to advocates is to consider carefully the context of the specific state in which you're operating. All of the possibilities discussed above seem to point towards state-specific and idiosyncratic factors playing a large role. For example, either judges have different conceptions of

good law or ideas about individual versus collective prestige, or the cases are shaped by the state-specific forum-shopping that precedes their appearance in court, or any number of other possibilities. The common factor is the importance of unique state-specific context and characteristics. Most adequacy advocates are likely to have a keen awareness of the specific political environment in which they operate, and are well-advised to use that knowledge to guide their actions, rather than theoretical arguments based upon data from other states.

Baker and Welner (2011), in the course of reviewing how courts are using education research in school finance litigation, discuss how lower-quality ideologically-driven research has found its way into court decisions, specifically Justice Alito's majority decision in *Horne v. Flores*. At the state supreme court level, advocates are likely to know, based on prior rulings, how many judges are likely to be ideologically driven. That kind of detailed knowledge about the individual court and judge in question, it now seems clear, will be more valuable for advocates in constructing winning cases than a less detailed analysis of many states.

ADDITIONAL SPECULATION ON ALTERNATIVE THEORIES AND EXPLANATIONS

Clearly, the current study fell short of its original goals. However, while conducting it, several alternative insights and avenues for exploration became apparent. While they are beyond the empirical reach of the current study, it is nonetheless important that they be considered here, so that they might motivate a follow-up study and provide some inspiration for future projects.

Are Conceptions of Adequacy State-Specific?

The legal reasoning behind adequacy lawsuits did not change much, if at all, from state to state. In all cases, plaintiffs were alleging that defendant states had not met their constitutional burden to provide an adequate education. And while researchers have acknowledged that there are different ways of *computing the cost* of adequacy from state to state, there seems to be an underlying assumption that we are talking about the same thing. In other words, the educational policy community seems to have assumed that an adequate education that prepares one for college or employment in, say, Idaho is substantially similar to that same concept in New York. And, when talking in the most general terms (as the court did in *Rose*), that assumption is likely correct. However, there has been between-state variability in standardized test scores and in percent of graduates who attend college for quite some time, as the wide range of SAT and ACT participation rates demonstrate. It is at least conceivable, then, that conceptions of adequacy in each state (or region) are fundamentally different, based upon that state's educational history and the local economy for which the state's schools is preparing workers. In other words, perhaps, taken in context, an adequate education in a given state is related to that state's predominant economic activity. Some detailed exploration of this, either historical or survey-based, might be useful. If my reasoning is correct, adequacy would mean different things in different states, and an adequacy lawsuit might ensue when the gap between citizens' expectations became wide enough, as opposed to when any objective standard was reached.

The Role of the Media in Framing Adequacy Issues

The current study neglected the role of the media as an institution, and its possible effect on how adequacy issues are initially perceived by the public, and then acted upon

in various forums. Assuming that elected politicians, including judges, take the desires of their constituents into account (to greater or lesser degrees, depending upon the frequency and intensity of the elections), the media could have a role in shaping those desires and the pressures under which politicians operate. It might be instructive to further examine media local media framing of adequacy and inequality. Something as simple as how the mainstream media presents a costing out study, as either well-done or outrageous, could be important in shaping not only the actions of litigants and therefore judges, but other actors as well.

Do Some Institutions Benefit From Inadequacy?

Implicit in the current study was the assumption that political actors value good education. There might be disagreements over what “adequate” or “equitable” means, but I never seriously considered the very real possibility that some political actors don’t place a very high priority on good schools, and perhaps even reap some kind of institutional benefit from continuing poor school performance. However, I am reminded of the concept of the “iron triangle,” typically used to explain the long-lasting relationships between bureaucrats, interest groups, and congressional committees, and the consequent difficulties that elected officials have with bureaucratic oversight. Essentially, three different actors derive some private benefit from continuing the status quo, even though the larger public would probably benefit from reform. It is at least conceivably, if disheartening, that a similar scenario could have developed because state-level education bureaucrats, teachers unions, credit-claiming legislators, and, perhaps, private foundations dedicated to school reform. Each party could actually derive some benefit from the status quo, and might stand to lose something of value in an adequacy

lawsuit. Private foundations, for example, dedicated to improving inner-city schools might lose their purpose. Unions would lose the ability to use “under-funding” as an excuse for poor student achievement, and might find attention directed at teachers and the possibility of value-added assessments.

An adequacy lawsuit would disrupt the status quo, and would introduce some uncertainty. Some actors, then, might find it useful to either push harder for adequate funding in other forums, where, perhaps, the results might be different (or at least more certain). Other actors might attempt to shape public opinion or file *amicus* briefs or take a number of other actions. Essentially, the point here is that not all actors are necessarily in favor of a court ruling for the plaintiffs, even if they nominally support improved education. It could perhaps be useful to look at the role that unions and interest groups play in school financing and in adequacy lawsuits, and to engage in a more detailed case-study type comparison.

Do Lawyers Play a More Important Role Than Political Scientists Give Them Credit For?

A focus on the politics and the institutional environment surrounding the filing and decision of a lawsuit essentially treats the lawyers themselves as interchangeable widgets. The implicit assumption, not only in the current study but in most other political science studies of this type, is that the case was properly argued, evidence was properly presented, and that all lawyers are essentially created equal. Of course, that’s quite an assumption. In any profession, even one with barriers to entry such as specific schooling and licensure requirements, there are bound to be those who are merely competent and those who are excellent. It is certainly possible, in light of the lack of findings here, that

the current study paid too little attention to what actually happened in courtroom. Indeed, much like educational research that neglects actual classroom practice, legal research that neglects the courtroom is leaving an important variable unexamined. Michael Paris (2009, p. 3) emphasized the importance of “legal translation,” the manner in which advocates use legal authority and select and represent the facts and evidence of a case. If cases are won and lost in that translation period, then the current study has missed something critically important.

Given the constraints of QCA, however, it would have been difficult to code courtroom performance or legal maneuvers in anything but the most superficial way. However, subsequent research could focus on the specifics of a case, perhaps comparing that across states. For instance, the costing-out methodology that the plaintiffs employ, and how they explain it, could go a long way towards explaining judges’ decisions. Trying to score or code these arguments would be difficult, and a researcher with legal training would probably be necessary, but it would still be a worthwhile endeavor.

DIRECTIONS FOR FUTURE RESEARCH

Given the discussion above, it seems clear that future research attempting to examine the causes of plaintiffs’ victory in adequacy lawsuits ought to use a more qualitative case-study approach. Jared Diamond and James A. Robinson (2010), in the introduction to their edited volume *Natural Experiments of History*, lay out a method for leveraging the depth of case-study research while incorporating the systematic comparisons that characterize most other research designs. Their methods, which would treat the adequacy decision as a kind of natural experiment and compare states that were

similar or different in some important way before or after the decision, might provide a useful starting point for future research.

It also seems clear that, as many states have had some kind of adequacy decision already, future research should examine the way that courts respond to follow-up lawsuits about compliance or lack thereof. As the number of states that have experienced a first adequacy decision decreases, the importance of these compliance lawsuits increases. Future research should also focus on investigating empirically how adequacy lawsuits have fit into each state's broader efforts at education reform. For example, it would be interesting to ask such questions as whether states in which an adequacy lawsuit was successful were any more or less likely to adopt rigorous standards, or to consider proposals for teacher merit pay, or other such reforms.

In any event, it is evident that more research on adequacy lawsuits is necessary to provide a comprehensive picture of both their causes and their effects.

REFERENCES

- Allison, G., & Zelikow, P. (1999). *Essence of decision: explaining the Cuban Missile Crisis*. New York, NY: Longman.
- Amenta, E., & Poulsen, J.D. (1996). Social politics in context: the institutional politics theory and social spending at the end of the new deal. *Social Forces*, 75(1), 33-60.
- Baker, B.D., & Green, P.C. (2008). Politics, empirical evidence, and policy design: the case of school finance and the costs of educational adequacy. In B. Cooper, J. Cibulka, & L. Fusarelli (Eds.), *Handbook of educational politics and policy* (pp. 311-37). New York: Routledge.
- Baker, B.D., & Wellner, K.G. (2011). School finance and courts: does reform matter, and how can we tell?. *Teachers College Record*, 113(11), Retrieved from <http://www.tcrecord.org> doi: 16106
- Baum, H.S. (2010). *Brown in Baltimore*. Ithaca, NY: Cornell University Press.
- Berkson, L.C. (1980). Judicial selection in the United States: a special report. *Judicature*, 64(4), 176-93.
- Bird, R.C and Smythe, D.J. (2006). The Role of Legal Precedents in the Diffusion of New Employment Doctrines, 1978-99. Available at SSRN: <http://ssrn.com/abstract=90924>
- Biggart, R. (1997). Why labor wins, why labor loses: a test of two theories. *The Sociological Quarterly*, 38(1), 205-24.
- Brace, P., Langer, L., & Hall, M.G. (2000). Measuring the preferences of state supreme court judges. *Journal of Politics*, 62(2), 387-413.
- Brady, H.E., & Collier, D. (Eds.). (2004). *Rethinking social inquiry*. New York, NY: Rowman and Littlefield.
- Braumoeller, B.F., & Goertz, G (2003). The statistical methodology of necessary conditions. In G. Goertz and H. Starr (Eds.), *Necessary Conditions: theory, methodology, and applications* (pp. 197-224). New York, NY: Rowman and Littlefield.
- Braumoeller, B.F., & Goertz, G. (2002). Watching your posterior: comment on Seawright. *Political Analysis*, 10(2), 198-203.
- Braumoeller, B.F., & Goertz, G. (2000). The methodology of necessary conditions. *American Journal of Political Science*, 44(4), 844-58.

- Brisbin, Jr., R.A. (1996). Slaying the dragon: Segal, Spaeth and the function of law in supreme court decision making. *American Journal of Political Science*, 40(4), 1004-17.
- Caldeira, G.A. (1983). On the reputation of state supreme courts. *Political Behavior*, 5(1), 83-108.
- Cann, D. (2007). Beyond accountability and independence: judicial selection and state court performance. *Judicature*, 90(5), 226-32.
- Canon, B.C., & Baum, L. (1981). Patterns of adoption of tort law innovations: an application of diffusion theory to judicial doctrines. *American Political Science Review*, 75, 975-87.
- Chia, J., & Seo, S.A. (2007). Battle of the branches: the separation of powers doctrine in state education funding suits. *Columbia Journal of Law and Social Problems*, 41, 125-46.
- Choi, A., Gulati, M., & Posner, E.A. (2008). *Which states have the best and worst high courts? (Paper # 405)*. Chicago, IL: John M. Olin Program In Law and Economics Working Paper Series.
- Cohen, D.K., & Moffitt, S.L. (2009). *The ordeal of equality: Did federal regulation fix the schools?*. Cambridge, MA: Harvard University Press.
- Cohn, J. (1999, October 25). Irrational exuberance. *The New Republic*, 25-31.
- Collier, D., Brady, H.E., & Seawright, J. (2004). Refocusing the discussion of methodology. In H.E. Brady and D. Collier (Eds.), *Rethinking Social Inquiry* (pp. 3-20). Lanham, MD: Rowman and Littlefield.
- Cress, D.M., & Snow, D.A. (2000). The outcomes of homeless mobilization: the influence of organization, disruption, political mediation, and framing. *The American Journal of Sociology*, 105(4), 1063-1104.
- Diamond, J., & Robinson, J.A. (2010). Prologue: natural experiments of history. In J. Diamond and J.A. Robinson (Eds.), *Natural Experiments of History* (pp. 1-14). Cambridge, MA: Harvard University Press.
- Dear, J., & Jessen, E.W. (2008). "Followed rates" and leading state cases, 1940-2005. *U.C. Davis Law Review*, 41, 683-711.
- Dion, D. (2003). Evidence and inference in the comparative case study. In G. Goertz and H. Starr (Eds.), *Necessary Conditions: theory, methodology, and applications* (pp. 95-112). New York, NY: Rowman and Littlefield.

- Dixon, M., Roscigno, V.J., & Hodson, R. (2004). Unions, solidarity, and striking. *Social Forces*, 83(1), 3-33.
- Duncombe, W., and Lukemeyer, A (2002). "Estimating the Cost of Educational Adequacy: A Comparison of Approaches." Paper Presented at the 2002 AEFA Annual Meeting.
- Dunn, J. & Derthick, M. (2007). Adequacy litigation and the separation of powers. In M. West & P. Peterson (Eds.), *School Money Trials* (pp. 195-212). Washington, DC: Brookings Institution Press.
- Eastman, J.C. (2007). Reinterpreting the education clauses in state constitutions. In M. West & P. Peterson (Eds.), *School Money Trials* (pp. 55-76). Washington, DC: Brookings Institution Press.
- Edwards, H.T. (1991). The judicial function and the elusive goal of principled decisionmaking. *Wisconsin Law Review*, 837-49.
- Elazar, D. (1972). *American federalism: a view from the states*. New York, NY: Thomas Y. Crowell.
- Emmert, C.F. (1992). An integrated case-related model of judicial decision making: explaining state supreme court decisions in judicial review cases. *Journal of Politics*, 54, 543-52.
- Epstein, L., & Knight, J. (1998). *The choices justices make*. Washington, DC: CQ Press.
- Evers, W.M., & Clopton, P. (2006). High spending, low performing school districts. In E. Hanushek (Ed.), *Courting Failure* (pp. 235-55). Stanford, CA: Education Next Books.
- Feldman, SM. (2005). The rule of law or the rule of politics? Harmonizing the internal and external views of supreme court decision making. *Law and Social Inquiry*, 30(1), 89-135.
- Galanter, M. (1974). Why the 'haves' come out ahead: speculations on the limits of legal change. *Law and Society Review*, 9, 95-160.
- Gardner, J.A. (2005). *Interpreting state constitutions: a jurisprudence of function in a federal system*. Chicago, IL: University of Chicago Press.
- Ginsberg, B. & M. Shefter (1990). *Politics by other means*. New York, NY: Basic Books.
- Glick, H.R., & Emmert, C.F. (1987). Selection systems and judicial characteristics: the recruitment of state supreme court judges. *Judicature*, 70(4), 228-35.

- Goertz, G. (2003). Cause, correlation, and necessary conditions. In G. Goertz and H. Starr (Eds.), *Necessary Conditions: theory, methodology, and applications* (pp. 47-64). New York, NY: Rowman and Littlefield.
- Goertz, G., & Starr, H. (2003). Introduction: necessary condition logics, research design, and theory. In G. Goertz and H. Starr (Eds.), *Necessary Conditions: theory, methodology, and applications* (pp. 1-24). New York, NY: Rowman and Littlefield.
- Guthrie, J.W., & Rothstein, R. (1999). Enabling adequacy to achieve reality: translating adequacy into state school finance arrangements. In H.F. Ladd, R. Chalk, & J.S. Hansen (Eds.), *Equity and Adequacy In Education Finance* (pp. 209-59). Washington, DC: National Academy Press.
- Hall, M.G. (2001). State supreme courts in American democracy: probing the myths of judicial reform. *American Political Science Review*, 95(2), 315-30.
- Hanushek, E.A. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18(4), 45-62.
- Hanushek, E.A. (2003). The failure of input-based schooling policies. *The Economic Journal*, 113, F64-F98.
- Hanushek, E.A. (Ed.). (2006). *Courting failure: How school finance lawsuits exploit judges' good intentions and harm our children*. Stanford, CA: Education Next Books.
- Harkreader, S., & Imershein, A.W. (1999). The conditions for state action in Florida's health-care market. *Journal of Health and Social Behavior*, 40(2), 159-74.
- Harris, P. (1985). Ecology and culture in the communication of precedent among state supreme courts, 1870-1970. *Law and Society Review*, 19(3), 449-86.
- Hedges, L., Laine, R., & Greenwald, R. (1994). An exchange: part I: does money matter? a meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(3), 5-14.
- Hedges, L., Laine, R., & Greenwald, R. (1994a). Money does matter somewhere: A reply to Hanushek. *Educational Researcher*, 23(4), 9-10.
- Heise, M. (2005). Educational adequacy as legal theory: implications from equal educational opportunity doctrine. *Proceedings of the Conference on "adequacy lawsuits: their growing impact on American education"* Cambridge, MA: <http://www.hks.harvard.edu/pepg/research.htm>

- Heise, M. (2002). Equal educational opportunity by the numbers: the warren court's empirical legacy. *Washington and Lee Law Review*, 59, 1309-1340.
- Hunter, M.A. (1999). All eyes forward: public engagement and educational reform in Kentucky. *Journal of Law and Education*, 28(4), 485-516.
- King, G., Keohane, R., & Verba, S. (1994). *Designing social inquiry*. Princeton, NJ: Princeton University Press.
- Kingdon, J. (1995). *Agendas, alternatives, and public policies*. New York, NY: Longman.
- Komesar, N.K. (1994). *Imperfect alternatives: choosing institutions in law economics and public policy*. Chicago, IL: University of Chicago Press.
- Koski, W.S., & Levin, H.M. (2000). Twenty-five years after Rodriguez: what have we learned?. *Teachers College Record*, 102(3), 480-513.
- Kovera, M.B., & McAuliff, B.D. (2000). The effects of peer review and evidence quality on judge evaluations of psychological science: are judges effective gatekeepers?. *Journal of Applied Psychology*, 85(4), 574-86.
- Kozinski, A. (1993). What I ate for breakfast and other mysteries of judicial decision making. *Loyola of Los Angeles Law Review*, 26, 993.
- Kritzer, H.M., & Beckstrom, D.C. (2007). Daubert in the states: diffusion of a new approach to expert evidence in the courts. *Journal of Empirical Legal Studies*, 4(4), 983-1006.
- Lindseth, A. (2006). The legal backdrop to adequacy. In E. Hanushek (Ed.), *Courting Failure* (pp. 33-78). Stanford, CA: Education Next Books.
- Lundberg, P.J. (2000). State courts and school funding: a fifty-state analysis. *Albany Law Review*, 63, 1101-46.
- Mayhew, D. (2005). *Divided we govern: party control, lawmaking, and investigations, 1846-2002 (2nd ed.)*. New Haven, CT: Yale University Press.
- Mayhew, D. (1974). *Congress: the electoral connection*. New Haven, CT: Yale University Press.
- McCammon, H.J., Chaudhuri, S., Hewitt, L., Muse, C.S., Newman, H.D., Smith, C.L. & Terrell, T.M. (2008). Becoming full citizens: the u.s. women's jury rights campaigns, the pace of reform, and strategic adaptation. *The American Journal of Sociology*, 113(4), 1104-47.

- McClurg, S.D., & Comparato, S.M. (2004). Explaining state supreme court compliance with the united states supreme court: causal heterogeneity among legal and political factors. *Proceedings of the Annual Meeting of the American Political Science Association*, http://www.allacademic.com/meta/p_mla_apa_research_citation/0/6/0/4/5/pages60455/p60455-1.php
- McCubbins, M. D., and Schwartz, T. (1984). 'Congressional oversight overlooked: police patrols versus fire alarms. *American Journal of Political Science*, 28, 165–79.
- McDonnell, L. (2005). No child left behind and the federal role in education: evolution or revolution?. *Peabody Journal of Education*, 80(2), 19-38.
- Monroe, B.L, & Gold, S.M. (2004). A close look at the qualitative comparative analysis (qca) family of methodologies. *Proceedings of the Annual Meeting of the Midwest Political Science Association*, http://www.allacademic.com/meta/p_mla_apa_research_citation/0/8/3/0/9/p83098_index.html
- Murphy, W. (1964). *The elements of judicial strategy*. Chicago, IL: University of Chicago Press.
- Murray, S.E., Evans, W.N., & Schwab, R.M. (1998). Education-finance reform and the distribution of education resources. *The American Economic Review*, 88(4), 789-812.
- Musheno, M.C., Gregware, P.R., & Drass, K.A. (1991). Court management of AIDS disputes: a sociolegal analysis. *Law & Social Inquiry*, 16(4), 737-74.
- Odden, A.R., Goetz, M.F., and Picus, L.O. (2007). *Paying for School Finance Adequacy with the National Average Expenditure Per-Pupil*. Prepared for the National School Finance Restructuring Project. Seattle, WA: University of Washington, Center for Research on Public Education. Working Paper #2.
- Paris, M. (2010). *Framing equal opportunity: law and the politics of school finance reform*. Stanford, CA: Stanford Univeristy Press.
- Pierson, P. (2004). *Politics in time*. Princeton, NJ: Princeton University Press.
- Pinello, D.R. (1995). *The impact of judicial selection method on state supreme court policy*. Westport, CT: Greenwood Press.
- Pullin, D. (2007). Ensuring an adequate education: opportunity to learn, law, and social science. *Boston College Third World Law Journal*, 27, 83-130.

- Ragin, C.C. (2008). *Redisigning social inquiry: fuzzy sets and beyond*. Chicago, IL: University of Chicago Press.
- Ragin, C.C. (2006). Set relations in social research: evaluating their consistency and coverage. *Political Analysis*, 14, 291-310.
- Ragin, C.C., Drass, K.A., & Davey, S. (2006). *Fuzzy-Set/Qualitative Comparative Analysis 2.0*. Tucson, Arizona: Department of Sociology, University of Arizona.
- Ragin, C.C. (2006). *User's guide to fuzzy-set/qualitative comparative analysis 2.0*. Tucson, Arizona: Department of Sociology, University of Arizona.
- Ragin, C.C. (2000). *Fuzzy set social science*. Chicago, IL: University of Chicago Press.
- Ragin, C.C. (1987). *The comparative method: moving beyond qualitative and quantitative strategies*. Berkeley, CA: University of California Press.
- Rebell, M.A., & Block, A.R. (1982). *Educational policy making and the courts: an empirical study of judicial activism*. Chicago, IL: University of Chicago Press.
- Redding, K., & Viterna, J.S. (1999). Political demands, political opportunities: explaining the differential success of left- libertarian parties. *Social Forces*, 78(2), 491-510.
- Reed, D.S. (2001). *On equal terms: the constitutional politics of educational opportunity*. Princeton, NJ: Princeton University Press.
- Roch, C.H., & Howard, R.M. (2008). State policy innovation in perspective: courts, legislatures, and education finance reform. *Political Research Quarterly*, 61, 333-44.
- Rose v. Council for Better Education*. 1989. 790 S.W. 2d 186 (KY Supreme Court).
- Rosenberg, G. (1991). *The hollow hope*. Chicago, IL: University of Chicago Press.
- Rothstein, R. (2004, December 2). Must schools fail?. *The New York Review of Books*, 51(19), Retrieved from <http://www.nybooks.com/articles/archives/2004/dec/02/must-schools-fail/>
- Roza, M., & Hill, P.T. (2006). How can anyone say what's adequate if nobody knows how money is spent now?. In E. Hanushek (Ed.), *Courting Failure* (pp. 235-55). Stanford, CA: Education Next Books.
- Rudalevige, A. (2005). Adequacy, accountability, and the impact of "No Child Left Behind". *Proceedings of the Conference on "adequacy lawsuits: their growing impact on American education"* Cambridge, MA:

<http://www.hks.harvard.edu/pepg//PDF/events/Adequacy/PEPG-05-27rudalevige.pdf>

- Rueschemeyer, D. (2003). Can one or a few cases yield theoretical gains. In J. Mahoney and D. Rueschemeyer (Eds.), *Comparative Historical Analysis in the Social Sciences* (pp. 305-36). New York, NY: Cambridge University Press.
- Sabatier, P.A. (Ed). (2006). *Theories of the policy process* (2nd Edition). Cambridge, MA: Westview Press.
- San Antonio Independent School District v. Rodriguez, 411 U.S. 1 (1973).
- Serrano v. Priest, 96 California Reporter 601 (1971).
- Schotland, R.A. (2007). New challenges to states' judicial selection. *Georgetown Law Journal*, 95, 1077-1105.
- Seawright, J. (2002). Testing for necessary and/or sufficient causation: which cases are relevant. *Political Analysis*, 10, 178-93.
- Shepsle, K, & Bonchek, M.S. (1997). *Analyzing politics: rationality behavior and institutions*. New York: WW Norton.
- Simon-Kerr, J.A and Sturm, R.K. (2009). Justiciability and the role of courts in adequacy litigation: preserving the constitutional right to education. *Stanford Journal of Civil Rights and Civil Liberties*. Available at SSRN: <http://ssrn.com/abstract=1312426>
- Springer, M.G., Liu, K., & Guthrie, J.W. (2009). The impact of school finance litigation on resource distribution: a comparison of court-mandated equity and adequacy reforms. *Education Economics*, 17(4), 421-44.
- Stern, S. (2006). Campaign for Fiscal Equity v. New York: the march of folly. In E. Hanushek (Ed.), *Courting Failure* (pp. 1-31). Stanford, CA: Education Next Books.
- Swenson, K. (1999). School finance reform litigation: why are some state supreme courts activist and others restrained. *Albany Law Review*, 63, 1147-82.
- Swinford, B. (1991). A predictive model of decision making in state supreme courts: the school financing cases. *American Politics Quarterly*, 19(3), 336-52.
- Tarr, G.A. (1998). *Understanding state constitutions*. Princeton, NJ: Princeton University Press.

- Tarr, G.A., & Porter, M.C. (1988). *State supreme courts in state and nation*. New Haven, CT: Yale University Press.
- Thomas, J., & Brady, K. (2005). The Elementary and Secondary Education Act at 40: Equity, accountability, and the evolving federal role in education. *Review of Research in Education*, 29, 51-68.
- Thro, W.E. (1989). To render them safe: the analysis of state constitutional provisions in public school finance reform litigation. *Virginia Law Review*, 75(8), 1639-79.
- Tropp, L.R., Smith, A.E., & Crosby, F.J. (2007). The use of research in the Seattle and Jefferson County desegregation cases: connecting social science and the law. *Analyses of Social Issues and Public Policy*, 7, 93-120.
- Walberg, H.J. (2006). High poverty, high performance schools, districts, and states. In E. Hanushek (Ed.), *Courting Failure* (pp. 235-55). Stanford, CA: Education Next Books.
- Walker, J.L. (1969). The diffusion of innovations among the American states. *American Political Science Review*, 63(3), 880-99.
- West, M.R., & Peterson, P.E. (Ed.). (2007). *School money trials*. Washington, DC: Brookings Institution Press.
- Williams, J. (2007). The non-implementation of New York's adequacy judgment. In M. West & P. Peterson (Eds.), *School Money Trials* (pp. 195-212). Washington, DC: Brookings Institution Press.
- Wood, B.D., & Theobald, N.A. (2003). Political responsiveness and equity in public education finance. *Journal of Politics*, 65(3), 718-38.

APPENDIX 1

Listing of Adequacy Cases

STATE	Case
ALABAMA	<i>Opinion of the Justices</i>
ALASKA	Kasayulie v. Alaska
ARIZONA	Roosevelt Elementary School District No. 66 v. Bishop
ARKANSAS	Lake View School District, No. 25 v. Huckabee
CALIFORNIA	Williams v. California
COLORADO	Haley v. Colorado Dept of Education
CONNECTICUT	Sheff V. O'Neil
DELAWARE	NO ADEQUACY LAWSUIT
FLORIDA	Coalition for Adequacy and Fairness in School Funding v. Chiles
GEORGIA	Consortium for Adequate School Funding In Georgia v. State of Georgia
HAWAII	NO ADEQUACY LAWSUIT
IDAHO	Idaho Schools for Equal Educational Opportunity v. Evans
ILLINOIS	Committee for Educational Rights v. Edgar
INDIANA	NO ADEQUACY LAWSUIT
IOWA	Coalition for a Common Cents Solution v. Iowa
KANSAS	Montoy v. Kansas
KENTUCKY	Rose v. Council for Better Schools
LOUISIANA	ADEQUACY CASE PENDING
MAINE	NO ADEQUACY LAWSUIT
MARYLAND	Bradford v. Maryland State Board of Education
MASSACHUSETTS	<i>McDuffy v. Secretary</i>
MICHIGAN	Durant v. Michigan
MINNESOTA	NO ADEQUACY DECISION - LAWSUIT SETTLED
MISSOURI	Committee for Educational Equality (CEE) v. Missouri
MONTANA	Columbia Falls Public Schools v. Montana
NEBRASKA	Gould v. Orr
NEVADA	NO ADEQUACY LAWSUIT
NEW HAMPSHIRE	Claremont School District v. Governor
NEW JERSEY	Abbott v. Burke
NEW MEXICO	Zuni School District v. State (NOTE: Cap. Funding only)
NEW YORK	CFE v. New York
NORTH CAROLINA	Leandro v. North Carolina
NORTH DAKOTA	ADEQUACY CASE PENDING
OHIO	DeRolph v. State
OKLAHOMA	Oklahoma Education Association v. State (after 2005, not included in QCA)
OREGON	Coalition For Equitable School Funding v. Oregon
PENNSYLVANIA	Marrero v. Commonwealth
RHODE ISLAND	City of Pawtucket v. Sundlun
SOUTH CAROLINA	Abbeville County School District v. State
SOUTH DAKOTA	NO ADEQUACY DECISION
TENNESSEE	Tennessee Small School Systems v. McWherter
TEXAS	West Orange-Cove Consolidated ISD v. Nelson
UTAH	NO ADEQUACY LAWSUIT
VERMONT	Brigham v. Vermont
VIRGINIA	NO ADEQUACY LAWSUIT

WASHINGTON	NO ADEQUACY LAWSUIT
WEST VIRGINIA	Tomblin v. Gainer
WISCONSIN	NO ADEQUACY LAWSUIT
WYOMING	Campbell County School District v. State

APPENDIX 2

QCA TRUTH TABLES

TRUTH TABLE FOR MODEL 1

Judicial Elections	Diffusion Through Circuit	Diffusion Through Constitutional Clause	Political Culture	Divided Government	High Performing Court	Liberal Court	Count	raw consist.	PRI consist.	product	
0	1	1	1	1	1	0	1	3	0.666667	0.666667	0.444444
0	0	0	0	1	0	0	0	2	0.5	0.5	0.25
0	1	1	1	1	0	0	1	2	1	1	1
0	1	1	1	1	1	0	0	2	1	1	1
1	0	0	1	1	1	0	1	2	0.5	0.5	0.25
0	0	0	0	0	0	0	1	1	1	1	1
0	0	0	0	1	1	1	1	1	1	1	1
0	0	1	0	0	0	0	1	1	1	1	1
0	0	1	0	1	1	1	1	1	1	1	1
0	0	1	1	1	0	0	0	1	0	0	0
0	1	1	1	0	0	0	0	1	1	1	1
0	1	1	0	1	1	1	1	1	0	0	0
1	0	0	0	0	0	0	1	1	1	1	1
1	0	0	0	1	0	1	1	1	0	0	0
1	0	1	0	1	0	0	0	1	1	1	1
1	0	1	0	1	0	0	1	1	0	0	0
1	0	1	1	1	0	0	0	1	1	1	1
1	0	1	1	1	1	0	1	1	1	1	1
1	1	0	1	1	1	0	1	1	1	1	1
1	1	1	1	1	0	0	1	1	0	0	0
1	1	1	1	1	0	1	0	1	1	1	1
1	1	1	1	1	1	0	1	1	1	1	1
1	1	1	1	1	1	1	0	1	1	1	1

TRUTH TABLE FOR MODEL 1A

Judicial Elections	Divided Government	High Performing Court	Liberal Court	Count	Plaintiffs Win?	raw consist.	PRI consist.	product
1	1	0	0	1	1	1	1	1
1	0	1	0	1	1	1	1	1
0	0	0	1	4	1	1	1	1
1	1	1	0	2	1	1	1	1
0	1	0	0	2	1	1	1	1
1	0	0	0	1	1	1	1	1
1	1	0	1	6	0	0.666667	0.666667	0.444444
0	1	0	1	3	0	0.666667	0.666667	0.444444
0	1	1	1	3	0	0.666667	0.666667	0.444444
0	0	0	0	4	0	0.5	0.5	0.25
1	0	0	1	2	0	0.5	0.5	0.25
1	0	1	1	1	0	0	0	0

APPENDIX 3:

MEDIAN AND STANDARD DEVIATIONS OF DISTRICT PER-PUPIL SPENDING

BY STATE

State	Year	Plaintiffs_Win	Median	Std Dev
Kentucky	1989	1		
New Jersey	1990	1		
Oregon	1991	0		
Alabama	1993	1		
Idaho	1993	1		
Mass	1993	1		
Nebraska	1993	0		
New Hamp	1993	1		
Arizona	1994	1		
Missouri	1994	0		
Rhode Island	1995	0	6834.5	1415.63
Wyoming	1995	1	6317	2837.54
New York	1995	1	8197	2937.28
Florida	1996	0	5008	402.95
Conn	1996	1	7919.5	1347.02
Illinois	1996	0	4612	1556.59
North				
Carolina	1997	1	5010	570.12
Vermont	1997	1	6095	1100.45
Ohio	1997	1	4879	1140.89
Michigan	1997	1	5718	1170.42
Penn	1998	0	6226.5	1044
South				
Carolina	1999	1	5701.5	672.64
New Mexico	1999	1	6419	1957.64
Alaska	1999	1	12459	4318
Maryland	2000	1	7048	626.54
Arkansas	2002	1	5806	1102.46
Kansas	2003	1	7927	1373.53
Texas	2003	1	7408	2866.96
Georgia	2005	0	7780.5	1062.35
Montana	2005	1	9008	4978.13
Oaklahoma	2007	0	7799.5	3176.8
Indiana	2009	0	8270	1192.9
Colorado	2009	1	8927	2676.95