



# PARDEE RAND GRADUATE SCHOOL

- CHILD POLICY
- CIVIL JUSTICE
- EDUCATION
- ENERGY AND ENVIRONMENT
- HEALTH AND HEALTH CARE
- INTERNATIONAL AFFAIRS
- NATIONAL SECURITY
- POPULATION AND AGING
- PUBLIC SAFETY
- SCIENCE AND TECHNOLOGY
- SUBSTANCE ABUSE
- TERRORISM AND HOMELAND SECURITY
- TRANSPORTATION AND INFRASTRUCTURE

This PDF document was made available from [www.rand.org](http://www.rand.org) as a public service of the RAND Corporation.

[Jump down to document](#) ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

## Support RAND

[Purchase this document](#)

[Browse Books & Publications](#)

[Make a charitable contribution](#)

## For More Information

Visit RAND at [www.rand.org](http://www.rand.org)

Explore [Pardee RAND Graduate School](#)

View [document details](#)

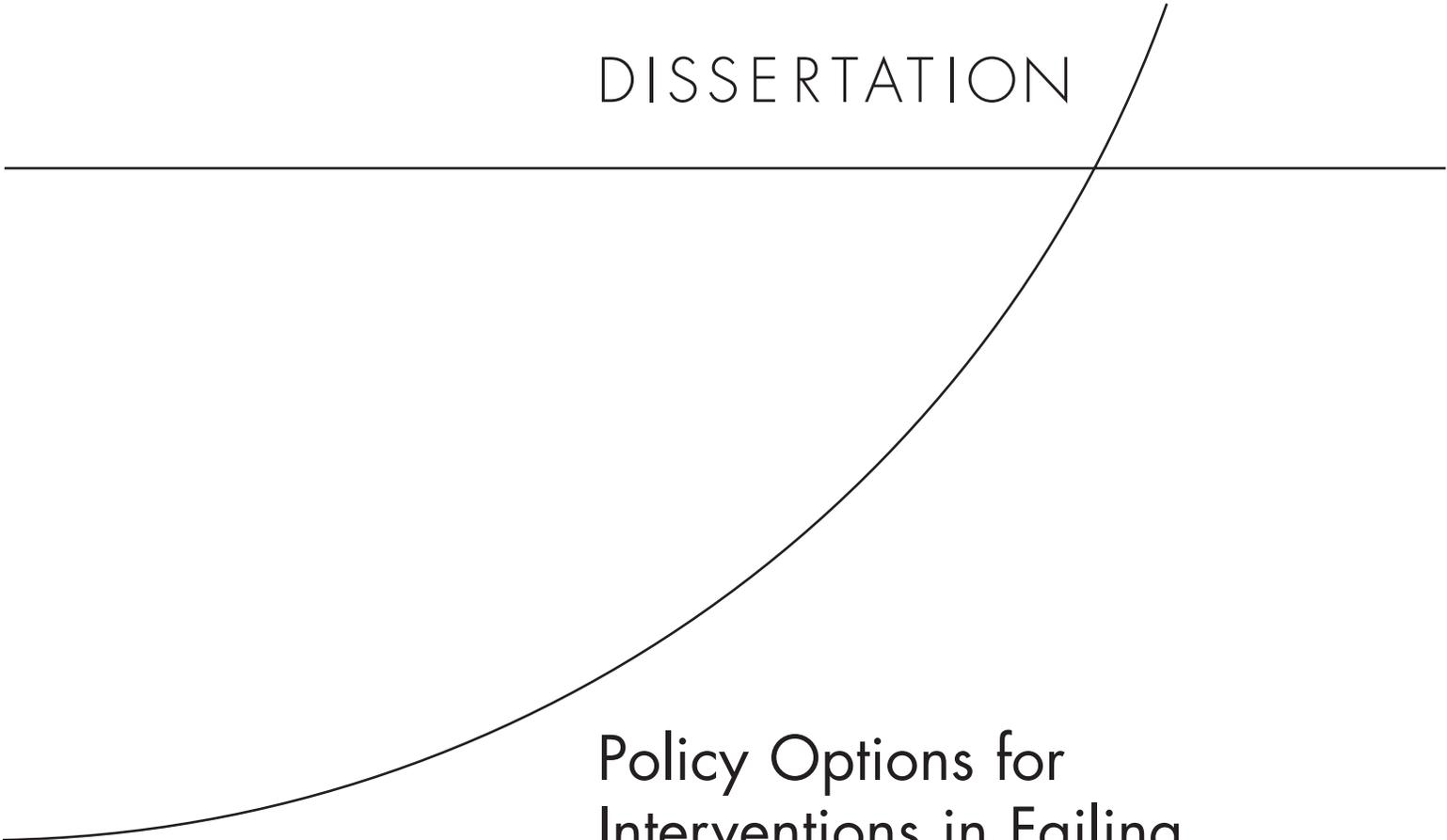
## Limited Electronic Distribution Rights

This document and trademark(s) contained herein are protected by law as indicated in a notice appearing later in this work. This electronic representation of RAND intellectual property is provided for non-commercial use only. Permission is required from RAND to reproduce, or reuse in another form, any of our research documents for commercial use.

This product is part of the Pardee RAND Graduate School (PRGS) dissertation series. PRGS dissertations are produced by graduate fellows of the Pardee RAND Graduate School, the world's leading producer of Ph.D.'s in policy analysis. The dissertation has been supervised, reviewed, and approved by the graduate fellow's faculty committee.

DISSERTATION

---



## Policy Options for Interventions in Failing Schools

Connor P. Spreng

This document was submitted as a dissertation in February 2005 in partial fulfillment of the requirements of the doctoral degree in public policy analysis at the Pardee RAND Graduate School. The faculty committee that supervised and approved the dissertation consisted of Charles A. Goldman (Chair), Catherine A. Augustine, and David Loughran.



PARDEE RAND GRADUATE SCHOOL

The Pardee RAND Graduate School dissertation series reproduces dissertations that have been approved by the student's dissertation committee.

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world. RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

**RAND**® is a registered trademark.

© Copyright 2005 RAND Corporation

All rights reserved. No part of this book may be reproduced in any form by any electronic or mechanical means (including photocopying, recording, or information storage and retrieval) without permission in writing from RAND.

Published 2005 by the RAND Corporation  
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138  
1200 South Hayes Street, Arlington, VA 22202-5050  
201 North Craig Street, Suite 202, Pittsburgh, PA 15213-1516  
RAND URL: <http://www.rand.org/>  
To order RAND documents or to obtain additional information, contact  
Distribution Services: Telephone: (310) 451-7002;  
Fax: (310) 451-6915; Email: [order@rand.org](mailto:order@rand.org)

# Policy Options for Interventions in Failing Schools

## TABLE OF CONTENTS:

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	OBJECTIVE .....	3
1.2	MOTIVATION.....	4
1.3	ORGANIZATION AND METHODS .....	5
<b>2</b>	<b>THEORY AND LITERATURE REVIEW .....</b>	<b>7</b>
2.1	RECENT TRENDS IN EDUCATION REFORM.....	7
2.1.1	<i>Performance of individual schools</i> .....	8
2.1.2	<i>Measures of school performance</i> .....	10
2.1.3	<i>Parental choice</i> .....	15
2.1.4	<i>Criticisms of the dominant trends</i> .....	19
2.2	SCHOOL FAILURE.....	21
2.2.1	<i>Definition of school failure</i> .....	21
2.2.2	<i>School failure and equity</i> .....	25
2.3	ARE INTERVENTIONS IN FAILING SCHOOLS REALLY NECESSARY? .....	28
2.3.1	<i>Improvement of schools</i> .....	29
2.3.2	<i>Exit of schools from the system</i> .....	34
2.4	INTERVENTIONS IN FAILING SCHOOLS .....	38
2.4.1	<i>Definition and classification of interventions</i> .....	38
2.4.2	<i>Prevalence of interventions</i> .....	42
2.5	IMPACT OF INTERVENTIONS .....	44
2.5.1	<i>Moderate and strong interventions</i> .....	44
2.5.2	<i>A shift in strategy for interventions under NCLB</i> .....	52
2.5.3	<i>Charter schools</i> .....	54
2.5.4	<i>Indirect impact of interventions</i> .....	57
<b>3</b>	<b>REFORM AND INTERVENTIONS IN NEW ZEALAND.....</b>	<b>59</b>
3.1	WHY NEW ZEALAND MATTERS .....	59
3.2	INTRODUCTION TO THE REFORMS OF 1989 AND 1991 .....	60
3.2.1	<i>History of education in New Zealand</i> .....	60
3.2.2	<i>Public sector management reform</i> .....	62
3.2.3	<i>Tomorrow's Schools</i> .....	65
3.2.4	<i>Design of the accountability framework</i> .....	67
3.2.5	<i>The reform's conception of failing schools</i> .....	69
3.3	DESIGN MEETS REALITY.....	70
3.3.1	<i>Demise of the charter</i> .....	71
3.3.2	<i>Erroneous assumptions</i> .....	74
3.3.3	<i>Equity issues</i> .....	75
3.3.4	<i>The problem emerges</i> .....	77
3.4	ADJUSTMENTS .....	79
3.4.1	<i>Interventions</i> .....	80

3.4.2	<i>Legal requirements</i> .....	85
3.4.3	<i>The role of the private sector</i> .....	87
3.5	RECENT DEVELOPMENTS IN MONITORING .....	88
3.5.1	<i>Consolidation of schools support and monitoring</i> .....	88
3.5.2	<i>Focus on data</i> .....	88
3.5.3	<i>Conclusions</i> .....	89
<b>4</b>	<b>INTERVENTIONS IN CALIFORNIA PUBLIC SCHOOLS</b> .....	<b>93</b>
4.1	WHY CALIFORNIA MATTERS .....	93
4.2	THE CALIFORNIA INTERVENTIONS .....	94
4.2.1	<i>Introduction to California's accountability framework</i> .....	94
4.2.2	<i>Details of II/USP</i> .....	96
4.2.3	<i>Details of CSR and HP interventions</i> .....	98
4.2.4	<i>II/USP Action Plans</i> .....	99
4.2.5	<i>Funding of interventions</i> .....	102
4.2.6	<i>How interventions were supposed to work</i> .....	103
4.3	DATA .....	106
4.4	II/USP AND COMPARISON SCHOOLS .....	107
4.4.1	<i>Selection of cohort 1</i> .....	108
4.4.2	<i>Characteristics of cohort 1 treatment and non-treatment schools</i> .....	109
4.4.3	<i>Selection of cohort 2 and 3</i> .....	117
4.4.4	<i>Characteristics of cohort 2 and comparison schools</i> .....	118
4.5	METHODS .....	120
4.5.1	<i>Key concepts for program evaluation methods</i> .....	120
4.5.2	<i>Basic models to estimate the interventions' impact</i> .....	121
4.5.3	<i>Feasible scope of the estimations</i> .....	123
4.6	LITERATURE REVIEW .....	126
4.6.1	<i>Betts &amp; Danenberg (2003)</i> .....	126
4.6.2	<i>O'Day and Bitter (2003)</i> .....	128
4.7	RESULTS .....	131
4.7.1	<i>Post-treatment comparison of means</i> .....	132
4.7.2	<i>Estimating the models for cohort 1</i> .....	135
4.7.3	<i>Additional corrections</i> .....	137
4.7.4	<i>Cohort 2</i> .....	139
4.7.5	<i>Conclusion</i> .....	140
4.8	DISCUSSION .....	140
4.8.1	<i>Heterogeneity of impact</i> .....	140
4.8.2	<i>Potential, remaining bias</i> .....	141
4.8.3	<i>Effect on non-treated units</i> .....	141
4.8.4	<i>The nature of school failure in California</i> .....	142
<b>5</b>	<b>CONCLUSIONS AND POLICY RECOMMENDATIONS</b> .....	<b>144</b>
5.1	RESPONSES TO SCHOOL FAILURE .....	144
5.1.1	<i>Common assumptions regarding responses to school failure</i> .....	144
5.1.2	<i>Erroneous assumptions in New Zealand</i> .....	145
5.2	EFFECTS OF INTERVENTIONS ON SCHOOL PERFORMANCE .....	146
5.2.1	<i>Existing studies</i> .....	147

5.2.2	<i>Impact of interventions in California</i> .....	148
5.3	POLICY RECOMMENDATIONS .....	150
5.3.1	<i>Interventions in general</i> .....	151
5.3.2	<i>Interventions in failing schools in California</i> .....	153
<b>6</b>	<b>EDUCATION REFORM IN THE EMIRATE OF QATAR .....</b>	<b>155</b>
6.1	WHY QATAR MATTERS.....	155
6.2	THE CONTEXT TO THE CURRENT EDUCATION REFORM IN QATAR .....	156
6.2.1	<i>Historical background</i> .....	156
6.2.2	<i>Reform efforts throughout the Qatari economy</i> .....	158
6.2.3	<i>K-12 education in Qatar</i> .....	159
6.3	OVERVIEW OF ‘EDUCATION FOR A NEW ERA’ .....	161
6.3.1	<i>New institutions</i> .....	162
6.3.2	<i>A series of reconstitutions</i> .....	163
6.3.3	<i>Remaining uncertainties</i> .....	164
6.4	ACCOUNTABILITY FRAMEWORK .....	165
6.4.1	<i>The contract</i> .....	165
6.4.2	<i>Reporting by the school</i> .....	167
6.4.3	<i>Reporting by the Evaluation Institute</i> .....	168
6.5	WHEN INDEPENDENT SCHOOLS FAIL.....	169
6.5.1	<i>Strategy for analyzing school failure</i> .....	170
6.5.2	<i>Strategy for intervening in failing schools</i> .....	174
6.5.3	<i>What would happen without a strategy?</i> .....	178
	<b>REFERENCES.....</b>	<b>179</b>



# 1 Introduction

Governments around the world are facing the challenge of public education reform. The policy problem is introduced by posing the following question:

What should be done about individual schools that are failing to perform?

The possible responses of the government, as the ultimate sponsor of public education,<sup>1</sup> faced with individual schools in the system of public education that are failing can be organized into three groups:

- I. Tolerate failing schools. School failure may be seen as a regrettable, but largely unavoidable part of the system.
- II. Change the system of public education. The system could be changed to prevent failure from becoming a problem by instituting mechanisms whereby failure is dealt with automatically, either by forcing schools to improve, or by ensuring that schools which are considered to be failing are forced to close.
- III. Devise a strategy of interventions. The government directs specific actions to be undertaken by government agencies or outside contractors, to actively deal with failing schools, either by instituting changes at the school that will lead to improvement, or by actively closing the school down.

The distinction between these three options, and especially between changing the system of public education and devising a strategy of interventions, is useful for conceptual purposes, even if in practice the line is not always clear-cut.<sup>2</sup> As a more detailed discussion in Chapter Two argues, a comprehensive strategy for interventions is needed for an effective system of public education.

Interventions are needed because tolerating failing schools is at odds with the role of the government in providing education to all citizens, and because changing the system of public education alone fails to solve the problems of individual failing schools. In the following section, these points are briefly explained and the terms are defined.

---

<sup>1</sup> The basic role of the government in financing public education is assumed. For a justification of this assumption, see Friedman (1955).

<sup>2</sup> These three options are stated here for the purpose of clarifying terms. The potential for classifying government responses in other ways is acknowledged.

Efforts to change the system of public education are referred to as education reforms. Most education reforms related to school governance address an identified, systemic failure in the delivery of education by granting all schools in the system more autonomy or by inviting providers other than the traditional, government-operated schools into the system. Most often structures are provided that encourage schools to achieve high performance. Such changes may well be similar to interventions in their mechanics (e.g., influencing the school's autonomy) and in their intent to help existing schools provide better education. The proponents of such broad reforms argue for a change in the school *system*, instead of intervening in individual, failing schools.<sup>3</sup> Education reforms aim to prevent the failure of individual schools or to institute mechanisms, to deal with failure automatically.

A strategy of interventions, on the other hand, is designed to address the failure of schools after they have been identified as being in need of change because, either they are doing worse than schools they are compared against, or their performance has fallen below an absolute threshold.<sup>4</sup> Interventions may provide extra funding or technical assistance, impose sanctions, or, in the extreme case, mandate that the school be closed.

This study addresses the policy problem of how interventions in individual schools should be structured, not only for the benefit of the affected school, but for the benefit of the entire system of public education in which they take place. The unit of analysis is therefore primarily the individual school, rather than the school system. And while education happens to a large extent in the classroom, between individual teachers and students, the close examination of these internal processes is not a main focus here. Analyzing the school as an individual unit corresponds to the recent trends in educational accountability that are described below.

What should be done about schools that are unable to succeed in the system of public education in which they're operating? A focus on interventions in individual failing schools makes sense for two main reasons.

---

<sup>3</sup> See, for example, Hanushek et al. (1994) and Ouchi (2003).

<sup>4</sup> A discussion of the definition of school failure follows in Chapter Two.

- a) The notion that education reforms, once instituted, will eliminate school failure is misplaced. The governance structures most frequently advocated for public education are precisely designed to make the decline and failure of individual schools more obvious. The introduction of market forces, and the establishment of autonomy and real accountability for schools, is no panacea against the need to have a policy in place for such instances of failure. Whether the strategy towards failing schools is ‘hands-off’, as it is currently with US charter schools,<sup>5</sup> or more involved, as it is with typical school districts, the government will not be able to completely abdicate the responsibility for intervening in failing public schools if the assumed responsibility for ensuring the provision of public education is not to be abandoned.
- b) A well thought-out strategy for interventions does not stand in conflict with systematic education reforms, but instead supports and lends credibility to them. By clarifying what happens when schools do not succeed, and by acknowledging the fact that failure will occur despite the implementation of reform, the chances for success of the reforms improve.

The need to respond to school failure exists regardless of the system of public education, and regardless of the education reform that has been implemented or planned. This study looks at three different school systems that each, in their own way, struggle with the issues described above.

## **1.1 Objective**

The objective of this research is to evaluate existing responses to school failure and provide recommendations for improved interventions in failing public schools.<sup>6</sup> This study develops answers to the following research questions:

1. What are the different responses to school failure?
2. What are the effects of interventions in failing schools on school performance?
3. What policy recommendations can be made regarding interventions in different policy contexts?

---

<sup>5</sup> The typical response of authorizers to charter school failure has been to close the school (Hassel and Batdorff, 2004). Imposing the closure of a school constitutes an intervention in the school’s operations.

<sup>6</sup> This study concerns itself exclusively with public education at the primary and secondary level (including Kindergarten), called K-12, excluding early childhood education and education after high school.

## **1.2 Motivation**

Despite nationwide expenditures on public elementary and secondary education of \$454 billion in 2001/02, or roughly 4.5% of GDP (NCES, 2004), there is a widespread impression that public schools in the United States are not serving the educational needs of children well. Therefore, the call for policy responses to failure in public education in the United States is not surprising. Nor is it a new concern. In the wake of the Soviet launch of the Sputnik satellite in 1957, American schools were challenged to close the perceived knowledge gap in science and mathematics with the Cold War foe (Rutherford, 1998), and federal legislation was enacted to help schools succeed in this effort (National Defense Education Act, 1958). Twenty-six years later, the formative report *A Nation At Risk* blamed the education system for a lack of competitiveness of US industries and technologies. The report noted “a growing impatience with shoddiness ... [that] is too often reflected in our schools and colleges” (Gardner et al., 1983). Since the 1990’s, the focus has turned to performance based accountability systems and, as a consequence, to individual schools that are failing to deliver the measurable results. School failure has become “increasingly ... an institutional problem that affects particular schools” (Connelly, 1999). The passage of the No Child Left Behind Act of 2001 (NCLB) once again brought the issue of school failure and interventions in response to school failure to the top of the policy agenda.

The concern with school failure is not unique to the United States. Around the globe, governments are struggling to ensure the provision of effective public education. Two related, but distinct, trends are becoming prominent in the proposals and implementation of education reform. First, the focus on the performance of individual schools is growing. Second, the introduction of market forces through parental choice is increasingly seen as a central accountability mechanism to force failing schools to improve or to force them to close (Plank and Sykes, 2003). Relatively little attention has been paid, however, to the assumptions underlying these mechanisms and whether they work equally well for all schools. There is still a lot to be learned about responding to failure in education delivery. In 1994, the Organization for Economic Cooperation and Development (OECD) launched a multi-year, international research effort called *Combating Failure at School*. While calling for a coordinated and transparent “strategy to

combat failure” in schools, the report notes that “very few of the policies to overcome failure have been systematically monitored or fully evaluated” (OECD, 1998).

The increased political will to place responsibility for education at the school level makes the investigation of individual school-level interventions an important and timely enterprise. While the number of studies that examine specific interventions of this sort has been growing in the past few years,<sup>7</sup> researchers still stress that “there is a dearth of evidence” on interventions (Ziebarth, 2002), and that “surprisingly little is known about what kinds of interventions are most likely to turn faltering schools into successful educational institutions” (Brady, 2003).

### **1.3 Organization and Methods**

Chapter Two uses existing studies, economic theory, and analogies with other fields of study to examine (a) the context in which the question regarding failing schools has become urgent, (b) how the concept of school failure itself can be best understood, (c) what the role of government can and should be in overseeing public education, (d) what is currently known about the impact of interventions on the performance of low-performing schools, and (e) how charter schools in the United States provide an interesting, though limited, case of handling school failure in a different manner.

Chapter Three provides a case study of the system of public education in New Zealand, with a focus on the changing approach towards failing schools in the 15 years since the dramatic education reform in 1989. The case study relies primarily on the analysis of official documents from the Ministry of Education and the Education Review Office, as well as published statements from leaders in the Ministry and existing studies by analysts from New Zealand and the United States. Conversations with Martin Connelly, Chief Policy Analyst at the Ministry of Education in Wellington, as well as with four former principals now working at Multi Serve Education Trust, have helped to set the context for and clarify the available information. While no formal interviews were conducted, the information drawn from these conversations, in person or in writing, are

---

<sup>7</sup> See, for example, Berends et al. (2002), Brady (2003), Department of Education (1998), ECS (2002), McCauley and Roddick (2001), Mintrop (2002), Stoll and Myers (1998).

cited in the text. The descriptive analysis offers evidence that interventions in failing schools are needed, even after the implementation of dramatic education reform.

The case study of New Zealand is followed by a quantitative analysis of interventions in Chapter Four, where the impact of interventions in failing schools on school performance will be estimated. To this end, data of interventions and school performance of K-12 schools in California are analyzed. California is a compelling choice for estimating the impact of interventions, because (a) failure in California's public schools is common, (b) existing studies have come up with conflicting estimates of the existing interventions' impact, (c) data are available, and (d) the proximity and direct contact to officials at government agencies at the state and district level provides access for the corroboration and explanation of the publicly available data. Similar to the case study of the reforms in New Zealand, a number of conversations with officials at the California Department of Education and the Los Angeles Unified School District have served to provide the details to publicly available information. Official documents and existing studies have provided the context for the chapter's quantitative analysis. A detailed description of the data itself and the methods used for their analysis is provided in Chapter Four.

In Chapter Five, the findings of the earlier chapters are summarized. Policy recommendations are developed based on the preceding analyses and presented for schools in general and public schools in California in particular.

Chapter Six provides an epilogue to the dissertation by introducing the policy environment in Qatar, the Emirate in the Persian Gulf, and its emerging education reform. The Qatari K-12 reform provides an ideal test-case for the consideration of interventions in failing schools, as the emerging institutions and contractual relationships between the schools and the government have been designed to provide answers to the ambiguities that are seen in other systems of public education. This descriptive analysis is based primarily on published documents available from the reform's website, as well as on studies and news articles published on the region.

## **2 Theory and Literature Review**

This chapter addresses the first research question regarding the different responses to school failure. The context of the emerging focus on the performance of individual schools is discussed first, followed by some conceptual background, including a discussion of the definition of school failure and the mechanisms that force failing schools to improve or decline further. The second part of the chapter presents a categorization of interventions and a review of the existing literature on interventions in regular public schools and in charter schools.

To understand the question about interventions in failing schools that motivates this study, it is useful to consider the context in which it has emerged. It is beyond the scope of this dissertation to give a detailed account of the arguments that drive this debate. Instead, the discussion is limited to representative arguments in favor of reform models that make this investigation particularly timely.

### ***2.1 Recent Trends in Education Reform***

Public schools not only teach literacy, math, arts, sciences, social studies, and other subjects, they also feed and protect the children, promote their health, encourage them to become life-long learners, and help impart the common values that will enable them to become engaged citizens. Public schools aim to promote the integration of different groups in society, including immigrants, and strive to provide the economically disadvantaged with an equal opportunity to succeed. To accomplish this important task, considerable resources are invested in public education (in the United States, nearly 5% of GDP, or more than \$450 billion in 2001/02). For a system of such enormous importance, it is not surprising that the calls for reform of public education keep recurring with varying degrees of intensity and urgency.

In 1647, when public education in the United States was first conceived through a decree of the General Court of the Massachusetts Bay Colony, all schools in towns of 100 or more families were mandated to teach Latin, as well as English. The goal then was to ensure that Puritan children would learn to read the Bible and receive basic information about their Calvinist religion (ARC, 2004). Since these early days of public education, reform efforts have continued to focus on the priorities of their times and have become

increasingly prominent in the public policy debate. After the launch of the Soviet satellite Sputnik in 1957, curricular reforms in math and science were implemented to make American students more competitive against their Soviet counterparts. In the 1960's and 70's, the rise of federal programs, such as Title I,<sup>8</sup> supported the civil rights movement and President Johnson's War on Poverty. Following the publication of the disparaging review, *A Nation At Risk*, in 1983, much of the reforms focused on preparing students to become more productive workers in an economy that was perceived to be insufficiently competitive with the economies of Japan and Germany.

A more recent development in the cycle of education reform plans is the emphasis on the performance of individual schools, and on the differences in performance between schools. The issue of failure among individual schools and the question of an appropriate response on the part of the government is driven by the integration of these concepts into systems of public education. The following discussion is based primarily on the education reform movement in the United States. Later in this section, however, the trends in education reform in other countries are shown to be similar in important aspects that relate to the issue of failing schools.

### *2.1.1 Performance of individual schools*

Systems of public education are under enormous budget pressures and yet must accomplish much. Among many other challenges, public schools need to integrate information technology into the classroom, maintain stability within the school despite increased student mobility, and integrate and accommodate increasing numbers of English language learners. As the particular difficulties vary widely among American public schools, reform efforts have begun to concentrate on the way schools are managed and held accountable.

How can public schools be improved without taking up more resources? In the past two decades, the answer of an increasing number of school reformers has been to clarify the mission of elementary and secondary schools. While the multiple missions of

---

<sup>8</sup> Title I refers to the first section of the Elementary and Secondary Education Act of 1965 (ESEA) which provides for additional funding from federal sources for students from low-income households. Later reauthorizations of ESEA, of which the No Child Left Behind Act of 2001 is the latest, have maintained or expanded the provisions of Title I.

public schools are acknowledged, a consensus is emerging in the United States that more attention needs to be focused on the cognitive learning of students.<sup>9</sup> Education reform along these lines begins with the statement of ambitious academic goals or standards for all children and curriculum frameworks that describe what all children should know and be able to do – that is, with a definition of the desired outcomes. The corresponding piece to defining the desirable outcomes or setting the standards is measuring the degree to which the students are able to meet them. Then, the school systems, schools, the adults in the school system, or the children can be held accountable for the progress towards the stated standards. This focus on educational outcomes is referred to as standards-based education reform.

Once the standards are set and the children’s progress towards them can be measured, the question remains who should be held accountable. As Helen Ladd (1996) states, “the ‘new state accountability’ generally starts from the view that the school is the most appropriate unit of accountability.” The choice of the school as the unit of accountability is not absolute, as adults in the system are held to account for their own performance as well. Teachers, in particular, have been subject to many attempts to attach performance incentives to their paychecks, based on the performance of the children in their classes.<sup>10</sup> Entire school systems, such as school districts, as well as the individual learners themselves, are sometimes held accountable for the measured performance relative to the stated goals. At the center of the emerging framework, however, is the individual school as the most important unit of accountability.

While a discussion of whether the school *should* be the unit of accountability is beyond the scope of this dissertation, it is useful to note that the choice is not inevitable. Indeed, holding the school accountable is somewhat at odds with a system that is based on measuring outcomes at the student level.

The features of what Ladd calls the “new state accountability” have important implications for the notion of school failure. While the interest in holding educators accountable is not new, the “current emphasis on and toward educational accountability

---

<sup>9</sup> For a summary of these trends, see Ladd (1996).

<sup>10</sup> One example is the use of teacher bonuses in California, as discussed by Kane and Staiger (2002b).

represents a departure, or evolution, from previous practice” (O’Day, 2002). The new accountability that has emerged in the past ten to fifteen years in the United States, and increasingly abroad, is marked by a reliance on “student outcomes as the measure of adult and system performance, a focus on the school as a basic unit of accountability, public reporting of student achievement, and the attachment of consequences to performance levels” (O’Day, 2002). While the details of the mechanisms vary across jurisdictions, based on the cited work by Ladd and O’Day, as well as other summaries on the subject (Elmore et al., 1996; Fuhrman, 1999), three key features can be identified that distinguish what is called the new accountability:

- i) Standards and tests. Educational standards are established as statements or goals of what the individual students should learn, alongside tests that measure the students’ progress towards those standards.
- ii) The school as the unit of accountability. A target level of student performance is established for every school. The target is aggregated across the school, though targets may include disaggregated benchmarks as well. The evaluation of the school’s performance is based, at least in part, on the average of the students’ performance. The results are made available to policy makers and often to parents.
- iii) Consequences. Based on its performance relative to the target, each individual school is faced with consequences, which may include awards, sanctions, or assistance.

The ‘new accountability’ refers hereafter to systems of accountability that have these features. As the cited authors have argued, the new accountability has grown out of the move towards standards-based reform in education and has become an integral part of it. Before the integration of market ideas into the delivery of public education is discussed (a related, but distinct, trend), the following section considers the implications of the way school performance is measured under the new accountability.

### ***2.1.2 Measures of school performance***

The reliance on student outcomes as a measure of school performance deserves a closer look. In the United States, the No Child Left Behind Act of 2001 (NCLB) codified

the trend, beginning in the late 80's, of setting statewide performance standards, and using scores from standardized tests to assess the differential performance among schools or groups of schools. NCLB has, in other words, turned the ongoing trend towards the new accountability into a federal law. The overriding importance of the measurement of student outcomes, and the school performance that is inferred from these outcomes, calls for a discussion of the two concepts of measurement: validity and reliability.

Validity is a key concept for any measurement, so, too, for student outcomes. Student outcome measures are said to have validity insofar as they are free from systematic error and are able to truly measure the construct of interest (e.g., student learning or school performance). Are the standardized tests used to assess performance of schools measuring all the things children should learn in school? And can a measure of children's learning capture everything a school should do? While most would agree that these two questions cannot – strictly speaking – be answered in the affirmative, the ability of standardized tests to assess the kind of learning children need to engage in and capture the public schools' central mission is one of the most contentious issues in education.

The seminal paper by Holmstrom and Milgrom (1991) illustrates the mechanisms of measuring a subset of the things that public schools should do, and attaching performance incentives to these measurements, as the new accountability mechanisms do. To follow Holmstrom and Milgrom's model, public schools can be thought of as having two distinct tasks. One of the tasks, let's say it is teaching math, is readily measurable. Students can be given a standardized test and the degree to which the school has succeeded in its task can be ascertained. Other than teaching math, the public schools have a second task, let's say it is either teaching art or teaching children to become good citizens. Because there are no good standardized tests for students in art or in their ability to be good citizens, the school's performance in this task cannot be readily assessed.

The intuitively obvious result from Holmstrom and Milgrom's model is that schools will do more of the task that is measured and for which they are rewarded. If consequences are based on the test results in math, while success in the other tasks has no consequences because it is not measured as diligently as math or not measured at all, then

schools will tend to focus on teaching math, rather than focusing on the other tasks. A strong focus on math is a desirable outcome if the second task of the school is to teach children to become good citizens and teaching math supports this activity (e.g., the discipline and rigor that goes into learning math has beneficial effects in other aspects of the students' lives). However, if the second task of the school is to teach art, and the schools do not consider the teaching of art and math to be complements but rather two subjects that compete for attention in the curriculum, then the Holmstrom-Milgrom model suggests that rewarding schools on the basis of test scores alone (i.e., test scores in math) will lead to an inappropriate prioritization of math in the schools' curricula, to the detriment of art education.

The predicted prioritization of the dimension that is measured is a general result of the model. An analogous situation is also to be expected within a single subject. For example, teachers might choose to have their students practice specific test formats that promise to make students achieve higher test scores. Teachers might do so even if they know that training specifically for the tests themselves does not further the students' learning, or at least not as much as other non-tested dimensions of teaching and learning might. For the case of education, this behavior predicted by the Holmstrom-Milgrom model – when incentives are attached to tasks that are measurable only in one dimension – is called 'teaching to the test'.

Roy Romer, the former governor of Colorado and current Superintendent of LAUSD, responds to those who criticize 'teaching to the test' by saying "that is what flight schools do, because we take flying seriously" (quoted in Will, 2004). He likens testing throughout the school year to what many football teams now do by studying photos during the game, when diagnosis is immediately useful. Romer's response makes a lot of sense if the tests are of sufficient quality and breadth to be worth teaching to. But is it really the case that student test scores have full validity as a measure of school performance? The Holmstrom-Milgrom model suggests that the validity of test scores is

important, when considering the appropriate schedule of consequences for schools based on these measures.<sup>11</sup>

Compared to the discussion about the validity of student test scores, the discussion about their reliability is less fraught with value judgments and emotions, but contentious nonetheless. For a measure to be reliable it should be free of random error. Are changes seen in the measure (averages of student test scores, in this case) really due to changes in the underlying construct, or are they due to something else? Kane and Staiger (2002b) show that the volatility in student test scores due to non-persistent factors accounts for a large portion of the variability. They use data from elementary schools in North Carolina to estimate that more than half of the variance in mean gain scores and over three-quarters of the variance in the annual change of mean test scores is a function of sampling variability (among cohorts of students) and other non-persistent factors. They note that “to date, school accountability systems have been designed with little recognition of the statistical properties of the measures upon which they are based.” Their research casts serious doubt on the reliability of performance measures as they are currently used in many accountability systems. Kane and Staiger conclude by warning that “an accountability system that seems reasonable in a world of persistent rates of change and easy-to-discern differences in value-added may generate weak or even perverse incentives when implemented in the world of volatile test scores.”

The authors further demonstrate the expected result that school-wide averages are more volatile for smaller schools than for larger schools due to the smaller sample that makes up the small schools’ averages. Since awards and sanctions go to the schools with the most extreme scores (or changes in scores), small schools should be expected to be overrepresented at either end of the distribution. Kane and Staiger find this to be the case. Reviewers of the Kane-Staiger study have largely supported the authors’ conclusions, and emphasize their importance for the design of accountability systems (Grissmer, 2002; Ladd, 2002). The question that the paper raises is whether the measures used to assess school performance (student test scores) are being used to identify the good versus the

---

<sup>11</sup> For empirical evidence of schools’ focus on increasing test scores to the detriment of other obligations, such as teaching non-tested subjects or attending to the broader developmental needs of the children (as predicted by the Holmstrom and Milgrom model) see, for example, O’Day and Bitter (2003).

bad schools or simply the lucky versus the unlucky schools. Kane and Staiger offer important evidence to suggest that the measures used are not as reliable as the incentives attached to them would warrant.

In a forceful critique of both the methodology, as well as the conclusions, David Rogosa argues that the reliability of the school performance measures is indeed much higher than Kane and Staiger suggest (Rogosa, 2002; Rogosa, 2003). He offers stylized examples of test scores which would yield contradictory results to the Kane-Staiger analysis, and also disagrees with the notion that accountability systems are partly about the relative standing among schools. This latter assertion seems strange, given that awards quite often go to the top schools and sanctions to the worst schools (in relative terms). The methodological disagreements about the estimation of the reliability of school performance measures aside, Rogosa's main concern seems to be the legitimacy of the new accountability in general, which he supports and which is challenged by what he feels is an excessively critical evaluation of the reliability of school performance indicators.

In the end, it should be understood that the exact use of student test scores, and the way school performance is inferred from them, matters a lot. This is where the question of validity and reliability of measures intersect. To a first order, school performance should be judged by estimating how much student test scores have improved, rather than how high they are in an absolute sense. In other words, the quantity of interest is the net effect, which means gain scores or differences from one year to the next, even if these are less reliably measured, according to Kane-Staiger, than absolute scores. However, even the more advanced techniques of value-added modeling are currently not reliable enough that they alone would support high-stakes consequences (McCaffrey et al., 2003). Instead, the measures are probably sufficiently valid and reliable that they do support "medium stakes" performance management, which implies less dramatic consequences for the schools (Klerman, 2005); consequences that take account of the fact that the schools' performance cannot be accurately measured. Alternatively, consequences could be based on a broader set of measures than test scores alone.

The validity and reliability of measurements of school performance will reappear in the following sections and chapters, as they have profound implications for the legitimacy of the new accountability, and for the way in which failing schools are handled in particular. As Hamilton and Stecher note (2002): “The specific features of an accountability system are likely to have a tremendous effect on the degree to which that system promotes desirable outcomes.”

In theory, the greater emphasis on individual schools as the unit of accountability is accompanied by granting schools greater autonomy in their decisions. This is arguably the crucial trend in public education reform: schools are subjected to greater accountability, in exchange for greater autonomy in achieving the established goals. Greater autonomy is the logical counterpart to a more tightly defined mission of primary and secondary schools, and it is also a complement to another important trend that often accompanies the new accountability, namely the move towards parental choice. As reform proponents are increasingly arguing in favor of harnessing the powers of the free-market to improve public education, parental choice has become a central theme. Though parental choice can be seen as a subset of the new accountability, or as a logical extension of it, parental choice is discussed separately here.<sup>12</sup>

### *2.1.3 Parental choice*

Among the proposals for reform of public education, the introduction of parental choice and market competition are gaining increasing acceptance not only in the United States, but also abroad. Governments around the globe have decided that giving parents more choices among schools is an appropriate policy response to the problems their systems of public education face. Naturally, the problems that parental choice policies are designed to address vary across countries, and so do the details of the policies. Nevertheless, as Plank and Sykes (2003) state, an “increasing reliance on markets is generally accepted as a plausible strategy for encouraging improvement in the educational opportunities that governments provide for their citizens.” Though the evidence that school choice policies have significant, beneficial effects remains

---

<sup>12</sup> Parental choice has also been introduced independently of, or at least prior to, the new accountability, as in the case of New Zealand (see Chapter Three). Similarly, implementation of the new accountability need not lead directly to parental choice, as in the case of the No Child Left Behind Act.

provisional and equivocal, the move towards choice and competition in national education systems appears inexorable. In no country is there a serious effort to expand the role of the state in educational provision or to restrict the educational choices of parents (Plank and Sykes, 2003).

The idea of parental choice in general, and educational vouchers as specifically, has been in discussion for some time (see Friedman, 1955; Levin, 1968). But it has been only in the 1990's that these policies have found their way into mainstream reform movements, to the top of the policy agenda, and into implementation.

School choice policies affect both the supply and the demand of public education, and the arguments in favor of expanding choice can be grouped accordingly. Choice impacts the demand for public education by giving parents more power to choose the school that their children will attend vis-à-vis the government which, in traditional school systems, assigns children to designated schools according to enrollment zones. The demand-side arguments are related to the importance of individual freedom and a concern for equal opportunity in education for the disadvantaged, as well as what might be described as communitarian ideas of parental involvement.

A fundamental idea in favor of parental choice suggests that if each individual is free to choose where to live and work, parents should be given similar freedoms when it comes to choosing their children's school. The issue of individual freedom is also used to support the idea that choice improves equal access to education, by providing options to economically disadvantaged families. Proponents argue that giving poor families the right to choose the public school their children attend provides them with options that richer households already enjoy. Richer parents have the option of sending their children to private schools or to move to a different neighborhood, where the local public school may provide a high-quality education. Howard Fuller, a prominent proponent of school choice policies in the United States and an advocate for the African American community argues this point forcefully. "This is a debate about power ... this is about whether parents of low-income African-American children should obtain a power that many critics of the choice movement exercise every day on behalf of their own children" (Fuller, quoted in Ladd, 2003). Finally, a set of arguments in favor of parental choice is

based on the notion that parents are more willing to participate in a school they have chosen. Thus, providing the parents with choice is a way to involve them in their children's education. This idea also pertains to the supply of education that is improved through the involvement of parents.

The arguments about the positive impact of parental choice on the demand of public education are not universally accepted. Empirical evidence suggests that parental choice leads to greater segregation of student populations across individual schools, and that choice policies favor families of higher socio-economic status (SES). As Mark Levin (1996) writes, "it appears that those who exercise the choice option are more likely to be of higher SES and to have higher achievement scores than those who continue to attend their assigned schools." Terry Moe, one of the most prominent knowledgeable supporters of school choice, concedes that unrestricted school choice "may well lead to selection effects with a class bias" (quoted in Levin, 1996).

The starting point for the arguments for parental choice that focus on the impact of choice on the *supply* of public education is the monopoly of the traditional education system, which is believed to have become excessively bureaucratic and wasteful.<sup>13</sup> Providing parents with the choice of where to send their children reduces or eliminates this monopoly power, since public schools can no longer rely on a captive student population. Parents can leave schools they dislike for schools they prefer, taking some or all of the public funding with them. Consequently, schools must make a more concerted effort to attract and retain students (Plank and Sykes, 2003). Typically public schools are given the autonomy to do so, which entails greater authority over staffing, the school budget, and parts of the curriculum. By allowing parents to send their children to the school of their choice, whether it's a traditional public school or an alternative (e.g., charter school), schools will be forced to compete for students and will, by analogy to the role of competition in the private sector, be induced to become more productive, providing higher-quality education at lower cost (Hoxby, 2002; Ladd, 2003). In theory, this mechanism works for the whole system, as schools that fail to respond to the

---

<sup>13</sup> See, for example, Chubb and Moe (1990).

competitive pressures will be driven out of business as they lose students. This, in turn, will gradually improve the educational system as a whole (Maranto et al., 2001).

Just as with the effects on demand, the effect of school choice on the supply of public education is not universally agreed upon. There has been strong opposition from vested interests against choice systems. The most vocal opponents are school personnel, who could potentially lose job security, particularly if the choice plan involves private schools (Hanushek et al., 1994). The mechanisms by which parental choice will help or force struggling schools to perform well or force failing schools to cease their operations is left implicit in most education reform proposals. These mechanisms will be explored later in this chapter and are discussed throughout this dissertation.

Practical experience has shown that the hope of choice proponents, in regards to school failure, is only partially fulfilled. The automatic elimination of all failing schools often does not happen, as school decline and failure do not always take the paths that theory would suggest. With market mechanisms at work in varying degrees across all schools and with parents given a greater role in determining the course of their children's education, the role of the government in this system is not simply smaller than before, but also – and importantly so – different than before. Early experiences have shown that the integration of parental choice into systems of public education works best when the role of the government with respect to failing schools is carefully considered and continually calibrated.

[The wide-spread implementation of choice policies] does not mean that the power of the state in the education system has been fully and finally eclipsed. The current enthusiasm for markets as the solution to a host of public and private problems will almost certainly wane over time, as the consequences of market failure begin to manifest themselves. By their nature, markets create both winners and losers. As the losses associated with the emerging market for schooling become more evident, the regulatory role of the state and educational professionals will once again increase. (Plank and Sykes, 2003)

Plank and Sykes predict the eventual increase of the regulatory role of the state in systems of public education that have incorporated parental choice. While parental choice is a way to hold schools accountable, by allowing parents to choose alternative schools for their children, the following sections and chapters will show that the mechanism does not work equally well for all schools.

Parental choice is not only a critical element of many education reform proposals, but it is also used as an intervention in individual schools. The No Child Left Behind Act of 2001 mandates that parental choice be offered in consistently low-performing schools.<sup>14</sup> Since this mandate has come into effect only during the past year, it is still too early to evaluate its impact. In the following section and throughout this inquiry, the focus is on parental choice as an element of education reform, rather than as an intervention, unless specified otherwise.

Effective choice relies, in either case, on school performance measures that enable parents to make informed decisions. The adequacy of the measures discussed earlier is thus of utmost importance. An educational marketplace, as any market, cannot function without valid and reliable performance information (Garn and Stout, 1999).

The new accountability and parental choice, whether separately or in concert, focus on the differential performance of individual schools, and each attach some set of consequences to school performance. Evers (2001) notes that “prior to the current accountability efforts public schools faced few consequences if they failed in the job of educating their students.” This dissertation is concerned with the impact and the nature of these “consequences.” More specifically, the following sections focus on the lower end of the distribution. How should public schools that are at the very end of the performance spectrum be dealt with? What can and should the consequences be, if any at all?

#### *2.1.4 Criticisms of the dominant trends*

Before continuing with the discussion of school failure in particular, it should be noted that there are fundamental criticisms of the emerging paradigm in public education (the new accountability and parental choice) as it is outlined above. That the described ideas for the improvement of public education should have become dominant in the past two decades is not inevitable and the way in which they continue to be challenged is useful to consider. While the purpose of this dissertation is not to contest the recent trends, a selected set of criticisms can contribute to the present inquiry and are introduced briefly in this section.

---

<sup>14</sup> The option to send their children to another, high-performing public school of choice within the same district is offered to parents of children that attend under-performing public schools (NCLB, 2001).

A fundamental challenge arises from the conception of education as an economic good. As Mitchell and Mitchell (2003) argue, a presumption underlying the introduction of market forces is that education is a durable good which has a residual and measurable value. The authors present a political economy perspective to argue that the diversity of policy proposals for the improvement of public education is best understood when education is not strictly thought of as being a durable good. Instead, education may be thought of as a service good, an investment good, or a cultural identity forming good.<sup>15</sup> The authors present their framework to direct the attention towards “the divergent interests and values embodied in competing policy proposals.” Policy proposals differ not only because they propose different paths to a common goal, but also because they are expressions of conflicting ideas about what education really does and should do. For example, a majority of parents may be motivated by the pursuit of schools with higher productivity, in the sense that education is a durable good. However, the possibility that a subset of parents thinks differently about education (i.e. they do not value productivity highly) is troubling, if these same parents are disproportionately represented in schools that have been identified as failing according to an accountability framework that equates failure with lesser productivity. Market forces through parental choice would, in such a case, do little to respond to the officially identified failure.

More broadly, not all observers believe that parents’ choice should be respected. The existence of truancy laws, for example, suggests that policymakers are willing to prescribe a certain level of education even if some might choose to consume less. From such a perspective, schools are perhaps less like private companies producing products for consumers and more like prisons that provide a service to the government or the society. Similar to the way that convicts’ preferences for any particular prison are

---

<sup>15</sup> Durable products can be retrospectively evaluated on the basis of the residual value of the product. An underlying assumption in choice based reforms is that parents can be relied upon to make these durable product type market decisions (at least often enough to dominate the thinking of education producers). If, however, education is seen as a service good, the evaluation takes place at the moment of consumption, and market choices rely much more on the reputation of the service provider than on the durability or resale value of the service itself. If education is more like an investment good, its perceived value would be based on judgments of the future regarding risk and the expected rate of return, rather than on the basis of current product price or value. Finally, some economic goods can be seen as cultural identity forming goods rather than durable products, services or investments. Church membership, for example, may be such a good. Going to a particular school might be more like going to a particular church, rather than buying any particular computer.

discounted, because the individual choices are likely to be in conflict with what furthers the common good, the preference of parents may also be out of step with the goals of the broader community.

Varenne and McDermott (1998) offer a critical examination of the language of failure and success that increasingly governs American schools. They present a fundamental challenge to the implication that, with the reduction of all experiences in public schools into a single scale, half of the schools are always below average. Since the performance of a school is based on the performance of its students, an accountability framework that labels so many children as ‘failures’ cannot be said to be working most effectively to enable and encourage all children. The authors propose a more context-dependent conception of public education and a more cautious use of the binary state of failure versus success.

The criticism of the language of failure aside, the new accountability that is emerging for public education does increase the focus on the differential performance of schools. A closer look at what failure means in this context is in order.

## **2.2 School Failure**

### **2.2.1 Definition of school failure**

Failure in general is relative and is most often defined in relation to some expected norm. According to the Oxford English Dictionary (2004), failure is defined as *omitting to effect one’s purpose; omitting to perform something due or required*. The problem in this context is that there is neither (a) a universally valid expectation for schools, nor (b) an agreement on how the performance should be measured, or (c) how low the performance has to be before the designation of failure is warranted. The use of the term ‘failure’ is itself controversial for describing the state of schools. Several studies have noted the negative impact on the morale and motivation of teachers within schools that have been given such a label (Mintrop, 2002; Myers and Goldstein, 1998).

Especially in the context of decentralization in the delivery of educational services, the notion of failure among individual providers is an important one. As Barber (1998) has argued, the difference between a bad school and a failing school lies “in the extent to which the school is capable of sustainable self-generated improvement.” Failure

implies a point at which the individual school is no longer able to generate sustained improvements in its operations on its own. As an organizational unit, the failed school is in need of interventions from the outside because, left to its own devices, it will not manage to provide adequate education, nor prevent its further decline.

This raises the question of whether failing schools are categorically different than bad schools that are not yet failing. Are schools at the lower end of the performance scale a homogenous group or a heterogeneous group, for example, in the degree to which they can make good use of additional funding or technical assistance? One way to think of this is through a medical analogy: obesity and asthma.

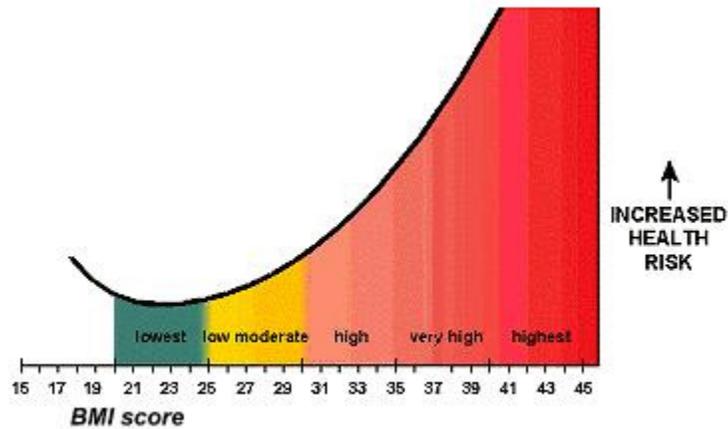
Obesity is a condition characterized by an excessive accumulation of fat tissue. It is usually determined using the body mass index (BMI), which is based on the ratio of an individual's weight over height-squared. Obesity is defined as a BMI equal to or greater than 30 kg/m<sup>2</sup>. Obesity correlates strongly with health problems and higher health care expenditures. About 60 million Americans are estimated to be obese, or roughly 30 percent of the adult population in the United States (NAASO, 2003).

Asthma is a chronic lung condition, characterized by difficulty in breathing. People with asthma have extra sensitive airways that react by narrowing or obstructing when they become irritated by factors in the environment, called triggers. Roughly 15 million Americans report having asthma (Asthma in America, 2004). The prevalence of both obesity and asthma in the United States has significantly increased in recent years.

There is a statistical relationship between the body mass index used to define obesity and the risk of health problems as a result of obesity. The curve is nonlinear. It is relatively flat for BMI indices around the range of non-overweight people. As the BMI increases the curve gets increasingly steep. But there is no discontinuity, even though obesity is defined as a BMI of 30 or above. While the cutoff at 30 may be useful heuristically, it may also be misleading, since a person with a BMI of 29 does not face a materially different level of risk than one with a BMI of 31.

**Figure 1 – Chart of BMI vs. health risk**

Source: [www.nutritionadvisor.com/BMI\\_Calculator.htm](http://www.nutritionadvisor.com/BMI_Calculator.htm)



With asthma, on the other hand, there is a difference between those who suffer from it and those who do not. It is not a smooth curve, since only a fraction of the population (though an increasing one) is affected by it. When it does exist, it can be more or less severe, but there is a categorical difference between those who actually have and those who don't have the disease.

So, the question is: Is school failure like obesity or like asthma?

Though Connelly (1999) notes that it is “unlikely that researchers will agree on ways of defining school failure,” Peter Mortimore’s definition of a successful school (cited in Barber, 1998) could be turned on its head. Thus, a failing school could be said to be one in which students progress less far than might be expected from a consideration of its intake. But such a concept, dependent as it is on the expectation of how far and in what way exactly children should progress when they are in school, sounds more like the somewhat arbitrary definition of obesity, and recalls the overuse of the term school failure in political rhetoric. The prevalence of failure would then depend on its definition; on the particular level of performance below which schools will be deemed to be failing.

Another possibility is to define failure by the intervention that is best suited to correct it. Depending on the severity of the intervention, failure may mean something quite different. But even with a steadily escalating schedule of interventions, it need not imply a smooth curve. According to Barber’s definition of school failure, a retained

capacity for self-generated improvement that a school either has or doesn't have implies a categorical difference. There may be a category of school failure where a standard set of interventions of varying scope and intensity is effective without significantly changing the school as an organization. In contrast, there may be a different category of school failure where these interventions and the specific support or pressure that comes along with them cannot be integrated effectively. The school as an entity may have lost (or never possessed) a capacity to respond adequately, even when given support from the outside. For such a category of schools, more severe interventions would be needed to achieve a significant effect and the label of failed school maybe appropriate. The interventions suitable for these schools would necessarily have to be very disruptive to the structures and processes within the school.

This dissertation does not attempt to answer this question conclusively. Instead, a failing school is simply one that has not met a threshold of performance specified by the accountability framework within which it operates. Failure thus implies that a school has been identified as performing below a given cutoff. As is discussed in greater detail in the following sections and chapters, the definition of failure varies widely across jurisdictions. Struggling, low-performing, or under-performing are all terms used interchangeably to designate schools that may be in danger of failing, but are not necessarily identified as such.

The identification of a school as failing does not suggest that the problem is of a specific type or that any particular response or intervention would be best to correct the failure. The origin of school failure may be with the management of the school, with the instructional program, or with other factors. The specifics of these different reasons for school failure are not a focus in this dissertation. Rather it is acknowledged that the nature of school failure is complex and the interventions used to address it need to be carefully planned and continually evaluated. The medical analogy is useful here, because the goal in either case, treatments for the afflicted or interventions in failing schools, will be to improve the outcomes for subjects. If the reason for bad outcomes is a condition like obesity, exercise and a healthy diet are generally good ideas, for all people and for those who are overweight especially. On the other hand, the possibility that the problem is a categorical one, like asthma, is important. In such a case, something different may

need to be done for a subset of schools, not just more of the prescription that works for all other schools. This possibility is important for the development of robust policies towards school failure.

### *2.2.2 School failure and equity*

After considering the definition of failure conceptually, a first look at school failure in practice is in order. A cursory examination of school failure will necessarily turn to the apparent positive correlation between the probability of a school being identified as failing and the proportion of students from low-income households attending the school.

One major reason school failure is a concern is that it tends to disproportionately affect economically disadvantaged children. To say that there is a strong connection between failing schools and the children most in need of good educational opportunities is in part a tautology, since the performance of schools is inferred from the analysis of student achievement data. But it also points to a deeper issue that school failure raises. The concern with school failure is almost invariably a concern with the equitable access to public education.

In addition to the difficulty failing schools have with using their resources productively, they may also have inadequate resources. In other words, the problem may not be only the efficiency with which the available inputs are used, but also insufficiently available inputs, compared to other schools that are doing better. As Hanushek et al. (1994) write, disadvantaged students may well require additional resources, even when all resources are being used effectively: children from disadvantaged households may simply be more expensive to educate. A funding mechanism for schools based on the assumption that all schools should get the same amount of money may leave schools serving low-income families scrambling, while their counterparts are doing fine. Treating these schools as if they were doing the same job, then, would not be adequate.

The effect of schools on student performance must be separated from the effects of families and other factors. Families with higher incomes tend to be more educated, smaller, and more likely to have both parents present. Raised with a greater emphasis on education and more encouragement to achieve, children from such families tend to be better prepared for school and to achieve at a higher level. Just the opposite holds for children from poorer families. Such

differences hide the true impacts of their schools. Some effective schools remain unnoticed because they cannot overcome all the adverse conditions that lower their students' performance. (Hanushek et al., 1994)

The authors argue for a public education system based on choice, but acknowledge that compensatory payments may have to be instituted in order to make the system fair for the disadvantaged children. Insofar as disadvantaged students are more expensive to educate, schools should receive more money when they accept such students.

William Ouchi's proposes (2003) to make choice policies work for everybody, including low-income families, by establishing a Weighted Student Formula. This was pioneered in Canada when the school district of Edmonton, Alberta, switched to parental choice. While the funding follows the students to the school, no matter which one they attend, the formula provides different amounts of funding depending on the learning needs of the child. For example, a student with a disability and English as a second language would receive extra funding on both counts compared to a native speaker with no disability. In this formula, low socio-economic status (SES) is treated as an additional learning need, and students from such families are funded at a higher rate than the children from wealthier households.

Hanushek and his coauthors, as well as Ouchi, emphasize that they do not believe in the often stated need for additional resources for public education generally. However, they do consider redistribution among schools according to the SES of their student populations appropriate. The prescriptions are quite radical compared to the way schools are funded in the United States. While the style in both cases suggests that the proposed change in relative funding was not meant to be radical, the recommendations of Ouchi and Hanushek et al. certainly stand in stark contrast to current funding practices, as the current practices are exactly the opposite from what is desirable, in consideration of the above arguments. Nearly half of funding for public schools is provided through local taxes, resulting in large, persistent differences in funding between wealthy and impoverished communities (Biddle and Berliner, 2003; Carey, 2004). As Biddle and Berliner note in a recent report (2003),

districts reporting higher levels of funding are more likely to come from communities where student poverty is minimal, whereas those reporting lower levels of funding more often come from communities where student poverty is sizable. To understand the magnitude of this problem it needs to be noted that America has by far the highest rate of poverty among children of any advanced, industrialized nation.

The skewed distribution among American schools is disconcerting, especially as the focus on the performance of individual schools intensifies. Though Hanushek himself is critical of input-based education policies more broadly (Hanushek, 2003), the arguments he and his coauthors present with regards to *relative* funding levels suggest that, given the unequal distribution of funding (unequal in the opposite way it's supposed to be!), a straightforward comparison between the performance of schools is inherently unfair.

Examples of high-performing public schools that serve low-income neighborhoods do exist. William Ouchi (2003) is fond of providing such stories as evidence that a low-income neighborhood need not imply a low performing school, despite funding differences as they may exist. The stories show that students from poor households can achieve high standards in regular public schools and that becoming a high-achieving school is not necessarily an issue of financial resources. The fact that the successes are few, however, also suggests that determined and innovative leadership and a committed staff that make such success stories possible are not the norm in schools in poor neighborhoods. In fact, available evidence suggests that schools serving low-income households find it harder than schools that serve wealthy neighborhoods to recruit and retain qualified teachers, quite aside from differences in funding (Guarino et al., 2004).

Funding from federal sources, through Title I of the ESEA or NCLB, is meant to level the playing field by providing additional funding to schools serving students from low-income households. NCLB in particular was passed to address the achievement gap that exists along SES lines (Stecher et al., 2003). But even within the population of Title I schools, those that serve a higher proportion of poor students are most likely to be failing to meet the established performance targets and thus be subject to federally mandated sanctions (Kirby et al., 2003).

It should be noted that the connection between school failure and equity is not simply an issue of money. The concern that the correlation between school failure and economic disadvantage raises can be extended to the political sphere. Low school performance tends to be prevalent among communities with marginalized political status. Whether the often proposed introduction of market forces would effectively empower the disenfranchised vis-à-vis the politically dominant communities seems doubtful in light of the evidence presented here. Indeed, existing research suggests that schools serving communities of higher socio-economic status respond more readily and more coherently to the demands of external performance-based accountability systems than their counterparts serving families of lower socio-economic status (O'Day, 2002). This suggests that schools serving the economically disadvantaged families actually lose ground relative to the well-positioned schools once an external accountability framework is instituted.

The issue of equity is clearly of great importance for failing schools. Some of the interventions that are discussed below entail additional funding that appears to correct for some of these observed inequalities in resource allocation. Since the additional funding is not allocated explicitly for that purpose, but to assist schools that are struggling, the incentives for schools arising from such allocations may produce unintended consequences and need to be considered critically.

### ***2.3 Are Interventions in Failing Schools Really Necessary?***

After an introduction of the trends in education reform that have resulted in a focus on the performance of individual schools and a discussion of school failure itself, the question remains what the government, as the ultimate sponsor of public education, should do about school failure, if anything. To restate from Chapter One, possible responses of a government faced with individual schools in the system of public education that are failing can be organized into three groups:

- IV. Tolerate failing schools. School failure may be seen as a regrettable, but largely unavoidable part of the system.
- V. Change the system of public education. The system could be changed to prevent failure from becoming a problem by instituting mechanisms whereby failure is

dealt with automatically, either by forcing schools to improve, or by ensuring that schools which are considered to be failing are forced to close.

- VI. Devise a strategy of interventions. The government directs specific actions to be undertaken by government agencies or outside contractors, to actively deal with failing schools, either by instituting changes at the school that will lead to improvement, or by actively closing the school down.

As was briefly stated in Chapter One, the first approach stands at odds with the goal of ensuring that all citizens have quality education. This option is not considered further, because it contradicts the role of the government as the sponsor of public education that is assumed, and because the definition of interventions is chosen to be deliberately broad. Options two and three are not mutually exclusive, of course. In fact, this research will present evidence that they are complements. Before starting the discussion of interventions, the mechanics of option two are examined to show why, and in what respect, interventions are indeed necessary.

### ***2.3.1 Improvement of schools***

The focus on the performance of individual schools, particularly on schools that are considered to be failing, is increasing, as discussed earlier. Parental choice is a mechanism for holding schools accountable, but also to subject failing schools to market forces. Ideally, parental choice in public education could be organized in a way that would provide schools with incentives to improve when they are struggling and would force schools to cease their operations when their failure persists. Interventions are indeed only necessary insofar as practice falls short of this ideal case. Below, the assumptions behind the mechanisms of built-in improvement and the exit of schools from the system of public education are considered. Understanding the basic assumptions helps explain why *policy* responses to failing schools have not featured prominently in the literature, and how interventions can be used most effectively.

How do the market forces that parental choice brings to bear on public schools help low-performing schools improve? Presuming there is a group of schools that have been identified as failures, how can these schools be brought back to a level of performance that is roughly in line with what other schools are doing?

A point Albert Hirschman (1970) makes is that failure of some magnitude is inevitable, due to the failure of some actors to live up to the behavior which is expected of them. Thus, organizations such as public schools are subject to “decline and decay, that is, to a gradual loss of rationality, efficiency, and surplus-producing energy, no matter how well the institutional framework within which they function is designed” (Hirschman, 1970). One key assumption for Hirschman is that the deterioration in the performance of the schools is due to unspecified reasons that are “neither so compelling, nor so durable as to prevent a return to previous performance levels.” This implies that the problem is primarily the failure to put existing, and generally sufficient, capability for the higher performance to effective use, rather than the absence of the necessary capability at the school level. Furthermore, Hirschman asserts that “it is likely that the very process of decline activates certain counterforces.”

In the case of public schools, Hirschman sees two possible responses by parents to the failure of public schools: Exit and Voice. When parents take their children to be educated elsewhere, they exercise their option to *exit*. The counterforce is the competitive pressure on the school in decline, forcing it to improve lest it lose more of its students to other schools. When parents express their dissatisfaction to the school’s leadership directly, they exercise their option to *voice* their opinions. In this case, the counterforce is the direct pressure on the leadership or staff of a school resulting from the demands of the parents.

The dynamic between exit and voice, the two possible responses by parents, has important implications for their impact on public education. If parents have the option to exit, they may be less likely to exert pressure through using their voice.

Suppose at some point, for whatever reason, the public schools deteriorate. Thereupon, increasing numbers of quality-education-conscious parents will send their children to private schools. This “exit” may occasion some impulse toward an improvement of the public schools; but here again this impulse is far less significant than the loss to the public schools of those member-customers who would be most motivated and determined to put up a fight against the deterioration if they did not have the alternative of the private schools. (Hirschman, 1970)

The contention is that the two mechanisms operate in opposite directions. By allowing the exit of some, the collective capacity for voicing grievances and forcing

change within the organization is weakened. Those customers who care most about the quality of the product and who, therefore, are those who would be the most active, reliable, and creative agents of voice are for that very reason also those who are likely to exit first in case of deterioration. This dynamic between exit and voice is especially bad for schools that are performing at a low level to begin with.

Deterioration of a product in the upper quality ranges has to be fairly substantial before the quality-conscious will exit and switch to the next better variety. Hence the scope for, and resort to, the voice option will be greatest in these ranges; it will be comparatively slight in the medium- and low-quality ranges. ... Since, in the case of [public education], resistance to deterioration requires voice and since voice will be forthcoming more readily at the upper than at the lower quality ranges, the cleavage between the quality of [public education] at the top and at the middle or lower levels will tend to become more marked. (Hirschman, 1970)

This argument relates directly to current rules about declining schools. One of the provisions of NCLB is to mandate parental choice in failing schools. Specifically, the law allows parents whose children are attending a Title I school in Year 1 of Program Improvement (PI)<sup>16</sup> or later stages of PI status to switch to another, non-PI school. Schools in PI status are by definition at the lower end of the quality spectrum because their performance history indicates a continued lack of progress. The federally mandated intervention allows the most vocal and quality conscious parents to move to another school, leaving the overall capacity among parents to oversee the school weakened. As Hirschman points out, a limited amount of competition may actually protect the failing school from having to institute difficult changes. With the most vocal parents taking the option to exit, the under-performing school is left with a set of more readily satisfied (or less vocal) customers, and it can continue to operate at a level that would not have been tolerated prior to the exit of some of the parents.

In light of this dynamic, Ouchi's example of the school district of Edmonton, Canada is especially noteworthy (Ouchi, 2003). A few years after the deregulation and move to full parental choice, when public schools started to make use of their freedom in administrative and curricular affairs, the fraction of parents sending their children to public schools *increased*. Collectively, public schools gained market share relative to

---

<sup>16</sup> Year 1 of PI means that a school has failed to make adequate progress for three years in a row, and is thus in the first year of being subject to a set of sanctions and support, of which parental choice is one.

private schools and, in response, some private schools applied to become public schools. What happened was that quality conscious parents had returned to the system they had exited previously (or chosen not to send their children to). This move is not only a sign of the current success of public schools in Edmonton, but is also an indication that future prospects have improved, as the parents' collective voice is strengthened. Unfortunately, such success stories are few.

It can be shown in theory that parental choice may have adverse effects on the overall delivery of education, as the quality conscious parents flee the low performing schools. This phenomenon corresponds to empirical observations, as noted by Levin (1996). However, Hirschman notes that exit and voice can work together to improve education, when exit is not too easy or not too attractive. When parents feel a greater loyalty toward the school their child is attending, the possibility of exit, of abandoning the school, may serve to increase the weight each parent's voice carries with the school. If the parents' voice is strengthened, their role in oversight of the schools and in pressuring declining schools back into delivering high-quality education will increase to the benefit of the entire school.

As Hirschman also points out, for economists the problem of improving an organization that is in decline is not a central concern. One reason for this neglect is that, in economics, one assumes that a firm that falls behind does so 'for a good reason'. The concept of a more or less easily repairable lapse, a failure that can be fixed, in other words, is typically not part of the equation. This insight is echoed in the way that reform proposals by economists have typically dealt with school failure. Eric Hanushek and his colleagues spelled out an education reform proposal of considerable scope ten years ago (Hanushek et al., 1994). The proposal is generally consistent with the recent trends in education reform discussed above. They propose a greater focus on student outcomes, the use of performance incentives based on outcomes, and a continuous movement towards greater efficiency and institutional learning at the school level. The question of what an appropriate response to failing schools might be, however, gets only a few sentences worth of attention. The authors suggest that students be given the option to change to another public or private school (with the help of vouchers) if their school is considered failing. While they discuss the pitfalls of choice policies in the book – for example when

neighboring schools are already above capacity or they are located too far apart from one another – there is no mention that the response to failing schools they propose might not be adequate.

Hanushek and his colleagues acknowledge (at least implicitly) they are bringing the viewpoint of a particular discipline to bear on a problem that has typically been handled by other disciplines. They argue that “although economic issues are central to the problems of education, economic ideas have been notably, and most unfortunately, absent from plans for reform,” and redress this balance by providing analyses and recommendations for reform. More recently, the economists Kane and Staiger, who have left their mark on the education literature with their research on the volatility of school performance measures, conclude a paper by stating that

firm-level evidence from other industries suggests that an important channel through which market reforms affect productivity growth is by shifting production to more productive firms and closing down less productive firms, rather than by gradual productivity improvements in every firm. (Kane and Staiger, 2002a)

That the lessons of failures in ‘other industries’ are presented in a rather casual manner to inform the reform of public education suggests that economists have managed to firmly establish their viewpoint in the education reform debate. As Hirschman points out, for economists, such as Kane and Staiger, the decline and collapse of individual firms or organizations is not necessarily a negative event. This view is based on the conception of the traditional model of the competitive economy, in which the recovery from any lapse is not really essential. As one firm loses out in the competitive struggle, its market share is taken up and its factors are hired by others, including newcomers: in the upshot, total resources may well be better allocated (Hirschman, 1970).

In public education the failure and closure of individual schools may well serve to improve the quality of education overall. But is the closure of firms and organizations that do not respond to competitive pressures really inevitable? Is it reasonable to assume that a public school that is failing and does not respond to the pressures that parental choice exerts will close down automatically? These questions are considered in the following section.

### 2.3.2 *Exit*<sup>17</sup> of schools from the system

For examining the process of how schools might exit from the system of public education, Hirschman provides again a useful starting point. He bases his discussion on the assumption that the reasons for decline were neither compelling nor durable. What happens to schools if this assumption does not hold? Perhaps the reasons for a school's decline are both compelling and durable. In other words, what happens if the economists' expectations, that the decline happens 'for a good reason,' is true? Is it reasonable to assume that the counterforces that are activated by the decline, as Hirschman assumes, will force the school to close down? The theory behind parental choice says that schools with the inherent capacity to operate successfully will be exposed to competitive pressures and will be forced to improve. Schools without the capacity for improvement, on the other hand, simply will be driven out of business. Kane and Staiger (2002a) refer to market reforms that have the effect of "closing down" firms that are not as productive as others. School closure is often discussed in terms of an automatic outcome, even among those who oppose the very concept of school closure as a market mechanism. As Ron Wolk (1998) states:

All of these alternatives are part of a market-driven reform strategy. Its premise is that public schools ... would work to improve themselves if forced to by fair and open competition. ... It is more likely that failing public schools in a competitive situation would simply continue to decline, lose students, and eventually go out of business-which free market advocates consider a proper result.

Wolk does clearly not think that school closure is a benign event, but he also seems to believe that the path of declining schools exposed to market forces will most likely end in the schools going out of business.

Exploring the analogy, explicitly used by Kane and Staiger (2002a), of firms in a competitive market further, imperfect as it may be, can provide important insight into how these assumptions have played out in practice. The literature on transition economies and on the issue of the soft budget constraint (SBC), as reviewed by Kornai et al. (2003), has a lot to offer to make the response from governmental and judicial institutions to failure of individual schools more tractable.

---

<sup>17</sup> Note that this 'exit' implies a school ceasing its operations, rather than the 'exit' of parents described above.

The so-called SBC syndrome was originally formulated for transition economies, but has since been shown to be relevant also to failures in more mature economies. The SBC syndrome describes the behavior that emerges within failing firms and also within state organizations that are trying to save the firms from total collapse. The SBC syndrome is based on microeconomics and emerges when the theoretically hard budget constraints are relaxed while the official rule or rhetoric is maintained. In other words, failing firms are given support even though the official rule suggests otherwise. Insisting on a hard budget constraint in this situation turns out to be a considerable challenge for the state organization (or other supporting entity):

Although the expense entailed in repeated bailouts may be high, the cost of economic and social disruption ensuing from the enterprise's collapse could well be even higher. And so ex post there may be an irresistible force for making the bailouts. Indeed, if the potential disruption from collapse is big enough, both parties will anticipate a continuing sequence of bailouts. (Kornai et al., 2003)

The key aspect of the SBC syndrome is that the inherent inconsistency is known to the firms that are failing, as well as to the supporting organizations that have the capacity to avert a collapse through a bailout of the firms. The ex ante and ex post perspectives of the supporting organization are radically different. Ex ante, it would wish to refrain from rescuing firms in order to keep the risk of failure low; but ex post, once a failure has occurred, it has strong reasons to undertake a bailout and support the firm. Had the declining enterprise truly expected that the promise of 'no bailouts' would be kept, it would have been motivated to reduce the chance of failure. It is this lost motivation – and, most important, the higher prospect of failure that comes with it – that is, as Kornai calls it, “the real tragedy of the SBC syndrome.” Existing studies on strong interventions reviewed later in this chapter confirm the applicability of this dilemma for public education.

What the studies have revealed in transition economies, alongside the prevalence of the SBC syndrome, is that the processes for bankruptcy and firm closure have to be established first, before decline and failure can take the path that textbook economics prescribes. Establishing the infrastructure that allows firms to be closed down in case of their failure has proven to be a trickier proposition than the reformers initially thought. Failure, especially among formerly state-owned enterprises, leads to the exit of the

organization only if bankruptcy laws and courts are in place, and if state organizations refrain from rescuing or bailing out any failing firms (Kornai et al., 2003).

The SBC concept complements Schumpeter's theory of creative destruction. As Schumpeter (1947) pointed out, the development of firms or economies, and the emergence of new technologies or forms of organization, is directly connected to the destruction of the status quo. Indeed it is the destruction of firms, organizations, and the technologies or organizations they use that enables the growth and creativity upon which the economy depends. "This process of Creative Destruction is the essential fact about capitalism. It is what capitalism consists of and what every capitalist concern has got to live with" (Schumpeter, 1947). While Schumpeter's main goal was to explain the birth of organizations and the role played by entrepreneurs in generating entry, he tacitly assumed that the market takes care of the organizations' death. His assumption many decades ago permeates the thinking not only about market economies, but now also about public education. The SBC syndrome formulated for transition economies illustrates that the assumption of automatic exit does not generally hold. Of course, public school systems differ in important ways from transition economies. But insofar as economists and reformers more generally have made assumptions (at least initially) that have proven to be both wrong and costly, the two realms are similar.

Reform proposals for public education that are based on the establishment of new schools, a focus on the differential performance of individual schools, and the role of educator-entrepreneurs in driving a creative marketplace remain incomplete as long as the process of school improvement or school closure is not more explicitly addressed. Though the economists' ingrained bias in favor of exit should be noted, whether it be the exit of individuals from the schools, or the exit of schools from the pool of education providers, the lessons that have been learned in transition economies are important here. The smooth exit of firms and organizations is dependent on both an infrastructure that enables the exit and an analogous expectation among the actors, both of which are difficult to establish in newly created markets. This has been noted for transition economies (Kornai et al., 2003), as well as for systems of public education, as examples throughout this dissertation and the case study in Chapter Three illustrate.

The discussion of the issues that the SBC syndrome raises falls short in one aspect. The hard budget constraint impacts not only the individual organizations as they are affected by it directly (for example, whether they have to declare bankruptcy or not), but it is also of paramount importance to the system itself. This point has been made by Klitgaard et al. (2005), with regards to performance indicators for the distribution of foreign aid. When the politically difficult consequences are not applied consistently, or not applied at all, the loss of trust in the accountability system and the reduced willingness to invest in it may be the most significant outcomes. The primary impact of the SBC syndrome may be not on the individual units, whether they are developing countries or struggling public schools, but instead on the health of the system.

Ouchi (2003) relates the case of the impressive turnaround of the infamous Goudy School in Chicago and proclaims with enviable optimism that “if the worst school in America can become one of the best, then every school can be a success.” This is certainly an attractive thought, and an important goal toward which public education systems should strive. However, the discussion above provides the background as to why such a viewpoint is inadequate for policy formulation. It is essential to acknowledge not only that failure in some form or another exists in all educational systems, but also that a discernible response to the failure of individual schools should not be assumed to be inevitable. Even after the introduction of market forces into public education through parental choice, failing schools may well continue to operate without being affected in the expected way. Available evidence suggests that market forces work the way they are intended to work for a certain group of schools, and perhaps for the majority of schools. There seems to remain, however, a significant group of schools which are unaffected, and on which the market forces may have little effect or the opposite of the one intended.

Since it is these marginal schools, at the lower end of the distribution of performance, that influence the perceived and actual effectiveness of the whole system most dramatically, the fact that these same schools may not react to market forces in the expected way matters a lot. The complete eradication of school failure, as it is often propagated in political rhetoric, is not necessarily a possible, nor even a desirable, goal. But knowing how to intervene effectively in failing schools is crucially important.

## **2.4 Interventions in Failing Schools**

After looking at the recent trends in education reform, school failure itself, its connection to equity, and the ways market forces deal with it imperfectly, the discussion now turns to examining specific interventions. The following sections introduce the categorization and listing of common interventions, followed by the studies that have examined specific interventions. Unless otherwise noted, the studies discussed in the remainder of this chapter deal with interventions in public schools in the United States.

### **2.4.1 Definition and classification of interventions**

Before analyzing the existing literature on interventions, the most common interventions should be categorized explicitly so that the use of terms is more consistent from this point forward. This grouping of interventions is loosely based on Ronald Brady's categories in his report on interventions in failing schools (2003) that is discussed below in detail.<sup>18</sup> The interventions listed here are not distinguished according to the entity that is doing the intervention, because the lines separating them are different from one state to the next, or from one district to the next; even for individual cases the line is not always clear. For example, in the Los Angeles Unified School District (LAUSD), audits for schools in Year 3 of Program Improvement have been imposed by the state, represented by the California Department of Education (CDE), according to a federal mandate regarding schools that receive federal aid through Title I. The audits were initially carried out by the CDE and by the Los Angeles County Office of Education (LACOE). After LAUSD developed the expertise to do such audits, the CDE allowed the district to do the audits directly, in conjunction with staff from LACOE. Nominally, the audits are still imposed by the state, according to a federal mandate, but the intervention itself is carried out by the district with assistance from the county. The point is that the intervention itself does not change significantly as the responsibilities for the intervention change. The mix and even the interchangeability of jurisdictions is not restricted to LAUSD. Across the United States, it can be the state, the county, or the district that is carrying out the very same intervention, based either on federal, state, or district mandate. Instead of distinguishing the interventions by the entity that is carrying them out, the

---

<sup>18</sup> Alongside the interventions in failing schools, Brady lists and briefly describes interventions in failing districts, such as district takeover by the state department of education. As the focus here is on individual schools, the issues related to failing districts are not considered further.

attention is on interventions by the government more generally, and it is left to the details of any particular case which entity is responsible for carrying out the intervention.

The clearest way to classify the different interventions is in terms of their intrusiveness: mild, moderate, and strong (Brady, 2003). The following paragraphs and tables provide brief descriptions of these categories and examples of interventions that generally fall into the categories. The tables use the same or similar examples of interventions as Brady does, and differences to Brady’s categorization are noted.

As the listed example of Parent Involvement as an intervention suggests, interventions are not limited to actions that take place at the school itself. The directed development and support of community involvement for the school (meeting with, involving, and actively seeking out parents, community organizations, or also commercial enterprises) would constitute an intervention in the school, even if the activity itself is not taking place on the school premises. If the government were to encourage groups of parents or other community members to start a brand-new school, however, this would no longer be included in the set of interventions as they are defined here.

Mild interventions are not disruptive to the existing operations in the school. These interventions attempt to add elements to the operations of the school that may promote the effectiveness of the existing structures and processes. The additional programs or initiatives are typically implemented by the existing staff.

**Table 1 – Mild Interventions**

<b>Mild Interventions</b>	
Identification	Public identification of failing schools. This serves to improve the monitoring of the schools by the parents and by the government, and is also intended to motivate the school or its community to respond to its poor performance.
Planning	Mandating that the school set up a plan for its improvement. This would typically include a needs assessment by the school (possibly with an outside consultant) and a description of specific steps that can be taken to address the identified shortcomings.
Technical Assistance	Supplying low-performing schools with technical assistance from an experienced educator or a contracted consultant.
Professional Development	Providing training to the school’s staff. The training is geared towards addressing the needs outlined in the school improvement plan.
Parent	Requiring increased parental involvement in the school. Given the

Involvement	initiative from the school or district/state, parents organize to assist in classroom activities, after-school opportunities, school management, or even in classes for themselves.
Tutoring	Providing supplemental tutoring for students. Tutoring services are provided after school, before school, or on Saturdays to the children most at risk of failing state tests, or as an option to the whole student body.
Change of Financing	Provision of additional funds or withdrawal of funding. The change in funding for a school is a part of many other interventions listed here (also moderate and strong interventions), but may be an isolated intervention as well.

While mild interventions are not intended to be disruptive, moderate interventions are necessarily disruptive to the school. Moderate interventions are initiated when some changes in the operations of the school are deemed to be needed. Some moderate interventions may be the result of the school’s initiative, while others may be mandated. Changes in some of the basic processes or structures in the school are the intent and the result of moderate interventions. Like mild interventions, moderate interventions are mostly implemented by or with the help of existing staff, but they often involve outside experts that assist with and enforce the required changes in the way the school operates.

**Table 2 – Moderate Interventions**

<b>Moderate Interventions</b>	
Increasing Instructional Time	Adding more mandatory instruction by changing the schedule for certain students or the entire school.
Audits	Auditing the school with a team of outside experts for a comprehensive needs assessment. The audit results in a report with specific recommendations for improvement. The audits imply that the school is unable to correctly identify its own shortcomings.
School-wide Action Plan / Comprehensive School Reform	Implementation of a plan for changing the processes or structures within the school. Such change may be driven by an Action Plan, a Comprehensive School Reform plan, or by the plan written by the Audit Team.
School Choice	Offering the students in failing schools the option to attend another, non-failing school. These interventions are often hampered by the availability of enrollment opportunities in neighboring schools. <sup>19</sup>
Restriction of Autonomy	Reducing the authority of the principal over matters such as the budget, curriculum, after school activities, or other matters.
Change of Principal	Replacing the current principal with a new leader. The importance of leadership is widely documented in turning a declining school around (Berends et al., 2002; Brady, 2003; Connelly, 1999; ECS, 2002; McRobbie, 1998; Ziebarth, 2002), and such a step, while not

<sup>19</sup> Brady lists school choice as a strong intervention. The categorization used here, however, appears to be more in line with the intervention’s disruptiveness.

	involving any structural changes per se, can be a highly disruptive intervention. It may lead to fundamental changes in the school's processes, culture, and organizational outlook, though not inevitably so.
--	--

Strong interventions are different from mild and moderate interventions insofar as the existing structures in the school are deemed beyond repair. Changes are mandated and carried out by people or entities outside the school. Strong interventions signal the government's belief that the school failed and cannot be improved in its current form. School takeovers and reconstitutions can also signal a more hopeful note, as a fresh start is initiated. Typically strong interventions are reserved for cases where moderate and mild interventions have failed, with the result that only the most broken schools are subject to the most severe interventions.

**Table 3 – Strong interventions**

<b>Strong Interventions</b>	
Reconstitution	Removing and replacing all, or almost all of the school's staff and leadership. The existing staff is typically required to reapply for a position at the same school. The student body remains. Most of the school's operations are rebuilt from the ground up, such as the curriculum, as well as other structures and processes within the school.
School Takeover	Handing over the governance of the school to either an agency from the state department of education, or to an outside provider, such as a school management company. This may include staff changes similar to reconstitutions.
School Closure	Closing the school's operation outright. All staff members are removed, and all students are assigned to other schools. The school may reopen after a few years, but not right away, and not in a form that resembles the old school. Closure is the strongest intervention possible.

The list of interventions is not meant to be exhaustive, but rather a description of the most common interventions that have been tried and for which there is, in various jurisdictions, a legislative mandate. The list also provides an illustration of the chosen categories and helps to clarify terms for the remainder of the dissertation. The interventions are not mutually exclusive, but are often combined concurrently with one another, or in sequence. Interventions in failing schools are typically applied in an escalating fashion from mild, to moderate, to strong. The mix of interventions, as well as the sequencing, makes the analysis of their impact on school performance difficult. An

additional complicating factor for any comprehensive analysis is that the frequency of use is distributed very unevenly across the listed interventions.

#### 2.4.2 *Prevalence of interventions*

Before discussing the impact of the interventions, this section discusses how often the interventions introduced and listed above are applied. In general, there is an inverse relationship between the interventions' disruptiveness at the school and the frequency with which it has been applied. Mild interventions have been common in many states, and all of the ones listed in Table 1 are now required interventions in underperforming schools in all states, as part of the new federal mandate for schools that have not met their improvement targets for two consecutive years (NCLB, 2001). Before the passage of No Child Left Behind, many state agencies and local school districts had been using these interventions for several years. These efforts are very much connected to the specific environment in which they have been applied. They are also considered, for the purposes of this dissertation, to fall within the realm of regular district policies and are not examined further.

The use of moderate interventions listed in Table 2 is less frequent and carries greater cost, both financially and politically, than the relatively unobtrusive mild interventions. The reauthorization of the ESEA prior to the passage of NCLB, the Improving America's Schools Act (1994), required the use of moderate interventions for schools receiving federal funding through Title I that had failed to meet improvement targets for five consecutive years.<sup>20</sup> The use of moderate interventions and their estimated impact is discussed below, and in Chapter Four the impact of a specific set of moderate interventions in California is examined in detail.

Strong interventions are reserved for schools to which the label *failing* applies unequivocally. Given the high costs that strong interventions impose, a set of observations are worth noting. Strong interventions are rarely tried because they are controversial and difficult to mount. The political costs for such interventions are particularly high. While 39 states had the authority to take strong actions prior to NCLB,

---

<sup>20</sup> State audits were required for schools in Year 3 of Program Improvement status, which schools would reach after 5 consecutive years without meeting any improvement targets.

and while these same 39 states contain dozens of failing schools that have not appreciably improved for years, strong interventions have remained extremely rare (Brady, 2003). The mechanics of the SBC syndrome correspond to the available evidence regarding strong interventions.

Just about everywhere it has been tried, reconstitution has raised the strong ire of local teachers unions.... New Jersey has avoided taking over more failing districts because, having already assumed responsibility for three of the state's largest districts for an average of over 10 years [each], there is scant evidence that it has transformed these districts as originally envisioned. ... In Staten Island, New York, parents and students even protested the closing of a school that was on the New York State SURR [Schools Under Registration Review] list for 13 years. (Brady, 2003)

Given the reluctance with which moderate and, in particular, strong interventions are imposed, it is not surprising to find that the number of studies examining the impact of interventions in failing schools is relatively small. While the literature on incentives systems in education is extensive, surprisingly little is known about what kinds of interventions are most likely to turn faltering schools into successful educational institutions (Brady, 2003). Regarding the impact of strong interventions in particular, there is a dearth of evidence. Given the available data, the number and types of strong interventions, and the uniqueness of the circumstances in which they were attempted, this omission is understandable. The only state that has had experience with school closures due to lack of academic progress is New York (Brady, 2003). From the case studies that are available on the effects of interventions, the most salient results that cut across various interventions are summarized in the section below. The fundamental mechanisms that govern the interventions in failing schools are also restated, as a number of authors have described such mechanisms based on their analyses.

Though interventions are typically applied in an escalating fashion, the full sequence from mild, to moderate, all the way to strong interventions is a rare occurrence. Interventions remain, more often than not, at the mild or moderate stage. With escalating interventions, the political costs increase in proportion to the intervention's disruptiveness. Strong interventions are seldom enforced because they carry large political costs and because their effectiveness is not beyond reproach. The hitch here is that strong interventions suffer from a lack of empirical evidence regarding their

effectiveness, so that they are only tried in the most extreme circumstances, which in itself limits the possibility of collecting more empirical evidence.

## **2.5 Impact of Interventions**

The possibility of doing a cost-benefit analysis, especially of strong interventions, has been doubted fundamentally (Brady, 2003), since it is the severity of the school's failure that determines which intervention is most appropriate. As a result, the mild and moderate strategies appear to offer the most cost effective measures, even though they are least likely to work on the severest school failures because the changes needed to turn the school around are greater than the less disruptive interventions allow for. With these caveats in mind, the following section reviews a selection of the literature that is available regarding the impact of moderate and strong interventions.

### **2.5.1 Moderate and strong interventions**

In a recent report on interventions in failing schools, Ronald Brady (2003) offers a comprehensive review. He describes and categorizes 17 different interventions that have been attempted by states or school districts since 1989. He examines three specific cases for a detailed look into moderate and strong interventions: the Schools Under Registration Review (SURR) in New York; the Comprehensive School Reform in Memphis, Tennessee; and the Reconstitution in Prince George's County, Maryland. In each of the three cases, the interventions resulted in roughly half the schools making positive gains subsequent to the intervention, while the other half did not appear to respond to the intervention in a significant, positive way. This simple examination of the schools' performance, however, was done without a careful selection of comparison schools, and the conclusions are anecdotal rather than the result of rigorous analysis. Nevertheless, even if based on anecdotal evidence, the raw result that moderate or strong interventions have a significant, positive impact roughly half the times they've been implemented is a theme that recurs in several other reports discussed below. From his review of interventions, Brady summarizes a set of conclusions. The lessons learned are that,

- (a) many decisionmakers are more inclined to ignore schools that are identified as failing than to intervene,

- (b) some turnaround efforts have improved some schools, but success is not the norm,
- (c) no particular intervention appears more successful than any other,
- (d) interventions are uneven in their implementation and always hard to sustain,
- (e) it is nearly impossible to determine which interventions are the most cost effective because they are attempted in very different situations, and
- (f) school leadership is a common characteristic to successful turnarounds.

Brady relies in his report on descriptions of interventions, case studies, interviews, and the average achievement scores of the schools in question, without more detailed analysis of those scores. Such an approach is a common element in this literature. A number of other reports build on experiences and theory to provide practitioners with advice and syntheses of research findings, while stating that the body of research they are synthesizing remains very small. Among a number of different reports along these lines (in methodology and target audience), a number of common conclusions regarding interventions (moderate and strong interventions in particular) can be drawn (Department of Education, 1998; ECS, 2002; McRobbie, 1998; Stecher et al., 2003; Ziebarth, 2002):

- a) The evidence of the interventions' impact is mixed throughout. Under the right circumstances, interventions of varying intensity and scope can be successful in having significant, positive impact on the performance of the school, as measured by student achievement, or measured qualitatively (school climate, student behavior, teacher expectations). Providing the right circumstances, however, is very hard to do.
- b) The two things that matter more than anything are (a) the capacity of the intervening body, and (b) the leadership in the school during or after the intervention.
- c) The balance between supporting the school and providing clear boundaries – and escalating sanctions if they are overstepped – is difficult, but crucial to get right.

- d) The criteria for the assessment should be legitimate and fair and should be seen as such by those who are affected by the interventions, in particular the teachers (i.e., validity and reliability should not be fundamentally questioned).
- e) Management of expectations is key – at the school, at the district, and in the community. Results should be expected and demanded, but not right away.
- f) Strong interventions always have high political costs. The preference for avoiding them demands the careful and realistic design of an escalating schedule of interventions that, once it is in place, should be adhered to.

These conclusions echo those offered in Brady’s report, and also what is known about interventions that attempt to reorganize the processes in the school in a comprehensive manner, so-called whole-school designs. The effectiveness of interventions that are based on a whole-school design has been extensively studied by RAND researchers, who were contracted to evaluate the effort to develop a set of whole school designs, called New American Schools. A summary of the numerous evaluation reports was released two years ago (Berends et al., 2002). These interventions, which have turned into the Comprehensive School Reform interventions, which are further discussed below, are different insofar as schools self-select to implement a particular whole-school design. RAND’s findings on these interventions are very much in line with the six points listed above. The evaluation emphasizes the political dimension of the interventions, as the consistency of the support of, and will for, the interventions was found to be a major determinant of their success. The problems that intervening organizations face are also discussed in a WestEd report as they relate to takeovers of failing schools by the state:

Academic takeover is largely a leap into the unknown. No body of research yet exists to provide guidance. Most state education officials admit being far from eager to step in and run an academically failed school ... The record of attempts is scant; of success even more so. The issue is bound up in questions of why schools fail, who is really responsible – for both the low performance and the turnaround – and, most fundamentally, how to effect a turnaround. Involved are issues of organizational behavior, community dysfunction, human psychology, legal precedents and larger, contextual problems of race, class and urban neglect. Yet the state ultimately bears responsibility for acting on behalf of the kids in such schools, almost invariably poor and minority children whose very life chances depend on the quality of their schools. (McRobbie, 1998)

The tension between the state's responsibility and its reluctance to engage schools in interventions with unknown impact is implicitly recognized in all studies of strong interventions. Federal law has attempted to force the states to act upon their responsibility as the ultimate guarantor of public education. A few years ago, in the official report released by the Department of Education (1998), President Clinton's directive spelled out that interventions would henceforth be expected in cases where low-performing schools were unable to improve on their own. "State and local education officials must step in and redesign failing schools, or close them down and reopen them with new, more effective leadership and staff." The report offers guidance in this regard with practical steps and suggestions for professionals who are engaged with failing schools: "Because low-performing schools rarely have the capacity to make the kinds of changes required to turn around on their own, persistently low-performing schools need technical assistance, encouragement, intervention, and hope."

In contrast to the Improving America's Schools Act of 1994, the No Child Left Behind Act of 2001 leaves states very little room for interpretation and imposes the requirement of using strong interventions in consistently low-performing schools (Department of Education, 2004). However, it is not yet clear how strictly these mandates are to be interpreted in practice. These issues are revisited in a separate section below.

A few studies have attempted to tackle the challenge of estimating the effects of strong interventions on school performance and in particular on student achievement. Analyzing the reconstitution efforts for a set of high schools in the city of Chicago, Alfred Hess (2003) offers rather glum conclusions:

Reconstitution in Chicago turned out to be a one-time event which was not particularly successful, either in creating significant change in the seven subject schools or in occasioning significant improvement in student achievement. [The observed, modest] improvements in student achievement more reflect the better preparation of entering freshmen than improvement in learning once students are enrolled in these seven high schools. The [district] has sought to improve on the process of reconstitution through two further sanctioning processes, reengineering and intervention, neither of which was any more effective in improving student achievement.

Striking a more positive note, Mac Iver and his colleagues (2003) followed a school on probation, and found positive effects of the efforts made at the school to avert

the school's reconstitution. Notably, the intervention that led to Cooke Middle School in Philadelphia being taken off probation was selected by the school itself. In a secret ballot 90% of the school's staff voted to begin implementation of the Talent Development Model, even before the district required the school to do so. The positive difference that self-selection into specific intervention makes is a recurring theme in the attempts to quantify the effects of interventions on school performance.

Among the available studies of interventions in failing schools, some of which are discussed above, a main theme emerges: looking at individual failing schools redirects the attention almost invariably to the policy environment in which they function. Despite a focus on individual schools, much of any identified failure can be traced back to challenges facing the entire districts, or state, in which they are located. The definition of which schools are identified as failing has often been a function of district capacity, as Ron Wolk (1998) documents in the case of Chicago. The capacity of the district determined the number of schools in which change could be reasonably attempted, which in turn influenced the district's decision for setting the performance standard below which a Chicago school was deemed to be failing and was put on probation. The implementation of Chicago's probation policy itself was examined by Finnigan and O'Day (2003). The authors applaud the district's strategy of offering long-term assistance to low-performing schools, but find that the assistance was less intense and less focused than would have been necessary to spark the hoped-for changes at the relevant schools.

The role of the responsible government agency in overseeing public schools is a topic of enormous importance for failing schools. And while proposals for providing schools with more autonomy often underemphasize the importance of oversight, it is useful to note that close oversight itself may carry hidden costs. In a study of high schools in the United Kingdom, Rosenthal (2004) analyzes the impact that periodic (i.e. roughly every four years) inspections by the Office of Standards in Education (Ofsted) had on test scores. The Ofsted inspections are mandatory for all secondary schools, are meant to examine all aspects of the schools' operations and are, for the inspected schools, a major event. A team of about a dozen inspectors visits the school for a full week. Schools may know up to a year in advance that an inspection will occur and have the time to prepare and assemble the necessary paperwork. While the inspection imposes

considerable costs on the school itself (disruption to the daily operations, preparation time, assembling documentation), the inspections are ultimately designed (and assumed by Ofsted) to lead to improvements in the quality of education that schools offer their students. Their impact on measurable student achievement is therefore a critical question to answer. Rosenthal finds that Ofsted inspections have a small, but significant, adverse effect on test scores for the year of the inspection, with no discernible offsetting effects in later years. The study is an important contribution to the discussion of oversight and the degree to which close inspections impact the schools before a determination of their performance has even been made.

The impact of determining and publicizing a school's performance can also have significant consequences on school operations. Heinrich Mintrop of UCLA (2002) analyzes the effects of putting schools on probation. He surveys teachers in Maryland and Kentucky schools to see how they reacted to the news that their school was officially put on probation. While he finds evidence that the notification "modestly energized" the school staff, there was also considerable antagonism generated among them by being judged by a system they considered unfair and not directly related to their work as educators. Mintrop makes the important observation that the accountability systems in both states, with the school as the designated unit of accountability, operated on the assumption of organizational stability. However, the reality in the affected schools was quite different. These schools educated a large proportion of children that were considered at risk, and faced considerable instability due to high turnover rates among the teachers and the school leadership.

The attempts of a particular district to provide effective interventions for its failing schools stands out. The Philadelphia School District attempted to reform its operations wholesale starting in 1994, under the *Children Achieving* platform introduced by the then newly hired superintendent David Hornbeck. After several years of expending considerable effort and financial resources, the first attempt by an urban district to test systemic reform in practice did not result in the desired improvements and was largely abandoned (Christman and Rhodes, 2002; Corcoran and Christman, 2002). The state took over jurisdiction of the district in December of 2001. Rather than seeking solutions for failing schools through a reform of the entire district, the state handed over

operational responsibility for the lowest performing schools to private contractors, so-called educational management organizations (EMO). Following a controversial assessment report by Edison Schools Inc. (Bradley, 2002), 45 of the city's 265 public schools were turned over to EMOs, with the largest contingent of 20 schools going to Edison itself (Mathews, 2003; Mathews, 2004). It is too early to properly assess the impact of these EMO takeovers, but initial reports suggest that the EMO schools show promising progress on the state exams. Here, as elsewhere, the search for solutions that deliver dramatic, immediate, and sustained improvements in failing schools is likely to be frustrated.

A set of interventions in Chile, which had been hailed for its dramatic success (Tokman, 2002; World Bank, 1999), turned out, upon more rigorous examination, to have been only marginally effective in improving the performance of underperforming schools (Chay et al., 2004). A discrepancy occurred between the results of this study and results of earlier studies because earlier studies did not properly account for the selection mechanism to the interventions and the volatility of test score averages, as studied by Kane and Staiger (2002b). On average, schools that experienced a negative shock to their test scores in the selection year were most likely to be selected for the interventions and, without the same negative shock the following year, appeared to be improving more rapidly than the schools that were not selected. What earlier studies had identified as the interventions' positive effect was thus primarily attributable to mean reversion in the outcome measure among selected schools.

A more positive conclusion regarding the impact of interventions is offered by Michael Stark (1998), who analyzes the United Kingdom's experience with interventions in failing schools, called special measures. Stark finds that the combination of inspections (by Ofsted, as introduced above), the requirement to develop an action plan in *all* schools, and a set of special measures for those schools who had been identified as failing, led failing schools to improve to the point of basic competency in the majority of cases. In a useful extension of the obesity-asthma analogy introduced earlier in this chapter, Stark offers a medical analogy to explain misguided approaches in intervention:

The treatment that a sick person needs to recover is different from the regime that will make an ordinary person fit. Indeed, a fitness regime imposed on invalids

may make them worse. The same seems to apply to schools; competence must precede excellence.

According to Stark, sustained leadership and oversight, which were institutionalized for all failing schools through the Ofsted inspections and the special measures, were the key to restoring them to health.

In the United States, there is no comprehensive system of inspections or a planning requirement for *all* schools, although a similar approach has been used for schools receiving federal funding through Title I when these schools consistently failed to meet their improvement targets and were not part of any other interventions by the state. Though interventions in failing schools in California are considered in detail in Chapter Four, a brief look at the federally mandated audits, especially for schools in the Los Angeles Unified School District (LAUSD), is useful.<sup>21</sup> In the school year 2001/02 the California Department of Education (CDE) placed 13 schools under State Corrective Action for having progressed to Year 3 of Program Improvement status.<sup>22</sup> Ten of these schools were in LAUSD. The state corrective actions were announced publicly and caused considerable shock in the schools, the communities they served, and the district (Ott, 2004). Audit teams were formed by staff from the CDE and the Los Angeles County Office of Education (LACOE). The audits themselves were based on the academic audits developed in Kentucky (Quinn et al., 2003) and resulted in binding recommendations for the schools and the district. There is some evidence that the interventions and the increased attention from the district had a positive effect, with all 10 schools in LAUSD managing to meet their targets in at least one of the past two years. As an illustration of how demanding such interventions can be for the people involved, it should be noted that new principals were assigned to all 10 schools within months of the interventions' start. Of those 10 new principals, 3 have already left their school, voluntarily or involuntarily, for reasons of deteriorating health (Ott, 2004). LAUSD Superintendent Romer points to the importance of supervision and support as the key lesson in the improvement process started within the schools:

---

<sup>21</sup> The interventions discussed here are governed by federal law and should not be confused with the interventions initiated by the state of California, as discussed in Chapter Four.

<sup>22</sup> Year 3 of Program Improvement is the designation for schools receiving federal funding through Title I that have failed to meet their improvement targets in any of the tested subjects for five years in a row.

One of the best illustrations of the benefit, or even the consequence if you don't adequately supervise, is [the ten schools under State Corrective Action]. We got into that circumstance because we failed to perform. ... There are some who would advocate: All you need to do is give schools the money and get out of their way. [The audited schools are] the best evidence that you just can't run a system that way. You just can't give the money to schools and get out of their way. ... We need to be aware that there has been substantial hands-on participation at the district level to bring this cure about. ... Here is real evidence where supervision makes a difference. (KLCS-TV, 2003)

Here Romer is also stressing the sustained oversight and willingness to intervene, similar to the examples Stark provided. What remains implicit in his remarks is that a withdrawal of the district's attention from these ten schools is not necessarily advisable, at least in the short-term.

While conclusions are offered in Chapter Five, there is a preliminary lesson that can be gleaned from the existing literature at this point. The results regarding the impact of interventions in failing schools remain mostly equivocal. Also, strong interventions, by themselves, are not a cure against failure. Instead, as Rudo (2001) states, "policies that incorporate the provision of material and human resources and establish a climate of support and leadership seem to have a much better chance of improving student performance." Above all, consistent and strong leadership at the school itself, as well as at the district or the relevant educational authority, reappears throughout the literature as the key ingredient to successful interventions in failing schools.

### ***2.5.2 A shift in strategy for interventions under NCLB***

The passage of the No Child Left Behind Act (NCLB) of 2001 is ringing in a new era for interventions in failing schools in the United States. It is useful to consider some of the issues that are involved in the implementation of these changes, as they speak to the question this dissertation seeks to address. It is still too early, however, for a proper evaluation of the impact of the interventions that NCLB now requires.

Given that states and districts have shown to be reluctant to intervene in failing schools, NCLB provides an important and dramatic change. NCLB mandates that states must progress steadily from mild, to moderate, to strong interventions, if the performance of the school does not improve. There is some doubt as to whether the legislative requirements are based on assumptions that hold. As Brady (2003) points out, NCLB

implicitly assumes that (a) the states and districts that are forced to intervene truly believe all schools can educate all of their students to high standards, that (b) they have the resources and the skills to add the missing elements to failing schools, and that (c) they have the will to take all of these steps.

Mintrop and Trujillo (2004) note in a comprehensive review of interventions, that “the more stringent corrective action requirements of [NCLB] are likely to create larger intervention burdens for states than many of the previous systems examined.” What the authors call the intervention burden is illustrated by the fact that in California alone, over 1,200 school will be identified as failing in the coming school year and will thus be in need of an intervention (Helfand and Smith, 2004). Already there is considerable debate whether the provisions, as they are set out in the law, can be implemented at all. There are increasing indications that some of the more rigid provisions will be eased (Dillon, 2004b).

The issue of parental choice as an intervention for underperforming schools is particularly contentious. Though early indications by the Bush administration suggested that parental choice would be strictly enforced as an intervention in all schools under Program Improvement (Schemo, 2002), it has become clear that the educational choices that NCLB mandates for parents will simply not exist without a concerted effort by the states (Hassel and Steiner, 2004). Regardless of the specific implementation of the law, the reality of feasible school choices is likely to fall short of the legislative intent. This past school year in LAUSD, for example, only 215 students switched out of a Program Improvement school (Helfand and Rubin, 2004). Significantly, this small number amounts to only about 0.1% of the more than 204,000 students who were eligible to switch schools! A similar pattern can be observed in other urban school districts across the United States. Aside from the considerable shortage of available seats in good schools, a lack of interest on the part of parents has contributed significantly to the weak response. As Tom Loveless, director of the Brown Center on Education Policy at Brookings explains, “the law does give real power to parents. It’s just not a power they are willing to use very often” (quoted in Helfand and Rubin, 2004).

An endorsement of NCLB is offered by the Columbia law professors Liebman and Sabel (2003), who argue that the federal law provides an opening for activists to engage the courts to mandate interventions in failing schools. In their view,

the Act authoritatively defines many of the worst existing disparities as avoidable and, therefore, invidious. The Act thus can trigger just the kind of locally, experientially, and consensually generated standards whose absence in the past has kept courts from carrying through with their initial commitments to desegregated, educationally effective schools.

While this vision is not likely to become relevant to failing schools in the short-term, it suggests that within the next few years interventions in failing schools may be enforced more frequently than they have been in the past through actions of the courts. In the past, reconstitutions or similarly strong interventions in several districts across the country were triggered by court decisions.

Even if the more dramatic aspects of the law are upheld, it seems clear that NCLB's provisions for interventions in failing schools will need some fine tuning in the next few years. As Finn and Hess (2004) remark in a recent paper, "ambitious federal statutes seldom succeed in changing behavior through good intentions or powerful sentiment. When they work, it is because they are feasible, plausible, and make measured use of mandates and incentives." For NCLB, adjustments towards greater feasibility and plausibility will certainly help in the implementation of the law's intent.

For analysts of interventions, the coming years will prove to be uniquely exciting. The evaluation of these interventions, however they may be carried out in practice, will provide an unprecedented opportunity to learn about the relative effectiveness of interventions.

### **2.5.3 Charter schools**

Interventions in failing charter schools are discussed separately from other schools because their organizational structure provides a useful discussion of failure, and because they are quite similar to the two education reform models discussed in detail in this dissertation: New Zealand and Qatar. The written charter allows a more straightforward definition of failure, namely as a failure to comply with the terms of the charter; in other words, a failure to reach the goals outlined therein. In addition to the

accountability to which all schools are subjected, the explicit charter allows the addition of specific criteria that can be weighed against the core goals measured with standardized test scores. How this could be used to set up a better accountability system and a way to detect and deal with school failure is discussed in Chapter Six, in the case of the Qatari Independent Schools. This section deals with charter schools in the United States.

Charter schools can be, and are, more readily closed down in the case of financial or educational mismanagement. However, the experience with charters also suggests that the infrastructure for dealing with failing schools is not as developed as it should be. Similar to the experience of the former transition economies (Kornai et al., 2003), the processes and infrastructures that are required for closing down schools are not easily established.

A report by the Center for Education Reform (2002) summarizes all charter school closures, from the opening of the first charter school in 1992 until October 2002. In total there were 194 closures, which is 6.7% of the 2,874 schools ever given a charter (up to October '02). The most common reasons for the closure of charter schools were budgetary problems, mismanagement, or loss of access to facilities. The report notes that many charter schools were closed due to “manufactured reasons,” because district personnel resistant to change were essentially opposed to the schools. At the same time, the closures themselves are seen as evidence that “the charter concept is succeeding.” Unfortunately, the school closures do not lend themselves to a proper evaluation of the impact of school closure. Aside from the absence of high-quality, longitudinal student data, a large proportion of the schools were closed within months of opening and could not have been considered fully operational schools. In California, for example, the total number of closures of charter schools from 1992 to 2002 is 14, though only 6 of those had been in operation for more than a year.

In August 2004, the California Charter Academy (CCA), an operator of 60 charter schools throughout California, collapsed amidst considerable attention from the media (Blume, 2004; Coleman, 2004; Dillon, 2004a). All 60 schools that it oversaw had to be closed, even though a number of them had been doing well in terms of community acceptance and student achievement. The timing was most unfortunate as over 10,000

students enrolled in the CCA schools had to look for alternative options at other schools only weeks before the start of the new school year, while the teachers and other personnel searched for new jobs. The episode calls into question the view that school closures are a sign of real and desirable accountability and are thus always to be applauded. It also points to the tricky connection between choice and accountability. The CCA schools were closed after the CDE refused to provide the operator with funding for the coming school year, citing the continued failure to comply with financial regulations as the main reason. The schools themselves, however, were doing well and enjoyed the continued support of parents, teachers, and students attending the various CCA schools.

In this charged debate about whether charter schools are being closed for the right reasons, whether they should be handled more firmly, or whether the oversight is too restrictive, Hassel and Batdorff (2004) provide some much needed evidence. They set out to empirically study the accountability questions that charter authorizers (i.e. the organization authorized to grant, but also to revoke charters) face, with an analysis of 50 high-stakes decisions selected at random from a total identified pool of 506 such decisions. Their findings emphasize the importance of building capacity on the part of the authorizers for oversight of charter schools and for dealing with failing schools in particular. While the study finds that authorizers have demonstrated their willingness to close failing schools, they tend to lack the basic systems needed to make an informed decision. Hassel and Batdorff also found that authorizers have learned from past mistakes and those that handle a higher volume of high-stakes decisions are more likely to base these decisions on the schools' true merits.

The importance of building the capacity of charter authorizers is repeated in most reports on charter schools and in some devoted specifically to the capacity of authorizers (California State Auditor, 2002; Hassel and Vergari, 1999; NACSA, 2004). This has been true especially insofar as charter schools have become a very diverse group of public schools. The variety and experimental energy that this implicates is a good thing, though from the authorizers' perspective, oversight is that much harder. This was echoed in the first national evaluation of charter schools in the US (Anderson et al., 2002), the comprehensive evaluation of California charter schools conducted by RAND (Zimmer et al., 2003), and a recent report on Arizona's charter schools (Hassel and Terrell, 2004).

As the attractiveness of charter schools hinges, in part, on the perceived ease with which failing schools can be dealt with, these cautions are important to keep in mind. The experience to date suggests that the proper execution of oversight responsibilities and charter school closures, where necessary, is more difficult than initially hoped. The infrastructure and the expertise for dealing with charter schools, particularly with charter schools that are failing in one way or another, take time and energy to establish. The realization that charters do not offer a quick and easy solution to underperforming public schools notwithstanding, the concept of charter schools is enjoying continued support. A recent study by WestEd, considering the possible reform alternatives for the Los Angeles Unified School District, essentially recommended the establishment of a system of charter schools to supplement and also to challenge the operations of the district in its current form (Koehler et al., 2003). Aside from recommending more choice, autonomy, and school-based management of operations, along with the student weighted formula discussed earlier, the report also emphasized the role of the district's superintendent and school board in providing "real and lasting leadership" for improvements. This emphasis is noteworthy for its recognition that the responsibility for public schools, especially when they are struggling, must ultimately lie with the political authority (educational agency, regional board, or similar institutions). Any reform program that does not account for the problems that have to be addressed by the political authority and its designated agencies is likely to provide incomplete recommendations for improvement.

#### ***2.5.4 Indirect impact of interventions***

The impact of interventions discussed above has referred primarily to the direct impact of interventions on the schools that are subjected to them. However, it should be emphasized that any set of interventions and the manner in which they are implemented has important effects also on schools that are not directly affected by them, as well as on the system of public education as a whole.

The question of the indirect impact of interventions on non-failing schools and on the entire system is a central theme of this dissertation. A direct estimation of such system-wide effects, however, is beyond the scope of this research. Whatever the chosen strategy for interventions in failing schools is, the deliberations here show that the

broader impact of the interventions is potentially the most important effect. In the discussion of the SBC syndrome earlier in this chapter, the parallel to the distribution of foreign aid was raised, as proposed by Klitgaard et al. (2005). Analogous to the authors' treatment of the impact of foreign aid on developing countries, an optimal strategy for intervening in failing schools would take into account (a) the interventions' impact on the failing schools themselves, (b) the incentives for all schools that would arise from any strategy depending on the costs and benefits of being subjected to the interventions (as perceived by the schools), and finally (c) the reputation of the entire system that partially depends on how school failure is handled. The reputation of the entire system would matter, for example, to a legislature considering the budget for public education or to education professionals considering whether or not to pursue a career in public education. The SBC syndrome and the effect of closures that are not actually executed according to the legislative intent speaks primarily to the second impact, namely the effect on organizations that now perceive the threat of closure to be an empty one.

The most prominent, real-life example of the interaction between a strategy for intervening in failing schools and the broader changes of education reform is the system of public education in New Zealand. This example is discussed in the following chapter.

### **3 Reform and Interventions in New Zealand**

Fifteen years ago, the system of public education in New Zealand went through “arguably the most thorough and dramatic transformation of a state system of compulsory education ever undertaken by an industrialized country” (Fiske and Ladd, 2000).<sup>23</sup> The reforms and the emergence of a strategy towards failing schools in the years since provide the ideal case study for the question this dissertation addresses.

#### **3.1 *Why New Zealand Matters***

This chapter’s narrative of the original reforms and the adjustments in the fifteen years since lends real-life evidence to the challenge of defining and responding to school failure.

The question that should be addressed at the outset is why the provided example of New Zealand’s experience is relevant to systems of public education elsewhere. After all, the reforms themselves and the context in which they took place are particular to New Zealand’s history, geography, and cultural heritage. However, despite the elements that make the experience unique to New Zealand, the central themes that emerged amidst the reform that impacted failing schools the most and that are discussed in this chapter, are driven by forces that are common to most, if not all, systems of public education. Public schools are providers of a service that the government is committed to secure for its citizens. The improvement of the delivery of this service is always deemed to be a worthwhile goal and one of the major policy challenges. Just as in New Zealand, public schools elsewhere serve communities that are marked by relative diversity in socio-economic status, in attitudes towards education, in willingness to shape the school that the children attend, and in countless other aspects that matter both for the demand and for the supply of education.

The particular context of the reforms is important in order to understand what happened and why, and to draw the right lessons from the experience. In the sections below, the necessary context and the reforms themselves are presented in a coarse overview, with aspects that were particular to this reform, such as curricular reforms or

---

<sup>23</sup> The only other country to implement a comparably radical education reform to date has been Chile, switching to a national voucher system in 1980. See, for example, Gauri (1998).

the particular challenge of rural schools, generally left aside. Instead, the troubling connection between failure and equity (the inability of market forces to reach those schools which are in most urgent need of being reached) is at the heart of this exploration. The issue of school failure among schools that are exposed to diversity and an uneven distribution of resources and challenges, just as they are in most other systems of public education, is emphasized. As is the importance of thinking the issue of failing schools through to the very end, including the institutions and expectations needed to make strong interventions a realistic option.

The discussion of these central issues that are not particular to New Zealand, but instead particular to the challenges of public education, make this chapter an important contribution to the present inquiry.

### **3.2 *Introduction to the Reforms of 1989 and 1991***

In the briefest introduction, the New Zealand education system looks like this: it serves approximately 740,000 students, who start school when they turn five years old and leave when they're 16 or 17 years old; there are about 2,700 schools, and all but 4% are public schools, many of which are small and serve remote, rural communities (Fancy, 2004). But to understand the development of education reform, a bit of historical background of public education in New Zealand is helpful.

#### **3.2.1 *History of education in New Zealand***

The state school system in New Zealand was formally established by the Education Act in 1877. Over the decades educational governance became increasingly centralized and hierarchical, as it did in other countries. The Department of Education grew in size and in the degree to which it controlled what was going on at individual schools. Primary schools were run through ten regional Boards of Education that had the authority over the schools' budget decisions, including hiring and firing of principals, teachers, and other staff. Secondary schools, originally not part of compulsory education, evolved separately, but in time ceded much of their independence in return for greater financial and other support. For both primary and secondary schools, the real influence over decisions in individual schools rested firmly with the Department of Education in Wellington. A national curriculum was prescribed as a set of curriculum statements, with

teachers given some latitude in how the material should be covered. The Department also negotiated national salary agreements with the national teachers' unions (Connelly, 2003).

**Figure 1 – map of New Zealand**



New Zealand's geography, as well as its particular history, are notable in connection to the education reform. Egalitarianism is an important element of the national character,<sup>24</sup> as are pragmatism and a penchant for innovation. Unlike other former British colonies, New Zealand was not conquered by force. The immigrants from Europe that arrived in the South Pacific from the early 19<sup>th</sup> century onward purchased land from the native Maori (literally translated: non-European; a term used by the native New Zealanders to distinguish themselves from the arriving Europeans). The status of New Zealand as a British colony and of the Maori as citizens of the Crown with equal rights was

formalized in the famous Treaty of Waitangi in 1840. Disputes after the treaty resulted in a series of wars, confiscation of large Maori-owned territories, and resentment among the Maori that continues to this day. The relationship between the Maori and the Pakeha (New Zealanders of European ancestry) has remained complex and at times strained. With more recent immigration from neighboring islands, the third major ethnic group, the Pacific Islanders, now account for roughly 7% of the population, while Pakeha make up 80% and the Maori 15% (Te Tari Tautau, 2004).<sup>25</sup> Though there have been considerable efforts made in recent years to integrate Maori culture better, both the Maori and Pacific Islanders share the plight of racial and ethnic minorities in other developed nations. They lag far behind the Pakeha majority in income, measures of well-being such as health, and – importantly – educational attainment.

<sup>24</sup> For example, New Zealand was the first nation to give women the right to vote, back in 1893.

<sup>25</sup> The percentages add to more than 100%, even if immigrants from other countries are not included here (another 5%), because people of mixed background are sometimes counted twice.

Prior to the reforms, students demonstrated a high standard of achievement when compared with students of a similar age in other countries. In reading ability, New Zealand's students scored consistently close to or at the top *on average* compared to students in other industrialized nations (Elley, 2004). Aside from high average educational attainment, however, the distribution of outcomes was alarmingly wide. While a subgroup of students achieved at very high levels in international comparison, the achievement of others, disproportionately Maori and Pacific Island students, was very low (Elley, 2004; PISA, 2001).<sup>26</sup>

Many parents were involved in the schools their children were attending and were happy with the provided education, even before the reforms began (Wylie, 1999). However, there were also growing concerns in the 1980's that some segments of the population were being left behind and a growing perception that the system of public education had become too bureaucratic, ponderous, and restrictive to individual schools wanting to respond to local conditions and needs. "The erosion of confidence and an undermining of effective cooperation between institutions and consumers" (Picot report, quoted in Connelly, 2005) was noted along with feelings of frustration in the face of a system that was perceived to be inflexible and unresponsive. The perception was reinforced by anecdotes of principals who had to spend days going through bureaucratic hurdles at the Department of Education in order to procure a brush to clean the school bus or an extra set of scissors for the front office (Fiske and Ladd, 2000). Reports from official commissions warned that the outdated structures of the Department of Education were undermining the established quality of teaching and learning.

Against this backdrop of New Zealand's public education in the early 1980's, a series of exogenous events sparked the initiation of dramatic reform.

### **3.2.2 *Public sector management reform***

The plans to reform the way public education was administered was part of a much larger effort to reform public sector management in New Zealand. These broad reforms were triggered by an economic crisis of unprecedented proportions. Up until the

---

<sup>26</sup> The average achievement of New Zealand's students, very high in reading and fairly high in math, has continued since the reforms (Elley, 2004; NCES, 2001; PISA, 2001). So has the skewed distribution of achievement, primarily along ethnic lines.

1970's, solid economic performance and unemployment consistently below 1% was bolstered by preferential trading agreements with the United Kingdom and allowed the support of an extensive welfare system (Fancy, 2004). Just as in the area of public education, the central government had taken on an increasingly direct role in the country's economic activity. By 1984, government expenditure amounted to 40% of GNP (Scott et al., 1997). After the UK joined the European Economic Community, the preferential trade agreements were cancelled. In addition, the worldwide economic downturn after the oil shocks of the 1970's meant less demand for New Zealand's exports and further economic contraction. As the budget deficit grew rapidly, along with unemployment, the urgency for reform became obvious. The government's role in the economy was now seen as a hindrance to structural change. And change was needed.

In 1984, the Labour party came to power and began to liberalize the economy. After Labour's reelection in 1987, "the agenda shifted to restructuring the state. ... By comparison with its all-out plunge into new and bracing waters, the Clinton/Gore attempt to reinvent the US federal government seems like a timid toe gingerly dipped into a tepid pool" (Nagel, 1997).

As Scott, Ball, and Dale, former leaders in the Treasury and co-architects of the reforms, write, "the purpose of these changes was to get the government out of activities it was inherently poor at managing and to improve those functions which remained the core responsibilities of government" (Scott et al., 1997). The reforms were remarkable in their scope and also insofar as the implementation throughout the government's activities closely followed the motivating theoretical framework. In part, this was due to the fact that these reforms were launched by the Labour party, who traditionally had taken the role of opposing a reduction of the government's role in the delivery of social services or economic activity. But, it was also due to the fact that in the middle of a huge crisis, a group of economists and public servants provided a coherent vision of what the appropriate response should be and how the government should do business more generally (Fiske and Ladd, 2000). When the National Party came back into power after the 1990 elections, the reforms were continued along the same lines and sometimes – as in the case of education – further deepened. The goal of the overall reforms was to increase the accountability and the efficiency of the government's activities, but also to

eliminate the budget deficit. A core set of concepts was applied in a reasonably uniform way across the provision of public services (Scott et al., 1997). The concepts included the separation of funding, purchasing, and provision of public services, the introduction of competition between service providers, and the implementation of private sector governance arrangements to strengthen incentives for efficiency of resource use.

As part of the restructuring of the government, the purpose of individual departments or government agencies was restated, turning them into providers of services to the minister requesting the delivery of a particular service. In the new framework, the minister, acting as the customer, has to set up a Purchase Agreement with the department or agency supplying the service prior to the fiscal cycle. In this agreement, the services that are to be delivered or produced have to be specified. At the end of a fiscal cycle, the department or agency delivers along with its financial statements (balance sheet, profit and loss statement, cash flow statement) a Statement of Service Performance, in which the services delivered are outlined and can be compared to the ex ante Purchase Agreement. According to Scott et al. (1997), the rationale for this reporting is to better measure the performance of the public service provider.

In the private sector the customer is in the best position to monitor the delivery of services – that is, monitor performance against the (implicit or explicit) contract. However, in the public sector the recipient of the service is frequently not the purchaser. Therefore, the recipient is less able to signal dissatisfaction through actual or threatened withdrawal of custom or legal action in relation to the existing contract. The Statement of Service Performance is an administrative attempt to substitute for the incentives associated with customer performance monitoring. For the Statement of Service Performance to have this effect, it is critical that there be an ex ante [Purchase Agreement] against which ex post performance can be measured.

This framework for performance measurement is emphasized here because it was a central concept to the overall reforms and contains an important idea. The argument of Scott et al. for having a Purchase Agreement, as well as the Statement of Service Performance, is that the recipient of the service and the purchaser are often not the same. Thus, the objections of the recipients of the service or their switch to another provider may not be as effective as if they were the purchasers themselves. In New Zealand's public education the recipients are really the children (arguably represented by their parents), whereas the purchaser is still the minister, representing the country. The

framework allows for the fact that some of the assumptions under which competitive markets make sense do not hold when it comes to the provision of public services. This framework for measuring performance has important implications for the notion of school failure. In this framework it is defined as the failure to demonstrate that the delivery of education (documented in the Statement of Service Performance) matches up with the education promised in the Purchase Agreement.

This, at least, was what the accountability framework for public education could have looked like. As discussed below, this framework from the broader reforms throughout New Zealand's government activities was not fully implemented in education. Though there is "an overwhelming view that the Treasury was one of the main drivers, if not the main driver," of the education reform (Young, 2003), this part of the puzzle was left out. The failure to carry this piece forward had important repercussions for the system of public education.

### 3.2.3 *Tomorrow's Schools*

The term Tomorrow's Schools was adopted from the title of the policy paper that outlined the planned reform of public education in the fall of 1988. Tomorrow's Schools refers here to the set of structural changes in the system of public education that were enacted by Parliament in 1989 and expanded upon in 1991. The central feature of the reform was the transfer of operational responsibility for each of the 2,700 primary and secondary schools to local Boards of every school, the introduction of full parental choice, and the encouragement of the development of a competitive culture in the state education system (Fiske and Ladd, 2000).<sup>27</sup>

The original design of the reforms was established by the Picot taskforce, named after its head, the businessman Brian Picot. The taskforce's recommendations were largely approved and fully incorporated into the subsequent policy paper *Tomorrow's Schools*. The taskforce concluded that the Department of Education was beyond repair. Instead of reform from within, they recommended the abolishment of the Department and

---

<sup>27</sup> These sweeping reforms were not enacted as a coherent whole, nor was the integration of market forces a driving force behind the initial design of Tomorrow's Schools in 1988 (Fancy, 2004). Yet, in regards to the management of schools and the issue of failure among individual schools, the simplified view presented here is more useful to analyze the experience with, and the adjustments to, school failure. Other aspects of the reforms, such as the important changes in the national curriculum, are not a focus.

of the regional Boards of Education that had been responsible for operating schools. Responsibility for operating schools was transferred to parent-elected School Boards of Trustees. These Boards became the legal employers of teachers, appointed principals and were made responsible for the overall management and running of public schools (Fancy, 2004).<sup>28</sup> The Department was completely replaced by a much smaller and much different Ministry of Education, whose primary responsibility was the provision of policy advice to the minister of education. The responsibility of operating the schools shifted from the central government to the Trustees. As the Picot taskforce's report declared at the outset: "Individual learning institutions will be the basic unit of education administration" (quoted in Fiske and Ladd, 2000).

A few short months after the release of the Picot report and the subsequent policy paper *Tomorrow's Schools*, legislation was passed by parliament to institute the majority of the recommendations. David Lange, Prime Minister and Minister of Education at the time, said it was "time for parents to take charge of their schools" and communities to define what they wanted their schools to do (quoted in Connelly, 2003).

Two years later, after the National Party had won the 1990 elections, the reforms were expanded. Enrollment zones were abolished for all schools. Schools that had more applications than capacity to accept students were free to set their own enrollment schemes or admission policies. This meant that children were no longer guaranteed a place at the nearest school in cases where the nearest school had more applicants than capacity.<sup>29</sup> Parents were encouraged to select the best educational situation for their children, and schools were encouraged to become competitive with one another in order to attract students to their school (Fiske and Ladd, 2000).

The details of the funding mechanism for schools have been changing somewhat and continue to be debated. Essentially, a school's Board or principal could select to have its teachers paid directly through the Ministry, or instead opt for bulk funding,<sup>30</sup> receiving all the money in one portion to be used in whichever way suited the school best. Bulk

---

<sup>28</sup> The role of Boards of Trustees is more directly involved in oversight and management than school boards typically are in the United States.

<sup>29</sup> It should be noted that many schools who received more applications than they had capacity gave preference to children living in the neighborhood (Connelly, 2005).

<sup>30</sup> The option for bulk funding was terminated in 2000.

funding was the initial intent of the reform, but the teachers' union objected, which led to the compromise of giving schools an option (Fancy, 2004). In either case, the school retained the freedom to dispense of other operational funds, for example by hiring additional teachers, or organizing maintenance and operational expenses in a better way. Schools have also been allowed to raise money in the community, within limits. The majority of the funding, however, schools receive through the Ministry on a per pupil basis. Since 1996, all schools are ranked by deciles according to the average socio-economic status (SES) of the student body. Decile 1 is for the student body with the lowest SES on average, and 10 is for the highest. Schools in the lower deciles, serving disproportionate numbers of students who are low-SES, and hence on average more costly to educate, receive additional funding compared to the schools in higher deciles. The funding is thus far more equitable than in the United States.

#### ***3.2.4 Design of the accountability framework***

Prior to the reforms, the Department of Education controlled the allocation of funds and used a cadre of inspectors to oversee, inspect, and work with individual schools. The inspectors offered advice and assistance to the schools they visited. They had considerable authority (e.g., they could get people fired, among other things), and they also were responsible for communicating official criticism (Fiske and Ladd, 2000). In doing so, they were asked to report on the performance that they helped to establish. This conflict of interest for the inspectors was seen by the reformers to reduce the integrity of the accountability framework. The new accountability system that was put into place with Tomorrow's Schools, changed this dynamic fundamentally.

The Education Review Office (ERO) was created to replace the system of inspectors from the Department of Education, who provided assistance as well as oversight. The ERO had been part of the Picot taskforce's recommendations, though under a different name, and is designed to be an auditing body largely independent of the Ministry and independent of any assistance that schools receive. The ERO is tasked with reviewing and reporting on the operations and performance of every school in the country (including private schools and early childhood education centers, though these are not

discussed further here).<sup>31</sup> ERO teams consisting of one to five members visit classes, review school documents, discuss a draft report with the school, and then issue a final, public report that lays out the school's strengths, as well as the deficiencies that need to be addressed. The focus of the reviews was initially on the school's compliance with legal requirements, but has more recently shifted towards educational outcomes. The purpose of the reports is to inform the school of the status of its own operations (so it may build on its strengths and improve on its weaknesses), to inform parents (so that they may respond to the performance of the school), and to alert the Ministry in case there are any grievous problems.

The regular cycle for ERO review is generally once every three years. If an ERO team feels there are substantial changes needed at the school, the team returns for a supplemental review within twelve months of the initial review to check on how the identified shortcomings have been addressed. In contrast to the inspectors prior to the reforms, the ERO teams are not permitted to become involved in the schools, so that their judgments remain as impartial as possible. ERO reports do not mandate that a particular intervention be taken, but rather outline the problems and possible solutions, while leaving it to the Ministry of Education to decide whether a particular intervention is warranted. Due to their public nature, the ERO inspections quickly evolved into a crucial event for all schools. The public reports resulting from the inspections do much to shape the perception of the school and the school's ability to attract parents who might consider sending their children there. In addition, the ERO also produces National Education Evaluation Reports on broader educational issues, based on the information compiled through the reviews of individual schools. The aim of the broader reports is to disseminate information about best practices and also to evaluate the state of education in clusters of schools, or nationally. The ERO reports on school clusters turned out to be quite influential in changing Ministry policies towards failing schools.

What makes the ERO's job especially challenging is the absence of quantitative performance data. Though students take exams at the end of high school, a notable

---

<sup>31</sup> The information in this and the following paragraph on the Education Review Office, its reports and the National Education Evaluation Reports, is from the ERO's Annual Report (2003a), the ERO's Statement of Intent (2003c), and the ERO's Evaluation Indicators for Education Reviews in Schools (2003b).

characteristic of the New Zealand school system has been the absence of a comprehensive system of national tests of student performance for students under age fifteen. New Zealand teachers are wary about the use of tests for reasons other than to organize classes or to diagnose the needs of particular students. A heavy reliance on student outcome data, such as is common in the United States, for example, is met with considerable concern and opposition among educators. In New Zealand's public schools, there is little tradition of aggregating test data or keeping them on a schoolwide, much less a national, basis. In conversation and informal presentations, ERO leaders have expressed their frustration with the absence of any objective performance measures that could supplement the qualitative judgments summarized in the reports of the inspection teams. While student achievement has to be addressed in the ERO reports, it is not a major focus and the schools are merely asked to have *some* system in place to collect, manage, and analyze the data (ERO, 2003b). At least in part as a consequence of the lack of data, there is no explicit definition of school failure for ERO, nor an attempt to measure failure (Woodfield and Gunby, 2003) other than the supplemental reviews. It has been argued that the supplemental reviews themselves, or discretionary assurance audits, are a definition of failure, as about 12% to 15% of all schools are affected by such additional scrutiny by ERO review teams (Connelly, 1999). Aside from recent efforts in the Ministry, there has been no clear definition of what performance level warrants either closer scrutiny or an intervention. The Ministry's attempts to remedy this situation are discussed later in this chapter. The implicit handling of failing schools in the original reform is discussed in the following section.

### 3.2.5 *The reform's conception of failing schools*

The reform was not explicit in its intention for failing schools,<sup>32</sup> though the need for raising achievement among underachieving students was clearly a concern. From the reviewed documents and commentaries, two implicit expectations of Tomorrow's Schools, as they relate to school failure, can be construed. The first goal was that inefficiencies in schools, and especially outright school failure, would be identified earlier and more clearly, and that such failures would be corrected. The correction would

---

<sup>32</sup> There's no mention of specific failing (or successful) schools in the Picot report or in the subsequent policy paper *Tomorrow's Schools* (Woodfield and Gunby, 2003).

be spurred by the market forces of unrestricted parental choice. Either failing schools would respond to the competitive pressures and improve themselves or the situation would be remedied through the exit of the failing schools from the system. In either case, the correction was supposed to happen primarily without government involvement. The accountability framework, the self-management of schools, and the encouragement of a competitive culture were geared towards this goal.

The second goal was that the new system of public education would serve the currently underachieving subgroup of the student population better. By allowing full parental choice, and greater parental involvement in schools, those subgroups who were not succeeding in the traditional system of public education would no longer be trapped in their bad schools. Parents could influence the school their children were attending or they could opt for a school of their own choosing, which would – on the whole – allow a leveling of the playing field and an eventual improvement of their educational attainment.

These hopes for the way in which Tomorrow's Schools would impact school failure are restated here because they are remarkably similar to the expectations of school reform proposals in the United States and elsewhere. The way in which these hopes were, or were not, fulfilled in the past 15 years contains valuable lessons for other systems of public education outside of New Zealand.

### ***3.3 Design Meets Reality***

For all the dramatic change Tomorrow's Schools imposed in such a short period of time, there were considerable successes to note. As trustees, principals, and teachers responded to the significant increase in the workload that Tomorrow's Schools required of them, the reforms were greeted with a mixture of caution and interest. Despite the wish of many to hold on to what they already had, rather than make changes, education practitioners believed that the reforms brought improvement to the school system. The vast majority believed that the pre-reform institutions should not be brought back (Wylie, 1999). The general consensus among professional and academic commentators has been that the 15,000 lay volunteers who govern New Zealand schools at any point in time are doing at least as well as the national and regional level officials who came before them

(Robinson and Timperley, 2004). As it pertains to failing schools, however, the reforms did not result in the realization of the hopes stated earlier.

The implementation of policies is always a messy business. Policies might fall short of their expectations for two reasons: on the one hand, it may prove difficult to put the policies into place the way they were designed. Even if the design was excellent, the policies can only work in the way their implementation is actually carried out. On the other hand, the assumptions that the policy design was based on may not fully hold, rendering the policies as they are implemented not as effective as they were expected to be. Even if the implementation was perfect, the policies might have been based on wrong assumptions, rendering the perfectly implemented policies quite imperfect.

As Secretary of Education Howard Fancy (2004) has recently pointed out,

while experience has proved different from what might have been originally envisaged and implementation more challenging than anticipated, it is important to record that the overall direction of the changes has not been challenged.

The next two sections concentrate on the two ways in which Tomorrow's Schools fell short of expectations: the failure to implement a coherent design for the schools' charters, and the erroneous assumptions upon which the hands-off policies of the Ministry of Education towards failing schools was based.

### ***3.3.1 Demise of the charter***

The recommendations of the Picot taskforce were in spirit quite similar to the performance measurement framework of the public sector management reform, as outlined by Scott et al. Specifically, the recommendations called for each school to propose a charter which would “act as a contract between the community and the institution, and the institution and the state.” The taskforce described these charters as the “lynchpin” of the new structure of compulsory education. Charters would provide each school with “clear and specific aims and objectives, expressed as outcomes” reflecting “both national requirements and local needs” (quoted in Fiske and Ladd, 2000).

Even though the parallel was not made explicit, the charter's description recalls the Purchase Agreement discussed earlier.<sup>33</sup> The charter was to be the individualized *ex ante* contract between the minister of education, the individual schools, and the community which they serve. The policy paper *Tomorrow's Schools* adopted the language of the Picot report and the prescription for school charters as a centerpiece of the accountability framework almost verbatim (Fiske and Ladd, 2000).

During the first few months of implementation, however, the concept of the charter was substantially weakened. First, it was turned from a three-way agreement between the Ministry, schools, and communities into a two-way agreement between the Ministry and the schools, thereby weakening the communities' role in formulating the objectives of the school. The change might have enabled a clarification of the bilateral contractual relationship between the Ministry and the schools. But instead of providing such clarification and guidelines for what charters should or should not include, the two-way contract was replaced by a one-way *undertaking* on the part of the schools' Boards of Trustees, which amounted to a contract that only the schools were legally obligated to adhere to, but not the Ministry (Fiske and Ladd, 2000). Finally, after opposition to this one-way obligation arose, the charters were effectively reduced to include only the centralized requirements on all schools, called the National Administration Guidelines, and were thus stripped of their role as school-specific contracts (Connelly, 2005).

The leadership in many schools (trustees and principals) across the country had expended considerable energy since June 1989 writing charters, while trying to understand and navigate the changing regulatory requirements. In the end, the charters were drawn up, but aside from the benefits to schools from the process itself (e.g. discussing mission statements, and coherent goals for the school),<sup>34</sup> the charters were not terribly meaningful as contracts.

---

<sup>33</sup> There is a subtle difference between the early conception of charters and the Purchase Agreements: whereas Purchase Agreements specify the outputs (i.e. the observable products and services produced) charters were to specify the outcomes (i.e. the goals) that are influenced by the outputs (Connelly, 2005).

<sup>34</sup> According to education leaders, this process was a useful introduction to the concept of self-managed schools, frustrations about changing regulation notwithstanding (Kilpin, 2004; Leslie, 2004; Mullins, 2004; Short, 2004).

The concept of the charter was never fully developed in a way that could be directly used. Its potential as an equivalent to a Purchase Agreement against which later performances could be evaluated was not communicated clearly. The charter wasn't consistently understood by those who were tasked with writing the charters (e.g., how much latitude each school had in writing them), nor by those in the Ministry who were supposed to evaluate and approve them. As a consequence, the effectiveness or the fairness of the reports issued by the ERO on individual schools was reduced. Without charters as actual contracts, there was no clear statement, *ex ante*, against which the actual performance of the school could be compared. The ERO teams could still use the National Education Guidelines and the National Administration Guidelines (Ministry of Education, 2004), which apply to all schools but are too broad to be effective in this regard. There was no way for the ERO to report on the degree to which the schools met their own "clear and specific aims and objectives."

Thus, school failure became, without the existence of a Purchase Agreement, a far trickier concept than it otherwise might have been. The differences of starting points for individual schools and the autonomy that schools enjoy are hard to account for in a consistent and fair manner when there is no school-specific *ex ante* agreement.

The ERO process was initially not welcomed by teachers and school administrators. The reports were often seen as unfair and the reviews as not being in touch with what the individual school was about (Fiske and Ladd, 2000). The discontent in the early years is partially a function of the change from the pre-reform inspectors that were given a far more collegial charge in reviewing *and* assisting the schools to a system in which inspectors were prohibited from offering direct assistance and were instead only at the school to write a critique of its operations. But a part of the discontent appears to have been a result of the process that lacked more individualized charters. More recently, however, the ERO process has adapted and has also started to be welcomed by principals, Boards, and teachers (Wylie, 1999).

Even though the charter had the potential to become the equivalent of a Purchase Agreement between the school, the community and the Ministry, this piece was largely lost during the implementation of Tomorrow's Schools. A second shortcoming in regards

to failing schools, not of the implementation, but rather of the design itself, is the initial strategy, or lack thereof, toward school failure.

### 3.3.2 *Erroneous assumptions*

The design of Tomorrow's Schools called for self-managing schools that would function independently from one another, assessed by ERO reports, funded by the Ministry, and held accountable primarily by the students' parents. If and when parents were not satisfied with the performance of the school, they could either intervene through the Board of Trustees, or else take advantage of the freedom that parental choice afforded them and send their children elsewhere for their education.

In case of gross misconduct or blatant incompetence on the part of the Board of Trustees, the minister could intervene in a school (Sinclair, 2001). The assumption was that, outside of such extreme circumstances, interventions were not something that the Ministry should be concerned with. When schools in decline would cease to be viable institutions, it was assumed they would simply close and their demise would be chalked up to the proper functioning of an educational marketplace. Specifically, the assumption about failing schools seems to have been, analogous to the discussion in Chapter Two, that in a system of public education, reformed along the lines of Tomorrow's Schools, failing schools would do one of the following:

- a) Improve. Compelled by competitive pressures, failing schools work harder and more efficiently in order to stave off collapse, which results in a return to a situation where the schools are providing education similar to the average education provided in the system.
- b) Close. If the schools are unable and/or unwilling to improve, they succumb to competitive pressures, lose their students and are forced to close.

The following sections illustrate that this central assumption about failing schools turned out to be wrong. Fiske and Ladd (2000) do not hesitate to point to the source of this assumption that they feel drove the lack of strategy of the government towards failing schools.

Treasury officials appear to be the most ... unsympathetic to the plight of struggling schools. If schools cannot compete effectively in the new educational

marketplace, there must be a management problem; any attempt to solve the problem by buttressing the school could prove counterproductive in the long run by promoting inefficiency in the provision of education.

It is true that the overall public sector management reforms in New Zealand were driven primarily by economists and policy analysts in the Treasury and the discussion of Hirschman in the previous chapter has indicated that an ingrained lack of concern with failure of individual providers might be expected in a design based primarily on economic principles. As the previous chapter's discussions have shown, the assumption that declining schools *should* go out of business, because their loss of productivity and of competitiveness must be happening for a good reason, does not always hold (Hirschman, 1970). Nor is it the most useful assumption, especially where providers of public services are concerned. Furthermore, for those schools where it may be assumed that they *should* be closed because they have become (and will remain) nonviable institutions, it is very far from automatic that they *will* close, even in a system of full parental choice.

However, it is also useful to recall the accountability framework suggested by Scott and his colleagues from the Treasury, presented earlier in this chapter. The elements necessary for an accountability framework to make the decentralized, competitive provision of public services work (Purchase Agreement and the Statement of Service Performance), as prescribed for the overall public sector management reform outside of education, were not implemented for Tomorrow's Schools. These crucial accountability mechanisms were left out during the implementation. The fact that the failure of individual public schools was left unattended is in part a result of this omission.

The combination of erroneous assumptions about school decline and failure, and an accountability mechanism that was not fully adequate for the situation, resulted in an incomplete strategy towards school failure.

### **3.3.3 Equity issues**

A few key issues that schools in New Zealand struggled with (and still do), that were insufficiently addressed in the initial reforms, are pointed out here. This has resulted in those subgroups of the population that were supposed to benefit the most from the reforms actually benefiting the least (Ladd and Fiske, 2001; Wylie, 1998).

Under Tomorrow's Schools, the responsibility to operate schools was passed on to the local communities. To organize the elections for the trustees, to train them, and to provide ongoing support, the School Trustees Association was formed and funded by the government. The association delivered training and support to all Boards across the country. Even with the provided training and support, the degree to which the communities are able to properly carry out this new function depends on the capacity of the people that are available and interested to serve on the Boards of Trustees. As the Boards have wide ranging responsibilities not only for curriculum delivery, but also for matters relating to personnel, property, and finances, any significant differences in the performance of the Board will affect the school.

Across all communities and newly formed Boards, the oversight over the operation of a school and having the responsibility for its proper functioning was something that all Boards found difficult to learn. In 1994, an ERO report (1994a) noted that the trustees, by and large, take their responsibility seriously. "In practice, they do not always know the extent of their responsibilities. Sometimes they make mistakes; sometimes they end up in Court. Most, however, manage their governance role with varying degrees of effectiveness." In regards to a more focused oversight role, including periodic self-review of the school and the appraisal of the principal, an ERO report in the same year was less positive.

Boards of Trustees and professional staff in schools have not yet accepted their responsibility to monitor, record, and report student achievement ... At the present time most school self-review information is not sufficiently comprehensive to provide a reliable base for external investigation. (ERO, 1994b)

Board capacity was particularly limited for schools that served the more disadvantaged students (i.e., lower decile schools) and a higher proportion of Maori and Pacific Islander students. Communities of low socio-economic status, which are often grouped along ethnic lines in New Zealand, found it more difficult to find parents among them who already had the experience, education, confidence, and interest that effective Board members would need. In many instances, the principal was left to instruct the Board about their responsibilities, rather than the other way around, even though principals are officially employed by the Boards (Fiske and Ladd, 2000).

One mechanism that was thought to deal with the differential capacity for good governance across schools was competition. But, in practice, competition does not always work the way that theory predicts. For New Zealand, there is some evidence that the degree to which elementary schools saw themselves in a competitive environment had negative impacts on the quality of student learning and other educational outcomes (Ladd and Fiske, 2001). What is more important, for the consideration of failing schools, is that competition by itself does not work evenly across all schools. Indeed competition may be least effective for schools that are failing, as the theoretical and empirical discussions in Chapter Two indicated. For Tomorrow's Schools, those schools which have been most likely to be serving economically disadvantaged students have been the very schools which have benefited the least in terms of student roll increases, stability, and lower student turnover within the school year (Wylie, 1999).

Schools that are at capacity rise above the competitive fray, while schools that are struggling clearly feel the pressure, but find it very difficult to improve on their own. The educational challenges they face are huge. As it has turned out, parental choice and competition, by themselves, are not the solution for the problems of the schools at the bottom.

#### ***3.3.4 The problem emerges***

A few years after the reforms, it became increasingly clear that the Ministry's approach towards failing schools could not be reconciled with the reform's goals with regards to equity. The lack of positive impact that the competitive pressures were having on the low-performing schools could no longer be ignored.

A series of ERO reports proved to be influential in triggering this realization. The 1996 ERO report on the schools in Mangere and Otara in South Auckland, two of the poorest communities in New Zealand, pointed out that a cluster of schools had failed to improve, even after a series of negative ERO reports and after repeated supplemental reviews by the ERO teams. The ERO chief review officer at the time, Judith Aitken, convinced her agency to make a public statement that would embarrass the Ministry of Education into action (Fiske and Ladd, 2000). The August 1996 report in effect described a cluster of schools in South Auckland as a disaster area characterized by rampant failure

in governance and management. As well as pointing to governance failures among the schools, the report called the appropriateness of the lay trustee model itself into question. The model seemed ill-suited to communities where it was difficult to elect or co-opt sufficient financial and professional expertise, and where cultural norms conflicted with the requirement that Boards act as the employer and appraiser of the principal. In some of the schools in question, the Boards gave principals both more discretion and less assistance than was intended by the legislative framework (ERO, 1996; Robinson and Timperley, 2004). Later reports on school clusters on the East Coast and in the Far North confirmed that Mangere and Otara were not isolated cases, but rather the most visible symptoms of fundamental difficulties in public education in New Zealand (ERO, 1997; ERO, 1998).

The reports were devastating to the reformers because school failure had been officially identified for several years running, and even though the negative reports were public and available to all parents, teachers, and the Ministry, nothing was happening that suggested that a correction of the untenable situation was underway. In Hirschman's language, the reasons for failure appeared to be both compelling and durable and any kind of counterforces to the school failure were woefully absent. Notwithstanding the competition from other schools to which these failing schools were said to be subject, the failing schools did not go out of business, nor did they improve. The schools continued to operate in the system, thereby providing evidence that the earlier stated assumption about failing schools was not accurate. It is worth noting that the schools in question were situated in the agglomeration of the large metropolitan area of Auckland, New Zealand's largest city. Thus, the failure of schools in Mangere and Otara could not be blamed on the particular geography in which these schools found themselves.

The system of public education was not supposed to deal with the persistent failures that the ERO reports had identified. Their existence questioned the assumptions of the reform and the formation of a strategy for dealing with them seemed to amount to a partial surrender of the hopes of the market forces' power. According to Bill Gavin, the principal at Otahuhu College in South Auckland (quoted in Fiske and Ladd, 2000), the Ministry was initially unrelenting: "The attitude was 'if they sink, they sink.' But now

there are too many sinking schools to ignore. You're talking about 10 percent of all schools and even some entire regions.”

The problem, raised in Chapter Two, is what ‘sinking’ really means within a system of government funded education. If it means neither exit of the school from the system, nor substantial improvement subsequent to the mobilization of counterforces, but instead the continued provision of education of an unacceptably low quality, then the government should realize its responsibility and intervene. In this context, the failure of individual schools implicates the entire school system for not performing as prescribed. These failing schools are not part of Schumpeter’s creative destruction because, left to their own devices, they will continue to operate. As Fiske and Ladd (2000) put it:

One might justify the deliberate creation of relative failures if competition served to enhance the overall quality of all schools, in effect raising the tide for all boats. Or such a policy might be defended if some sort of safety net were in place to catch the expected losers early on and to take the steps necessary to make them into viable schools. Neither condition appears to have been present in New Zealand.

A strong answer is provided by Woodfield and Gundy (2003), who review Fiske and Ladd’s work and do not agree with their conclusion that the reforms have been responsible for making the schools’ problems worse than they had been previously. Instead, they argue that the reforms’ “main impact was to make the problems more transparent, creating a great deal of (discomforting for some) public pressure in the process.” The lively discussion on either side of the argument does not fundamentally alter the contention that failing schools were not adequately planned for, regardless of whether the reforms made the failure worse or simply more visible. This omission in the overall reform strategy matters and the response of the Ministry is remarkable for its pragmatism and willingness to part with previous assumptions about school failure.

### **3.4 Adjustments**

When the reality of consistently failing schools could no longer be ignored, the need for a strategy including interventions directed by the Ministry became urgent. The Ministry began, reluctantly at first, and recently more deliberately, to put a strategy in place for intervening in failing schools. There have also been legal changes that have

partially corrected for some of the shortcuts taken during the implementation of Tomorrow's Schools.

### 3.4.1 Interventions

The initiation and the growth of intervention programs by the New Zealand Ministry of Education is an important experience to consider when evaluating approaches to failing schools. Tomorrow's Schools imposed a radically new governance mechanism on the entire system of public education; a governance mechanism that has been accepted by the education practitioners and has endured the test of the past 15 years. Embedded in the policy environment that was initially set up for the schools, the communities, and especially the Ministry, were a set of assumptions that turned out to be wrong. The adjustment over the past 15 years, driven by New Zealand's emphasis on egalitarianism, pragmatism, and innovation, is just as noteworthy as the original reforms.

After 1989, when the governance of schools passed from the Department of Education to the Boards of Trustees, the Ministry of Education had only two options for intervening in individual schools. The two statutory interventions were (a) to dissolve the Board of Trustees and replace it with a commissioner appointed by the minister, or (b) to initiate legal proceedings against the Board (Fiske and Ladd, 2000). The radical nature of these interventions, with the Ministry acting against the school's Board and thus against the representatives of the community which the school is serving, meant that they could be invoked only in the most dire circumstances. Any form of *support* from the Ministry was not envisioned. In effect, the Ministry had limited itself to intervening in schools after they had reached a point of unequivocal and abject failure. Before then, the Ministry would not get involved.<sup>35</sup>

After five years of this non-interventionist strategy, the Ministry of Education initiated a Schools Support Project<sup>36</sup> to establish a range of support strategies for schools that were unable to handle the new policy environment by themselves. It was expected

---

<sup>35</sup> It should be noted that decentralized support services, such as teacher training, were made available to schools from the beginning of the reform (Connelly, 2005). Schools could take the initiative themselves and get support and advice (some of which was provided at no cost to the schools), if they wanted.

<sup>36</sup> The program for interventions in the Ministry of Education was initially called *Schools Support Project*, later renamed *Schools Support*, and more recently renamed *Schools Monitoring and Support*. All three terms are used here, depending on the timeframe in question.

that the strategies would allow for interventions earlier in a school's struggles; earlier than a situation that would call for the full dissolution of the Board, for example. The interventions were to provide support for the resolution of difficulties that could impact students negatively, provide support for the development of self-management capabilities in specific communities, and provide support for sustained change and improvement in schools (McCauley and Roddick, 2001). There are two types of interventions pursued by Schools Support: first, Safety Net interventions for individual schools at various stages of decline and/or failure, second, Schooling Improvement strategies for clusters of schools that are consistently underperforming and are in need of a more sustained and collaborative process in order to improve their performance.

The Safety Net strategy is made up of four sets of interventions of increasing severity: Informal Action, Formal Action, Business Case, and Statutory Action.<sup>37</sup> The general idea is that the escalation to the next (more severe) intervention happens only when the prior intervention has been shown not to work. The three less severe interventions were added to support Statutory Action, the only intervention at the Ministry's disposal from the beginning. This was done because it became clear that 'letting the market work' until the situation was so bad that Board replacement or other statutory action as warranted was not the most effective or efficient way to deal with schools in trouble (Sinclair, 2001).

In fact, the interventions were at times supported by direct cost-benefit analyses. For example, plans were completed for two declining schools in South Auckland after a consulting firm advised the Ministry that it would not only be more politically sound, but also less expensive to put additional resources into the schools than it would be to close them down (Fiske and Ladd, 2000). As a result multi-million dollar intervention programs were targeted for these schools to help them improve over a few years span.

The Schooling Improvement Strategy was initiated in 1996, a year later than the Safety Net interventions. These interventions are not so much geared towards individual, failing schools, but are geared towards supporting clusters of schools and building

---

<sup>37</sup> The details of the strategy described here have been amended in recent years according to the legal framework discussed below. The key point is the rise of a more flexible strategy for interventions.

capacity across the cluster and between the schools and their communities. Specifically, these interventions were launched after the publication of the ERO reports on schools in South Auckland and later reports on clusters of schools on the East Coast and in the Far North. Both types of interventions began modestly, but have grown considerably in recent years. By the time the comprehensive evaluation of Schools Support was published by the Ministry in 2001, some 300 schools had taken part in one of 16 Schooling Improvement initiatives. For Safety Net interventions, the number is just shy of 250 schools as of October 2000. In all, roughly 10% of all schools have been in Schooling Improvement initiatives and another 10% in Safety Net interventions (McCauley and Roddick, 2001). By 2001, 60% of the funding for interventions went to school clusters and 40% to safety net interventions. This allocation of funds is an indication that the intention of those involved in Schools Support is to become more proactive and preventative with respect to support of struggling schools (Sinclair, 2001).

According to Mary Sinclair, the Manager for the Monitoring and Support Group, the growth and further consolidation of the group's activities is based on the following assumptions (Sinclair, 2001):

- The population of New Zealand is not homogenous and not all communities are equally capable (at all times) of assuming the stewardship responsibilities envisaged in the education changes made in 1989;
- It is desirable that communities and schools that are capable of being self-managing are provided with the opportunity to do so;
- Schools sited in lower socio-economic communities, where there are higher proportions of Maori and Pacific Islands students, are more likely to require external support to achieve community, school and Government expectations;
- The dissolution of boards of trustees and the public exposure of failing schools does not, on its own, necessarily result in improved student achievement or school performance;
- It is preferable for Government to provide for an integrated and efficient strategy to support schools where there are potential risks and also where risks are apparent rather than 'sit back' and 'wait' for failure to occur;
- Nowhere in the world has a 'formula' been developed that can be introduced to communities and schools to 'turn them around' as a 'quick fix';
- Interventions in individual schools and clusters of schools must be designed to meet the particular needs of the school/s;

- We need to continuously review, research and evaluate our New Zealand strategy for monitoring and support in order to learn what works well and is successful for different types of students and schools;
- Change cannot be effected ‘overnight’. The desired longer-term outcomes will require a focus on relationship management and task focus that is sustained and provides a healthy balance between ‘pressure’ and ‘support’.

The assumptions stated here by Mary Sinclair are applicable not only to New Zealand; the realizations and the formulated assumptions are just as relevant to systems of public education elsewhere. Also, with respect to the appropriate length of funding, the Ministry is beginning to change its view. In the early years of Schools Support, a maximum of three years of funding was provided to schools in either category of intervention. More recently, however, emerged “a realization, based on the international literature and on findings from other evaluations of Schools Support initiatives, that achieving sustainable improvements and genuine partnership can take significantly longer than three years” (McCauley and Roddick, 2001).

An overall evaluation of the Ministry’s efforts with interventions was published three years ago. McCauley and Roddick’s evaluation revealed that there was overall strong support for the Schools Support interventions, even if some of the processes, requirements, and expectations were not as clear as they could have been. The report established that

- (g) interventions had required, as in other countries, patience, resources, time, and much energy on the part of all participants,
- (h) the identification of failing schools in particular and the entry into intervention programs was often slow, did not inspire the confidence of the affected educators, and was not transparent,
- (i) needs analysis and development of action plans were perceived to be difficult processes, yet useful in providing impetus and focus for improvement, and
- (j) exit criteria and processes were far less thought through and transparent than they should be.

What is interesting in terms of the discussion above is the contractual framework that the Ministry, the school (or cluster of schools), and the community use in the course of an intervention. The framework that is set up, once an intervention is agreed to, closely resembles the contractual relationship that the school charter, as it was originally envisioned, was built on. For interventions, the Ministry and the school (and, to varying degrees, the community the school serves) agree to a plan with specified outcomes that have to be met, resources that will be made available for achieving the outcomes, and consequences for succeeding or failing to succeed to reach specified targets. The contractual relationship and a strategic planning procedure, including pre-specified outcomes, were part of the recommendations of the ERO reports on the failing performance of clusters of schools (ERO, 1996; ERO, 1997; ERO, 1998).

As Howard Fancy, the Secretary of Education, has conceded, the initiation and growth of Schools Support “represented a very major shift away from the initial hands off approach of Tomorrow’s Schools” (Fancy, 2001). Secretary Fancy also criticizes his own department in this regard, saying, “I think it is no credit to the Ministry whatsoever that it required an ERO report in a number of cases to trigger us to start intervening” (Fancy, 2001).

Unfortunately, a quantitative evaluation of the impact of the interventions on school performance is not feasible. The absence of student achievement data or other readily quantifiable measures of school performance render such an analysis impractical for present purposes. There is widespread agreement that the growth of interventions has resulted in system improvements politically, morally, and economically. Empirical evidence for this credo, however, remains elusive. The relative absence of strong interventions, or statutory interventions, should also be noted. Perhaps this is more true here than in other systems. Cathy Wylie (2002) notes in a comparison of the New Zealand system with three other decentralized systems, that “the punitive approach is less in evidence, and is not systematic.”

The Ministry’s willingness to approach the problems of failing schools, individually or as clusters of schools, is a testament both to the urgency of the issue of failing schools in a system that is based on the equitable provision of public education, as

well as to the pragmatism with which the challenging realities (original assumptions aside) were dealt.

### 3.4.2 *Legal requirements*

The growth of Schools Support was not the only change in the way the Ministry of Education chose to deal with school failure. In 2000, a significant revision of the legal framework for education policy took place. After the elections of 1999, Labour came back into power as the dominant party in a coalition government for the first time after 9 years as the opposition party. Under the new government the parliament passed the Education Standards Act of 2001, originally introduced as the Education Amendment Act 2000 (Ministry of Education, 2000; Ministry of Education, 2001). The Act introduced more flexible rules for Boards of Trustees regarding staggered elections, school mergers, combined Boards, eligibility of trustees, and Board constitutions. The act also included new rules for teacher registration, and other regulations. With regard to failing schools, the act imposed a number of important changes. Among them were:

- i) The requirement to give preference to children from the schools' enrollment zones was reintroduced.<sup>38</sup> Students living within a given enrollment zone are guaranteed a place at the school. For students outside the enrollment zone, preference is given to siblings of currently attending students and children of school staff. Beyond these rules, additional spots at an oversubscribed school have to be awarded through a supervised lottery.
- ii) Bulk funding for teachers' salaries was removed. Now the teachers will be paid directly through the Ministry. Schools still have the option of hiring additional teachers, or other staff, funded from their own sources.
- iii) Broader and more explicit powers were granted to the Minister and Secretary of Education for intervening in schools, without the need for the school's approval. If there is reason to believe that the education or the safety of the children is at risk, the Minister and the Secretary of Education are given explicit authorization to intervene in schools by mandating
  - a. provision of specific information to the Ministry,

---

<sup>38</sup> This change only affects schools that are at capacity (i.e. more applicants than available spots in the school), since only schools at capacity were allowed to turn away applicants.

- b. development and implementation of an Action Plan,
- c. appointment of a special advisor, or
- d. appointment of a statutory manager who takes on some of the Board's responsibilities.

iv) The Act introduced planning and reporting changes to school charters and charter requirements. Starting in 2003, in addition to statements of mission and school values, school charters are required to have a section that sets out the school's strategic plan, with an explicit focus on student achievement and informed by ongoing self-review. Each year, schools will be required to lodge a copy of their annual plans, with an analysis of the progress made towards the stated goals, and reports (along with their financial accounts) with their local Ministry of Education Office so that the government will have better information to inform its policy review and support strategies. The new National Administration Guidelines mandate the inclusion of the following into an annually updated charter:

- a. goals for improved student achievement for the next 3 to 5 years,
- b. an annually updated section that sets out the school's improvement targets for the current year, and
- c. the activities the school plans to help it reach its strategic goals.

The revision of the National Administration Guidelines (point V) imposed the establishment of a strategic planning framework. All schools are required to plan and communicate their planning within a specified format. The format is laid out in the Ministry documentation *Planning for Better Student Outcomes* and includes the requirements for an annually updated charter with pre-specified outcomes and student data systems to collect, store, and analyze student-level data (Ministry of Education, 2003a). It is debatable whether this change amounts to a belated implementation of the Purchase Agreement, as envisioned by Scott et al. What is undoubtedly true is that the charters have received a new lease on life with the requirement that they be revised to incorporate changes in the national education guidelines. It should be noted that an increasing portion of school leaders now considers the charters to be beneficial in planning (Wylie, 1999).

The coordination between the ERO and the Ministry has been strengthened even though it may run somewhat counter to the theoretical framework upon which much of the original reforms were based (McMahon, 2002). The perceived usefulness of ERO reviews by principals and teachers has increased. Schools are heeding national emphases through the ERO review process. Principals of schools in small towns and provincial cities were particularly likely to find the ERO reviews helpful to their work (Wylie, 1999).

### 3.4.3 *The role of the private sector*

A comprehensive review of the role of the private sector in New Zealand is beyond the scope of this chapter. However, with regards to interventions specifically, there are a couple of issues that should be emphasized.

The first is that the Ministry, even as it has reevaluated its role towards failing schools and has become more proactive in initiating interventions, has often hired private contractors to carry out the interventions. Private providers of educational services for schools and their Boards have been operating in New Zealand since the old Department of Education was replaced with the smaller Ministry (Mullins, 2003). For example, the training for Boards of Trustees and related work within the schools' communities, the evaluation and training at specific school sites, the provision of payroll services for teachers, and even the actual management of schools has been and is being carried out primarily by private contractors, rather than by the Ministry itself (Kilpin, 2004; Leslie, 2004; Mullins, 2004; Short, 2004). Other interventions, such as the school improvement initiatives, have relied on pre-existing community structures (Maori councils, for example) to provide support for schools and to provide an alternative to the strict Board of Trustees-model for community involvement in the schools (Robinson and Timperley, 2004).

The story of increased involvement of the government is not primarily one of the increased role that the government agencies are *directly* playing in the provision of education. Rather, it is a story of the government *ensuring* that some action is taken where the education provided in the school is no longer sufficient. It should be noted that despite the presence of private providers of educational services with plenty of expertise

and capacity, it has been the coordinated and coordinating role of the Ministry as the ultimate locus of responsibility for the provision of public education that has been the key to change in failing schools. Ultimately, the Ministry adjusted to its responsibility by designing and coordinating responses to school failure, rather than insisting on the build-up of its own capacity for carrying out interventions.

### **3.5 *Recent Developments in Monitoring***

In recent months, the Ministry, and Schools Monitoring and Support specifically, have attempted to further establish processes and consolidate the early experiences. Some broader trends that are relevant for the way the government conceives of and intervenes in failing schools can be also discerned.

#### **3.5.1 *Consolidation of schools support and monitoring***

As the evaluation of Schools Support has revealed, the criteria for entry and exit are unclear for schools, as well as for the Ministry itself. Monitoring and Support are currently implementing a more deliberate system of identifying schools that are “at risk” and schools that need interventions (Ministry of Education, 2003b). Though this effort is not without its detractors within the Ministry, it addresses some of the deficits identified by McCauley and Roddick (2001). Some of these efforts are revisited in greater detail in Chapter Five.

There is some understanding that the problem of school failure may be twofold. For some schools, the decline may well be temporary and with a sufficient support program, they can be brought back onto the right track. For other schools, however, the system may simply not work and not enable them to be consistently successful in the present policy environment. As Brian Donnelly, the associate minister of education with oversight over the ERO, said, “some schools will never work under this system, and for them we will have to have a different system. Some will have to be back under direct control of the ministry, and South Auckland will get a design for schooling that will be unique.” (quoted in Fiske and Ladd, 2000)

#### **3.5.2 *Focus on data***

Though there exists a vastly different attitude towards test scores than in the US, the increase of the scale and scope of Ministry interventions in failing schools has

brought with it increased efforts to collect, analyze, and rely on quantitative data (including student test scores). Throughout the public school system, the need for increased collection, management, and analysis of data is becoming obvious. Despite the initial reluctance of teachers and school administrators in New Zealand to embrace the uses of quantitative data, the calls for a more coherent approach towards data, and student outcome data in particular, are getting louder. Independent of the efforts of Schools Monitoring and Support, the ability to handle and interpret data is increasingly emphasized as a key component of successful planning in schools. The need to build such capacity at the school level was noted by the evaluation of Ministry interventions (McCauley and Roddick, 2001), and has been noted in countless ERO reports on individual schools. The strictly qualitative nature of the ERO reports appears to be falling short of the kind of rigor that is desirable and possible in an accountability system. The ERO reports have increased the pressure on schools and on the Ministry to promote and establish more quantitative performance measures, especially for comparisons across schools. A new mandate for information gathering was also included in the Education Standards Act 2001 (Ministry of Education, 2001). As the mandate is only for individual schools, the issue of comparability across schools is not yet addressed satisfactorily.

Recent efforts by the Ministry include the development of a widely used internet portal ([www.tki.org.nz](http://www.tki.org.nz)) for the dissemination, but also the collection and management of information relevant to education, including school performance.

### **3.5.3 Conclusions**

Insofar as the adequate provision of public education for all citizens is the purview of the government, the New Zealand reforms offer an important narrative. The introduction of market forces into a system of public education has to take careful account of the fact that these forces do not operate equally across the system. Insofar as consistently failing schools are a contradiction to the adequate provision of education for all, a deliberate strategy of what to do with such schools must be included in reform programs. While it has been true – at times – that competitive forces dealt with failing schools, where parents withdrew support from schools and the schools were forced to transform themselves, in other instances the schools were too far gone to pick themselves

up or there was simply no competition (Connelly, 2004). The new system of public education had not made contingencies for what happened if a school could not improve (or would not) and if there were no alternatives for parents.

Handing over the governance of public schools to the communities which they serve was the central innovation in the education reforms, and was intended to relieve the central government of the need of involvement. As Connelly (2003) writes, “it is difficult to find any other education system that has taken local decision making as seriously as New Zealand.” But as the reforms, and the cases of school failure in particular, have shown, local control also has serious costs. As Cathy Wylie (2002) argues, the exclusively local influence and control of schools “precludes contributions to the area or the system as a whole. This is not simply a matter of attitude, or workload: there are no systemic processes in which either boards of trustees or educators can or must take on wider responsibilities.”

The New Zealand reforms have provided an experiment for the hypothesis that the individual units (i.e. schools) of a system of public education will be guided by the invisible hand towards the best outcome. Such a hypothesis does not stand up to the available evidence in this case. As Secretary Fancy (2004) has said

I think we have learned that schools cannot be “isolated islands” in themselves but also need to be seen as archipelagos with measure of both independence and interdependence ... The degree of devolution did not happen – largely, I think, as a result of a pragmatic appreciation over time that the original models were too simple and did not recognize the nature of interdependency and complexity within an education system.

It has been argued that the increased prescriptions for planning and reporting, as well as the Ministry’s increased role in low-performing schools, amount to a return to centralization, and a return to the situation that Tomorrow’s Schools was intended to correct (Young, 2003). On closer examination, however, the legal changes are more complex than a simple increase of the Ministry’s role. The latitude for schools to influence capital investments, for example, was also expanded in the same legislation. Overall, it appears more apt to characterize the legal changes as a fine-tuning of the system, as well as a correction for some of the non-interventionist ideas that were put into place but didn’t work out as intended (Connelly, 2003; Fancy, 2004).

A common comment regarding the schools' charter is that the good schools don't need the Ministry to prescribe the format of the strategic planning, because they are doing this already and the regulation is burdensome to them, while it is instead only the struggling schools that need help in terms of their strategic planning (Leslie, 2004; Mullins, 2004). From a policy perspective, the combination of the two sentiments results in something akin to what New Zealand came around to doing. While the good schools continue to find the prescription of the planning format cumbersome and unnecessary, it is well within their capacity to adopt the mandated format and adhere to the regulations. On the other hand, the schools that are struggling can benefit from such a framework, and the Ministry itself, or the ERO, is able to compare the approaches across schools more easily.

Graham Young (2003) asserts that the new, tighter regulations for accountability present an unnecessary encroachment of the central government into the affairs of individual schools and are "the enemy of creativity, for in such a contrived and controlled culture the opportunity for future school design is considerably reduced." However, in light of the real struggles that many schools had and continue having, a mandated format of reporting, within which the content can and should be specific to the school, seems very much in line with the original intent of the reforms, if not in line with some of the assumptions the initial design was built upon.

To assuage Young's valid concerns about the decrease in allowed creativity, it seems natural that the flexibility of the Purchase Agreement should be proportional to the school's autonomy. But autonomy and individuality for schools are not contradictions to concise, binding Purchase Agreements. An example of this is presented in detail in Chapter Six. The contractual agreements between the schools, the Ministry, and the communities, that have been reintroduced with the interventions, offer an interesting glimpse of what an *ex ante* agreement might look like.

As the introductory section of this chapter has argued, New Zealand's experience with radical education reform and with the issue of failing schools within the reform is important for systems of public education more generally. As Woodfield and Gunby (2003) note,

a key lesson from Tomorrow's Schools is that prior to educational reforms it is imperative to pre-formulate an objective function, implementable measures, and definitions of success and failure, and to have disinterested parties monitor and evaluate outcomes.

The experience provides powerful evidence against the proposition that failing schools will not become an issue in a well-designed system that makes use of market forces.

## **4 Interventions in California Public Schools**

This chapter addresses the second research question: What is the impact of interventions on school performance? To this end, interventions in low-performing public schools in California are examined.

After a short section motivating California's experience with failing schools as a useful and important case for analysis, the interventions are introduced. A brief introduction to California's accountability framework, including the way school performance is measured, is followed by a discussion of the interventions' structure, their implementation, and the underlying notion of failure that they are supposed to correct or compensate for. Next, a brief description of the data itself is followed by the details of how the cohorts of schools were selected in practice, with a comparison of the characteristics between intervention and comparison schools. Based on these comparisons, the models used to estimate the impact of the interventions are proposed. Two existing studies of the impact of these interventions are discussed thereafter, followed by the results that the proposed models return. A final section concludes the chapter with a discussion of the presented results.

### **4.1 Why California Matters**

There are a number of reasons why California is a compelling choice for estimating the impact of interventions in a quantitative manner. First, it is the sheer size of the intervention program, involving large expenditures of scarce public funds, that makes a proper estimation important. Since a large number of schools are involved, the size of the program makes an analysis also more interesting. Particular aspects of the interventions, above all a random assignment to the interventions, provide a useful basis for an analysis. The quantitative data are available from official data files, and the proximity to officials in state and district offices has proven to be invaluable in furthering the understanding of the interventions' details.

Most significantly, the evaluation of the impact of the interventions described below is important for this dissertation, because the program constitutes a distinct effort by the California legislature to address the issue of failing schools in a deliberate and comprehensive manner. The interventions are an integral part of a broader accountability

framework introduced in recent years along the lines of the new accountability described in Chapter Two. The ways in which California's strategy of interventions in failing schools did or did not succeed is thus of broader interest.

## **4.2 The California Interventions**

### **4.2.1 Introduction to California's accountability framework**

The starting point for the interventions this chapter examines is the Public Schools Accountability Act (PSAA) which was passed by the California legislature in 1999. PSAA, or SB1X (the acronym for Senate Bill 1X) as it is more commonly known and as it is referred to in this dissertation, establishes an accountability framework for all public schools in California (SB1X, 1999).

The centerpiece of SB1X is the Academic Performance Index (API), a summary measure of public schools' yearly academic performance. The API is a weighted index of student performance measures and other school indicators.<sup>39</sup> It is calculated for every K-12 public school in California on the basis of scores from tests taken by students in the spring of every year. The API is used to determine a school's decile rank between 1 and 10, annually. The schools' API scores are ranked in each grade span (elementary, middle, high) and divided into deciles, ranging from decile 1 for the 10% of schools with the lowest API scores to decile 10 for the 10% of schools with the highest API scores.

Since the relative weights of test scores for the calculation of the API have changed over the years, the comparison of API level scores from one year to the next is not straightforward.<sup>40</sup> To remedy this, the California Department of Education (CDE) calculates each year, in addition to the new API base (level score) with additional components or differently weighted ones compared to the API base from the previous

---

<sup>39</sup> The API has evolved over the years and will likely continue to do so. Initially (for 1999/00 and 2000/01), it took into account only student test scores on the Stanford Achievement Test Version 9 (SAT-9). In later versions of the API, an increasing proportion of student test scores on the California Standards Tests (CST), the California High School Exit Examination, and the California Achievement Test Sixth Edition (CAT/6, replacing the SAT-9 in 2003) were factored into the calculation of the API. For further information, see the explanatory notes to the yearly API reports (CDE, 2000b; CDE, 2001; CDE, 2002; CDE, 2003b; CDE, 2004). Though the law allows the inclusion of additional school performance measures, the API has so far been calculated using only student test scores.

<sup>40</sup> The *average* level of API scores for each grade span (elementary, middle, high) is adjusted annually using a Scale Calibration Factor (SCF). This correction of the API average across all schools in each grade span does not, however, render API scores for individual schools readily (or accurately) comparable across multiple years.

year, the API growth score. The API growth score is based on the same testing period as the new API base, but is used as a comparison to last year's API base using last year's components and weights. The difference between the API growth score and last year's API base, called the API growth, is thus the official measure, as well as the most comprehensive measure available, of annual school performance (CDE, 2004). This annual change in API, the API growth, is used primarily as the dependent variable and the measure for school performance in this chapter.

Schools that tested fewer than 100 students or return valid scores for a non-representative proportion of the students, as well as schools that are under the jurisdiction of county boards of education, community day schools, and alternative schools are not included in the API rankings. An alternative accountability system has been developed for such schools (CDE, 2000a). For schools that changed substantially in enrollment, demographics or attendance boundaries from one school year to the next, no API growth is calculated, because of the non-comparability of student bodies (Ed-Data, 2004).

The API is calculated to have a low score of 200 and a maximum of 1000 (SB1X, 1999). The State Board of Education has set the performance target for all schools at 800. Each school with a valid API growth score has yearly growth targets that they are required to meet. The schoolwide growth target is equal to 5% of the difference between the school's API and 800, or one API point, whichever is greater. Schools with an API above 800 must maintain their API above 800 in order to meet their schoolwide target. In addition to the schoolwide target, each numerically significant student subgroup must also make 80% of the school's target for a school to have met its API target. Student subgroups by ethnicity and socio-economically disadvantaged students are considered numerically significant if they constitute 15% of the student population, or if the number of students is 100 or more (CDE, 2000a; Ed-Data, 2004).

California's accountability framework, with the API as its cornerstone, is an example of the new accountability introduced in Chapter Two. The assessment of the schools' performance is based squarely on student test scores. The idea of holding *schools* accountable as a means towards raising *student* achievement is made explicit in the legislative intent expressed at the outset of SB1X. It notes that assessments have

indicated “that many pupils in California are not now, generally, progressing at a satisfactory rate to achieve a high quality education,” and that to remedy the situation of underachieving students the law is formulated to “hold each of the state’s public schools accountable for the academic progress and achievement of its pupils within the resources available to schools” (SB1X, 1999). The problems that are associated with making this connection between the performance of schools and the performance of its students as direct as SB1X are discussed in the discussion section at the end of this chapter.

Along with the API, SB1X also introduced two programs designed to focus schools’ efforts on increasing their API score: a rewards program for schools that are able to demonstrate high performance on the API and a set of interventions in schools where the API indicates low performance. The latter part of the new accountability framework is the subject of the remainder of this chapter. The set of interventions established in SB1X to address the issue of underperforming schools is called the Immediate Intervention/Underperforming Schools Program (II/USP).

Alongside II/USP, the California Department of Education (CDE) also implemented an intervention that was funded through federal sources (Comprehensive School Reform, CSR), as well as an intervention that was not part of the initial law (High Priority Schools Grant Program, HP). These two additional interventions are briefly described below as well. The focus of the analysis is on the impact of II/USP interventions and CSR interventions are included insofar as they are similar to II/USP and where their inclusion furthers the understanding of the interventions’ effect.

#### *4.2.2 Details of II/USP*

The general idea of II/USP, as stipulated in the law (SB1X, 1999), is to provide schools in the lower 5 deciles of statewide API performance who have not met their API growth targets in the previous year with additional funding. The funding is to support the development of an Action Plan specific to the individual school in the first year and the implementation of the Action Plan over the next two years, with an option for a third year of implementation funding. In return for the additional funding and technical expertise to help with the establishment and implementation of the Action Plan, the schools subject themselves to a stricter accountability regime including sanctions for failure to meet API

growth targets by the end of the implementation period. II/USP is funded at \$50,000 for each selected school during the planning year. This planning grant supports a needs assessment at the school and the development of the Action Plan in cooperation with an outside expert. Once it is developed, the Action Plan has to be approved by the CDE,<sup>41</sup> and is implemented in the two, or possibly three, subsequent years. The funding during implementation is set at \$200 per student, or a minimum of \$50,000 for schools with less than 250 students. To carry out the Immediate Intervention/Underperforming Schools Program, SB1X instructs the Superintendent of Public Instruction to select 430 schools per year to participate in II/USP. The criteria for eligibility and selection are important for the quantitative analysis below, in particular for the choice of an appropriate control group.

The eligibility criteria are broad, so that even though the number of participating schools is quite high at 430 annually, the number of schools applying for the program has been higher still, resulting in a sizable group of non-selected schools among those that applied for the program. What is more, the selection of schools for the program is not competitive or based on specific qualifications that the applicant schools need to demonstrate. Instead, the assignment to II/USP is done by lottery. It is, however, subject to a number of specific requirements, which turn out to be very important. SB1X stipulates that the schools selected for II/USP should be spread evenly across the 5 lower deciles. This means that, ideally, 86 schools are selected from each decile to make up a cohort of 430 schools. The number of elementary, middle, and high schools that should make up each cohort is specified in the law as well. No more than 301 elementary schools, 78 middle schools, and 52 high schools should be in each cohort, with the 86 schools selected within each decile range representing each grade span proportionately (SB1X, 1999). How these restrictions played out in practice is discussed below.

---

<sup>41</sup> By law the CDE can reject the Action Plan outright. In practice, this has not happened, according to CDE officials (Colton, 2004b). Instead, the CDE sent plans that failed to meet the requirements back to the school for a rewrite. In the end, all Action Plans have been approved by the CDE, even as some have been considered to be of questionable quality (CSU, 2001).

#### 4.2.3 *Details of CSR and HP interventions*

The two interventions that are closely tied to II/USP are the Comprehensive School Reform (CSR)<sup>42</sup> and the High Priority Schools Grant Program (HP). In contrast to the state-funded II/USP, CSR is federally funded. The operational requirements and the selection criteria differ somewhat between II/USP and CSR, though the level of funding and the support by outside experts is similar. Differences in funding and selection procedure are discussed in greater detail below, along with the implementation of II/USP. The main substantive difference to the II/USP interventions is that CSR interventions do not develop and implement a plan for specific improvements at a school, but rather seek to implement a pre-existing, comprehensive model for school improvement, a so-called whole-school design. Without the need for a planning year, CSR is instead designed to have three years of implementation. Where the II/USP Action Plans seek to correct shortcomings that were identified at the school and build on existing strengths, the CSR models are generally more expansive providing prescriptions for most aspects of school operations that enable the school's transformation into an improved educational organization following a particular design (Department of Education, 2002).

The High Priority Schools Grant Program (HP) was added by the California legislature to the accountability framework in 2001 (CDE, 2003a). Like II/USP, the program is also funded by the state, but focuses more narrowly on the lowest performing schools (only schools from decile 1 and 2 are selected, with priority given to the lowest performing schools within decile 1). HP is explicitly an extension of the II/USP program and also requires participating schools to develop and implement a specific Action Plan. Though differences in the requirements for the Action Plan exist between HP and II/USP (HP Action Plans are, generally speaking, less detailed and can be based on existing II/USP Action Plans), efforts are underway to align the requirements. Schools in HP are required to participate in a number of pre-specified programs of professional development and to acquire of specific educational material for students. Though the impact of HP on school performance is not a part of the quantitative analysis, its addition

---

<sup>42</sup> Initially, this intervention was called Comprehensive School Reform Demonstration (CSR). It was renamed with the adoption of the No Child Left Behind Act of 2001 (NCLB).

to California's strategy for intervening in underperforming schools matters as HP is an intervention program with a different focus from II/USP.

Common to all three interventions is the involvement of external evaluators, the implementation of a specific improvement plan, and the requirement to involve the community in the school's improvement. More importantly, the official measure of success for all interventions is the API growth; student achievement as measured in the yearly growth of test score averages on standardized tests.

#### *4.2.4 II/USP Action Plans*

The central aspect of II/USP is the development of an Action Plan and the subsequent implementation of this plan. While parallels to contractual agreements in other settings are explored further in the discussion section at the end of this chapter, the following section gives a brief overview of how the implementation of these plans has worked in practice.

SB1X (1999) specifies the requirements for the Action Plans that the schools must develop, together with external consultants, and submit to the CDE for approval. The Action Plans are required to focus on

improving pupil academic performance, improving the involvement of parents and guardians, improving the effective and efficient allocation of resources and management of the school, and identifying and developing solutions that take into account the underlying causes for low performance by pupils. (SB1X, 1999)

As these broad and inclusive themes should be included in the Action Plan, it is useful to note that schools are, strictly speaking, not held accountable for the actual implementation of these goals. The success of the Action Plan's implementation and of the intervention more generally is measured exclusively by the API growth that the school manages to achieve. The Action Plan is also required to "set short-term academic objectives ... for a two-year period that will allow the school to make adequate progress toward the growth targets established for each participating school" (SB1X, 1999). The focus on meeting API targets by the end of the implementation period is thus an explicit part of the Action Plan together with the broader themes.

John Mockler remarked in the fall of 2000, serving at the time as the interim Secretary for Education for Governor Gray Davis, that many Action Plans submitted to

the CDE for approval by prospective cohort 1 schools were not very good (cited in CSU, 2001). The plans were not geared towards California's new accountability framework, made very little use of the available data, did little to identify challenges specific to the teaching and learning of the schools' students, and did not refer back to the standards as much as the CDE administrators had hoped. Instead many plans focused primarily on processes within the school. Not a single Action Plan, according to Mockler, asked for a waiver of state regulations, as allowed under SB1X, or suggested a new allocation of existing resources at the school.

A number of external evaluators, speaking at the same event, took stock of their experiences after having developed the first round of Action Plans (CSU, 2001). The evaluators remarked that many schools they had been working with had difficulty adapting to the key aspects of the new accountability that California had put in place. Curriculum instruction was not aligned to the standards, and curricula within the school were not coordinated to the degree they should be. The evaluators' job turned out to be one of basic professional development and consultation, rather than identifying weaknesses or improvements that were specific to the school or to its students in most need of help.

What the discussion of the Action Plans' implementation brings to the fore is that there is substantial change going on across all California's schools with the development of new standards, the use of new tests, and an increased focus on the information drawn from these tests. In practice, it seems that the II/USP interventions have had to focus on schools managing this change. Rather than correcting any substantial problem that was there all along, the Action Plans are directed towards enabling schools to adapt to the environment of the new accountability framework.

How did the schools experience this change? And what are the lessons that can be drawn from the way the Action Plans were implemented? Commissioned by the CDE, Anne Just and her colleagues have produced three research reports on the implementation of the II/USP and CSR programs with schools in cohort 1 (2002; 2001; 2003). The reports include simple analyses of the impact of the intervention programs on schools, but focus mainly on implementation and on preliminary lessons that can be drawn from the

early experiences with these interventions. Overall, the reports find that the implementation of the Action Plans has varied across districts and across participating schools. In surveys of the principals and teaching staff in the participating schools, the external evaluators received generally high marks for their work in the development of the Action Plan, with a few schools being less than pleased about the support they got. The role of school districts in supporting and developing the capacity to implement the necessary changes was often unclear and continued to vary considerably throughout the implementation. Though the capacity for change among the participating schools seemed to be in doubt after the first planning year, the conclusions from survey responses and site visits by the research team were quite hopeful by the end of the third year. The most recent report (Just and Wong, 2003) notes that the primary barriers to Action Plan implementation were high student mobility and teacher resistance to change, but that many positive changes had occurred in the schools. Positive changes in curriculum and instruction, in school dynamics and other indicators of school performance, as well as beneficial organizational and structural changes all served as evidence of the effectiveness of Action Plan implementation. Staff satisfaction and willingness to recommend participation in II/USP to other schools were high. In addition, the staff in schools that had institutionalized school reform and built capacity were confident about their ability to sustain reform without continued II/USP funding. The evidence in terms of the interventions' impact on the API is less positive, however. After the second year of Plan implementation, only 43% of Action Plan schools met their API growth targets (Just and Wong, 2003), compared to roughly 70% and 50% of these schools that met their API growth targets after the first year of planning and the first year of implementation, respectively.

A recent case study of the changes Greenfield Middle School undertook as part of II/USP point to concrete changes that can foster success in the school that can be measured with the API (Bushman and Goodman, 2003). The school administrators, together with the external evaluators, forced the teachers at Greenfield to teach to the standards by making the standards more explicit and prominent than they had been. Each teacher received a laminated poster on which the standards were listed, so that not only the teacher, but also the students could refer back to the actual standards and understand

what sort of learning was supposed to be happening at any given time. While this focus may elicit concerns about excessively narrow instruction, it follows the path that the California accountability framework has laid out. In the case of Greenfield Middle School, the API scores indicate that the chosen strategy has indeed been a successful approach.

To make the implementation of the Action Plans possible, the CDE provided the funding as indicated in SB1X. Before addressing the nature of the school failure that the interventions under SB1X were designed to correct, some context regarding the amount of funding that the interventions entail is provided.

#### *4.2.5 Funding of interventions*

As part of the discussion of the interventions' implementation, the following section discusses the amount of funding that the interventions entail. Is \$200 per student a large or a small number? Is it big enough to have a significant impact on a struggling public school in California? Schools in the Los Angeles Unified School District (LAUSD), the largest school district in California and the second largest district in the United States, can serve as an example. According to William Ouchi (2003), the annual operating budget per student (excluding capital costs) for schools in LAUSD is on average just under \$10,000.<sup>43</sup> The II/USP funding thus represents a mere 2% average increase in the operating budget of the school. However, as Ouchi points out, principals of LAUSD schools directly control only 6.7% of the total operating budget, while the vast majority is controlled by the district in the form of teachers' salaries or transportation costs, for example. The average amount that principals can work with to make individualized changes at the school level is thus approximately \$670 per student. An additional \$200 per student represents an increase of about 30% in discretionary spending that is available to the principal. Though this numeric example is specific to LAUSD, and principals in other districts may have somewhat more budgetary control, it illustrates the approximate magnitude of II/USP funding. While the additional funding is only a small increase compared to a school's overall operating budget, it is a significant increase in the budget controlled by the principal and offers more leeway for programs designed for a

---

<sup>43</sup> Based on numbers from the 2001/02 school year.

particular school and its students' most urgent educational needs. These are the programs that the Action Plan is intended to promote and coordinate.

During the first year of implementation for cohort 1, budget shortfalls meant that the full \$200 per student was not available. According to the responsible CDE officials, funding for II/USP implementation in the school year 2000/01 amounted to roughly \$167 per student (Colton, 2004b). In subsequent implementation years and cohorts, the full \$200 was dispersed to participating schools.

SB1X requires schools to match the funding they receive from so-called categorical funds, which include Title I funding for example, or other funding that is available to the school outside of SB1X interventions. According to the director of the SB1X office at LAUSD, the purpose of these matching funds is two-fold (Gerst, 2004b). On the one hand, they leverage the funding received from the interventions, so that the necessary improvements described in the Action Plans are also being addressed with existing funds and are thus more likely to be addressed effectively. On the other hand, the matching funds are an indication that the school is willing and able to commit the resources necessary for change beyond the duration of the SB1X intervention itself and to continue to support the programs set out in the Action Plan beyond the availability of the additional funding from the CDE.

For schools in the HP program, an additional \$200 per student was made available, a total of \$400 (matching requirement still at \$200) per student. These funds have only been available in the school years 2002/03 and 2003/04, i.e. the last two years for which data are available (CDE, 2003a).

#### ***4.2.6 How interventions were supposed to work***

Before focusing on the data used in this analysis and the selection of schools to the II/USP interventions, it is worth considering the ways in which the interventions were meant to impact schools. Not only for the design of the interventions themselves, but also for the current analysis of the interventions' impact, the nature of the identified underperformance is an important question. This issue ties back to the earlier discussion about school failure in Chapter Two. The question that needs to be asked is: What is the nature of the shortcoming that the interventions are meant to correct? Before launching

into an analysis of what impacts can be detected in the data, it is worth considering what the interventions are *supposed to* accomplish.

Two contrasting views are proposed here, and are revisited later on, for the consideration of the model that ought to be used for detecting the interventions' effects. A first view of the nature of the problem at individual public schools is that there is something specific, identifiably wrong in the way that they operate. Once this shortcoming in the school has been identified, possibly with the help of an outside expert, it can be addressed and corrected over the course of a couple of years, given additional resources and incentives to make the necessary changes (e.g., incentives based on the threat of sanctions if the provided funding does not lead to a measurable increase in student achievement). If this view of the problem is valid, and the goal of the interventions is to help the schools fix the problem, then the schools will be able to return to a status without interventions once the problem is fixed. To examine whether the interventions have adequately addressed this problem, an analysis of the interventions' real impact should look for evidence of permanent shifts in the performance of schools after they have had the intervention.

An alternative view of the problem of schools suggests a different analysis. Perhaps the real problem at the struggling schools is that they face challenges that are different and more difficult to handle than the challenges faced by other public schools. Their failure may not be a matter of any specific thing that they are doing wrong, but a matter of requiring greater support and more specialized planning to deal with difficult-to-educate student populations, for example, or with parents that are less likely to be supportive of the children's learning process. Interventions in schools facing difficulties of this nature will aim to provide the support and funding that the schools need, even if the schools need the support on an ongoing basis. Once the intervention stops and the funding is withdrawn, the difficulties remain. To examine whether interventions are beneficial to schools faced with such systemic problems, the analysis of the interventions' impact should seek to identify shifts in the performance of schools that are transitory, lasting for the duration of the intervention, possibly including a time-lag for the interventions to impact the school and for its effects to wear off.

Of course, reality is unlikely to match up perfectly with either one of these stylized views of the problems that underperforming schools are facing. And yet, the choice of a model to detect the interventions' impacts implies a judgment about which of the two views of the failure and the likely impact of the interventions comes closest in describing the true nature of the schools' dynamics.

The adjustments initiated in recent years, such as the establishment of the HP interventions and its partial replacement of II/USP, suggest that the second view is gaining in influence. However, the way II/USP was set up originally, i.e. with its limited time horizon and the distribution of funds across the lower half of California's schools, suggests that the former view has been more dominant in the establishment of SB1X. The schools undergoing the interventions are required to demonstrate at least "significant growth" (meaning: growth in the API, even if all targets are not met) within two years and required to meet the API targets within a maximum of three years. In case the interventions do not succeed, SB1X provides for an escalating series of sanctions.

It is unrealistic to expect that an econometric estimation of the interventions' impact can correctly identify the long term impact of the interventions. Constraints posed by the available data (only 5 years of API data are available to date), as well as the more fundamental difficulty of isolating changes in test scores that are attributable to the interventions, especially in the longer-term, suggest that only the short term impact can plausibly be estimated here.

One aspect of the interventions that was designed to be an important incentive for change was the threat of sanctions for schools that did not manage to change. SB1X requires the State Superintendent to reassign the principal and either (a) provide parental choice,<sup>44</sup> (b) allow the parents to turn the school into a charter school, (c) outsource the management of the school, (d) restructure the school staff's contracts, including reassignment to other schools, (e) reconstitute the school, or (f) close the school (CDE, 2000a; SB1X, 1999). In practice, none of these strong interventions have been carried out as a consequence of a school being in the II/USP program (Wagner, 2004). As one

---

<sup>44</sup> The requirement for parental choice explicitly excludes the provision of any transportation for children attending other schools (SB1X, 1999), rendering the option for school choice far less likely to be exercised.

external evaluator who worked with a number of II/USP schools during the first planning year stated: “We found that most schools don’t believe anything is going to happen to them if they don’t meet their growth targets. Three years is a long time. We really need to have a discussion about what is going to happen to schools and it needs to be based on reality.” (quoted in CSU, 2001)

Instead of the strong interventions that SB1X allows for, the schools that do not show any growth during the implementation period have come under so-called state monitoring. According to CDE officials, state-monitoring implies that a School Assistance and Intervention Team (SAIT) conducts an audit of the school and makes recommendations for improvement (Wagner, 2004). The implementation of SAIT recommendations is not voluntary, as the Action Plan was, and is imposed as a binding requirement on the school and its district. Since the fall of 2002, a total of 128 schools have come under state monitoring.

### **4.3 Data**

Before discussing the selection of schools to the II/USP cohorts, the data used in the analyses below are briefly described here. The API data used in the analysis were downloaded from the CDE website (<http://api.cde.ca.gov/datafiles.asp>) in DBase format and transferred into Stata7 format via StatTransfer (Circle Systems, 2003). The five datasets were the API growth datasets, which include the API base data from the previous year, as well as all the relevant covariates. Statewide API rankings, and similar-schools rankings were downloaded separately from the five corresponding base datasets and merged. All calculations were done with Stata, Version 7 (StataCorp., 2001), while individual graphs were produced in Excel (Microsoft, 2000).

The information on eligibility, application, and selection for II/USP and CSR programs comes from two sources. The first and primarily used dataset was provided by Pat McCabe, the policy and evaluation administrator at the CDE, and the person responsible for the selection of individual schools to the SB1X intervention programs. This dataset also contains decile rankings for 1999, 2000, and 2001 and designations for grade spans (elementary, middle, high). These designations for grade span and decile rankings are the ones used in most calculations. A second dataset was provided by Cliff

Rudnick from the School Reform Assistance Office at the CDE. This second dataset contained longitudinal data for whether individual schools received any kind of funding as part of one of the three SB1X interventions, as well as cohort designations for II/USP and CSR and a binary variable for schools in HP.

SAT-9 test results in reading and math were accessed from a dataset that had been downloaded from the website of the Standardized Testing and Reporting (STAR - <http://star.cde.ca.gov>).<sup>45</sup>

#### **4.4 II/USP and Comparison Schools**

This section looks at the selection of the II/USP cohorts, as well as at the baseline characteristics of the selected schools and the group of schools that are used as a comparison. There have been three cohorts of II/USP schools to date.<sup>46</sup> Cohorts 1, 2, and 3 were selected in the fall of 1999, 2000, and 2001, respectively. What are termed the treatment, intervention, or II/USP schools in the remainder of this chapter are the schools that have been selected for II/USP in one of the three cohorts. The non-treatment, comparison, or control schools are the set of schools that applied for selection, but were not selected and whose school performance is used as a comparison to the school performance of the II/USP schools. This chapter is primarily focused on cohort 1 of II/USP, since the circumstances of cohort 1 allow for the most meaningful analysis. An additional less detailed analysis of cohort 2 is included also, along with a description of the implications of the changes under cohort 3.

While the law is quite detailed about the eligibility of schools for the interventions and the process of selection among those who applied, the exact details of the selection in practice turns out to be important for understanding the way in which the interventions' impact should be analyzed.

---

<sup>45</sup> The SAT-9 dataset was downloaded and prepared by PRGS Doctoral Fellow Bryce Mason, as part of a PRGS course, and made available for this dissertation also.

<sup>46</sup> The continuation of II/USP beyond the three cohorts is unclear at this time, as other programs, such as HP have become more important and the difficult adjustment to the federal mandates under NCLB have taken up much time and financial resources.

#### 4.4.1 Selection of cohort 1

For the selection of cohort 1 of II/USP at the beginning of the 1999/00 school year, the API had not yet been fully established. As a result, the eligibility for II/USP for cohort 1 was based on the average student scores by school from the spring 1998 and spring 1999 Stanford Achievement Test Version 9 (SAT-9). Based on the combined SAT-9 scores, schools were ranked into 10 deciles and all schools in the lower 5 deciles were eligible to apply (growth was not considered for cohort 1). The CDE notified all districts in which any of the 3,145 eligible schools were situated. The districts returned lists of schools that volunteered for inclusion in the II/USP program.<sup>47</sup> From the pool of 1,423 schools that applied, 350 schools were selected randomly, subject to the restrictions laid out in the law. The number of selected schools was only 350, because 80 schools had been pre-selected to the federally funded CSR program. The process whereby the CSR schools were selected is “shrouded in mystery” (McCabe, 2004b) and therefore this group of schools does not enter the analysis, even though they constitute, together with the 350 selected II/USP schools in cohort 1, the legally mandated first cohort of 430 schools.

The restrictions for the selection of II/USP schools from the 1,423 applicant schools included the selection of an even number of schools across all five lower deciles and a specified number of schools by grade span. But while there was an equal number of possible II/USP selections per decile and an equal number of schools eligible per decile, schools from deciles 1 and 2 turned out to be far more likely to apply. The implications of this difference in the probability of selection are explored further in the next section.

According to Pat McCabe (2004b), who was the person at the CDE carrying out the selection of schools to the program, one particular administrative hiccup led to a significant underrepresentation in cohort 1 of schools from LAUSD. When LAUSD returned the list of schools applying for II/USP in late fall of 1999, the list was sent in five separate batches spread across the five days that the CDE was accepting the application lists from the districts. LAUSD officials had done this to indicate a prioritization among the applicant schools from LAUSD, giving the schools on the list

---

<sup>47</sup> Not all schools were actively involved in this choice, but some were simply put on the list by the district (Just et al., 2001). The number of times this occurred is unclear, however.

that was submitted on the first application day the greatest chance of being selected, in response to the CDE announcement that earlier submissions would get priority. But since other districts did not do the same thing, the number of applications received on the first day was already significantly more than the 350 spots that were left to be allocated in the II/USP program. As a result, the second, third, fourth, and fifth parts of the list of schools that LAUSD submitted to the CDE were never considered for selection, even though the schools applied for II/USP. This misplaced strategy on the part of LAUSD influenced the proportion of LAUSD schools that were selected from the group of applicants. It should be noted that this underrepresentation influences the analysis of the interventions' impact only if schools from LAUSD are significantly different from schools with the same grade span and decile ranking from other California school districts. This issue is also discussed further below.

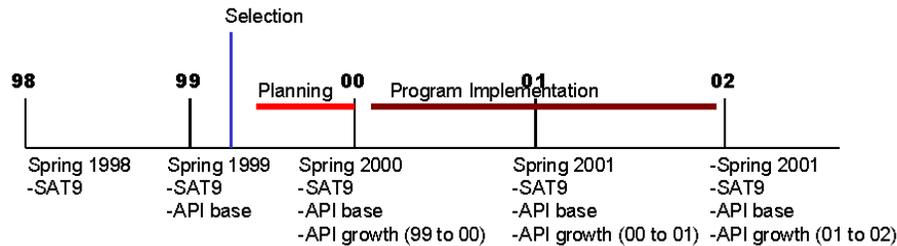
The process whereby schools applied for II/USP is not uniform across districts or time. At LAUSD, for example, the SB1X office was established in 2000, and has actively been involved in assisting its schools in the process of applying to, planning for, and implementing II/USP. Since the establishment of the SB1X office, all eligible schools in LAUSD are *required* to apply for II/USP (Gerst, 2004a). Despite the shifting process and unclear decision rule that has led to the schools' applications, the process of selection from the list of schools that applied for II/USP is centralized at the CDE and the decision rules are uniform and transparent, if complicated, given the legal restrictions placed on the schools' selection.

#### ***4.4.2 Characteristics of cohort 1 treatment and non-treatment schools***

Given the selection criteria laid out in the law and the details of how the selection to cohort 1 worked out in practice, a first observation can be made regarding the random assignment: the selection was not random for the entire group, but only within the restrictions set out by law. Therefore, it is important to examine the baseline characteristics of the treatment and non-treatment schools with particular care, so that any differences between the two groups can be properly accounted for in the models of the treatment's effect. For clarification, Figure 2 provides an illustration of the timeline for

cohort 1. The hash marks indicate the available test score data from the spring of the respective year.

**Figure 2 - Timeline for cohort 1**



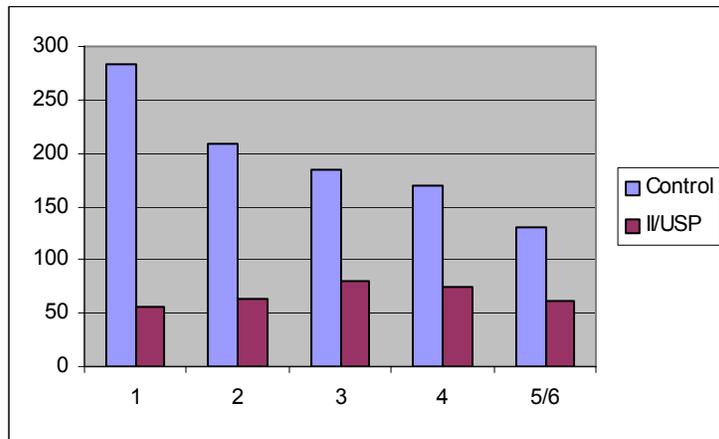
The 1999 API base scores, as well as 1999 SAT-9 scores, are based on tests conducted in the spring of 1999, months before the eligibility, application, or selection of schools for II/USP. In Table 4 below, the t-statistic of 4.75 for the hypothesis test whether the two means of API scores between treated and non-treated schools are from the same distribution shows that the two groups are significantly different in terms of baseline student performance. The schools in the II/USP program have a higher level of performance, as measured by the API, than the comparison schools. *Within* deciles the API scores of the treatment and non-treatment groups are not statistically different from one another. With a higher proportion of low-decile schools among the applicants and the number of II/USP spots equal across the deciles, the probability of being selected for the II/USP program was at 32% almost twice as high for decile 5 schools than the 17% probability for decile 1 schools. It should be noted that the numbers provided in all the tables are based on the schools for which there is data available (API or SAT-9 scores are reported, including API decile rankings) which leads to a slightly lower number of observations included in the tables.

**Table 4 – Base-line breakdown by decile**

<b>1999 API base</b>	<b>No-treat. – mean (st. error) – #obs</b>	<b>Treatment – mean (st. error) – #obs</b>	<b>t-stat</b>	<b>Probability of selection</b>
<i>Overall</i>	500.7 (2.3) – 976	522.0 (3.7) – 335	4.75**	0.26
decile 1	412.7 (1.8) – 283	413.3 (4.2) – 56	0.14	0.17
decile 2	477.4 (1.2) – 208	477.8 (2.4) – 64	0.15	0.24
decile 3	524.6 (1.1) – 184	526.4 (1.7) – 80	0.84	0.30
decile 4	567.8 (1.0) – 170	567.5 (1.5) – 74	- 0.19	0.30
decile 5	607.4 (1.1) – 131	607.6 (1.9) – 61	0.19	0.32

It is the differential probability of selection among the schools with different API levels that drives the overall difference in the API between the two groups. The histogram in Figure 3 shows the schools per API decile ranking based on the numbers in Table 4 above.

**Figure 3 – Number of cohort 1 schools per API decile ranking**



Since the deciles are set up according to API scores, Figure 3 indicates that the mean API scores between the two groups should be expected to be quite different, with the group labeled ‘Control’ (i.e. applied-but-not-selected) having a lower API mean on account of the greater proportion of schools from lower deciles in the group. As Table 4 above indicates, the mean for schools in cohort 1 of II/USP is 522 or 21 points higher from the control group mean of 501.

Note that, as in Table 4, the 80 CSR schools are not included here. Their inclusion would not alter the picture dramatically.<sup>48</sup> Thus the selection was not random overall and the non-treatment group as a whole does not provide an accurate comparison for the treatment group. Instead, a proper comparison will have to control for the differential probability of selection across the decile rankings.<sup>49</sup>

A second dimension that is a restriction in the selection of schools for II/USP is the grade span or school type (elementary, middle, high). The breakdown of 1999 API base scores by school type reveals that the treatment and non-treatment groups are

<sup>48</sup> Since CSR schools are primarily from decile 1 and 2, their exclusion explains the lower numbers for II/USP schools shown in the lower two deciles compared to the higher deciles.

<sup>49</sup> Decile 5 includes individual schools which are in decile 6 according to the official API rankings, because the selections were based on earlier calculations of the rankings (SAT-9 scores for cohort 1).

significantly different from one another, even though the probability of selection is similar for elementary, middle, and high schools.<sup>50</sup>

**Table 5 – Base-line breakdown by grade span**

<b>1999 API base</b>	<b>No-treat. – mean (st. error) – #obs</b>	<b>Treatment – mean (st. error) – #obs</b>	<b>t-stat</b>	<b>Probability of selection</b>
<i>Overall</i>	500.7 (2.3) – 976	522.0 (3.7) – 335	4.75**	0.26
high schools	519.4 (6.1) – 118	543.8 (8.5) – 41	2.12*	0.26
middle schools	508.6 (5.9) – 164	527.6 (8.0) – 64	1.78	0.28
elem. schools	495.7 (2.7) – 694	516.6 (4.6) – 230	3.86**	0.25

To make the comparison more tractable, the following tables show only the breakdown by API deciles. The fact that the three school types are different from one another in terms of the baseline for school performance is accounted for in the models introduced later on.

To further analyze the difference in baseline student performance between the treatment and non-treatment schools, the SAT-9 scores are listed analogously to Table 4. Between 1999 and 2001, the API was based exclusively on scores from the SAT-9. Math and reading were the two main components across elementary, middle, and high schools. A breakdown of the weights that SAT-9 math and reading scores were given in the respective APIs is provided in Table 6.

**Table 6 – Weighting of SAT-9 scores in API**

<b>Year</b>	<b>Grade span</b>	<b>Reading</b>	<b>Math</b>
1999	Middle/Elementary	30%	40%
1999	High	20%	20%
2000	All	20%	20%
2001	Middle/Elementary	12%	40%
2001	High	8%	20%

Table 7 and Table 8 show the SAT-9 scaled scores<sup>51</sup> in math and reading. These are school-level averages weighted by the number of students in each grade of average scaled scores by grade (as reported by STAR).

<sup>50</sup> SB1X also mandates an even geographic and rural/urban distribution. Though these conditions were looked at after the selection and found to have been met, they did not enter the selection rule explicitly (McCabe, 2004a).

<sup>51</sup> Each student's raw score (i.e. number of correct answers) in any test is scaled according to the grade level which allows the scores to be averaged for each grade and across grades for each school.

**Table 7 – SAT-9 math scaled score by decile grouping**

<b>1999 SAT-9 math</b>	<b>No-treatment mean (st. error)</b>	<b>Treatment mean (st. error)</b>	<b>t-stat</b>
<i>Overall</i>	614.8 (1.1)	618.7 (1.9)	1.71
decile 1	602.4 (2.0)	601.1 (4.4)	-0.27
decile 2	609.6 (2.3)	616.1 (4.5)	1.37
decile 3	618.1 (2.6)	621.4 (4.0)	0.70
decile 4	624.1 (2.6)	621.3 (3.7)	-0.60
decile 5	633.1 (2.9)	632.1 (4.3)	-0.21

For both math and reading, the differences between treatment and non-treatment group generally confirm the results from the API base scores reported earlier. The schools selected for II/USP have a significantly higher level of student performance (though for math, the difference is only significant at the 10% level), but there are no significant differences within decile groupings.

**Table 8 – SAT-9 reading scaled score by decile grouping**

<b>1999 SAT-9 reading</b>	<b>No-treatment mean (st. error)</b>	<b>Treatment mean (st. error)</b>	<b>t-stat</b>
<i>Overall</i>	616.8 (1.1)	621.7 (1.8)	2.31*
decile 1	602.8 (1.9)	601.6 (4.0)	-0.27
decile 2	611.3 (2.1)	617.2 (4.2)	1.34
decile 3	619.6 (2.5)	624.3 (3.7)	1.05
decile 4	628.3 (2.4)	626.2 (3.5)	-0.51
decile 5	636.9 (2.7)	637.3 (4.2)	0.09

Since SAT-9 data from the spring of 1998 are available, it is possible to analyze a baseline growth prior to the selection into II/USP. The growth in scaled scores for school  $i$ ,  $\Delta y_i$ , is calculated simply by subtracting the two level scores:  $\Delta y_i = (y_i^{99} - y_i^{98})$ . Table 9 and Table 10 list the baseline growth in SAT-9 math and reading, respectively.

**Table 9 – SAT-9 math growth by decile grouping**

<b>98-99 SAT-9 math growth</b>	<b>No-treatment mean (st. error)</b>	<b>Treatment mean (st. error)</b>	<b>t-stat</b>
<i>Overall</i>	4.6 (0.2)	5.0 (0.3)	0.82
decile 1	3.1 (0.4)	4.1 (0.7)	1.02
decile 2	5.4 (0.5)	3.7 (0.8)	-1.88
decile 3	4.5 (0.5)	5.3 (0.7)	1.02
decile 4	5.1 (0.4)	4.5 (0.7)	-0.69
decile 5	6.2 (0.5)	7.2 (0.9)	0.99

In terms of math and reading scaled score growth in the year prior to the selection for II/USP, the treatment and non-treatment groups are not significantly different overall. There are differences in deciles 1 and 2, especially in reading.

**Table 10 – SAT-9 reading growth by decile grouping**

<b>98-99 reading growth</b>	<b>No-treatment mean (st. error)</b>	<b>Treatment mean (st. error)</b>	<b>t-stat</b>
<i>Overall</i>	2.6 (0.2)	2.9 (0.3)	0.84
decile 1	1.0 (0.4)	3.0 (0.7)	2.07*
decile 2	3.5 (0.5)	1.7 (0.9)	-1.85
decile 3	2.6 (0.4)	3.2 (0.6)	0.70
decile 4	3.0 (0.4)	2.0 (0.7)	-1.31
decile 5	3.9 (0.4)	4.7 (0.8)	1.05

These differences should be kept in mind and call for a further examination of the more detailed characteristics between the two groups. To that effect, nine covariates are considered. The variables are listed and explained in Table 11.

**Table 11 – additional variables listed as *controls***

<b>Variable name</b>	<b>Description</b>
meals	percentage of students tested that are participants in the free or reduced price lunch program
ell	percentage of english language learners
full	percentage of teachers with full credentials
valid_num	number of students tested
pct_hi	percentage of hispanic or latino students
avg_ed	average parental education level
mobility	percentage of students who first attended the school in the present year
pct_aa	percentage of african american students
pct_as	percentage of asian students

Table 12 below lists the nine covariates in each of the API deciles that are included explicitly in the model. For the three variables listed at the end, *mobility*, *pct\_aa*, and *pct\_as*, the patterns between the treatment and non-treatment schools, as well as the patterns across the deciles are not uniform and are therefore not described further.

For the six remaining variables, it can be readily seen that characteristics which would be expected to impact a school’s average student achievement differ across the deciles, with schools in the lower deciles having a higher percentage of students from low socio-economic background, a higher percentage of English language learners, and a lower percentage of fully credentialed teachers. Schools in lower deciles also appear to be

larger and have a higher proportion of students from Hispanic backgrounds with a lower average education level of their parents. These differences across the deciles appear as expected and are similar within the treatment and the non-treatment group.

Between the two groups, the differences for all variables (except the three listed at the end) are in the direction that should be expected, if the treatment schools were from a higher API decile. The differences do not appear to be significant for all deciles, but they are uniform across most decile groups. The t-statistic listed indicates the likelihood of the two groups being from the same underlying distribution.

**Table 12 – School characteristics for II/USP cohort 1 schools**

<b>1999/00 school year</b>	<b>No-treatment mean (st. error)</b>	<b>Treatment mean (st. error)</b>	<b>t-stat that true means are equal</b>
D1 - meals	88.8 (0.8)	88.0 (2.0)	-0.40
D2 - meals	81.8 (1.1)	79.2 (2.1)	-1.16
D3 - meals	74.9 (1.3)	67.5 (2.2)	<b>-2.97**</b>
D4 - meals	61.2 (1.5)	62.2 (2.0)	0.40
D5 - meals	55.6 (1.7)	52.9 (2.2)	-0.95
D1 - ell	54.6 (1.2)	51.4 (2.8)	-1.05
D2 - ell	45.7 (1.4)	43.2 (2.4)	-0.90
D3 - ell	36.5 (1.3)	32.7 (2.1)	-1.62
D4 - ell	26.4 (1.3)	25.0 (1.7)	-0.61
D5 - ell	21.8 (1.3)	20.5 (1.9)	-0.57
D1 - full	69.1 (0.9)	71.4 (2.4)	1.0
D2 - full	76.8 (0.9)	79.6 (2.2)	1.37
D3 - full	79.3 (1.1)	82.9 (1.4)	<b>1.97*</b>
D4 - full	85.1 (0.9)	88.1 (1.4)	1.76
D5 - full	85.6 (1.0)	90.9 (1.3)	<b>3.07**</b>
D1 – valid_num	718.0 (35.9)	497.8 (43.8)	<b>-2.65**</b>
D2 – valid_num	590.3 (35.2)	473.3 (30.9)	-1.78
D3 – valid_num	585.9 (36.5)	453.7 (23.7)	<b>-2.30*</b>
D4 – valid_num	502.8 (28.1)	423.0 (25.1)	-1.75
D5 – valid_num	522.2 (34.9)	437.3 (44.7)	-1.43
D1 – pct_hi	74.0 (1.5)	67.1 (3.6)	-1.90
D2 – pct_hi	65.3 (1.8)	59.5 (3.1)	-1.56
D3 – pct_hi	57.5 (1.6)	53.7 (2.7)	-1.28
D4 – pct_hi	43.7 (1.8)	43.2 (2.5)	-0.16
D5 – pct_hi	40.2 (1.6)	36.1 (2.6)	-1.38
D1 – avg_ed	1.97 (0.03)	1.89 (0.06)	-1.04
D2 – avg_ed	2.15 (0.03)	2.11 (0.04)	-0.85
D3 – avg_ed	2.33 (0.03)	2.27 (0.04)	-1.13
D4 – avg_ed	2.45 (0.03)	2.41 (0.04)	-0.85
D5 – avg_ed	2.63 (0.03)	2.56 (0.04)	-1.22

D1 – mobility	20.0 (0.8)	21.4 (0.7)	0.71
D2 – mobility	18.5 (0.8)	19.3 (1.6)	0.47
D3 – mobility	18.8 (0.9)	21.4 (1.9)	1.44
D4 – mobility	19.2 (1.0)	19.6 (1.5)	0.18
D5 – mobility	18.4 (0.9)	16.5 (0.9)	-1.34
D1 – pct_aa	15.3 (1.2)	17.1 (3.0)	0.59
D2 – pct_aa	14.5 (1.5)	16.1 (2.4)	0.54
D3 – pct_aa	14.7 (1.4)	10.8 (1.4)	-1.71
D4 – pct_aa	11.7 (1.2)	13.7 (1.8)	-0.88
D5 – pct_aa	12.6 (1.4)	11.8 (1.6)	-0.31
D1 – pct_as	3.7 (0.5)	6.9 (1.6)	2.57*
D2 – pct_as	6.6 (0.8)	6.3 (1.0)	-0.22
D3 – pct_as	6.6 (0.7)	8.3 (1.3)	1.27
D4 – pct_as	6.8 (0.6)	5.1 (0.8)	-1.53
D5 – pct_as	7.1 (0.9)	6.7 (1.0)	-0.23

In terms of the schools' percentage of fully credentialed teachers, as well as the schools' size, the treatment group is significantly different from the non-treatment group, again in the direction expected if the II/USP schools were from higher API deciles.<sup>52</sup> The t-statistics in bold and with double asterisk indicate that the II/USP cohort 1 schools are smaller and have a greater proportion of fully credentialed teachers, at a 1% confidence level. Such a systematic difference would be highly unlikely under the assumption of randomization within the identified decile rankings. A possible source of this difference is the underrepresentation of schools from LAUSD, as the description of the selection to cohort 1 above indicated. Though the 350 spots for II/USP in cohort 1 were filled according to the restrictions set out by law, only one fifth of the applicant schools from LAUSD were considered for selection. This systematic difference should be controlled for, if there is reason to believe that LAUSD schools are significantly different from other applicant schools with the same grade span and API decile ranking. Table 13 below shows the comparison for elementary schools in decile 1 as a representative group which II/USP schools were selected to. The table lists not only the baseline for the six variables described above, but also for the measures of school performance in levels and pre-

<sup>52</sup> The total number of students enrolled at the school, as well as the percentage of students in the school that were tested, was also examined. This was to confirm that *valid\_num* is indeed a good proxy for school size and that the difference seen between the treatment and non-treatment schools is not a function of the schools in the treatment group being more likely to exclude low performing children from taking the year-end tests. As the additional analyses have indicated, there is no significant difference in the schools' propensity to test their students and *valid\_num* appears to be a good measure for school size, as well as for the differential variance in the test score averages that make up the API.

selection growth as they were discussed earlier. The comparison shows clearly that the 89 decile 1 elementary schools from LAUSD are significantly different from the 164 decile 1 elementary schools from other districts.

**Table 13 – Decile 1 elementary schools from LAUSD and other applicants**

<b>1999/00 school year</b>	<b>Decile 1 elementary schools, not LAUSD</b>	<b>Decile 1 elementary schools from LAUSD</b>	<b>t-stat that true means are equal</b>
99 – meals	91.1 (0.8)	96.8 (0.4)	<b>5.51**</b>
99 – ell	56.5 (1.6)	65.7 (1.7)	<b>3.70**</b>
99 – full	74.4 (1.4)	62.6 (1.0)	<b>-5.87**</b>
99 – valid_num	393.9 (12.0)	617.1 (26.0)	<b>8.82**</b>
99 – pct_hi	71.2 (2.1)	83.1 (2.0)	<b>3.85**</b>
99 – avg_ed	1.84 (0.0)	2.0 (0.1)	<b>2.96**</b>
API base 1999	411.8 (2.3)	402.3 (3.4)	<b>-2.368**</b>
MathSS 1999	584.8 (1.1)	580.1 (0.6)	<b>-3.13**</b>
ReadingSS 1999	587.4 (1.1)	579.9 (0.6)	<b>-5.02**</b>
MathGrowth 98-99	4.6 (0.7)	1.6 (0.6)	<b>-2.96**</b>
Read.Growth 98-99	2.9 (0.6)	-1.9 (0.6)	<b>-4.99**</b>

The same comparison across all grade spans and API decile (not listed separately) shows that these differences are representative of the overall trend in the difference of schools between LAUSD and other districts. Among the 305 LAUSD schools that applied for II/USP cohort 1 (for which data are available), only 11 schools were selected, or roughly 3%.

An appropriate model for the estimation of the interventions’ impact will thus have to take into account that the schools in the treatment group are, on average, larger (due to higher proportion of urban schools) and are faced with more serious resource constraints (fewer fully credentialed teacher, while a higher proportion of the student body is poor, English language learners, from Hispanic background, and with less educated parents).

#### **4.4.3 Selection of cohort 2 and 3**

For the selection of cohorts 2 and 3, the API decile rankings were available, as were the API growth numbers. Eligibility for II/USP was thus based on schools having an API score in the lower 5 deciles *and* schools failing to meet their API growth targets.<sup>53</sup> Given these two criteria for eligibility and the fact that 430 schools were no longer

<sup>53</sup> This group includes schools that met their schoolwide target, but failed to meet targets for their subgroups.

eligible because they had been selected for II/USP and CSR programs in cohort 1, the number of eligible schools and number of schools that applied for inclusion in the programs was smaller for the second and third cohorts. For cohort 2, there were 936 eligible schools of which 528 applied for the II/USP funding. For cohort 3, there were 1266 schools that were deemed eligible by the CDE and 751 of those applied. For each cohort, the CDE selected 430 schools by lottery, but again only within the restrictions set out by law regarding an even distribution of treatment schools across deciles and a specific breakdown of elementary, middle, and high schools.

The federally funded CSR program became an option for schools in cohort 2 only as the planning year was underway during which all 430 schools received the \$50,000 planning grant. In the fall of 2001, 47 schools out of the 430 schools in cohort 2 switched to the federally funded CSR program. As there were more CSR spots available than applicants, the switch to CSR was a non-competitive process (Colton, 2004b). Since all schools in cohort 2 were selected in the same way and CSR and II/USP are quite similar, the CSR schools are included in the analysis for cohort 2.

The selection to cohort 3 is far more complicated, because a number of factors combined to make the identification of suitable comparison groups to the treatment schools very difficult (Colton, 2004a). The selection was tied up with the funding available from different sources, the application behavior of the schools eligible for the interventions, as well as the introduction of HP as a program with different eligibility and selection criteria. Given these analytical challenges at the outset, the estimations below are focused on the impact of the interventions on schools in cohort 1 and 2.

#### *4.4.4 Characteristics of cohort 2 and comparison schools*

In this section the basic characteristics of schools in cohort 2 of II/USP are described, analogous to the description of cohort 1 schools and their comparison schools above, albeit in less detail.

An initial breakdown by grade span between schools that were selected for cohort 2 and schools that applied, but were not selected, shows that there are no elementary schools in the second group. The reason for this is that fewer elementary schools applied (252) than SB1X stipulates to be the appropriate number (301). As a result, all

elementary school applicants were selected, and only middle and high schools are among the applicants that were not selected, as Table 14 shows.

**Table 14 – Base-line breakdown by grade span for cohort 2**

<b>2000 API base</b>	<b>No-treat. – mean (st. error) – #obs</b>	<b>Treatment – mean (st. error) – #obs</b>	<b>t-stat</b>	<b>Probability of selection</b>
<i>Overall</i>	499.9 (6.7) – 92	524.5 (3.6) – 423	2.94	0.82
high schools	512.1 (8.6) – 54	541.3 (6.8) – 69	2.70	0.56
middle schools	482.6 (10.2) – 38	533.0 (7.1) – 102	3.80	0.72
elem. schools	--	516.42 (4.9) – 252	--	1

Since all elementary schools that applied were selected for cohort 2, the sample is restricted to middle and high schools. Elementary schools are excluded from the analysis, since the elementary schools in cohort 2 do not have any non-treated applicant elementary schools to be compared against. From the overall average of the baseline API score (since cohort 2 was selected in the fall of 2000, the 2000 API base is the appropriate baseline), a difference in average student performance among the schools in the two groups can be made out, similar to the difference in cohort 1. Table 15 shows the breakdown by API decile ranking only for middle and high schools, without elementary schools. From the calculated probability of selection, the same mechanics as for cohort 1 can be observed: schools from lower deciles have a lower probability of selection to the program due to the higher number of applicants among them. As a result, the overall average student performance between the two groups as a whole is different. Within the decile rankings, the differences are significant only for decile 2 schools.

**Table 15 – Base-line breakdown by decile for cohort 2**

<b>2000 API base</b>	<b>No-treat. – mean (st. error) – #obs</b>	<b>Treatment – mean (st. error) – #obs</b>	<b>t-stat</b>	<b>Probability of selection</b>
<i>Overall</i>	500.6 (6.8) – 91	536.4 (5.1) – 171	4.20	0.65
decile 1	443.1 (5.5) – 40	438.3 (5.2) – 38	-0.63	0.49
decile 2	504.7 (3.1) – 22	513.5 (2.6) – 37	2.14	0.63
decile 3	554 (4.2) – 14	553.9 (2.2) – 40	-0.02	0.74
decile 4	588.7 (3.2) – 11	591.9 (2.1) – 34	0.75	0.76
decile 5	623.5 (6.6) – 4	626.4 (3.0) – 22	0.38	0.85

With some knowledge of how the interventions were implemented, how the schools were chosen to participate in the intervention programs, and what the issues are in picking the appropriate group of schools for comparison, the following sections

address the actual analysis of the available data. The next section proposes specific models for estimating the interventions' impact.

## **4.5 Methods**

Here, the models for an appropriate estimation of the interventions' impact are proposed and discussed, preceded by a brief review of the key concepts in program evaluation.<sup>54</sup>

### **4.5.1 Key concepts for program evaluation methods**

The primary goal of any program evaluation is to estimate the difference in outcomes between a world where the unit (e.g., an individual school) participates in the program and where it does not. The evaluation problem then consists of estimating whichever outcome is not observed in the data.<sup>55</sup> Estimating the impact of the treatment (e.g., II/USP intervention) requires finding a good approximation for the outcome that the treated units would have experienced had they not received the treatment. Finding a proxy for this counterfactual remains the central dilemma in program evaluation. Using the average outcomes of the non-treated units as such a proxy raises the question of whether the two groups (treated and non-treated) differ in ways that affect the outcome other than through the treatment and that are not accounted for with covariates. Any systematic difference between the two groups that is not accounted for explicitly in the model might lead the estimation to attribute differences in the data to the treatment, the known difference between the groups, when they are really a result of unknown differences leading to omitted variable bias in the parameter estimate.

As the detailed presentation of the selection procedure for the II/USP interventions and the comparison of the composition of cohort 1 schools and the applicant pool suggest, a random selection is not a valid simplification of how the schools were chosen to participate overall, even if within the given restrictions the choice was random. Failing to account for these restrictions explicitly in the model would introduce a bias in the estimate of the treatment effect.

---

<sup>54</sup> See, for example, Smith (2004) for a comprehensive review of the program evaluation literature.

<sup>55</sup> Though indirect effects on non-treated units or on the entire system may be important also, these are addressed only in the concluding discussion section.

It should also be noted that any estimated effect has external validity only for the applicant schools. Since the decision to apply for the intervention remains somewhat unclear, with individual schools being volunteered by their district (Just et al., 2001), the extent to which the schools' motivation played a role in the application cannot be accurately assessed.

The following section proposes a set of models that are employed to estimate the impact of the interventions or to identify any change in outcomes between the two groups that cannot be attributed to any other differences.

#### 4.5.2 *Basic models to estimate the interventions' impact*

If the assignment to the II/USP intervention, or the treatment, had been truly random, the estimated impact of the treatment would simply be the difference in the average outcome between treatment and non-treatment schools at a point in time after the schools have received the treatment. Any difference in the outcome that would be too large to have happened by chance (i.e. less than 5% probability) could be said to be the impact of the treatment. Given that there were restrictions within which the assignment was random, the difference in mean achievement between II/USP schools and comparison schools can be assessed taking into account the selection bins that SB1X prescribes by using a simple model to compare the means of school performance after the intervention has begun:

$$(1) \quad y_{it} = \beta_1 T_{it} + bins_i \beta_2 + X_{it} \beta_3 + \lambda_i + \varepsilon_{it}$$

where  $y_{it}$  is the API (level) of school  $i$  at a time  $t$  after the intervention has taken effect (in the calculations below, this is either 1 or 2 years after the selection into II/USP);  $T_{it}$  is a dummy variable equal to one if the school was selected for II/USP;  $bins_i$  is a vector of dummy variables that completely exhausts the set of selection bins (deciles x grade span);  $X_{it}$  is a vector of time-varying covariates listed in Table 11; and  $\lambda_i + \varepsilon_{it}$  is the error term consisting of a school-specific component,  $\lambda_i$ , and an idiosyncratic component  $\varepsilon_{it}$ . The parameter of interest is  $\beta_1$ , which measures the difference in average school level student performance between the treated and the non-treated schools. If  $y_{it}$

is set to be the SAT-9 scaled scores in reading or math used in the tables earlier, this same model applies, with an analogous interpretation of  $\beta_1$ .

If the treatment was truly random within bins, then we can assume  $\text{cov}(\lambda_i + \varepsilon_{it}, T_{it}) = 0$ , and  $\beta_1$  is an unbiased estimate of the effect of the II/USP intervention on average test scores. Note that while the inclusion of  $\text{bins}_i$  is necessary to produce an unbiased estimate of  $\beta_1$ , the inclusion of the time-varying covariates  $X_{it}$  is not. If, however,  $\text{cov}(\lambda_i + \varepsilon_{it}, T_{it}) \neq 0$ , then  $\beta_1$  is biased. This potential bias can be addressed by estimating (1) in first differences:

$$(2) \quad y_{it} - y_{i,t-1} = \beta_1 T_{it} + (X_{it} - X_{i,t-1})\beta_2 + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

where  $y_{it} - y_{i,t-1}$  is the difference between the pre-treatment or baseline test score average,  $y_{i,t-1}$ , and the school's test score average after the treatment,  $y_{it}$ . The parameter of interest,  $\beta_1$ , now measures the impact of the treatment on within-school changes in test scores. Note that  $\lambda_i$  and any other fixed school characteristic, such as  $\text{bins}_i$ , drops out of the model.

If the school-specific component of the error term is time-varying (as opposed to the time-invariant component,  $\lambda_i$ ), it would not drop out of the first-differenced model. There are two approaches to dealing with time-varying heterogeneity. The first is to assume that the inclusion of time-varying covariates in the model,  $X_{it}$ , or their first differences,  $(X_{it} - X_{i,t-1})$ , solves the problem. A second approach is to first difference (2) in order to control for the influence of school-level unobserved factors correlated with both the growth in test scores and the propensity to receive the treatment:

$$(3) \quad \Delta_t y_i - \Delta_{t-1} y_i = \beta_1 T_{it} + (\Delta_t X_i - \Delta_{t-1} X_i)\beta_2 + (\Delta_t \varepsilon_i - \Delta_{t-1} \varepsilon_i)$$

where  $\Delta_t y_i - \Delta_{t-1} y_i$  is the difference in the pre-treatment test score growth and the test score growth at time  $t$  after the treatment; and, analogously,  $(\Delta_t X_i - \Delta_{t-1} X_i)$  is the difference between the pre- and post-treatment change in the time-varying covariates.

Pre-treatment growth for cohort 1,  $\Delta_{t-1}y_i$ , is available only with SAT-9 data, rather than API growth. For cohort 2, the pre-treatment growth between 1999 and 2000 is available for SAT-9 scores and API growth. It should also be noted that the differencing of API scores over multiple years is not straightforward, as described earlier. Therefore, the dependent variable,  $y_{it} - y_{i,t-1}$ , used for estimations with the API is the API growth supplied by the CDE for either a single year or the sum of the two relevant, yearly API growths over two years.

The overrepresentation of schools from LAUSD in the comparison group has not been addressed specifically. Aside from estimating the models *with* the covariates, a second option is to estimate the interventions' impact on schools from all districts *other than* LAUSD by excluding all LAUSD schools from the sample. This would eliminate 305 schools, or roughly 30%, from the non-treatment group and 11 schools, or roughly 3%, from the treatment group. Such an elimination of specific schools from the sample introduces a non-random variation itself. Since most of the 305 schools from the non-treatment group were not effectively considered for selection, it may provide an improved set of schools for comparison to the remaining treatment schools.

Since all models presented here involve school level averages of outcome variables, all estimations of these models use a weighted least squares approach with weights according to the number of valid test scores that were used to calculate the averages.

Below, the results are presented from estimating models (2) and (3) using the available API and SAT-9 data. Results for model (1) are not reported, since the above analysis of the characteristics of treatment and control groups has shown that the selection, even within the selection bins, was probably not random.

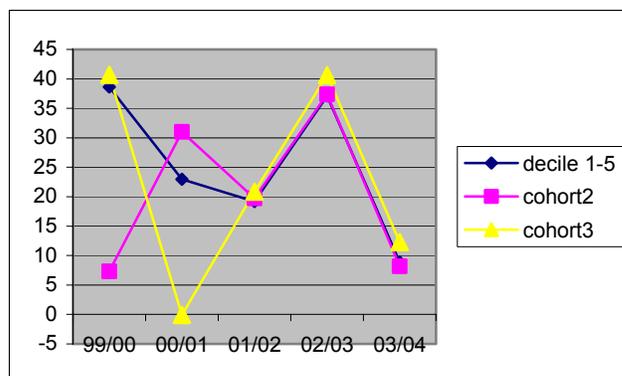
#### **4.5.3 Feasible scope of the estimations**

The models are estimated primarily for cohort 1 using the data of the first two years after the selection, 1999-2001. The focus is on cohort 1, because the selection procedure and the larger number of eligible and applicant schools gives the estimations

greater statistical power. Selected additional estimations are conducted with cohort 2 schools, also using the first two years after the selection, 2000-2002.

The comparison of the outcomes of the treated and non-treated group in later years, after 2001 for cohort 1, becomes less tractable, because the two additional II/USP cohorts were selected in the fall of 2000 and 2001, respectively, and many of the schools in cohort 2 and 3 applied, but were not selected for the intervention in 1999. This means that the non-treatment group from 2001 onward contains schools that are implementing II/USP in the new cohorts.<sup>56</sup> The presence of the two new cohorts leaves the evaluation with two basic choices both of which are problematic. First, the additional cohorts might be disregarded as new interventions, and an evaluation could be seen as a lower bound estimate for an impact of II/USP on cohort 1 schools. A second option that appears more attractive at first glance is to eliminate any schools that were selected for cohort 2 and 3 from the non-treatment group. This second option is itself problematic, however, because the selection criteria changed after cohort 1 was selected. Schools in cohort 2 and 3 became eligible for II/USP in 2000 and 2001, because they failed to meet their API growth targets in 1999/00 and 2000/01, respectively. As the graph in Figure 4 shows, the development of the 5-year API growth is quite different for schools from cohorts 2 and 3 than for the other schools in the lower five deciles.

**Figure 4 - Graph of API growth**



As expected, cohorts 2 and 3 show a dramatically lower API growth in 1999/00 and 2000/01, respectively. Thus, the exclusion of cohort 2 and 3 from the non-treatment

<sup>56</sup> In the school year 2000/01, the second school year which is analyzed for cohort 1, cohort 2 is already developing an Action Plan. The challenge of selecting appropriate comparison groups beyond two years after the selection to the intervention is analogous for cohort 2.

group for the analysis of the interventions' effects on cohort 1 would directly influence the group's average outcome of interest (i.e. API growth). The report by the American Institutes for Research (AIR) that is reviewed below reasons that

excluding comparison schools that subsequently participated in II/USP Cohort 2 from the comparison group for Cohort 1 would bias the apparent "II/USP effect" downwards in 2000, since it would remove unusually low-performing schools from the comparison group that year – schools that failed to meet their achievement targets and were thus eligible for participation in II/USP that year. Excluding comparison schools that subsequently entered II/USP Cohort 3 would remove unusually low-performing schools from the comparison group for both Cohort 1 and 2 in 2001, for similar reasons. On the other hand, retaining comparison schools for Cohort 1 that subsequently participated in II/USP Cohort 2 would likely bias the apparent "II/USP effect" downwards in 2001 and 2002, since the comparison group would include some schools that received the achievement benefits, if any, of participation in II/USP in these years. Similarly, retaining comparison schools for Cohort 1 or 2 that participated in II/USP Cohort 3 would bias the "II/USP effect" downward in 2002. (O'Day and Bitter, 2003)

With these limitations in mind, the estimations, for which the results are presented and discussed below, use only the first two years of post-selection data for cohort 1 and cohort 2.

Given the available data and the fact that the parameter of interest, i.e. the impact of participation in II/USP, should theoretically be the same for schools in different cohorts, it is tempting to employ an additional model of estimation: a Difference-of-Differences (DoD) model using the full panel of data. With a single treatment effect to be estimated, across all three cohorts, and the treatment spread across many units, beginning and ending at different times, with some units never treated, a DoD model seems at first like the logical approach. However, the same problem as discussed above renders such a model inappropriate. Since the initiation of the treatment for cohort 2 and 3 is directly related to a temporary change in the outcome variable (i.e. not having made the API growth targets the previous year), the DoD approach would return biased estimates with a dramatically overestimated impact of the interventions. The results of such a model do not provide much information on the actual impact of the interventions and could only be interpreted as a confirmation of the methodological critique by Chay et al. (2004) described in Chapter Two. Interventions for which schools with lower growth in the

selection year become eligible appear to have large positive effects when mean reversion of the outcome measure is not properly accounted for in the estimated models.

## **4.6 Literature Review**

Two different sets of researchers have attempted to quantify the effect of II/USP interventions on the performance of California public schools, as measured by student test scores: Julian Betts and Anne Danenberg in a conference paper that was recently published as a chapter in a book and the American Institutes of Research (AIR) under the leadership of Jennifer O'Day and Catherine Bitter. These two analyses are reviewed in this section.

### **4.6.1 Betts & Danenberg (2003)**

A number of authors have contributed to the book *No Child Left Behind* a collection of papers on education systems that have implemented the new accountability embraced in the new federal law. Julian Betts and Anne Danenberg (2003) offer a chapter on the accountability framework in California and on the impact of II/USP on low-performing schools in particular.

A first question that Betts and Danenberg address is whether the introduction of the new accountability framework under SB1X has impacted the distribution of school resources. As proxies for school resources they examine the changes in teacher characteristics, such as the percentage of teachers with full credentials and the percentage of teachers with teaching experience. While they find strong evidence that the gaps in resource allocation or teacher qualification between the schools with the highest levels of achievement and schools with low achievement have widened between 1995 and 2001, they do not find evidence that the introduction of SB1X exacerbated this trend. Along with a widening gap in teacher qualifications, the gap in the percentage of students qualifying for reduced price or free lunch is similarly increasing between the highest scoring schools and the schools with the lowest average achievement.

These findings are particularly important in light of the earlier question regarding the nature of the schools' failure that the II/USP program is designed to correct. If resources to schools in lower deciles in terms of teacher qualifications and experience are *decreasing* and the challenges that these schools face in terms of the relative socio-

economic background of the student population are *increasing* at the same time, the schools in the lower deciles will, all else equal, struggle to meet performance expectations that are set independent of these underlying trends. At least part of the failure of the lower decile schools will, as these trends suggest, be caused by structural problems that will be impossible to remedy at the schools in an individualized intervention over three years.

Betts and Danenberg also examine a problem of evaluations of other interventions, namely the self-selection of schools. Whenever schools can directly influence their inclusion in an intervention program, unobservable aspects of motivation or other school characteristics make it difficult to arrive at an unbiased estimate of the intervention's impact. To provide some answers to this problem, Betts and Danenberg examine the incentives that eligible schools face when deciding whether to apply for II/USP intervention.

While the question they are asking is relevant to accountability frameworks more broadly, it does not take into account the reality of II/USP applications which were submitted by the districts rather than by the schools themselves. As discussed earlier, the districts have been noted to submit schools as applicants without giving the schools the choice to make the decision themselves (Gerst, 2004b; Just et al., 2001).

The authors note that the applicants to II/USP are disproportionately low-performing schools, but that the group of selected schools is spread evenly across the lower five deciles. In noting this pattern the authors confirm the findings offered above. They conclude that “the lottery system through which schools were selected effectively evened out the distribution across deciles as shown,” without referring to the fact that SBIX imposes the even distribution as a legal mandate (Betts and Danenberg, 2003).

Despite their careful examination of the selection problems, the assumptions for the model they choose to estimate the impact of II/USP interventions on underperforming schools remain somewhat unclear. Betts and Danenberg examine the two-year growth in API scores between spring 1999 and spring 2001.<sup>57</sup> They compare the scores between the

---

<sup>57</sup> The earlier and more expanded version of their paper, prepared for the June 2002 conference “Taking Account of Accountability: Assessing Politics and Policy,” indicates that the actual scores used were SAT-

cohort 1 II/USP schools and cohort 1 CSR schools (treatment) with the group of schools that applied, but was not selected for the II/USP intervention (non-treatment). The authors acknowledge that “it seems almost unfair to expect that II/USP reforms could bear fruit in two years, given that participating schools spend their first year planning the reforms that take root only in the second year.” Despite these reservations the paper’s results provide a significant positive estimation of the intervention’s effect over these two years. The model that the authors use includes dummy variables for grade span and a binary variable for whether the school was selected for cohort 1. The coefficient of this latter variable provides the estimate for the parameter of interest. The estimated impact of being selected for cohort 1 of either II/USP or CSR is a difference of 17.5 points in the combined API growth over the two years (99/00 and 00/01). As the effect is significant at the 1% level and represents roughly one-third of the average API growth over the two years, this is a noteworthy result.

A problem with the model that Betts and Danenberg use follows directly from the discussion above regarding the selection of the cohorts. Treating the CSR and II/USP schools, on the one hand, and the schools that applied but were not selected, on the other hand, as being the same except for the treatment is not supported by the available evidence. The assertion that the remaining applicants provide “a valid comparison group because these schools exhibited similar motivation in that they applied but did not win the randomized selection process” is incorrect, because the selection process was decidedly non-random for the group as a whole. The selection of CSR schools followed a process that is no longer available for analytical purposes. For the remaining 350 II/USP schools, the restrictions set out by law established a treatment group with a higher proportion of high-performing schools compared to the rest of the applicants.

#### **4.6.2 *O’Day and Bitter (2003)***

The second analysis of the II/USP interventions’ effect was commissioned by the CDE and conducted by the American Institutes for Research (AIR), in partnership with Policy Analysis for California Education (PACE) and EdSource, a team of researchers led by Jennifer O’Day and Catherine Bitter (2003). Their report offers a comprehensive

---

9 reading and math scores, as a proxy for API scores. SAT-9 scores in reading and math constitute a large part of the API score for the two years in question, see Table 6, but they are not identical.

evaluation of the II/USP interventions and of the impact of SB1X on public education in California more broadly. Overall, the authors find that the implementation of the new accountability system has focused attention on student achievement and on the issue of low performing schools in particular. As a consequence of this focus, they also note a reported tendency in schools to neglect non-tested subjects and developmental needs of students not directly related to measurable achievement. With respect to school performance, as measured by student achievement, the report states:

Against the backdrop of very large increases in [student achievement] scores in the state, the direct additional contributions of II/USP ... to mean achievement across participating schools has been negligible. Small planning year “bumps” in achievement growth during the planning year tend to dissipate in subsequent years, being washed out by substantial variation among II/USP schools and by powerful district effects on student performance.

The emphasis on the dramatic increase in student test scores provides an important context within which any II/USP effects might be detected. The report notes that the familiarization with the standardized tests, an increased focus on the results of these tests, and additional effort devoted to achieving high scores have all contributed to the broader trend of rising test scores.

Regarding the effect that II/USP interventions specifically had on test scores, the report’s findings provide a challenge to the estimates offered by Betts and Danenberg. The AIR report uses SAT-9 test scores at the student level directly, as well as analyzing API data at the school level. In order to take account of the hierarchical structure of the data, the authors employ Hierarchical Linear Modeling (HLM).<sup>58</sup> As the summarizing quote above indicates, they estimate a small positive and significant effect of participation in II/USP during the first year (planning) and no significant impact of II/USP on school performance in later years. Like the analysis by Betts and Danenberg, the AIR report takes account of the fact that the API growth differs among the grade spans. Instead of controlling for different levels of API growth within a single model, as Betts and Danenberg did, they analyze elementary, middle, and high schools in separate models.

---

<sup>58</sup> For a basic introduction to HLM, see, for example, Osborne (2000).

The report examines not just the effect of II/USP on test scores, but offers an in-depth discussion of the interventions more generally, including case studies and participants' surveys. Given their thorough analysis of the II/USP process, one particular similarity to the analysis by Betts and Danenberg stands out: the differences of schools across the deciles is assumed to be negligible. In the case of the AIR report, this view is all the more surprising as the importance of picking an appropriate comparison group is clearly recognized and the selection of the II/USP schools, as well as the properties of the group of schools that applied but did not get selected, is discussed in detail. O'Day and her colleagues (2003) concede the problem addressed earlier:

Although Cohort 1 II/USP schools were selected at random, the selection process was conducted separately within SAT-9 deciles, and the selection rates apparently differed somewhat across deciles. In particular, the proportion of schools selected for participation was somewhat lower in the first decile than in the fifth, resulting in a somewhat higher overall mean API score among schools selected for participation than among schools not selected.

Despite the recognition of the problem, the authors conclude that the difference in the probability of selection (e.g., 17% for decile 1 vs. 32% for decile 5) differed only "somewhat," and that the difference in mean API score of 21 points between the treatment and non-treatment group is negligible. It may be true that the systematically different composition of treatment and non-treatment group according to the differential probability of selection has no significant impact on the difference in outcomes. From the results provided in the AIR report alone, however, that determination cannot be made with confidence.

For cohort 1 II/USP schools, the AIR report estimates a positive impact of being selected on the API of 8.7, 10.8, and 7.8 for elementary, middle, and high schools, respectively. These impacts are for the first year only, however, with no significant impacts detected in the second year (i.e. the first year of II/USP implementation) in either one of the three grade spans. Also, the authors note that the first-year increases (all significant at the 1% level) amount to only 1/8 or 1/9 of a standard deviation of API growth. Though the impact of the first two years combined is not separately calculated, it amounts clearly to less than the Betts and Danenberg estimate of the impact on API growth over the first two years. In contrast to the relatively simple model that Betts and

Danenberg employed, the HLM model in the AIR report controls for the effects of student characteristics, either with averages at the school level for the analysis of API scores (percent Black, Asian, Hispanic; percent of students qualifying for free or reduced lunch; percent English language learners; percent of student first attending the school in the current year; average education level of the parents) or at the level of the individual student for the analysis of SAT-9 test scores. The model also controls for the percentage of teachers with full credentials.

The first-year bump in test scores may be due to the treatment or due to the fact that the two groups are not the same in terms of their average API decile ranking. In addition, the overrepresentation of schools from LAUSD in the comparison group is not considered in the report, although the impact of district effects on the results is found to be much larger than any effects that were attributed to the treatment itself.

The choices by O’Day and her colleagues in choosing appropriate comparison groups are somewhat contentious. The authors acknowledge this for cohort 2 and 3 when stating that their choices of comparison groups for cohort 2 and 3 “are more prone to selection bias” (O’Day and Bitter, 2003). They choose to expand the comparison group to the full group of eligible schools. The issue that matters for cohort 2 is the disproportionate number of schools from LAUSD that were not selected for cohort 1. These schools remained eligible for selection in cohort 2, if they missed the API targets, and were required to apply for II/USP following the establishment of the SB1X office at LAUSD in 2000. This resulted in a disproportionate number of applicant schools from LAUSD among the eligible schools for cohort 2. No LAUSD schools, in other words, are in the group that was added to the comparison that O’Day et al. used (eligible, but did not apply). For cohort 3, the selection of HP schools is not mentioned in the AIR report. The omission matters, because more than 130 schools that were eligible, but were not selected for II/USP were instead selected to HP with an identical planning year to II/USP and funded at \$400 per student thereafter.

#### **4.7 Results**

Using the available API and SAT-9 data, this section presents the results of the estimations of the models introduced earlier. In a first step, the mean values of the

relevant outcome variables for cohort 1 are compared between the treatment and non-treatment group. After these simple comparison of means the models (2) and (3) are estimated for cohort 1, followed by the estimations with the proposed correction for the LAUSD mishap (LAUSD exclusion). A selected set of estimations for cohort 2 are presented next, prior to a brief summary of the results.

#### 4.7.1 Post-treatment comparison of means

The following tables show the post-treatment difference of the average student achievement between the treatment and the non-treatment schools after one and two years, as measured by the API level and growth. The tables list the differences between the two groups of API and SAT-9 levels after 1 and 2 years of the intervention, as well as the difference of one-year and two-year growth in API and SAT-9.

Note that comparison of score levels (first half of the table) is similar to model (1), but are only broken down by API deciles rather than by the exhaustive selection bins. In contrast, the comparison of score *growth* is equivalent to the estimation of model (2) without the covariates. The comparison of the overall score growth (i.e. the first line in the lower half and lower quarter of the table) returns the identical results as the regression results in Table 19.

**Table 16 – API comparisons**

<b>API</b>	<b>Mean differences Treatment – No-treatment</b>	<b>t-stat</b>
<i>2000 API level</i>	27.1 (4.9)	5.54**
decile 1	5.2 (6.1)	0.86
decile 2	6.2 (4.7)	1.33
decile 3	5.3 (4.7)	1.11
decile 4	4.4 (4.7)	0.93
decile 5	14.3 (5.0)	2.84**
<i>2001 API level</i>	23.3 (4.6)	5.09**
decile 1	12.1 (6.9)	1.75
decile 2	4.0 (5.6)	0.71
decile 3	3.8 (5.6)	0.68
decile 4	-1.08 (5.2)	-0.21
decile 5	12.9 (6.1)	2.10*
<i>99-00 growth</i>	6.2 (2.1)	3.00**
decile 1	5.6 (4.5)	1.25
decile 2	3.7 (4.5)	0.82
decile 3	3.4 (4.8)	0.72
decile 4	3.5 (4.8)	0.74

decile 5	16.6 (4.8)	3.43**
<i>99-01 growth</i>	3.2 (2.8)	1.15
decile 1	9.7 (7.7)	1.44
decile 2	4.4 (6.1)	0.72
decile 3	1.3 (6.4)	0.21
decile 4	-3.64 (5.8)	-0.63
decile 5	15.1 (6.3)	2.40*

The comparison of API level and growth show a consistent pattern of slight score increases for the treatment group. For the individual deciles, the increases are significant only for schools in decile 5. The implications of this uneven impact across the deciles that seems to drive the significance of the overall impact is discussed in greater detail below.

The results for SAT-9 math and reading are listed in Table 17 and Table 18 below, also for 1- and 2-year level and growth. The pattern for SAT-9 math is similar to the API comparison, with generally higher levels and growth across the decile groups, though not significant, and with strongly significant differences for the groups overall.

**Table 17 – SAT-9 math post-treatment comparisons of means**

<b>SAT-9 math</b>	<b>Mean differences Treatment – No-treatment</b>	<b>t-stat</b>
<i>2000 level</i>	4.9 (2.1)	2.32*
decile 1	0.1 (4.5)	0.02
decile 2	7.3 (4.5)	1.60
decile 3	4.4 (4.4)	0.99
decile 4	-2.0 (4.3)	-0.45
decile 5	-0.5 (4.9)	-0.09
<i>2001 level</i>	5.2 (2.0)	2.54*
decile 1	0.7 (4.3)	0.15
decile 2	6.6 (4.4)	1.51
decile 3	4.5 (4.3)	1.05
decile 4	-1.8 (4.1)	-0.44
decile 5	1.7 (4.8)	0.35
<i>99-00 growth</i>	1.1 (0.4)	2.57*
decile 1	1.3 (0.9)	1.40
decile 2	1.2 (0.9)	1.25
decile 3	1.0 (0.9)	1.09
decile 4	0.8 (1.0)	0.84
decile 5	0.9 (1.1)	0.81
<i>99-01 growth</i>	1.4 (0.5)	2.60**
decile 1	1.8 (1.1)	1.65
decile 2	0.3 (1.1)	0.31
decile 3	1.1 (1.1)	0.97
decile 4	1.0 (1.3)	0.77

decile 5	3.0 (1.2)	2.49*
----------	-----------	-------

For the comparison of SAT-9 reading scores the pattern is less clear and even the overall growth does not appear to be significantly different between the treatment and the non-treatment group. The baseline of SAT-9 reading scores Table 8 above revealed an approximate difference of 5 points between the average reading scaled score in the treatment and the non-treatment group. This difference is roughly maintained, with no differential gains for treatment schools compared to the comparison schools indicated by these results.

**Table 18 – SAT-9 reading post-treatment comparisons of means**

<b>SAT-9 reading</b>	<b>Mean differences Treatment – No-treatment</b>	<b>t-stat</b>
<i>2000 level</i>	4.9 (2.0)	2.38*
decile 1	-0.1 (4.3)	-0.02
decile 2	6.5 (4.3)	1.51
decile 3	3.5 (4.2)	0.82
decile 4	-2.0 (4.2)	-0.48
decile 5	0.1 (4.7)	0.01
<i>2001 level</i>	4.3 (2.0)	2.19*
decile 1	-0.2 (4.1)	-0.05
decile 2	5.5 (4.1)	1.32
decile 3	2.8 (4.1)	0.68
decile 4	-2.47 (4.0)	-0.62
decile 5	0.6 (4.6)	0.13
<i>99-00 growth</i>	0.1 (0.4)	0.30
decile 1	1.0 (0.8)	1.27
decile 2	1.0 (1.0)	1.03
decile 3	-1.3 (0.7)	-1.84
decile 4	0.2 (0.8)	0.19
decile 5	0.0 (1.0)	0.03
<i>99-01 growth</i>	-0.5 (0.4)	-1.12
decile 1	0.8 (0.9)	0.87
decile 2	-0.2 (1.0)	-0.25
decile 3	-2.0 (0.9)	-2.20*
decile 4	-0.3 (1.0)	-0.31
decile 5	0.5 (1.0)	0.56

After having reviewed the straightforward differences of means, broken down by API deciles, the available data are used to estimate model (2) and (3) in the following section.

#### 4.7.2 Estimating the models for cohort 1

Using the available API and SAT-9 data, results of the models described above are estimated here.

The results, provided in Table 19 below, show the estimated coefficient for the treatment dummy (i.e. the effect of having been selected into the treatment),  $\beta_1$ , for different dependent variables of model (2) without the inclusion of the covariates. The results show that the increase in API and SAT-9 math scores was significantly higher for the schools in II/USP than the comparison schools. The results suggest that the API growth during the first year of planning for schools in II/USP was 6 points higher than the API growth of schools that were not selected for the intervention. Over two years, the API growth is only 3 points higher for II/USP schools, suggesting no positive impact of II/USP in the second year. While the impact on the post-treatment increase in SAT-9 math scaled scores is highly significant, no such significant impact can be detected for reading.

It should be noted that the following tables present regression results that are not weighted. Since the outcome variables are school-level averages of student test scores, an alternative approach to the one presented here would weight the outcome measures by the number of students that took the test. In additional regressions (not listed here) the impact of using the weights was ascertained. Since the weighting does not change the results substantially, the more straightforward, non-weighted regression results are presented.

**Table 19 – Estimating (2), without covariates**

<b>Dependent variable</b>	<b>Coefficient (st. error)</b>	<b>t-statistic</b>
API growth 1999-2000	6.25 (2.08)	<b>3.00**</b>
API growth 1999-2001	3.23 (2.81)	1.15
SAT-9 math growth 1999-2000	1.10 (0.43)	<b>2.57*</b>
SAT-9 math growth 1999-2001	1.35 (0.52)	<b>2.60*</b>
SAT-9 reading growth 1999-2000	0.11 (0.37)	0.30
SAT-9 reading growth 1999-2001	-0.48 (0.43)	-1.12

As an additional check of the difference of SAT-9 math increases in the second year (not listed here) confirms, the results in the table are driven by a particularly strong increase in SAT-9 math during the first year. In an additional set of estimations the regressions were run by taking account

It is worth considering the plausibility of a significant impact of the II/USP intervention during the first year in greater detail. Schools in cohort 1 were notified of their selection to II/USP in the fall of 1999. In winter of that same school year, the schools (or responsible districts) selected the external evaluator to work with them. Together, the schools and the external evaluator conducted a needs assessment, leading to the development of an Action Plan and its submission to the CDE for approval. Despite a rushed timeline, the Action Plans were completed and approved only in April and May of 2000, at the time the students were tested to provide the results for what accounts, in the regression above, for the growth in the first year (CDE, 2000a; CSU, 2001). Though the needs assessment and the work in the school that fed into the development of the Action Plan may have had immediate beneficial effects, it would be surprising to see the higher student test scores only weeks afterwards. Also, the positive impact is not carried forward, and in the second year, after the Action Plan had been implemented for a full year, there is no additional significant impact on test scores of students. This issue is discussed at length in the AIR report (2003). Its authors conclude that

the planning year bump in growth is the result of the increased attention to student outcomes and instruction engendered by the selection of a school into the program, public scrutiny through the press, the activity of the External Evaluators, and the planning process itself – all concentrated in the first year.

This conclusion is based on the result of the quantitative analysis. Based on the qualitative data, the authors state, in contrast to the above conclusion, that they were not able to observe “any planning year activities that were either strong enough or focused enough to directly account for even the small planning year bump in achievement of II/USP schools relative to comparison schools.”

In any case, the data suggest that something positive happened with student achievement in the first planning year, even though the impact did not carry forward into the implementation years. But as the earlier discussion of program evaluation indicated, this conclusion is only valid, if there is no other systematic difference between the treatment and the non-treatment group that is left unaccounted for.

Using SAT-9 data in reading and math, model (3) can be estimated to test for the interventions’ impact on the change in achievement growth. Recall that the dependent

variable,  $\Delta_t y_i - \Delta_{t-1} y_i$ , is the difference between the change in test score averages prior to the treatment, 1998-1999, and the change after the treatment, 1999-2000 or 1999-2001.

**Table 20 – Estimating (3), without covariates**

<b>Dependent variable</b>	<b>Coefficient (st. error)</b>	<b>t-statistic</b>
Change in SAT-9 math growth 1999-2000	0.79 (0.63)	1.26
Change in SAT-9 math growth 1999-2001	1.03 (0.68)	1.51
Change in SAT-9 reading growth 1999-2000	-0.19 (0.59)	-0.33
Change in SAT-9 reading growth 1999-2001	-0.79 (0.64)	-1.24

The results in Table 20 show that the estimation of model (3) indicate that the impact of II/USP on SAT-9 math scores is not significant. The zero results for SAT-9 reading are consistent with the results in Table 19. These results suggest that time-varying errors may be correlated with the treatment effect that have not been accounted for in the estimation of model (2). Additional checks of the results in Table 19 are warranted.

#### 4.7.3 Additional corrections

To test whether the results listed in Table 19 are driven by the changing characteristics of the schools and their student bodies, the vector of covariates is included. The results remain basically unchanged with the inclusion of the differences of the covariates.

**Table 21 – Estimating (2), with covariates**

<b>Dependent variable</b>	<b>Coefficient (st. error)</b>	<b>t-statistic</b>
API growth 1999-2000	7.02 (2.16)	<b>3.25**</b>
API growth 1999-2001	3.84 (2.81)	1.37
SAT-9 math growth 1999-2000	1.11 (0.44)	<b>2.49*</b>
SAT-9 math growth 1999-2001	1.18 (0.52)	<b>2.26*</b>
SAT-9 reading growth 1999-2000	0.11 (0.38)	0.29
SAT-9 reading growth 1999-2001	-0.33 (0.43)	-0.78

Since further estimations of model (3) and similar estimations for cohort 2 confirm that the inclusion of the covariates does not alter the results significantly, the rest of the results presented are estimations from the models (2) and (3) without the covariates.

An additional correction proposed above addresses the issue of LAUSD schools in particular. The comparison of the treatment and non-treatment group established that

schools from LAUSD are significantly overrepresented in the comparison group and that schools from LAUSD are significantly different from schools from other districts. With only 11 schools in the treatment group and 305 schools in the comparison group, an alternative estimation of the models is offered by eliminating LAUSD schools from both the treatment and non-treatment group. The results in Table 22 below shows the results analogous to Table 19 above, after eliminating the schools from LAUSD from the sample.

**Table 22 – Estimating (2), without LAUSD schools and without covariates**

<b>Dependent variable</b>	<b>Coefficient (st. error)</b>	<b>t-statistic</b>
API growth 1999-2000	5.71 (2.30)	<b>2.48*</b>
API growth 1999-2001	6.20 (2.92)	<b>2.12*</b>
SAT-9 math growth 1999-2000	0.69 (0.49)	1.41
SAT-9 math growth 1999-2001	0.81 (0.57)	1.41
SAT-9 reading growth 1999-2000	0.49 (0.42)	1.18
SAT-9 reading growth 1999-2001	0.31 (0.45)	0.70

The results show that the results from Table 19 are somewhat sensitive to the exclusion of LAUSD schools. The estimated impact of the intervention on API growth during the first year is at 6 points essentially the same as in Table 19 and can be interpreted as a confirmation of the above results. The impact on SAT-9 math scaled scores, however, is no longer significant. This result confirms the difference between the estimations of model (2) and (3) for SAT-9 scores, as discussed for the results in Table 20. Since the SAT-9 math scores make up only between 20% and 40% of the API (the proportion differs according to school year and grade span, as detailed in Table 6 above), these results are not inconsistent.

For an additional check of the above results a set of regressions was run without decile 5 schools. As the breakdown by deciles in Table 16 has indicated, the impact of II/USP was much larger for decile 5 schools and the overall results driven, in part, by the dramatic impact of II/USP on these schools. While the estimate’s significance was reduced (only significant at the 10% level), the exclusion of decile 5 schools did not eliminate the estimated impact completely.

In the summary below, the focus on the interventions’ impact on API growth. Results from the estimation of model (3) and the estimated impact on SAT-9 scores are

not discussed further, since these additional estimations have primarily served to confirm the estimations of model (2). With API growth as the primary measure of performance for California’s public schools, the interventions’ impact can thus be more readily interpreted from a policy perspective.

#### 4.7.4 Cohort 2

The data used for estimating model (2) and (3) for cohort 2 do not include elementary schools, as indicated earlier. Instead, the impact of II/USP is estimated using the data from middle and high schools only. The results in Table 23 show a significant impact of II/USP on API growth in the planning year. The increase is at just over 6 points similar in size to the estimated impact of II/USP on cohort 1, with positive, but non-significant impacts on SAT-9 scores.

**Table 23 – Estimating (2), without covariates**

<b>Dependent variable</b>	<b>Coefficient (st. error)</b>	<b>t-statistic</b>
API growth 2000-2001	6.27 (3.15)	<b>1.99*</b>
API growth 2000-2002	2.15 (3.72)	0.58
SAT-9 math growth 2000-2001	0.99 (0.52)	1.90
SAT-9 math growth 2000-2002	0.83 (0.59)	1.40
SAT-9 reading growth 2000-2001	0.90 (0.54)	1.66
SAT-9 reading growth 2000-2002	0.40 (0.57)	0.70

Table 24 below shows the results of estimating model (3) with data from cohort 2, as a further check of the above results. The estimates of model (3) show a positive, non-significant impact of II/USP on the difference in API and SAT-9 growth between the pre-intervention year and the first year of planning.

**Table 24 – Estimating (3), without covariates**

<b>Dependent variable</b>	<b>Coefficient (st. error)</b>	<b>t-statistic</b>
Change in API growth 2000-2001	6.14 (4.07)	1.51
Change in API growth 2000-2002	2.40 (4.65)	0.52
Change in SAT-9 math growth 2000-2001	1.15 (0.74)	1.55
Change in SAT-9 math growth 2000-2002	1.04 (0.79)	1.31
Change in SAT-9 reading growth 2000-2001	1.12 (0.79)	1.42
Change in SAT-9 reading growth 2000-2002	0.64 (0.81)	0.79

The results in Table 23 and Table 24 can be interpreted as a tentative confirmation of the results obtained for cohort 1. The impact on the first-year API growth is significant and also estimated to be 6 API points, while the results in Table 24 are more tenuous.

#### **4.7.5 Conclusion**

The models and the above estimations have attempted to identify an effect of II/USP on the performance of underperforming schools in California. The three main findings are summarized here.

- i) II/USP interventions have had a positive impact in the first year of planning on schools that applied for the intervention. The impact on API growth during the planning year was approximately 6 points.
- ii) Relative to the average API growth among all schools in the respective samples (40 points in 1999/00; 14 points in 2000/01) this impact of II/USP represents a 16% and 43% increase, respectively.
- iii) The implementation of II/USP interventions has had no significant impact beyond the first planning year.
- iv) The impact of II/USP is uneven across API deciles and highest for the higher performing schools.

#### **4.8 Discussion**

While policy recommendations are provided in Chapter Five, the following sections discuss the above results and raise some of the more specific questions that remain with regards to the interventions under SB1X.

##### **4.8.1 Heterogeneity of impact**

The result of a first-year increase in test scores correspond to the findings by O'Day and Bitter (2003), notwithstanding the differences in the selection of the comparison group and methodology. The size of the impact estimated here, however, is smaller than the estimate by O'Day and Bitter.

The interventions set out by SB1X are remarkable insofar as they are applied across the full lower half of the performance distribution of California's public schools. It is not clear, especially in light of the discussion earlier in this chapter and in Chapter Two, that any set of interventions spread across such a wide performance spectrum should be expected to result in a single identifiable effect. Even the impact that was identified is not uniform across all schools. The impact for schools in cohort 1 was highest among decile 5 schools, as shown in Table 16 above, and drives some of the

significance of the results for cohort 1. The differential impact across API deciles notwithstanding, the results offer evidence that the planning forced upon schools participating in the program has had a beneficial effect across all schools, even as this aspect of the intervention has taken up only a fraction of the interventions' total cost.

#### *4.8.2 Potential, remaining bias*

Since the interventions are not all coordinated at the same agency, the possibility for interventions that compensate for non-selection to II/USP has to be taken into consideration. In addition to interventions that are targeted directly at the underperforming students, the district offices seem intent on providing all their schools with support, regardless of whether or not they were selected for II/USP. Thus, implicit or even explicit compensatory support from the district may flow to the schools that are in the control group. Such additional support from the districts provided to schools that do not receive the state intervention would bias the estimate of the impact of II/USP downward.

#### *4.8.3 Effect on non-treated units*

The AIR report points out that the emphasis on intervening in underperforming schools in SB1X managed to substantially increase the focus on the issue of underperformance in the public school system more generally (O'Day and Bitter, 2003). Insofar as compensatory support from the districts has been given to non-selected schools, it would constitute an indirect impact of the intervention on non-treated units.

Though the above analysis has not found evidence that the \$200 per student for II/USP implementation had much of a direct effect on the schools' performance, the willingness of the state legislature to devote considerable amount of money towards seeking solutions to the problems in the state's schools has likely had an impact on the perceived possibilities of positive change in the future. The change of such a perception may be important for people considering a career in public education or others considering contributions of other kinds. If the perception is that shortcomings are acknowledged and that something is being done, even if the "something" is not very effective (yet), the general willingness to invest in the search for solutions may increase. On the other hand, as long as harsher sanctions are stipulated in the law for schools under

II/USP that fail to meet their API growth targets by the end of the intervention and, at the same time, such sanctions are never escalated beyond putting schools under state monitoring, the trust in the system may deteriorate. As comments from outside experts have indicated, schools in II/USP have not considered the threat of strong interventions as set out in SB1X to be a real possibility, regardless of their performance (CSU, 2001). The direct estimation of such system-wide effects, however, is beyond the scope of this dissertation.

#### *4.8.4 The nature of school failure in California*

Earlier in this chapter two hypotheses were offered regarding the nature of school failure in California that the II/USP interventions were supposed to correct. In one view of failure, the problem is systemic and requires long-term interventions. Temporary interventions such as II/USP would thus lead only to temporary shifts in performance. In another view of school failure, the problem at the individual schools is largely managerial and is fundamentally ‘fixable’. II/USP interventions should, if this latter view of the problem is more accurate, lead to permanent shifts in performance.

Given the large population of schools eligible for II/USP, temporary lapses in performance are likely to be an issue for many schools. The structure of II/USP strongly suggests that the interventions would work best for such ‘fixable’ problems. As Betts and Danenberg (2003) have shown, problems in California public schools are also, and perhaps primarily so, of a more systemic nature. This is true especially among low performing schools, as the resources that are available are *decreasing*, while the challenges in terms of the relative needs of the student body are *increasing*. This is not merely a matter of funding, but also a question of each school’s capacity of the teaching staff, insofar as low-performing schools find it more difficult to recruit and retain qualified teaching staff (Guarino et al., 2004). II/USP interventions do not offer a solution for problems of such a fundamental nature. There are indications that there is some recognition of the need for a more equitable mechanism for funding. The addition of HP interventions, for example, appears to be aimed at helping especially low decile schools to deal with some of their ongoing challenges. In addition, the underlying need for

instituting more equitable funding mechanisms is gaining recognition among policy makers (Archer, 2004).

## **5 Conclusions and Policy Recommendations**

This fifth chapter draws conclusions from the above analyses and provides a set of recommendations for different policy contexts. The objective of this dissertation has been to answer the following research questions:

What are the different responses to school failure?

What are the effects of interventions in failing schools on school performance?

What policy recommendations can be made regarding interventions in different policy contexts?

The first two research questions were addressed in previous chapters. These chapters' results are summarized here, organized by the research questions posed. The third research question is answered by providing policy recommendations based on the findings of this dissertation. In the sixth and final chapter, the policy recommendations are applied to the system of public education emerging in Qatar.

### **5.1 Responses to School Failure**

As governments around the globe are struggling to improve public education, the question of what to do about failing schools has become urgent as a consequence of the increasing focus on the performance of individual schools. This focus results from the emergence of the so-called new accountability. The new accountability is based on the prescription of curriculum standards and standardized tests that measure how far the students have progressed in relation to the standards, the designation of the individual school as the unit of accountability, and the establishment of consequences for individual schools based on the measured performance. As a related, but distinct trend, the introduction of parental choice is increasingly advocated as an important mechanism to improve public education, in particular by enabling parents to hold schools accountable for underperformance.

#### **5.1.1 Common assumptions regarding responses to school failure**

By instituting the new accountability and parental choice in particular, automatic responses to failing schools are expected to work across the whole system. The assumptions underlying this expectation are that all schools are exposed to competitive

pressures and either underperforming schools have the capacity to respond to these pressures, or schools lacking the capacity to respond are forced to close.

However, as the discussion in Chapter Two has indicated, these assumptions are unlikely to be valid for all schools. Not all schools will be under competitive pressures, as some parents may not voice their disagreement, may not choose according to the policymakers' intent, or may not choose at all. Granting parental choice may indeed lead quality-conscious parents to flee underperforming schools leaving the remaining parents with less overall capacity to influence the school's operations. For schools that do come under competitive pressure, the capacity to respond in a positive way to such pressure may be missing, even if they are given additional funding or the help of outside experts. As the discussion in Chapter Two illustrated, school failure is not of a single type and may require responses that are different from prescriptions for other, non-failing schools. And finally, underperforming schools that do not improve are not automatically closed. As existing studies of strong interventions in public schools and charter schools as well as related studies of transition economies have shown, closing schools is hard to do and requires relevant capacity and infrastructure on the part of the responsible political authority.

Chapter Two pointed also to the troubling connection between socio-economic or political disadvantage and school failure. While the political aspects have not been considered in detail, proposals for providing financial resources to schools on a more equitable basis (by taking into account the learning needs of the student body) exist and are, from a technical standpoint, feasible. Though such proposals have been implemented in individual districts and on a basic level in New Zealand, the funding for public schools in the United States more generally and in California in particular is far from equitable. How the inverse relationship between the individual schools' educational challenges and the resources available to them affects the discussion of interventions is addressed below.

### *5.1.2 Erroneous assumptions in New Zealand*

The case study of New Zealand's experience with education reform and failing schools in Chapter Three added empirical evidence of a system-wide test of the stated assumptions. New Zealand's education reforms instituted parental choice, passed

governance responsibilities on to local boards, and encouraged a culture of competition among public schools. Implicitly, the reforms subscribed to the classic assumptions regarding school failure. Initially, the Ministry of Education did not develop a strategy for interventions, as underperforming schools were expected to improve or be forced to close due to the competitive pressure from parental choice.

However, competitive pressure failed to impact a subset of underperforming schools in any positive way. School failure endured and remained concentrated among the schools serving socio-economically disadvantaged communities. After reports from the independent auditing agency called attention to the situation, the Ministry of Education was forced to reverse its policy towards failing schools. In a pragmatic approach to the reality of persistent failures in some communities, the Ministry developed a strategy for intervening in schools that were identified as failures. In addition, a strategy for reducing the likelihood of future failures included the deliberate creation of support structures for the locally governed schools in cases where these were missing. Using existing capacity outside the government bureaucracy, interventions have typically been initiated and overseen by the Ministry, but carried out by non-governmental organizations. Alongside the emerging strategy for intervening actively in cases where the market forces did not work, legal changes instituted more prescriptive requirements for planning and reporting on an annual basis. These added requirements have been deemed successful in forcing all schools to adopt a more conscious approach to the coordination of their activities.

## **5.2 *Effects of Interventions on School Performance***

Since automatic responses to school failure are not likely to work equally well for all schools, the examination of the available options for interventions is necessary. In Chapter Two the possible interventions in failing schools were listed and classified as mild, moderate, and strong interventions. These categories are based on the degree to which the interventions are disruptive to the structures and processes of the school. As the interventions increase in their disruptiveness to the school from mild, to moderate, to strong, the frequency with which they are used decreases. As a consequence, the least amount is known about interventions with the highest degree of disruptiveness.

A brief summary of what is known about the impact of interventions on school performance is presented here based on the review of the literature, followed by a summary of the quantitative analysis of the II/USP interventions' impact on school performance of underperforming schools in California.

### *5.2.1 Existing studies*

The review of existing studies shows that relatively little is known about the impact of interventions on school performance. In particular, the comparative effectiveness of interventions, for example between moderate and strong interventions, is almost impossible to estimate with any degree of confidence. Studies of interventions indicate that turning around underperforming schools is very hard to do well. Success is not the norm and there is no particular intervention that is generally more successful than any other intervention. The two elements that have consistently been found to be the key determinants in whether an intervention is successful are the capacity of the intervening body, including the consistency of the political support when hard decisions have to be made, and the quality of the leadership at the school itself. For interventions in New Zealand, the detailed needs assessment and planning phase of the intervention itself has been noted to be the most beneficial part of the process (McCauley and Roddick, 2001). In particular, the inclusion of outside organizations in failing schools seems to have enabled the establishment of additional support and ongoing pressure.

The impact of planning itself on school performance estimated in Chapter Four is consistent with its use in most interventions and the recently imposed, legislative requirement for strategic planning in all New Zealand public schools.<sup>59</sup> While planning is not required in public schools in the United States, interventions in failing schools often impose such a requirement for the duration of the intervention. Extensive intervention agreements, for example for Program Improvement schools under federal sanctions in California, as discussed in Chapter Two, documented the current state of the school and the attainable goals for the next year and future years. The use of strategic planning in this way, even if it is only within the intervention, has proven to be an important tool for guiding the intervention and, ultimately, for school improvement. This is the type of

---

<sup>59</sup> Strategic plans are required for all public schools in the UK (Stark, 1998), as described in Chapter Two.

performance agreement that has reappeared throughout the provided examples and recalls also the framework outlined for the initial reforms of public sector management in New Zealand (Purchase Agreement and Statement of Service Provision).

School closure as a strong intervention is especially important to consider. Though school closure has been an explicit option in many jurisdictions, the only state in the United States to ever close a public school due to academic failure is New York (Brady, 2003). The available evidence from the education sector, especially from charter schools, as well as from transition economies, has shown that failing schools do not “go out of business” automatically. The process of closing operational schools effectively depends on an existing infrastructure in terms of human capital (experience within the executing agencies, for example) and also in terms of expectations on the part of all parties involved. Closing schools is very expensive to do, if it can be done at all, given the political ramifications involved.

Aside from school closure, what is common to most interventions is that they are typically set to be of a fixed duration. This implies an *ex ante* interpretation of what the nature of the failure is that the interventions are supposed to correct. The way interventions are commonly organized presumes that the failure is a temporary problem that can be fixed in a limited amount of time. Once the problem is fixed, the school can return to its natural state of operating without failure. Since school failure is frequently a problem that takes more time to address adequately, interventions with such pre-defined duration are likely to be inadequate.

### *5.2.2 Impact of interventions in California*

To contribute directly to an answer for the second research question, Chapter Four has provided original estimates of the impact of the II/USP interventions on school performance of underperforming schools in California. What is particularly attractive and also unusual about II/USP interventions from an analytical standpoint is that the selection to the interventions was only loosely based on the perceived need. Instead the selection to II/USP was done by lottery, within specified limits or assignment bins as specified by law.

Using school level data of API levels and growth, as well as data of SAT-9 scores in reading and math, the interventions' impact on the within-school growth of test score averages among applicant schools was estimated. Only the first two years after the selection to the intervention for cohort 1 and 2 could be reliably estimated, due to selection bias introduced by the addition of subsequent cohorts. Unlike previous studies done on these interventions, the details of the selection procedure were taken into account explicitly. A small, but statistically significant impact of the interventions on the growth of student achievement during the first year of planning was detected. The impact was 6 points difference in one-year API growth, which amounts to an effect size of less than 1/10 of a standard deviation of API scores. The differential growth was not sustained in later years, however.

The small impact during the planning year was driven primarily by a larger and more significant impact among the highest performing II/USP schools in decile 5, with smaller and less significant planning year increases estimated for lower deciles. The heterogeneity of the interventions' impact illustrates two key issues that have been discussed before. First, the use of a single type of intervention to remedy the challenges that such a diverse group of schools face seems too simplistic in light of the complexity and heterogeneity of school failure itself. Second, California's low performing public schools are faced with an increasingly unequal distribution of resources, relative to the higher performing counterparts serving predominantly student populations that are less expensive to educate. Schools in the lower deciles are likely to need a more experienced and effective teaching staff or educational support programs more urgently than a two-year program that attempts to fix something in the short-term.<sup>60</sup>

---

<sup>60</sup> On a more fundamental and more controversial note, it may also be true that peer effects are so strong that it is unrealistic to expect schools serving low-income communities to perform at the same level as other public schools. However, Ouchi's example of the Goudy school in Chicago (Ouchi, 2003) and similar success stories indicate: given the resources commensurate with the educational challenges that schools face (financially, and in particular in terms of an effective teaching staff and school administration) high performance among low-income schools is possible.

### 5.3 Policy Recommendations

This section attempts to answer the third research question by providing policy recommendations that follow from the work presented in this dissertation. A first set of policy recommendations discusses the policy implications of this dissertation's findings in general. An additional subsection offers the policy implications for California. These policy recommendations are applied in greater detail to the Qatari education reform in the sixth and final chapter.<sup>61</sup>

Though the responsibility of the government in providing public education has been assumed throughout the dissertation, it should be emphasized that this does not imply a responsibility to provide public education directly. Instead, the responsibility implies a guarantee that an adequate level of public education is provided by someone and is made available to all. This distinction is important also for interventions themselves. What has been loosely termed *intervention by the government* throughout this dissertation involves, analogously to the broader responsibility of the government, the assurance that an intervention, if and when it is needed, is financed and carried out. Whether the intervention is actually implemented by a government agency or contracted-out to non-governmental experts does not, for the purposes of this dissertation, matter much. Available evidence from New Zealand, as well as from California, suggests that the development of private sector expertise in this area is possible. Whether such a development of service providers could be engendered elsewhere, and in what time frame, remains to be seen. Thus, a general conclusion that overall more involvement by the government is needed would not be a correct conclusion of the evidence presented here. What this dissertation has shown is that current education reform proposals would be improved by openly confronting the issue of school failure. Failing schools not only challenge and undermine the potential of such proposals, regardless of whether the failures are the result of an increased identification of pre-existing failures or the result of the introduction of market forces themselves. The neglect to acknowledge that some schools will fail and that such failures present a problem that must be dealt with carefully and deliberately unnecessarily limits the usefulness of the debate on how public

---

<sup>61</sup> Of the three policy contexts discussed in this chapter, New Zealand is not included again specifically, since the adjustments for developing a comprehensive framework for dealing with failing schools are currently underway.

education is best to be organized in the future. As Michael Barber (1998) noted: “A policy for tackling failure is a necessary element of any credible strategy designed to promote successful schools for all.”

### *5.3.1 Interventions in general*

A primary policy recommendation is offered along with recommendations that are associated with and flow from this first recommendation.

The primary policy recommendation that follows from the presented research is that interventions in failing schools should be specifically designed for the failure that they are meant to correct.

Along with a relative dearth of discussion about school failure in education reform proposals, most current intervention strategies and the way in which they are implemented suggest that school failure is seen as a generic problem that can be addressed with generic interventions. The evidence presented here indicates that this is not the case. Instead of applying a standard format of interventions to failing schools, the fact that several different types of school failure may exist should be taken into account. Currently, the analysis of school failure in the individual case seems to be insufficient. The effectiveness of any intervention strategy will depend heavily on the degree to which the problem is understood prior to the initiation of the intervention itself.

It is worth pointing out explicitly that this recommendation implies that interventions with a pre-defined time limit, as they are most commonly used, should not continue to be implemented in all cases. The government, or its designated agency, should be prepared to initiate interventions without time limits in persistently failing schools. This recommendation, and the prospect of indefinite interventions specifically, is an important departure from the idea underlying most interventions, namely that any individual school can be made successful, once the identified shortcoming is fixed. However, in contrast to problems that can be fixed in the short- or medium term with interventions as they are currently used by school districts and education departments, particularly challenging situations for schools may simply not be ‘fixable’ in a defined time frame. Schools need ongoing supervision, support, and pressure. The government must be ready to proxy for such forces where these are absent, even while assisting the

development of such forces outside of the government. While a hands-off approach may be possible, and perhaps desirable, for a majority of schools, the emergence of persistent failure among individual schools and the inadequacy of quick-fix remedies will drive the need for a deliberate and involved role of the government in a minority of schools. The need arises from the absence of automatic responses to school failure, as well as from the prohibitively large costs and fundamental difficulties that closing individual schools entails.

The primary policy recommendation formulated above implies that the correct and early identification of school failure are of utmost importance. Two related, secondary recommendations are offered in this regard. First, the prospects for correctly identifying school failure, on the one hand, and for preventing avoidable failures, on the other hand, are enhanced by making strategic planning a requirement for all public schools. As discussed earlier, strategic planning has been shown to be an effective tool for all schools and for underperforming schools in particular. If appropriate templates are provided through the oversight agency, along with professional development akin to the assistance offered during the II/USP planning year, for example, such a planning requirement would not put undue pressure on schools, but rather force a careful evaluation of each public school's situation and its goals in relation to it.

Second, the optimal approach to identifying failing schools is to institute a so-called two-tier model of monitoring. On a first tier, a level of performance is specified as a minimum requirement using inexpensive and less accurate measures of performance (e.g. student test score averages, school attendance or dropout rates). As long as the noisy measures of performance remain above the specified level, no further investigations are warranted. For schools whose performance measures drop below the specified level, a more detailed investigation is warranted to ascertain whether the school is indeed at risk of failing and if so what should be done about it. In a formal principal-agent model of this problem Ronald Dye (1986) has shown that the described approach is the optimal strategy.

Implicitly, this approach is currently used by many education agencies. Making the two-tier system an *explicit* strategy is useful for regulating and reducing the degree of

direct control that schools are subjected to. As long as the performance of the school remains above a specific threshold, the school is considered to be succeeding and any involvement by the oversight agency is not needed. An explicit two-tier approach is currently being implemented in New Zealand by the Monitoring and Support group (Miller, 2004; Ministry of Education, 2003c).

### *5.3.2 Interventions in failing schools in California*

In addition to the above recommendation and its implications for an effective strategy of interventions, an additional policy recommendation is offered for public schools in California. Based on the analysis of interventions in California public schools in Chapter Four, and additional discussions of interventions in Chapter Two, the following change in policy is recommended.

II/USP as a system of interventions should be discontinued and the existing strategies for interventions in the various jurisdictions should be streamlined into a single and more transparent accountability framework.

The strategy of interventions that California's accountability framework introduced in 1999 has not had a significant effect on participating schools. While small significant gains from planning for the interventions were estimated, the principal and most expensive part of the intervention did not appear to have any significant positive impact. While a general focus on the performance of individual schools and on the problem of underperforming schools in particular has been documented to be a consequence of the new accountability framework (O'Day and Bitter, 2003), the interventions themselves have had little additional effect. Due to the multiple levels of jurisdictions (district, state, federal), all of which have their own criteria for intervening, the accountability framework that presents itself to failing schools is unclear and the effects of any given intervention are harder to evaluate. A central aspect of the new accountability, of which California's SB1X is an example, is that specific consequences follow the measurement of a school's performance. While the API provides a strong focus on student outcomes, the consequences of any given performance on the API have so far been insufficiently transparent. It should be noted that, with the implementation of the No Child Left Behind Act (NCLB) in California, changes akin to the stated policy

recommendation are underway. California is bringing its accountability framework in line with NCLB, as are the individual districts. II/USP is unlikely to be continued and instead the focus will turn towards compliance with federal mandates.

A final addition to the policy recommendations in general, and to California in particular, relates to the issue of equity. As the evidence throughout the dissertation has indicated, the strong correlation between failing schools and the degree of socio-economic disadvantage of the communities they serve is indeed troubling. While creative and individualized solutions should be found to deal with school failure on a case-by-case basis, some of the systematic disadvantage that entire sections of the public education system face is avoidable. Technically, closing the gap in terms of funding is relatively easy, as the differential funding instituted in New Zealand and also the weighted student funding introduced in Chapter Two have indicated. In California the *political* hurdles toward such a step appear to be considerable, even as a more equitable distribution of funds among California's schools is undoubtedly overdue.<sup>62</sup> For the time being, the resource gap between schools is not only not narrowing, it is widening considerably (Betts and Danenberg, 2003). An argument might be made in favor of II/USP, insofar as it directs additional funding to the schools in the lower half of the performance distribution. However, masking the efforts for more equality in funding in the language of interventions in failing schools creates an unnecessary and unhelpful mix between temporary cash injections, permanent funding equality, and the need for specifically designed, long-term interventions where real school failure persists.

---

<sup>62</sup> Schools whose students are on average more expensive to educate, because they come from socio-economically disadvantaged communities, are also funded at a significantly lower level (Biddle and Berliner, 2003; Carey, 2004).

## **6 Education Reform in the Emirate of Qatar**

This chapter introduces the education reform in Qatar and applies the findings of the presented research to the particular case. The Emirate of Qatar is currently implementing a comprehensive education reform or *Education for a New Era* (ENE) as it is referred to officially and throughout the remainder of this chapter. The reform is not just dramatic relative to the systems of public education elsewhere in the region, but also unique in its scope and its integration of the most advanced ideas in educational delivery and management.

### **6.1 Why Qatar Matters**

Before describing the outlines of ENE, this section will consider why an exploration of the education reform in Qatar is useful for this dissertation. The importance of the example of Qatar is based primarily on the type of education reform that is being developed and implemented, thanks to Qatar's particular political economy.

For the purposes of this dissertation, ENE matters because the planned accountability framework provides a unique perspective into the possibilities and the fundamental limitations of educational accountability and interventions in failing schools. As the description of the reform in the following sections will show, ENE is arguably more comprehensive and more thought-through with regards to accountability than any other national system of public education today. As the discussions in this dissertation have alluded to, the consideration of policy options for interventions in failing schools often runs into the issues of existing institutional barriers. With the emergence of a focus on the performance of individual schools, they constrain the range of options for dealing with failing schools. Examples of such barriers are data limitations, the political power of unions, or a set of reform principles that is insufficiently grounded in empirical evidence.

In exploring policy options for interventions in failing schools, the planning of a strategy for interventions in an idealized setting is a helpful exercise. What would policy options for interventions be, if the typical institutional constraints did not exist? What would a strategy look like in a system of public education that was built from the ground up, with an advanced accountability framework including full parental choice, with the availability of a wealth of information on all participants in the system and on education

providers in particular, with few resource constraints, and with solid political support? Would a strategy for interventions in failing schools still be needed in such a system?

The ongoing implementation of ENE in Qatar offers an example of an emerging system of public education that comes as close to the described situation as is likely to be possible. The discussion of how interventions in failing schools would be organized in this particular case is therefore useful and contributes to the understanding of how a strategy for intervening in failing schools can complement a modern system of public education.

The success or failure of education reform in Qatar matters also because of its signaling effect as to what may or may not be possible in a region that is desperate for development and desperate for education reform (see UNDP, 2002; UNDP, 2003). Successful reform of public education in the Middle East may promote, in addition to direct benefits to the citizens and the reforming country, tolerance and modernization more broadly and thus provide a counter-balance to the destructive and extremist voices. Even some of these extremist voices have taken note of the implications of this possibility for the region (Bin Laden, cited in United Press International, 2004). While these broader themes are not the focus here, they illustrate further why the Qatari reform is a fitting example to use in this research.

The background of the country itself and the structures that are relevant for public education in Qatar are introduced in the following section, before the education reform and the accountability framework are presented in detail.

## **6.2 *The Context to the Current Education Reform in Qatar***

This section explores the particular context that has sparked and is continuing to drive the education reform. The political economy that arises from Qatar's history and political system is considered, as well as the status quo of K-12 education and broader reform efforts in the Emirate, of which ENE is one central component.

### **6.2.1 *Historical background***

Qatar, a small, wealthy Emirate in the Persian Gulf, looks back on a relatively short history as a modern nation, gaining independence from Britain in 1971. Less than a century ago, in 1907, British official sources stated that Qatar's total resources consisted

**Figure 4 - Map of Qatar**



of 1,430 camels, 250 horses, and 817 pearl boats, with a population just shy of 30,000 (Crystal, 1995). Pearls were the mainstay of the Qatari economy, as the only export to speak of. Agriculture was practically impossible and even the production of dates was limited. Preliminary oil drillings were halted during the second World War, and the market for natural pearls, which had come under considerable competitive pressure from new Japanese cultured pearls during the 1920's, dried up completely. As a result, the economy collapsed and roughly half the population left the peninsula. Those that stayed did not live well. Rupert Hay, the British Political Resident, described Doha at the end of the 1940's as "little more than a miserable fishing village straggling along the coast for several miles and more than half in ruins ... there was no

electricity, and the people had to fetch their water in skins and cans from wells two or three miles outside of town" (quoted in Crystal, 1995). It is a remarkable starting point, little more than five decades ago, from which Qatar has emerged into a wealthy, dynamic, and innovative country. The key to this transformation, of course, was the income from the exploration of fossil fuel reserves in Qatari territory.

After the development of oil extraction began in earnest in the 1950s, the domestic economy turned its focus almost exclusively towards oil and gas exports. From 1955 onward, no pearling ships set sail (Crystal, 1995). Compared to other countries in the Gulf, Qatar's oil reserves are modest. The production rate that Qatar Petroleum is aiming for, well in excess of its current OPEC quota, is roughly one million barrels per day (current production is 2/3 that amount). At that rate of production, known oil reserves will last for another 35 years (Qatar Special Report, 2002). In contrast, the reserves of

liquid natural gas are the third largest in the world. Gas reserves are projected to last well over two centuries, depending on the projected rate of extraction (EIU, 2004).

Politically, Qatar has been dominated by the Al-Thani family since the mid 19<sup>th</sup> century (EIU, 2004). The current ruler, His Highness The Emir Sheikh Hamad Bin Khalifah Al-Thani, came to power during a bloodless coup in 1995, in which he deposed his father. The Emir has pursued a path of openness towards the West, for example by allowing the United States military to establish bases in Qatar, and openness for political change as well. A new constitution was proposed by the Emir and approved by popular vote in April 2003, granting consultative powers to a new advisory council. As the ultimate power still rests with the Emir, however, Qatar remains an absolute monarchy for the foreseeable future (EIU, 2004).

The Emir's political authority is not contested domestically, and the presence of a large number of American military forces provides a powerful security guarantee internationally (EIU, 2004). The Doha-based television station Al-Jazeera is renowned for being a voice for radical and controversial opinions that are otherwise muted or censored in the region and is partially sponsored by the Qatari government. Among all journalists operating within Qatar itself, however, self-censorship is imposed without exception when it comes to Qatari political issues. Any stories regarding national security or the royal family in particular are strictly off limits (EIU, 2004). Qatar's remaining disputes with its neighbors have been settled peacefully in recent years. Unlike other countries in the region, Qatar has not experienced a significant level of violence. The only incident in recent years was a car bombing in February 2004 that killed the former Chechen leader, Zelimkhan Yanderbiyev; an act for which two Russian intelligence agents were apprehended and charged (EIU, 2004).

### ***6.2.2 Reform efforts throughout the Qatari economy***

Under the leadership of the current Emir, Qatar has pursued an aggressive course of economic modernization. Most notable are the partial privatizations of the telecommunications monopoly Q-tel, the Qatar General Electricity and Water Corporation (including separate privatization of water and electricity production, distribution, and retail), the Qatar National Bank (including a broader liberalization of the

financial sector), and the breakup of other major industries into quasi-private companies (EIU, 2004).

At the heart of these reforms lies the attempt to eliminate some of the excesses of the welfare state that have developed in the past five decades, and are now perceived to be a hindrance to further development. As Crystal and Zahlan both note, the role of the state bureaucracy in oil exporting countries has traditionally not been one of resource extraction, as is the case in other countries, but instead one of resource distribution (Crystal, 1995; Zahlan, 1998). High-paying jobs continue to be available in the vast, overstaffed government bureaucracy. Despite a lot of changes toward privatization and a more dynamic economic marketplace, a recent survey found that more than 86% of all Qataris prefer to work for their government, while private companies prefer to hire foreign nationals (Peninsula, 2003).

The modernization of education has been a central part of these efforts. This has happened under the leadership of the Emir's second wife, Her Highness Sheikha Mouza bint Nasser al-Misnad, who also serves as the chairperson of Qatar Foundation (the sponsoring and coordinating body of broad reform efforts in education), and is the UNESCO Special Envoy for Basic and Higher Education. To modernize tertiary education and turn Qatar into a regional center for knowledge production, the government expects to spend \$5 billion over the next five years on the establishment of outposts of Western colleges outside Doha (EIU, 2004). Among Western university campuses already up and running at the specifically constructed *Education City* are: Weill Cornell Medical College; Texas A&M University; Virginia Commonwealth University College of Fine Arts and Design; and Canada North Atlantic College of Technology; as well as offices of the American think tank RAND Corporation, the RAND Qatar Policy Institute. In line with the program of expanding education facilities in Qatar, the budgeted allocation for education and youth welfare more than doubled from \$115 million in fiscal year 2002/03 to \$264 million in 2003/04 (EIU, 2004).

### 6.2.3 *K-12 education in Qatar*

The reform of K-12 public education in Qatar is a central component of these broader reform efforts, even as the existing system of public education has been

established only relatively recently. Public education was introduced in 1952,<sup>63</sup> and the Ministry of Education was formed four years later (Al-Mulla, 2003). Since its inception, the Ministry of Education has grown into a top-heavy bureaucracy with an annual budget of over \$350 million (Qatari Ministry of Education, 2000), operating just over 200 schools for roughly 70,000 students, resulting in a per-pupil spending of over US\$5,000 per year. Another 47,000 students attend private schools. The vast majority of these schools and students are situated in the metropolitan area of the capital, Doha, where roughly 90% of Qatar's population resides. The high number of non-Qatari residents whose children are not eligible for government-funded public education is one reason why there is a relatively large proportion of students in private schools. The other major reason, however, is the perception among Qataris that the public schools are failing to provide an adequate level of education (Al-Mulla, 2003). Partly in recognition of this situation, and to ease the pressure on overcrowded public schools, the government provides financial support for private schools, such as land grants, textbooks, healthcare and utilities (EIU, 2004). Though private schools enjoy greater autonomy compared to the public schools, they are also monitored by the Ministry inspectors and are subject to rules and regulations of remarkable scope and detail (Qatari Ministry of Education, 1986).<sup>64</sup> Qatari K-12 schools regulated by the Ministry, whether public or private, educate boys and girls separately in all grades, and teach religious studies as one of the curriculum's centerpieces. In addition, from 5<sup>th</sup> grade onwards, Qatari boys can only be taught by men.<sup>65</sup> It should be noted that foreign schools, such as the American School of Doha or the Doha English Speaking School, are exempt from most regulations, and teach an entirely Western curriculum. These schools are considered to deliver an education of significantly higher quality and are far more expensive than private Arabic schools. They educate, aside from the children of Western expatriate workers, a number of local students from the wealthiest and most prominent Qatari families.

---

<sup>63</sup> First royalties from oil exports were received in 1950.

<sup>64</sup> The Ministry rules for private schools include, for example, the requirement for a specific number of restrooms per student, the prohibition against teaching anything not explicitly approved by the Ministry, or the suggestion of offering expensive watches to students with high achievement.

<sup>65</sup> This is due to Islamic tradition and is significant insofar as there is a surplus of Qatari female teachers.

To summarize, the Qatari system of public education is small, concentrated in one major metropolitan area, overly regulated, and considered to provide low-quality education. The political situation in the country is stable and marked by an unusual concentration of power in a leadership that is firmly committed to reform. The considerable wealth that is directed to the broader reforms, and to the reforms in education in particular, means that the designs for, and implementation of, reforms are less burdened by resource constraints than they typically are in other contexts. Furthermore, also in a wealthy Emirate like Qatar, public schools remain the foundation for the educational experience of most citizens. Driven by this underlying motivation, the reforms that this chapter discusses were recently launched. A broad overview of the reform is provided in the following section, before the specifics of the accountability framework are discussed.

### **6.3 Overview of 'Education for a New Era'**

With the Emiri Decree of November 2002, the education reform, termed Education for a New Era (ENE) was officially launched (Office of Communications, 2004a). ENE is still only in the initial stage and will be phased-in over several years, unlike the overnight decentralization in New Zealand. The ENE reforms do not rely on any existing structures, but instead develop an entirely new system for K-12 public schools parallel to the Ministry of Education and to the traditional public schools it runs. The first 12 schools opened in September 2004.

At its most fundamental level, ENE is designed to establish a system of autonomous schools that operate similarly to US charter schools, as privately operated organizations under contract with the government. Though the physical infrastructure (school buildings, etc.) will remain government property in most cases, the operational responsibility for the school lies with an incorporated school operator, consisting of at least two people (one of which may also be, but need not be, the principal). Any existing organization or corporation organized under Qatari law may also apply to become a school operator (Education Institute, 2004a). Since the word *charter* has no useful translation in Arabic, the schools are called Independent Schools to emphasize their difference from the schools operating under the direct control of the Ministry. The details

of how these schools operate and the relevant policy context is explained in the following sections.

### **6.3.1 *New institutions***

The aforementioned Emiri Decree established the Supreme Education Council (SEC) to oversee the entire education sector. The SEC is designed as the locus of ultimate authority for public and private education in Qatar and operates in a hands-off manner, as a corporation's board of directors would. Two institutes have been established to carry out the operational responsibility of guiding the K-12 system: the Education Institute and the Evaluation Institute.

The Education Institute in effect replaces the current Ministry in its role of financing and overseeing the individual schools. Unlike the Ministry, however, the Education Institute does not operate any Independent Schools directly, but takes a role similar to that of a charter authorizer in the United States. The Education Institute is responsible for overseeing the Independent Schools from the planning stage onwards, and encouraging and aiding the development of relevant capacity among applicants and Independent Schools in operation. The institute helps applicants develop a contract,<sup>66</sup> reviews completed contracts prior to their official approval by the SEC, and oversees Independent Schools once they are up and running (Education Institute, 2004a). As ENE is a standards-based reform, the Education Institute is responsible for the establishment of curriculum standards in four ENE defined core subjects: Arabic, English, math, and science. Professional Development for the type of instruction needed to meet the standards, but also for the establishment and operation of an Independent School, is organized through the Education Institute as well.

The Evaluation Institute, on the other hand, does not have any regulatory authority or responsibility over Independent Schools. Instead, the Evaluation Institute's job is to collect, manage, store, and analyze data on all K-12 schools and to disseminate this information in the most appropriate formats to all stakeholders, including the schools themselves, teachers, students, parents, the broader community, the Education Institute,

---

<sup>66</sup> Analogous to charters. They are named contracts in this chapter because, without an Arabic translation for charter, that is what they are called throughout ENE.

and the SEC (Evaluation Institute, 2004). The scope and level of detail of the data collection and analysis should be emphasized. On a first level, similar to systems of the new accountability elsewhere, the progress towards the standards in the four core subjects is measured with standardized tests developed specifically for ENE. Every student in every grade takes such a standardized test in each of the four subjects every year. Aside from these national tests, all parents and students are surveyed annually, along with school administrators and teachers. Finally, the Evaluation Institute collects additional information at the school sites in order to be able to present a rich picture of each school. The implications of this detailed reporting for the accountability framework are discussed further below.

### **6.3.2 *A series of reconstitutions***

While ENE invites applications from individuals wanting to develop new schools, the first 12 contracts that were approved in June 2004 suggest that, at least initially, the largest number of Independent Schools is to originate from the conversion of public schools currently being operated by the Ministry of Education (Education Institute, 2004a; Office of Communications, 2004a; Office of Communications, 2004b). The transformation of the schools from being tightly controlled by the Ministry to being operated as autonomous organizations by an incorporated school operator, imposes dramatic changes at each individual school site. The change to becoming an Independent School is comparable to a reconstitution that is imposed on failing schools in the United States. All current students are allowed to remain at the school, though they have the option to transfer to another school. The school's management, all administrators and teaching staff, the curriculum for all subjects, the school's general policies, and especially the approach to teaching are all new. To aid this transition, external experts provide on-site training and guidance. Unlike reconstitutions in the United States, in the case of Independent Schools the policy context in which the 'reconstituted' schools find themselves has also been radically changed from the context prior to its reconstitution.

In the change of existing Ministry schools to newly incorporated Independent Schools, the difference in the policy context gives the definition of school failure an interesting twist. From the viewpoint of the Education Institute, the existing schools have

to be considered failing to a degree where radical reconstitution is the only option. This fact is, after all, what motivates the reform in the first place. On the other hand, from the perspective of the old system, or the Ministry that currently operates the schools, there is nothing inherently wrong with its schools, and there is no standard up to which the schools are failing to perform. Thus, the very same schools are failing from the viewpoint of the new policy context, or are fine as they are from the viewpoint of the existing policy context. The difficulties that such a radical shift in perspective will create within the individual school sites, and especially among the students, remain to be seen.

Two additional comments are worth noting in this regard. First, as the reform is still in the process of being implemented, it is too early to properly evaluate the success of the individual reconstitutions, or of the reform more generally. Second, the 12 schools that opened as Independent Schools in September 2004 were not selected as schools most in need, or least in need, of reconstitution. Rather, political considerations determined the selection of an initial pool of schools eligible for reconstitution, such as specific schools that had been operating with a higher degree of autonomy from the Ministry, as well as selected elementary and middle schools that would result in a reasonably even distribution across Doha's metropolitan area. This diffuse mechanism for selection will make an evaluation of the individual reconstitutions more difficult, even as the ongoing implementation necessitates such an approach.

The interventions that are of primary interest for this dissertation and that are discussed here are future interventions in Independent Schools that will have been operating under ENE for some time and that, in the new policy environment, might fail to deliver an adequate level of education.

### ***6.3.3 Remaining uncertainties***

With ENE establishing a system of public education without the burden, but also without the support, of existing structures, many aspects of the reform are still being worked out. In particular, the concern with providing equitable funding to individual schools, given the differences in student characteristics as they exist even in a relatively homogenous society as Qatar, have not been fully resolved. The funding mechanism has been designed to be flexible and to allow for a learning period in the Education Institute,

by allowing the possibility of grants to be awarded to individual schools if particular circumstances justify it. At the moment, schools are funded with per-student allotments of \$4,400, \$5,300, and \$6,000, for elementary, middle, and high schools, respectively (Premack, 2004). Though there is a 20% supplement written-in for special education needs, no further details are established at this point. Whether a more detailed funding mechanism, for example using a weighted student formula, will be implemented at a later stage is not clear.

#### **6.4 Accountability Framework**

The accountability framework is described here along three key features: the contract between the Independent Schools and the SEC, the reporting that the Independent Schools themselves are required to do, and the reports issued by the Evaluation Institute. Arguably the most central feature of the accountability framework, not discussed in a separate section, is parental choice. Though there isn't effective choice in the initial years (certainly not now, with 12, mostly oversubscribed, schools operating under ENE), the design of ENE is set up to include and foster competition among the schools, as a result of parental choice.

##### **6.4.1 The contract**

Analogous to charters in the United States, the contracts between the individual operators of Independent Schools and the Supreme Education Council are the centerpiece of ENE's accountability framework. The SEC is represented by the Education Institute, which is responsible for promoting and assisting the development of contracts for potential school operators, as well as for reviewing them and recommending their approval to the SEC. Once the schools are operating in the new system, the Education Institute is responsible for overseeing the schools and the degree to which they remain true to the terms of their contract.

The guidelines for the contracts spell out the format of the contract in detail and prescribe flexibility for operators to make their own determinations regarding the education that their Independent School is going to offer (Education Institute, 2004b; Education Institute, 2004c; Education Institute, 2004d). The difference between these two aspects should be emphasized: While the guidelines leave no room for interpretation

regarding the format that the contracts should take and spell out not only the general areas to be covered (e.g., educational plan), but also specific details that need to be addressed (e.g., strategy for involving parents), the guidelines are remarkably non-prescriptive regarding the specifics of the plans that are required to be in the contract. The only pre-specified content, aside from legal obligations common to all contracts, is the annual participation in the national tests in the four core subjects, data collection and reporting to the Evaluation Institute, and the publication of an annual report (Education Institute, 2004b). The vast majority of what has to be specified in the contract, including, for example, the school's mission, its curriculum, the school calendar, and the school's governance and management structure, is not pre-specified by the guidelines. What matters most for the accountability is that schools are free to suggest how they would like their success to be measured. Typically, this would involve the average performance on the national tests to some degree. But it is not specified how much weight the test scores should have and the inclusion of other performance measures in the contract is highly encouraged. Giving schools the liberty to propose their own measures of success can be said to follow directly from the liberty to propose a specific mission. Since Independent Schools are free to propose the type of education that they want to offer, they should also propose a plan with specified goals and measures to roughly define success or failure in their undertaking. With regard to school performance and the question of school failure, this element becomes the key component of the contract. Given the specific, agreed-upon measures of success in each contract, school failure can be defined as a lack of compliance with the terms of the contract.

The contracts represent an opportunity for schools to present evidence of their performance according to the measures that have been initially agreed upon by both the Education Institute and the Independent School. This role of the contract, and the lack of pre-specified content in particular, puts tremendous pressure on the Education Institute for the process of assisting the development, the review, and the eventual approval of the contracts themselves. All judgment and central planning on the part of the Education Institute has to be brought to bear on this particular stage of the process. Some of the issues that require at least a degree of central planning are the schools' locations, the grade distribution, the schools' admissions policies, but also less easily defined issues,

such as the organization of student learning and the goals that the school sets for itself and its students. The flexible nature of the contracts makes the review and approval of them all the more difficult. As a consequence, there will be considerable temptation to include an increasing number of restrictions into the guidelines, so that the process of review and approval of the contracts can be simplified for the Education Institute.

A related challenge is to communicate not only to the school operators and their staff, but also to the public at large, how Independent Schools will be different from the current public schools run by the Ministry. For example, the frequently asked questions of the ENE specify that no more than 25 students should be in a single classroom (Education Institute, 2004a), even as this restriction is not further explained and is not a part of the original guidelines (Education Institute, 2004b). Aside from a possible instrument of a communication strategy, this change might well also be an attempt to simplify the daunting process of reviewing and approving contracts.

The importance of strategic planning has been emphasized in previous chapters and in the recommendations in Chapter Five in particular. The establishment of a detailed contract encompasses the first element of the requirement for strategic planning. The counterpart to the contract is the Annual Report, where amendments to the contract are required to be included. This possibility for adjustments allows an ongoing effort by the schools to refine their contract.

#### ***6.4.2 Reporting by the school***

One of the few requirements for Independent Schools is that each has to produce a report every year according to a pre-specified outline (Education Institute, 2004b). The purpose of the annual report is to inform the parents and the broader community about what has been going on in the school and to provide evidence of the school's strategic planning. In addition, the same report is used for the official reporting by each Independent School to the Education Institute. In particular, as it pertains to the measures of success, the annual report provides evidence on how the goals set out in the contract have or have not been achieved and the implications for the next planning cycle .

The combination of the contract and the annual report is analogous to the accountability framework outlined by Scott et al. (1997) for the public sector

management reform in New Zealand. Instead of contract and annual report, Scott et al. called these documents the Performance Agreement and the Statement of Service Provision. It should be noted that the school's contracts are between the SEC and the operator of the Independent School. However, with the requirement to publish a detailed report annually, the broader community and the parents in particular have the possibility to become directly involved as recipients and overseers of the school's activities and its reporting. How the capacity for assuming this role effectively is distributed among the Independent Schools' communities remains to be seen.

#### *6.4.3 Reporting by the Evaluation Institute*

The contracts and the annual reports, which are specific to every school, are not the only tools offered in the accountability framework. Designed as the objective provider of comparable information across all K-12 schools in Qatar, the Evaluation Institute collects, analyzes, and reports quantitative and qualitative information to all stakeholders. The provided information "will assist parents in selecting the best schools for their children and allow school systems to assess the effectiveness of each individual school" (Evaluation Institute, 2004). Though the details of the reporting to the various stakeholders are not fully settled, the main theme of the Evaluation Institute's activities is clear. The competition that ENE is set up to engender will become more meaningful through the availability of high-quality information on each school with a level of detail and breadth that arguably surpasses most any other system of public education. Beyond the role that quality information will play in improving the effectiveness of competition between schools, the Evaluation Institute's role in providing information to the Education Institute is crucial.

With the Evaluation Institute as an independent entity, the regulatory responsibilities of the Education Institute are made far more tractable. In addition to the test results that provide information on student achievement in each school in the four core subjects, the results from surveys and school site information for each school allow the Evaluation Institute to provide a rich picture of the individual schools' effectiveness and potential problems.

The role of the Evaluation Institute is to provide information that is applicable to most, if not all, schools in the form of standardized reporting. This is in the same spirit as the API reported for California's public schools. Despite the planned inclusion of a rich set of information on each school in the reporting done by the Evaluation Institute, the comparison of test score averages by school in the four core subjects is going to be a central, if not the dominant, aspect of the school report cards.

On the one hand, the goal is to allow for a system of schools to emerge that have a variety of educational approaches and a variety of ways to pursue and define success. Independent Schools are generally free to define the weight that the results on the national tests have for their performance in their contract. In their annual reports, individual schools may focus and are even encouraged to focus on student achievements outside the test results. Among this variety of programs that are hoped to emerge, the Qatari parents can choose the school they want their children to attend. On the other hand, ENE imposes a common yardstick in the form of student achievement measures and disseminates schools' relative achievements in this particular dimension widely to make the comparison easier and the competition more effective. How ENE's dual goals of variety and accountability through effective competition will shape the Independent Schools in the coming years will provide an interesting case for further research.

### ***6.5 When Independent Schools fail***

Though ENE is reconstituting existing schools at an ambitious pace and is expanding rapidly, a strategy for intervening in Independent Schools that fail after being opened (as opposed to schools that never open as planned) has not been developed. Yet an intervention strategy for failing schools is necessary for a system that emphasizes the performance of individual schools and promotes competition among them. In the Education Institute's official publication of frequently asked questions, school failure is only briefly addressed and school closure is mentioned as the only specific potential intervention (Education Institute, 2004a). The explicit communication of potential school closure helps differentiate reconstituted Independent Schools from those operated by the Ministry.

In this section, the findings and formulated policy recommendations of the earlier chapters are applied to the specific case of Qatar. Specifically, the question is: What strategies of the Education Institute and the Evaluation Institute are necessary for success? Chapter Five recommends that interventions in failing schools should be specifically designed to address the particular failure to be corrected. This recommendation is applied to the Qatari case in two steps: the analysis of school failure and the implementation of the intervention. A concluding section offers possible outcomes for ENE if a strategy towards failing schools is not formulated.

### *6.5.1 Strategy for analyzing school failure*

Strategic planning is a requirement for all Independent Schools. For the Annual Report, the outline is pre-specified. This forces each school to reflect on its own performance, to provide an outlook and a plan for the future, and also to highlight aspects of the school's operations that may be unique or otherwise not captured in the quantitative student achievement and survey data (Education Institute, 2004b). The Annual Report is intended to be used for general communication and advertisement and needs to meet reporting requirements set by the Education Institute. In addition, as interventions in New Zealand (McCauley and Roddick, 2001) and planning have shown,<sup>67</sup> the requirement to produce a plan about a schools' operations, successes, and failures can be an beneficial process in itself.

The identification and analysis of school failure is relatively straightforward in cases where malfeasance by school operators or the school's management is detected. If there is a willful breach of contract, the Education Institute must immediately revoke the schools' contract and further analysis is not necessary to determine an appropriate response. In charter schools in the United States, for example, financial mismanagement is one of the most frequent reasons cited for charter revocation (Center for Education Reform, 2002). In Qatar, periodic financial reports analyzed by the Education Institute and annual reports by external auditors enables the early detection of management missteps. Aside from financial mismanagement, there may also be other forms of

---

<sup>67</sup> See, for example, planning as part of the federally mandated interventions in California (Ott, 2004) or the impact of planning on schools in the United Kingdom (Stark, 1998), as described in Chapter Two, as well as the direct impact of planning on student test scores, as estimated in Chapter Four.

malfeasance that would lead to the immediate revocation of the contract. The appropriate next step for the Education Institute after any contract revocation is discussed in the next section.

The identification of failure for educational reasons is less straightforward. As discussed above, the Evaluation Institute is charged with collecting, storing, analyzing, and disseminating information on all Independent Schools.<sup>68</sup> In doing so, the Evaluation Institute considers a rich set of information, such as

- i) student achievement data from annual tests in four core subjects for every student in every grade,
- ii) results from annual surveys of students, parents, and school staff,
- iii) school profiles from on-site inspections,
- iv) Annual Reports for each Independent School,
- v) communication/complaints from parents, community members and/or the school (if applicable).

With substantial high-quality information available, the implementation of a two-tier monitoring strategy is uniquely promising for ENE. For the first tier, the Evaluation Institute monitors the available information on all Independent Schools and sets a cutoff for the minimum level of performance. Independent Schools that fall below the cutoff are considered to be failing schools (though a less dramatic term might be used for communication purposes). All failing schools would require an in-depth analysis of the reasons for failure. In the first tier, therefore, the Evaluation Institute identifies failing schools based on available data. This identification process undoubtedly needs ongoing calibration. In addition to the student test scores in four subjects, the relative weighting of the results from the annual surveys, on-site inspection reports, and school-specific information is not clear and should be a subject of ongoing discussions. In addition, specialty schools may need to be assessed with an adjusted test score cutoff. As such ongoing adjustments will be necessary, the independent Evaluation Institute is well

---

<sup>68</sup> In addition to the Independent Schools, the Evaluation Institute is also charged with evaluating performance in public and private Arabic schools (all schools regulated by the Ministry of Education). Since these are outside of ENE, the Evaluation Institute's responsibilities with respect to those schools are not considered further.

positioned to impartially assess the Independent Schools' performance and set an appropriate cutoff.

Initially, the Evaluation Institute should set a relative cutoff to enable easy adjustments over time and avoid the possibility of identifying too many or too few schools. For example, the Evaluation Institute could use available data to identify the lowest performing 15% of all schools. Such a relative cutoff could take account of the institutional constraints of intervening in the failed schools. The appropriate choice of a cutoff should be coordinated with the Education Institute. In later years, an absolute measure of performance cutoff should be developed in order to make the first-tier monitoring more transparent for schools.

After the Evaluation Institute identifies the failing schools in the first-tier, the second-tier involves the in-depth analysis of the reasons for failure. Again the Evaluation Institute is the appropriate agency to oversee this analysis for three reasons: First, the Evaluation Institute leads school assessment. It is therefore appropriate that the in-depth analysis of the operations of failing schools should also be a part of its mandate.

The second reason is that analysis of school failure has typically been insufficient or absent in other school systems and inappropriately influenced by preconceptions about the nature of failure. The potential for biased second-tier analysis of school failure would be somewhat reduced if the independent Evaluation Institute were in charge. The Education Institute, in contrast to the Evaluation Institute, is more directly involved with the schools' management and is also in charge of detecting outright malfeasance or otherwise blatant breaches of contract on the part of the school operator. The Education Institute thus might view school failure more generally as a problem of management limiting the value of the second-tier analysis.

A third and related reason is that the separation between the analysis of failure (Evaluation Institute) and the coordination and oversight of interventions (Education Institute, as described below) helps to maintain the impartiality of the analysis. As described in Chapter Three, the system of inspectors from the Department of Education in New Zealand prior to the education reforms in 1989 was seen to be a hindrance to the impartial assessment of the schools' situation. The inspectors were responsible for both

the assessment of the schools' operations and the coordination of assistance. They were therefore responsible for reporting on a situation they had helped set up compromising their impartiality. The establishment of the Education Review Office, with responsibilities only in assessment but not in assistance, created just such a separation. The separation allowed the strong criticism of failing clusters of schools (ERO, 1996), and ultimately led to a fundamental shift in policy at the Ministry of Education. Similarly, dividing the responsibilities for the analysis of failure and the provision of support between the Evaluation Institute and the Education Institute ensures that failing schools are looked at from a relatively impartial perspective and that the analysis is not influenced by the amount of support that the Education Institute may have provided to the schools already.

One drawback of making the Evaluation Institute responsible for the second-tier analysis is a potential conflict within the Evaluation Institute due to the evaluation of schools on a common scale in the first-tier and of failing schools individually in the second-tier. The contracts allow for the formulation of school-specific goals and an individual weighting of the results of the student achievement scores. For example, the Evaluation Institute also reports on the performance of all schools independent of the school-specific goals or the specific situation. The tension between a standardized analysis in the first-tier and a detailed, school-specific analysis in the second-tier will require ongoing attention and, periodically, a critical review of the respective processes.

The first-tier analysis of school failure identification occurs annually. The second-tier school failure analysis occurs in the months following the first-tier analysis. To conduct the second-tier analysis, the Evaluation Institute should contract with education experts, such as former principals, to visit the schools and do an on-site review of the school's operations and the community involvement. In addition, a more in-depth analysis of the available data test scores, survey results, and possibly other data should support the on-site evaluation. Finally, to ensure impartiality, the outside expert contracted to do the analysis should not be permitted to be concurrently employed in a support function for any Independent School.

The end product of the detailed data analysis and the on-site inspection of the Independent School by an outside expert is a report on the school's operations, and analysis of the underlying reasons for its underperformance. The report should be provided to the school (possibly giving the school a chance to respond) and to the Education Institute so that appropriate steps may be taken to remedy the identified shortcomings.

Following the submission of the results from the second-tier analysis, the Education Institute must initiate an appropriate intervention at each of the identified schools. On the part of the Evaluation Institute, there should be a follow-up analysis possibly including a follow-up on-site inspection by the outside expert at least once a year, depending on the type of intervention and the expected changes.

#### **6.5.2 *Strategy for intervening in failing schools***

What should be done, once failing schools have been identified and the reasons for failure have been analyzed? The Education Institute must initiate and coordinate an appropriate response to the identified failure. In this section, a set of examples is described to indicate the kind of school failure that could be expected in Qatar and the kind of interventions that could be used to address them.

In instances of outright breaches of contract or willful malfeasance on the part of the school's management, the Independent School's contract must be revoked, as described earlier. For such instances of school failure a second-tier analysis is not necessary and the relationship between the school operator and the SEC, represented by the Education Institute, is terminated. The school itself, however, may well be in a position to continue operating, albeit under a different management or school operator. For such cases, the contract for the school should be offered to existing operators with a good track record. Among the interested operators the Education Institute should select the operator with the highest capacity for taking on the additional school and help the operator develop a new contract for the school. This may include the condition of firing key management personnel at the school.

Management takeover is likely to work better in Qatar than it has worked elsewhere. As ENE is a system built primarily on the serial reconstitution of existing

schools all participants in the system (Education Institute, school operators, teachers, parents, students, and the broader community) are becoming accustomed to management takeovers, are gaining experience in how to do it well, and are developing a realistic set of expectations for such an intervention.

As management takeover is a feasible strong intervention, school closure should only be contemplated in cases where the students' well-being is at risk (i.e. school's culture is poisoned in some way, school building is falling apart, or the school's reputation is beyond repair).<sup>69</sup>

For the cases where the school failure does not require an immediate response, the second-tier analysis should provide suggestions of how the causes of low performance should be addressed. The goal of the Education Institute's interventions should always be to sustainably address the main reasons for failure and restore the Independent School to a satisfactory level of performance.

In cases where the second-tier analysis has identified shortcomings in the school's governance structure, the intervention should seek to improve it. As was briefly discussed in Chapter Two, interventions could focus on activities outside the school itself as well as inside. In order to strengthen a school's oversight mechanisms, the intervention by the Education Institute might seek to involve local organizations more heavily. Such attempts have been shown to be useful in New Zealand (McCauley and Roddick, 2001) and could include the active partnering of the school with community organizations, groups of parents, or even corporations. In a radically decentralized system of public education it should be expected that the degree to which each school is connected to its community and the degree to which the community effectively carries out its implicit oversight responsibilities will not be adequate in all cases. At least in the beginning of ENE a number of schools, as well as their respective communities, are likely to struggle with these new roles. Involving the community to a greater degree in the school presents an opportunity to infuse the school's management not only with an oversight structure, but

---

<sup>69</sup> It should be emphasized that school closure on account of being unable to attract enough students is *not* a process that should be prevented. Rather, the interventions discussed here (including school closure) are for Independent Schools that continue to serve a sufficient number of students to be able to operate.

also with a potential for support and particular expertise that individual parents or organizations may be willing to provide to the school.

The purpose of intervention aimed at strengthening a school's oversight and support structure should always be to coordinate capacity outside of the Education Institute. Though the responsibility for the intervention ultimately rests with, and should be coordinated by, the Education Institute, intervention would ideally manage to mobilize resources outside the Institute.

Under ENE, each school is responsible for developing and implementing its own educational program. Here too, it is likely that not all schools will be equally able to continue to develop and strengthen their in-class instruction. The reasons for such difficulties may be an insufficient coordination of the curriculum across subjects and grades, ineffective teaching methods, lack of student motivation, or a poorly designed educational program. An intervention by the Education Institute should be to identify educational difficulty within the school and sustainably address the problem. As before, partnerships with existing capacity are preferred though not always the most appropriate solution. Periodically, Independent Schools may have to work with outside consultants to realign their instructional program and coordinate all activities among the teachers and school administration. The Education Institute's intervention may then consist of suggesting a particular program to the school and overseeing its progress. In other instances partnerships with other Independent Schools could be established to mutually benefit from aspects of the curriculum that may have been particularly well established at any one school.

The funding mechanism for Independent Schools, as described in the presentation to prospective school operators (Premack, 2004), allows for the establishment of grants for special educational purposes. Interventions addressing the educational needs of a school's student body will be able to take advantage of such grants by arguing that a specialized investment in equipment or the establishment of a particular program within the school could address the identified shortcoming within the boundaries set by the available grants.

The timeline for such interventions, as well as the particular funding arrangements for each intervention, will have to be determined at the outset. As examples throughout the dissertation have shown, a pre-defined length for interventions is not likely to fully address the various shortcomings equally well. Indeed, some interventions may evolve into ongoing projects for strengthening particular educational programs in individual schools or for strengthening community involvement in individual Independent Schools. As the Education Institute only initiates and coordinates the interventions, such an ongoing role is not a contradiction to the Institute's general charge of not being directly responsible for schools' operations.

In cases where the closer investigation *does* reveal that the shortcoming is fixable, placing the school on probation and providing a specified time frame for improvement might be an appropriate measure. In general, this would depend on the assessment of the problem. The formulation of a specific plan for such an intervention would always need to include the criteria under which the intervention can and must be terminated. This is necessary to avoid unnecessary involvement by or school dependency on the Education Institute.

In Qatar, even after ENE is fully implemented, the capacity of schools and the communities they serve will vary. While this variation is likely to emerge in the metropolitan area of Doha also, it is useful to consider that remote schools are likely to start operating under ENE within a few years. Communities in the west of the peninsula, serving mostly a lower socio-economic Bedouin population as well as new communities by the northern gas fields in Ras Laffan are likely to be served by remote Independent Schools. The Education Institute's oversight responsibility becomes particularly important in these cases due to the lack of educational alternatives in these locations. The uneven distribution of pressure and support facing the Independent Schools may necessitate prolonged interventions by the Education Institute for some schools. Strengthened oversight mechanisms and a more active role in day-to-day operations may be necessary in such schools. Special training for community involvement, including the schools' governing boards, may help to develop the communities' capacity in the medium and long term.

### 6.5.3 *What would happen without a strategy?*

In order to successfully motivate the need for formulating an intervention strategy, two possible outcomes are considered here if no strategy were developed.

Without a strategy for interventions in failing schools, the Education Institute might get too heavily involved in Independent Schools. If the specific criteria for initiating an intervention are lacking, the Education Institute might be inclined to intervene in all cases where it perceived outcomes or processes that did not match its expectations. Once interventions were initiated, it would be difficult for the Education Institute to withdraw from the school without having processes and criteria for ending interventions. In an attempt to avoid further school failures, the Education Institute would be tempted to impose greater control and more regulations on Independent Schools on the basis of perceived needs at individual failing schools. Considering the history of Qatar's bureaucracy described earlier and the particularly centralized approach of the Ministry of Education, such a response by the Education Institute would not be surprising. Without a strategy for interventions, emerging school failures might thus lead to a creeping return of centralized control over Qatar's public schools.

Alternatively, without a strategy for interventions, the Education Institute might not get involved in Independent Schools at all. Without an expectation that intervening in schools should be done at all and without any strategies for doing so, the Education Institute might choose to abstain from taking any active role in individual schools. The failure of individual Independent Schools would emerge and become visible, in part due to the reporting by the Evaluation Institute. The continued existence of such failing schools would undermine the potential of ENE as an improved system of public education. Without a strategy for interventions, the affected Qatari children and Qatar more generally would thus not be served as well as they could and should.

## REFERENCES

Al-Mulla, Yousuf A. Rahman Ali, *Scholastic Achievement in Private and Public Elementary Schools in the State of Qatar: A Multilevel Modelling Approach*, Manchester: University of Manchester, 2003.

Anderson, Lee, Nancy Adelman, Kara Finnigan, Lynnyonne Cotton, Mary Beth Donnelly, and Tiffany Price, *A Decade of Public Charter Schools*, Washington DC: SRI International, 2002.

ARC, "Historical Timeline of Public Education in the US," *Applied Research Center*, 2004. Online at [www.arc.org/erase/timeline.html](http://www.arc.org/erase/timeline.html) (as of September 10, 2004).

Archer, Jeff, "'Weighted' Funding of Schools Gains Favor," *Education Week*, 2004. Online at [www.edweek.org](http://www.edweek.org) (as of November 3, 2004).

Asthma in America, "Asthma Statistics," *GlaxoSmithKline*, 2004. Online at [www.asthmainamerica.com/statistics.htm](http://www.asthmainamerica.com/statistics.htm) (as of September 13, 2004).

Barber, Michael, "The Dark Side of the Moon: Imagining an End to Failure in Urban Education," in Louise Stoll and Kate Myers, eds., *No Quick Fixes: Perspectives on Schools in Difficulty*, London: The Falmer Press, 1998.

Berends, Mark, Susan J. Bodilly, and Sheila Nataraj Kirby, *Facing the Challenges of Whole-School Reform: New American Schools After A Decade*, Santa Monica: RAND Corporation, MR-1498-EDU, 2002.

Betts, Julian R., and Anne Danenberg, "The Effects of Accountability in California," in Paul E. Peterson and Martin R. West, eds., *No Child Left Behind?*, Washington DC: Brookings Institution Press, 2003.

Biddle, Bruce J., and David C. Berliner, "What research says about Unequal Funding for Schools in America," *WestEd*, 2003. Online at [www.wested.org/online\\_pubs/pp-03-01.pdf](http://www.wested.org/online_pubs/pp-03-01.pdf) (as of August 18, 2004).

Blume, Howard, "Sixty Charter Schools Fall, With a Little State Shove," *Los Angeles Times*, 2004. Online at [www.latimes.com/news/opinion/sunday/commentary/la-op-blume29aug29\\_1\\_1469759.story](http://www.latimes.com/news/opinion/sunday/commentary/la-op-blume29aug29_1_1469759.story) (as of September 25, 2004).

Bradley, David H., "Fuzzy Math in Philadelphia: Edison Overstates Savings Possible Through Outsourcing School District Maintenance and Operations Work," *Keystone Research Center*, 2002. Online at [www.keystoneresearch.org/pdf/edison%20report.pdf](http://www.keystoneresearch.org/pdf/edison%20report.pdf) (as of June 11, 2004).

Brady, Ronald C., *Can Failing Schools Be Fixed?*, Washington DC: Thomas B. Fordham Foundation, 2003.

Bushman, James, and Greg S Goodman, "What Teaching the Standards Really Means," *Leadership*, Vol. 33, No. 2, 2003, pp. 26-29.

California State Auditor, *California's Charter Schools*, Sacramento: California Bureau of State Audits, 2002.

Carey, Kevin, "The Funding Gap 2004: Many States Still Shortchange Low-Income and Minority Students," *The Education Trust*, 2004. Online at <http://www2.edtrust.org/NR/rdonlyres/30B3C1B3-3DA6-4809-AFB9-2DAACF11CF88/0/funding2004.pdf> (as of October 26, 2004).

CDE, "Detailed Summary of SB1552 Changes to PSAA (SB 1X)," *California Department of Education*, 2000a. Online at [www.cde.ca.gov/ta/ac/ap/documents/sb1xand15522.pdf](http://www.cde.ca.gov/ta/ac/ap/documents/sb1xand15522.pdf) (as of November 10, 2004).

---, "Explanatory Notes for 1999-2000 API Growth Report," *California Department of Education*, 2000b. Online at [www.cde.ca.gov/ta/ac/ap/expnotes99g.asp](http://www.cde.ca.gov/ta/ac/ap/expnotes99g.asp) (as of November 13, 2004).

---, "Explanatory Notes for 2000-2001 API Growth Report," *California Department of Education*, 2001. Online at [www.cde.ca.gov/ta/ac/ap/expnotes00g.asp](http://www.cde.ca.gov/ta/ac/ap/expnotes00g.asp) (as of November 13, 2004).

---, "Explanatory Notes for 2001-2002 API Growth Report," *California Department of Education*, 2002. Online at [www.cde.ca.gov/ta/ac/ap/expnotes01g.asp](http://www.cde.ca.gov/ta/ac/ap/expnotes01g.asp) (as of November 13, 2004).

---, "Application Information for Schools Applying in October, 2003 to the High Priority Schools Grant Program," *California Department of Education*, 2003a. Online at [www.cde.ca.gov/ta/lp/hp/documents/guidance2.pdf](http://www.cde.ca.gov/ta/lp/hp/documents/guidance2.pdf) (as of November 16, 2004).

---, "Explanatory Notes for 2002-2003 API Growth Report," *California Department of Education*, 2003b. Online at [www.cde.ca.gov/ta/ac/ap/expnotes02g.asp](http://www.cde.ca.gov/ta/ac/ap/expnotes02g.asp) (as of November 13, 2004).

---, "2003-04 Academic Performance Index Growth Report: Information Guide," *California Department of Education*, 2004. Online at [www.cde.ca.gov/ta/ac/ap/documents/infoguide03g.pdf](http://www.cde.ca.gov/ta/ac/ap/documents/infoguide03g.pdf) (as of October 27, 2004).

Center for Education Reform, *Charter School Closures: The Opportunity for Accountability*, Washington DC: The Center for Education Reform, 2002.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiola, *The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools*, unpublished research, 2004.

Christman, Jolley Bruce, and Amy Rhodes, "Civic Engagement and Urban School Improvement: Hard-to-Learn Lessons from Philadelphia," *Consortium for Policy Research in Education*, 2002. Online at [www.cpre.org/Publications/children07.pdf](http://www.cpre.org/Publications/children07.pdf) (as of June 28, 2004).

Chubb, John E., and Terry M. Moe, *Politics, markets, and America's schools*, Washington DC: Brookings Institution, 1990.

Circle Systems, Stat/Transfer, v. 7.0.02, Seattle, WA, 2003.

Coleman, Jennifer, "California Charter Academy closes remaining two schools," *The Mercury News*, 2004. Online at [www.mercurynews.com/mld/mercurynews/news/local/states/california/the\\_valley/9375691.htm](http://www.mercurynews.com/mld/mercurynews/news/local/states/california/the_valley/9375691.htm) (as of September 25, 2004).

Colton, Barbara, telephone communication with the author on November 29, 2004a.

---, telephone communication with the author on September 28, 2004b.

Connelly, Martin, *Leadership and School Failure*, unpublished research, 1999.

---, "Creeping Centralism: Fact or Fabrication," *SPANZ Journal*, 2003.

---, email communication with the author, dated February 19, 2004.

---, email communication with the author, dated January 12, 2005.

Corcoran, Tom, and Jolley Bruce Christman, "The Limits and Contradictions of Systemic Reform: The Philadelphia Story," *Consortium for Policy Research in Education*, 2002. Online at [www.cpre.org/Publications/children08.pdf](http://www.cpre.org/Publications/children08.pdf) (as of June 28, 2004).

Crystal, Jill, *Oil and politics in the Gulf: Rulers and merchants in Kuwait and Qatar*, New York: Cambridge University Press, 1995.

CSU, "Immediate Intervention/Underperforming Schools Program," *California State University/California Education Policy Seminar*, 2001. Online at [www.calstate.edu/ier/reports/iiups.pdf](http://www.calstate.edu/ier/reports/iiups.pdf) (as of October 24, 2004).

Department of Education, "Turning Around Low-Performing Schools," *US Department of Education*, 1998. Online at [www.ed.gov/PDFDocs/turning.pdf](http://www.ed.gov/PDFDocs/turning.pdf) (as of December 30, 2004).

---, "Overview of the Comprehensive School Reform Program (CSR)," *US Department of Education*, 2002. Online at [www.ed.gov/programs/compreform/2pager.html](http://www.ed.gov/programs/compreform/2pager.html) (as of November 16, 2004).

---, "LEA and School Improvement," *US Department of Education*, 2004. Online at [www.ed.gov/policy/elsec/guid/schoolimprovementguid.doc](http://www.ed.gov/policy/elsec/guid/schoolimprovementguid.doc) (as of June 21, 2004).

Dillon, Sam, "Collapse of 60 Charter Schools Leaves Californians Scrambling," *New York Times*, 2004a. Online at [www.nytimes.com/2004/09/17/education/17charter.html?hp](http://www.nytimes.com/2004/09/17/education/17charter.html?hp) (as of September 17, 2004).

---, "U.S. Set to Ease Some Provisions of School Law," *The New York Times*, March 14, 2004.

Dye, Ronald A., "Optimal Monitoring Policies in Agencies," *Rand Journal of Economics*, Vol. 17, No. 3, 1986, pp. 339-350.

ECS, "State interventions in low performing schools and school districts," *Education Commission of the States*, 2002. Online at [www.ecs.org/clearinghouse/41/52/4152.pdf](http://www.ecs.org/clearinghouse/41/52/4152.pdf) (as of May 2004).

Ed-Data, "Understanding the Academic Performance Index," *The Education Data Partnership*, 2004. Online at [www.ed-data.k12.ca.us/articles/Article.asp?title=understanding+the+api](http://www.ed-data.k12.ca.us/articles/Article.asp?title=understanding+the+api) (as of November 13, 2004).

Education Institute, "Frequently Asked Questions for Operators," *Supreme Education Council, Qatar*, 2004a. Online at [www.english.education.gov.qa/files/1187\\_FAQwebsite\\_version.pdf](http://www.english.education.gov.qa/files/1187_FAQwebsite_version.pdf) (as of December 22, 2004).

---, "Guidelines for Preparing a Contract to Operate an Independent School in Qatar," *Supreme Education Council, Qatar*, 2004b. Online at [www.english.education.gov.qa/files/763\\_Guidelines\\_eng.pdf](http://www.english.education.gov.qa/files/763_Guidelines_eng.pdf) (as of December 22, 2004).

---, "Independent School Application," *Supreme Education Council, Qatar*, 2004c. Online at [www.english.education.gov.qa/files/1149\\_application\\_in\\_English.pdf](http://www.english.education.gov.qa/files/1149_application_in_English.pdf) (as of December 22, 2004).

---, "Independent School Application Guide," *Supreme Education Council, Qatar*, 2004d. Online at [www.english.education.gov.qa/files/1150\\_English\\_guidelines.pdf](http://www.english.education.gov.qa/files/1150_English_guidelines.pdf) (as of December 22, 2004).

EIU, *Qatar Country Profile*, London, UK: Economist Intelligence Unit, 2004.

Elley, Warwick, "New Zealand literacy standards in a global context: The uses and abuses of international literacy surveys," *University of Canterbury, NZ*, 2004. Online at [www.tmc.waikato.ac.nz/english/ETPC/article/pdf/2004v3n1art3.pdf](http://www.tmc.waikato.ac.nz/english/ETPC/article/pdf/2004v3n1art3.pdf) (as of October 1, 2004).

Elmore, Richard, Charles Abelman, and Susan H. Fuhrman, "The New Accountability in State Education Reform: From Process to Performance," in Helen Ladd, ed., *Holding Schools Accountable*, Washington DC: The Brookings Institution, 1996.

ERO, "Effective Governance," *Education Review Office*, 1994a. Online at [www.ero.govt.nz/Publications/eers1994/94no7hl.htm](http://www.ero.govt.nz/Publications/eers1994/94no7hl.htm) (as of October 5, 2004).

---, "Self Review in Schools," *Education Review Office*, 1994b. Online at [www.ero.govt.nz/Publications/eers1994/94no3hl.htm](http://www.ero.govt.nz/Publications/eers1994/94no3hl.htm) (as of October 4, 2004).

---, "Improving Schooling in Mangere and Otara," *Education Review Office*, 1996. Online at [www.ero.govt.nz/Publications/eers1996/mangere.htm](http://www.ero.govt.nz/Publications/eers1996/mangere.htm) (as of October 5, 2004).

---, "Improving Schooling on the East Coast," *Education Review Office*, 1997. Online at [www.ero.govt.nz/Publications/eers1997/eastcoast.htm](http://www.ero.govt.nz/Publications/eers1997/eastcoast.htm) (as of October 5, 2004).

---, "Schooling in the Far North," *Education Review Office*, 1998. Online at [www.ero.govt.nz/Publications/eers1998/far\\_north/farnorth1.htm](http://www.ero.govt.nz/Publications/eers1998/far_north/farnorth1.htm) (as of October 5, 2004).

---, *Annual Report*, Wellington: Education Review Office, 2003a.

---, *Evaluation Indicators for Education Reviews in Schools*, Wellington: Education Review Office, 2003b.

---, *Statement of Intent 2003/04*, Wellington: Education Review Office, 2003c.

Evaluation Institute, "About the Evaluation Institute," *Supreme Education Council, Qatar*, 2004. Online at [www.english.education.gov.qa/section/sec/evaluation\\_institute/](http://www.english.education.gov.qa/section/sec/evaluation_institute/) (as of December 25, 2004).

Evers, Williamson M., "Standards and Accountability," in Terry M. Moe, ed., *A Primer on America's Schools*, Stanford: Hoover Institution Press, 2001.

Fancy, Howard, "National Conference Schools Monitoring and Support: Opening address by the Secretary for Education," *Ministry of Education*, 2001. Online at [www.tki.org.nz/r/school\\_improvement/resources/conf\\_papers/conference\\_opening.doc](http://www.tki.org.nz/r/school_improvement/resources/conf_papers/conference_opening.doc) (as of October 5, 2004).

---, "Education Reform: Reflections on New Zealand Experience," *New Zealand Ministry of Education*, 2004. Online at

[www.minedu.govt.nz/index.cfm?layout=document&documentid=9750&data=1](http://www.minedu.govt.nz/index.cfm?layout=document&documentid=9750&data=1) (as of October 19, 2004).

Finn, Chester E., and Frederick Hess, "On leaving no child behind," *The Public Interest*, No. 157, 2004, pp. 35-36.

Finnigan, Kara, and Jennifer O'Day, "External Support to Schools on Probation: Getting a Leg Up?," *Consortium for Policy Research in Education*, 2003. Online at [www.cpre.org/Publications/Chicago.pdf](http://www.cpre.org/Publications/Chicago.pdf) (as of June 28, 2004).

Fiske, Edward B., and Helen F. Ladd, *When Schools Compete: A cautionary tale*, Washington DC: Brookings Institution Press, 2000.

Friedman, Milton, "The Role of Government in Education," in Robert A. Solo, ed., *Economics and the Public Interest*, Piscataway, NJ: Rutgers University Press, 1955, pp. 123-144.

Fuhrman, Susan H., "The New Accountability," *Consortium for Policy Research in Education*, 1999. Online at [www.cpre.org/Publications/rb27.pdf](http://www.cpre.org/Publications/rb27.pdf) (as of October 12, 2004).

Gardner, David P., Yvonne W. Larsen, William O. Baker, Anne Campbell, Emeral A. Crosby, Charles A. Foster, Norman C. Francis, A. Bartlett Giamatti, Shirley Gordon, Robert V. Haderlein, Gerald Holton, Annette Y. Kirk, Margaret S. Marston, Albert H. Quie, Francisco D. Sanchez, Glenn T. Seaborg, Jay Sommer, and Richard Wallace, "A Nation at Risk: The Imperative for Educational Reform - A Report to the Nation and the Secretary of Education," *The National Commission on Excellence in Education*, 1983. Online at [www.ed.gov/pubs/NatAtRisk](http://www.ed.gov/pubs/NatAtRisk) (as of May 25, 2004).

Garn, Gregg A., and Robert T. Stout, "Closing Charters: How a Good Theory Failed in Practice," in Robert Maranto, Scott Milliman, Frederick Hess and April Gresham, eds., *School Choice in the Real World: Lessons from Arizona Charter Schools*, Boulder: Westview Press, 1999.

Gauri, Varun, *School Choice in Chile: Two Decades of Educational Reform*, Pittsburgh: University of Pittsburgh Press, 1998.

Gerst, Harry, telephone communication with the author on May 3, 2004a.

---, telephone communication with the author on November 19, 2004b.

Grissmer, David, "Comment to Kane-Staiger's Volatility in School Test Scores," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2002*, Washington DC: Brookings Institute Press, 2002, pp. 269-272.

Guarino, Cassandra, Lucrecia Santibanez, Glenn Daley, and Dominic Brewer, *A Review of the Research Literature on Teacher Recruitment and Retention*, Santa Monica: RAND Corporation, 2004.

Hamilton, Laura, and Brian Stecher, "Improving Test-Based Accountability," in Laura Hamilton, Brian Stecher and Stephen Klein, eds., *Making Sense of Test-Based Accountability in Education*, Santa Monica: RAND, 2002, pp. 121-144.

Hanushek, Eric A., "The Failure of Input-Based Schooling Policies," *The Economic Journal*, Vol. 113, No. 485, 2003, pp. F64-F98.

Hanushek, Eric A., Charles S. Benson, Richard B. Freeman, Dean T. Jamison, Henry M. Levin, Rebecca A. Maynard, Richard J. Murnane, Steven G. Rivkin, Richard H. Sabot, Lewis C. Solmon, Anita A. Summers, Finis Welch, and Barbara L. Wolfe, *Making schools work: improving performance and controlling costs*, Washington, DC: The Brookings Institution, 1994.

Hassel, Bryan C., and Meagan Batdorff, "High Stakes: Findings from a National Study of Life-or-Death Decisions by Charter School Authorizers," *Public Impact*, 2004. Online at <http://bluweb.qblue.com/publicimpact/high-stakes/highstakes.pdf> (as of May 3, 2004).

Hassel, Bryan C., and Michelle Godard Terrell, "The Rugged Frontier: A decade of public charter schools in Arizona," *Progressive Policy Institute*, 2004. Online at [www.ppionline.org/documents/AZ\\_Charters\\_0604.pdf](http://www.ppionline.org/documents/AZ_Charters_0604.pdf) (as of June 9, 2004).

Hassel, Bryan C., and Sandra Vergari, "Charter Granting Agencies: The Challenges of Oversight in a Deregulated System," *Education and Urban Society*, Vol. 31, No. 4, 1999, pp. 406-428.

Hassel, Bryan, and Lucy Steiner, "Stimulating the Supply of New Choices for Families in Light of NCLB: The Role of the State," *Education Commission of the States*, 2004. Online at [www.ecs.org/clearinghouse/54/26/5426.doc](http://www.ecs.org/clearinghouse/54/26/5426.doc) (as of September 29, 2004).

Helfand, Duke, and Joel Rubin, "Few Parents Move Their Children Out of Failing Schools," *Los Angeles Times*, 2004. Online at [www.latimes.com/news/education/la-me-transfer8nov08,1,5475412.story?coll=la-news-learning](http://www.latimes.com/news/education/la-me-transfer8nov08,1,5475412.story?coll=la-news-learning) (as of November 10, 2004).

Helfand, Duke, and Doug Smith, "1,200 Schools in State Could Face Federal Penalties," *Los Angeles Times*, 2004. Online at [www.latimes.com/news/education/la-me-failure13oct13,0,4798082.story?coll=la-news-learning](http://www.latimes.com/news/education/la-me-failure13oct13,0,4798082.story?coll=la-news-learning) (as of October 26, 2004).

Hess, G. Alfred, Jr., "Reconstitution - Three Years Later: Monitoring the Effect of Sanctions on Chicago High Schools," *Education and Urban Society*, Vol. 35, No. 3, 2003, pp. 300-327.

Hirschman, Albert O., *Exit, voice, and loyalty*, Cambridge, Massachusetts: Harvard University Press, 1970.

Holmstrom, Bengt, and Paul Milgrom, "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics and Organization*, Vol. 7, 1991, pp. 24-52.

Hoxby, Caroline M., "School Choice and School Productivity (or could School Choice be a tide that lifts all boats?)," *NBER Working Paper 8873*, 2002. Online at <http://papers.nber.org/papers/w8873.pdf> (as of May 3, 2004).

IASA, "Improving America's Schools Act of 1994," *103rd Congress of the United States*, 1994. Online at [www.ed.gov/legislation/ESEA/toc.html](http://www.ed.gov/legislation/ESEA/toc.html) (as of December 2, 2004).

Just, Anne E., and Larry E. Boese, "Public School Accountability (2000-2001)," *CDE Division of Policy and Evaluation*, 2002. Online at [www.cde.ca.gov/ta/lp/iu/documents/psaa0001.pdf](http://www.cde.ca.gov/ta/lp/iu/documents/psaa0001.pdf) (as of June 21, 2004).

Just, Anne E., Larry E. Boese, and et al., "Public School Accountability (1999-2000)," *CDE Division of Policy and Evaluation*, 2001. Online at [www.cde.ca.gov/ta/lp/iu/documents/psaa9900.pdf](http://www.cde.ca.gov/ta/lp/iu/documents/psaa9900.pdf) (as of June 21, 2004).

Just, Anne E., and Teri Wong, "Public School Accountability (2001-2002)," *CDE Division of Policy and Evaluation*, 2003. Online at [www.cde.ca.gov/ta/lp/iu/documents/psaa0102.pdf](http://www.cde.ca.gov/ta/lp/iu/documents/psaa0102.pdf) (as of June 21, 2004).

Kane, Thomas J., and Douglas O. Staiger, "The Promise and Pitfalls of Using Imprecise School Accountability Measures," *Journal of Economic Perspectives*, Vol. 16, No. 4, 2002a, pp. 91-114.

---, "Volatility in School Test Scores: Implications for Test-Based Accountability Systems," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2002*, Washington DC: Brookings Institute Press, 2002b, pp. 235-283.

Kilpin, Trevor, personal communication with the author on January 16, 2004.

Kirby, Sheila Nataraj, Scott Naftel, Mark Berends, and Jennifer Sloan McCombs, *Schools Identified as in Need of Improvement under Title I: Recent Evidence from the National Longitudinal Survey of Schools*, Arlington: RAND, 2003.

KLCS-TV, "LAUSD Board Meeting: Addressing Accountability with High Priority Schools," December 18, 2003, Los Angeles, 2003.

Klerman, Jacob, "Measuring Performance," in Robert Klitgaard and Paul Light, eds., *High-Performance Government: Structure, Leadership, Incentives*, Santa Monica: RAND Corporation, 2005, pp. 343-379.

Klitgaard, Robert, Johannes Fedderke, and Kamil Akramov, "Choosing and Using Performance Criteria," in Robert Klitgaard and Paul Light, eds., *High-Performance Government: Structure, Leadership, Incentives*, Santa Monica: RAND Corporation, 2005, pp. 407-446.

Koehler, Paul, Catherine Jovicich, Joan McRobbie, Malia Villegas, June Lee-Bayha, and Paul Hood, "Transforming Education in Los Angeles: Creating Excellence for All Students," *WestEd*, 2003. Online at [www.wested.org/online\\_pubs/LA-Alliance-Report.pdf](http://www.wested.org/online_pubs/LA-Alliance-Report.pdf) (as of June 8, 2004).

Kornai, Janos, Eric Maskin, and Gerard Roland, "Understanding the Soft Budget Constraint," *Journal of Economic Literature*, Vol. 41, No. 4, 2003, pp. 1095-1136.

Ladd, Helen F., "Introduction," in Helen F. Ladd, ed., *Holding Schools Accountable*, Washington DC: The Brookings Institution, 1996.

---, "Comment to Kane-Staiger's Volatility in School Test Scores," in Diane Ravitch, ed., *Brookings Papers on Education Policy 2002*, Washington DC: Brookings Institute Press, 2002, pp. 273-280.

---, "Introduction," in David N. Plank and Gary Sykes, eds., *Choosing Choice: School Choice in International Perspective*, New York: Teachers College Press, 2003.

Ladd, Helen F., and Edward B. Fiske, "Does Competition Generate Better Schools? Evidence from New Zealand," *Terry Sanford Institute Working Paper Series SAN01-16*, 2001. Online at [www.pubpol.duke.edu/people/faculty/ladd/SAN01-16.pdf](http://www.pubpol.duke.edu/people/faculty/ladd/SAN01-16.pdf) (as of September 2003).

Leslie, Sue, personal communication with the author on January 16, 2004.

Levin, Henry M., "The Failure of the Public Schools and the Free Market Remedy," *The Urban Review*, Vol. 2, No. 1, 1968, pp. 32-37.

---, "Educational Vouchers: Effectiveness, Choice, and Costs," *Stanford University*, 1996. Online at [https://secure1.sc.netnation.com/~russells/working\\_papers/levin.doc](https://secure1.sc.netnation.com/~russells/working_papers/levin.doc) (as of August 20, 2004).

Liebman, James S., and Charles F. Sabel, "The Federal No Child Left Behind Act and the Post-Desegregation Civil Rights Agenda," *North Carolina Law Review*, Vol. 81, No. 4, 2003.

Mac Iver, Douglas J., Allen Ruby, Robert Balfanz, and Vaughan Byrnes, "Removed from the List: A Comparative Longitudinal Case Study of a Reconstitution-Eligible School," *Journal of Curriculum and Supervision*, Vol. 18, No. 3, 2003, pp. 259-289.

Maranto, Robert, Scott Milliman, Frederick Hess, and April Gresham, "Real World School Choice: Arizona Charter Schools," in Robert Maranto, Scott Milliman, Frederick Hess and April Gresham, eds., *School Choice in the Real World: Lessons from Arizona Charter Schools*, Boulder: Westview Press, 2001.

Mathews, Jay, "The Philadelphia Experiment," *EducationNext*, 2003. Online at [www.educationnext.org/20031/50.html](http://www.educationnext.org/20031/50.html) (as of June 11, 2004).

---, "Philadelphia Shows Progress in Schools Run by Companies," *Washington Post*, 2004. Online at [www.washingtonpost.com/wp-dyn/articles/A35074-2004Nov8.html?sub=AR](http://www.washingtonpost.com/wp-dyn/articles/A35074-2004Nov8.html?sub=AR) (as of November 18, 2004).

McCabe, Pat, telephone communication with the author on November 19, 2004a.

---, telephone communication with the author on October 1, 2004b.

McCaffrey, Dan, J.R. Lockwood, Daniel M. Koretz, and Laura Hamilton, *Evaluating Value-Added Models for Teacher Accountability*, Santa Monica: RAND Corporation, 2003.

McCauley, Lisa, and Susannah Roddick, "An Evaluation of Schools Support," *Research and Evaluation Unit - New Zealand Ministry of Education*, 2001. Online at [www.minedu.govt.nz/web/downloadable/dl6653\\_v1/schools-support.pdf](http://www.minedu.govt.nz/web/downloadable/dl6653_v1/schools-support.pdf) (as of October 4, 2004).

McMahon, Tim, "Schools Planning and Reporting - Evidence to Action," *New Zealand Ministry of Education*, 2002. Online at [www.minedu.govt.nz/index.cfm?layout=document&documentid=6927&data=1](http://www.minedu.govt.nz/index.cfm?layout=document&documentid=6927&data=1) (as of October 4, 2004).

McRobbie, Joan, "Can State Intervention Spur Academic Turnaround?," *WestEd*, 1998. Online at [www.wested.org/pub/docs/policy/intervention.htm](http://www.wested.org/pub/docs/policy/intervention.htm) (as of June 8, 2004).

Microsoft, Excel, v. 2000, Redmond, WA, 2000.

Miller, Doug, personal communication with the author on January 27, 2004.

Ministry of Education, "Circular: Education Amendment Act 2000," *New Zealand Ministry of Education*, 2000. Online at [www.minedu.govt.nz/index.cfm?layout=document&documentid=5107&indexid=1011&ndexparentid=1010](http://www.minedu.govt.nz/index.cfm?layout=document&documentid=5107&indexid=1011&ndexparentid=1010) (as of October 7, 2004).

---, "Circular: Education Standards Act 2001," *New Zealand Ministry of Education*, 2001. Online at [www.minedu.govt.nz/index.cfm?layout=document&documentid=6426&indexid=1011&ndexparentid=1010](http://www.minedu.govt.nz/index.cfm?layout=document&documentid=6426&indexid=1011&ndexparentid=1010) (as of October 7, 2004).

---, "Planning for Better Student Outcomes - quarterly updates and blue brochure," *New Zealand Ministry of Education*, 2003a. Online at [www.minedu.govt.nz/index.cfm?layout=document&documentid=9812&data=1](http://www.minedu.govt.nz/index.cfm?layout=document&documentid=9812&data=1) (as of October 4, 2004).

---, *Schools Monitoring and Support - presentation*, Wellington: New Zealand Ministry of Education, 2003b.

---, *Schools Monitoring and Support: Introductory Presentation*, Wellington: New Zealand Ministry of Education, 2003c.

---, "Relevant Legislation, NEGs and NAGs," *New Zealand Ministry of Education*, 2004. Online at [www.minedu.govt.nz/index.cfm?layout=index&indexID=8186&indexparentid=1012](http://www.minedu.govt.nz/index.cfm?layout=index&indexID=8186&indexparentid=1012) (as of October 3, 2004).

Mintrop, Heinrich, "The limits of sanctions in low-performing schools: A study of Maryland and Kentucky schools on probation," *Education Policy Analysis*, Vol. 11, No. 3, 2002.

Mintrop, Heinrich, and Tina Trujillo, "Corrective Action in Low-Performing Schools: Lessons for NCLB Implementation from State and District Strategies in First-Generation Accountability Systems," *National Center for Research on Evaluation, Standards, and Student Testing*, 2004. Online at <http://cresst96.cse.ucla.edu/reports/r641.pdf> (as of December 30, 2004).

Mitchell, Douglas E., and Ross E. Mitchell, "The Political Economy of Education Policy: The Case of Class Size Reduction," *Peabody Journal of Education*, Vol. 78, No. 4, 2003, pp. 120-152.

Mullins, Jo, telephone communication with the author on November 9, 2003.

---, personal communication with the author on January 16, 2004.

Myers, Kate, and Harvey Goldstein, "Who's Failing?," in Louise Stoll and Kate Myers, eds., *No Quick Fixes: Perspectives on Schools in Difficulty*, London: The Falmer Press, 1998.

NAASO, "What is Obesity?," *North American Association for the Study of Obesity*, 2003. Online at [www.naaso.org/information/what\\_is\\_obesity.asp](http://www.naaso.org/information/what_is_obesity.asp) (as of September 13, 2004).

NACSA, "Built for Quality: The Capacity Needed to Oversee Charter Schools," *National Association of Charter School Authorizers*, 2004. Online at [www.charterauthorizers.org/files/nacsa/BECSA/IssueBriefNo3.pdf](http://www.charterauthorizers.org/files/nacsa/BECSA/IssueBriefNo3.pdf) (as of August 18, 2004).

Nagel, Jack H., "Editor's Introduction," *Journal of Policy Analysis and Management*, Vol. 16, No. 3, 1997, pp. 349-356.

National Defense Education Act, "Summary of Major Provisions," *Center for Studies in Higher Education*, 1958. Online at <http://ishi.lib.berkeley.edu/cshe/ndea/ndea.html> (as of May 26, 2004).

NCES, "Highlights from the Third International Mathematics and Science Study-Repeat," *National Center for Education Statistics*, 2001. Online at <http://nces.ed.gov/pubs2001/2001027.pdf> (as of October 1, 2004).

---, "Public School Expenditures," *National Center for Education Statistics*, 2004. Online at [www.nces.ed.gov/fastfacts/display.asp?id=66](http://www.nces.ed.gov/fastfacts/display.asp?id=66) (as of May 26, 2004).

NCLB, "No Child Left Behind Act of 2001," *107th Congress of the United States*, 2001. Online at [www.ed.gov/policy/elsec/leg/esea02/index.html](http://www.ed.gov/policy/elsec/leg/esea02/index.html) (as of December 2, 2004).

O'Day, Jennifer, "Complexity, Accountability, and School Improvement," *Harvard Educational Review*, Vol. 72, No. 3, 2002, pp. 293-329.

O'Day, Jennifer, and Catherine Bitter, "Evaluation Study of the Immediate Intervention/Underperforming Schools Program and the High Achieving/Improving Schools Program of the Public Schools Accountability Act of 1999," *American Institutes for Research*, 2003. Online at [www.air.org/pubs/PhaseIIfinalReportentire.pdf](http://www.air.org/pubs/PhaseIIfinalReportentire.pdf) (as of December 2, 2004).

OECD, "Overcoming Failure At School," *Organization for Economic Co-operation and Development*, 1998. Online at [www.mszs.si/eurydice/pub/oecd/failure.pdf](http://www.mszs.si/eurydice/pub/oecd/failure.pdf) (as of May 25, 2004).

OED, "Oxford English Dictionary," *Oxford University Press*, 2004. Online at <http://dictionary.oed.com/> (as of August 24, 2004).

Office of Communications, "Fact Sheet: Education for a New Era," *Supreme Education Council, Qatar*, 2004a. Online at [www.english.education.gov.qa/files/838\\_factsheetfinal.pdf](http://www.english.education.gov.qa/files/838_factsheetfinal.pdf) (as of December 22, 2004).

---, "Press Release: SEC Signs Operators to Run New Independent Schools," *Supreme Education Council, Qatar*, 2004b. Online at [www.english.education.gov.qa/files/967\\_Operator\\_ContractsJune\\_6.pdf](http://www.english.education.gov.qa/files/967_Operator_ContractsJune_6.pdf) (as of December 25, 2004).

Osborne, Jason W., "Advantages of hierarchical linear modeling," *Practical Assessment, Research & Evaluation*, 2000. Online at <http://pareonline.net/getvn.asp?v=7&n=1> (as of December 3, 2004).

Ott, Maria, personal communication with the author on September 27, 2004.

Ouchi, William G., *Making Schools Work: A Revolutionary Plan to Get Your Children the Education They Need*, New York: Simon & Schuster, 2003.

Peninsula, "Over 86% Qataris favour govt jobs: Study," *The Peninsula*, February 20, 2003.

PISA, "Summary of International Report: New Zealand," *Programme for International Student Assessment*, 2001. Online at [www.pisa.oecd.org/NatReports/PISA2000/NewZealandSummaryInternationalResults.pdf](http://www.pisa.oecd.org/NatReports/PISA2000/NewZealandSummaryInternationalResults.pdf) (as of October 1, 2004).

Plank, David N., and Gary Sykes, "Why School Choice?," in David N. Plank and Gary Sykes, eds., *Choosing Choice: School Choice in International Perspective*, New York: Teachers College Press, 2003.

Premack, Eric, "CSDC Presentation for the Education Institute, October 26-27, 2004," *Supreme Education Council, Qatar*, 2004. Online at [www.english.education.gov.qa/files/1216\\_Financial\\_Planning\\_v5\\_updated\\_Read\\_Only\\_.pdf](http://www.english.education.gov.qa/files/1216_Financial_Planning_v5_updated_Read_Only_.pdf) (as of December 22, 2004).

Qatar Special Report, "A 21st century success story," *The Middle East Magazine*, No. 324, June 2002.

Qatari Ministry of Education, *System of Internal Regulations of Private (non-governmental) Education*, Doha: State of Qatar, 1986.

---, *Annual Report*, Doha: State of Qatar, 2000.

Quinn, Juli, Linda Reed, and Bo Vitolo, "Building a Local Assistance Model For School Improvement," *California County Superintendents Educational Services Association, CCSESA*, 2003. Online at <http://wwwstatic.kern.org/gems/ccsesaCisc/Quinn.pdf> (as of May 5, 2004).

Robinson, Viviane, and Helen Timperley, "Strengthening Education in Mangere and Otaara Evaluation: Final Evaluation Report," *Ministry of Education*, 2004. Online at [www.minedu.govt.nz/web/downloadable/dl9200\\_v1/semopdf](http://www.minedu.govt.nz/web/downloadable/dl9200_v1/semopdf) (as of April 2004).

Rogosa, David, "Irrelevance of Reliability Coefficients to Accountability Systems: Statistical Disconnect in Kane-Staiger "Volatility in School Test Scores"," *Stanford University*, 2002. Online at <http://www-stat.stanford.edu/%7Erag/api/ksresst.pdf> (as of May 20, 2004).

---, "Confusions about Consistency in Improvement," *Stanford University*, 2003. Online at <http://www-stat.stanford.edu/%7Erag/api/consist.pdf> (as of May 12, 2004).

Rosenthal, Leslie, "Do school inspections improve school quality? Ofsted inspections and school examination results in the UK," *Economics of Education Review*, Vol. 23, No. 2, 2004, pp. 143-151.

Rudo, Zena H., "Corrective Action in Low-Performing Schools and School Districts," *Southwest Educational Development Laboratory*, 2001. Online at [www.sedl.org/pubs/policy91/policy91.pdf](http://www.sedl.org/pubs/policy91/policy91.pdf) (as of December 30, 2004).

Rutherford, F. James, "Sputnik and Science Education," *American Association for the Advancement of Science*, 1998. Online at [www.nas.edu/sputnik/rutherfo.doc](http://www.nas.edu/sputnik/rutherfo.doc) (as of May 27, 2004).

SB1X, "Public Schools Accountability Act of 1999," *California State Legislature*, 1999. Online at [www.leginfo.ca.gov/pub/99-00/bill/sen/sb\\_0001-0050/sbx1\\_1\\_bill\\_19990405\\_chaptered.html](http://www.leginfo.ca.gov/pub/99-00/bill/sen/sb_0001-0050/sbx1_1_bill_19990405_chaptered.html) (as of December 2, 2004).

Schemo, Diana Jean, "New Federal Rule Tightens Demands on Failing Schools," *The New York Times*, November 27, 2002.

Schumpeter, Joseph A., *Capitalism, Socialism and Democracy*, New York: Harper & Brothers, 1947.

Scott, Graham, Ian Ball, and Tony Dale, "New Zealand's Public Sector Management Reform: Implications for the United States," *Journal of Policy Analysis and Management*, Vol. 16, No. 3, 1997, pp. 357-381.

Short, Andrew, personal communication with the author on January, 2004.

Sinclair, Mary, "Making a Difference Together: An Overview," *Schools Monitoring and Support - Ministry of Education*, 2001. Online at [www.tki.org.nz/r/school\\_improvement/resources/conf\\_papers/sch\\_mon\\_maryS.doc](http://www.tki.org.nz/r/school_improvement/resources/conf_papers/sch_mon_maryS.doc) (as of October 5, 2004).

Smith, Jeffrey, "Evaluating Local Economic Development Policies: Theory and Practice (version 1/12/04)," *University of Maryland*, 2004. Online at [www.glue.umd.edu/~jsmithz/papers/ledeval011204.pdf](http://www.glue.umd.edu/~jsmithz/papers/ledeval011204.pdf) (as of November 10, 2004).

Stark, Michael, "No Slow Fixes Either: How Failing Schools in England are Being Restored to Health," in Louise Stoll and Kate Myers, eds., *No Quick Fixes: Perspectives on Schools in Difficulty*, London: The Falmer Press, 1998.

StataCorp., *Stata Statistical Software*, v. 7.0, College Station, TX, 2001.

Stecher, Brian, Laura Hamilton, and Gabriella Gonzalez, *Working Smarter to Leave No Child Behind*, Santa Monica: RAND Corporation, 2003.

Stoll, Louise, and Kate Myers, "No Quick Fixes: An Introduction," in Louise Stoll and Kate Myers, eds., *No Quick Fixes: Perspectives on Schools in Difficulty*, London: The Falmer Press, 1998.

Te Tari Tautau, "Quick Facts - People," *Statistics New Zealand*, 2004. Online at [www.stats.govt.nz/domino/external/web/nzstories.nsf/htmldocs/Quick+Facts+-+People](http://www.stats.govt.nz/domino/external/web/nzstories.nsf/htmldocs/Quick+Facts+-+People) (as of September 28, 2004).

Tokman, Andrea, "Evaluation of the P900 program: A targeted education program for underperforming schools," *Central Bank of Chile*, 2002. Online at [www.bcentral.cl/esp/estpub/estudios/dtbc/pdf/dtbc170.pdf](http://www.bcentral.cl/esp/estpub/estudios/dtbc/pdf/dtbc170.pdf) (as of December 14, 2004).

UNDP, *Arab Human Development Report 2002: Creating Opportunities for Future Generations*, New York: United Nations Development Programme, 2002.

---, *Arab Human Development Report 2003: Building a Knowledge Society*, New York: United Nations Development Programme, 2003.

United Press International, "Bin Laden tape blasts reform, Saudi royals," *GulfInTheMedia*, 2004. Online at [www.gulfinthemedia.com/gulf\\_media/full.php?module=view\\_article\\_en\\_same&enc=1&free=1&id=93787](http://www.gulfinthemedia.com/gulf_media/full.php?module=view_article_en_same&enc=1&free=1&id=93787) (as of December 22, 2004).

Varenne, Hervé, and Ray McDermott, *Successful Failure: The School America Builds*, Boulder: Westview Press, 1998.

Wagner, Laura, telephone communication with the author on November 29, 2004.

Will, George F., "A Radical Approach To a Big City's Schools," *Washington Post*, 2004. Online at [www.townhall.com/columnists/georgewill/printgw20040314.shtml](http://www.townhall.com/columnists/georgewill/printgw20040314.shtml) (as of September 12, 2004).

Wolk, Ron, "Strategies for Fixing Failing Schools," *Education Week*, 1998. Online at [www.edweek.org/ew/articles/1998/11/18/12pew.h18.html](http://www.edweek.org/ew/articles/1998/11/18/12pew.h18.html) (as of October 18, 2004).

Woodfield, Alan, and Philip Gunby, "The Marketization of New Zealand Schools: Assessing Fiske and Ladd," *Journal of Economic Literature*, Vol. 41, 2003, pp. 863-884.

World Bank, "Educational change in Latin America and the Caribbean: A World Bank strategy paper," *World Bank*, 1999. Online at [www.fmmeducacion.com.ar/Sisteduc/Banco/bmeducamerica.pdf](http://www.fmmeducacion.com.ar/Sisteduc/Banco/bmeducamerica.pdf) (as of December 14, 2004).

Wylie, Cathy, "School Self-Management in New Zealand - How can it make a real difference?," NZCER Annual Conference, Wellington, 1998.

---, *Ten Years On: How Schools View Educational Reform*, Wellington: New Zealand Council for Educational Research, 1999.

---, "The local and systemic roles of school trustees," NZARE conference, Palmerston North, 2002.

Young, Graham, "Creeping Centralism," *SPANZ Journal*, 2003.

Zahlan, Rosemarie Said, *The making of the modern Gulf states: Kuwait, Bahrain, Qatar, the United Arab Emirates and Oman*, Reading, UK: Ithaca Press, 1998.

Ziebarth, Todd, "State Takeovers and Reconstitutions," *Education Commission of the States*, 2002. Online at [www.ecs.org/clearinghouse/13/59/1359.pdf](http://www.ecs.org/clearinghouse/13/59/1359.pdf) (as of May 20, 2004).

Zimmer, Ron, Richard Buddin, Derrick Chau, Brian Gill, Cassandra Guarino, Laura Hamilton, Cathy Krop, Dan McCaffrey, Melinda Sandler, and Dominic Brewer, *Charter School Operations and Performance: Evidence from California*, Santa Monica: RAND Education, 2003.