

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Mark Ronald Saliba entitled The Response of School Leaders to the Demands of Equity in the Environment of Accountability and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Education

Kris Bosworth

Date: 11/13/07

Robert Hendricks

Date: 11/13/07

John Cheslock

Date: 11/13/07

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College. I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director: Kris Bosworth

Date: 11/13/07

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Mark Ronald Saliba

ACKNOWLEDGEMENTS

I would like to acknowledge the following for their generous assistance and support:

Dr. Kris Bosworth, my committee chair

Dr. John Cheslock, committee member

Dr. Robert Hendricks, committee member

The following individuals reviewed and vetted the interview questions and provided helpful guidance:

Mr. Ray Chavez

Dr. Stephanie Fryberg

Mr. Steve Holmes

Mr. Roger Owen

Mr. Rick Hahn provided valuable background on accountability systems.

Dr. Matthias Mehl graciously ran the quantitative word analysis and provided guidance on its interpretation.

Ms. Deleana OtherBull and Ms. Rabiah Muhammad read the interview transcripts and provided helpful insights.

DEDICATION

This dissertation is dedicated to the inspiring women who participated in the study and to my wife, Stephanie, and my parents, Ron and Helen, for their constant love and support.

TABLE OF CONTENTS

	Page
LIST OF FIGURES.....	11
LIST OF TABLES.....	12
ABSTRACT.....	13
CHAPTER 1. INTRODUCTION AND PROBLEM STATEMENT.....	14
The Requirement to Educate All Children.....	14
Contextual Inequalities That May Hinder the Goal of Equitable Education.....	16
<i>Segregation and Resegregation</i>	17
<i>School Funding Mechanisms</i>	17
<i>Differences in Performance</i>	19
<i>Inequalities within Schools</i>	20
NCLB and the Shift of Accountability.....	21
The Challenge for School Leaders.....	22
Previous Practices Related to Achieving Educational Equity.....	23
Requirements of Accountability Systems.....	27
The Influence of School Profile.....	29
Research Questions.....	30
Description of the Research.....	36
Professional Significance of the Study.....	36
Limitations of the Study.....	37
Definitions.....	38
CHAPTER 2. REVIEW OF THE LITERATURE.....	42
Search Process.....	42
Inequalities that Hinder Educational Equity.....	44
<i>Segregation</i>	45
<i>Desegregation and resegregation trends</i>	46
<i>The effect of desegregation on student outcomes</i>	49
<i>The future of desegregation</i>	54
<i>School Funding</i>	56
<i>How funding inequalities are structured</i>	57
<i>Three waves of litigation</i>	59
<i>Theoretical considerations</i>	62
<i>The cost of adequacy</i>	64
<i>Inequalities of Performance</i>	66

TABLE OF CONTENTS – *Continued*

	Page
<i>Theoretical literature</i>	66
<i>Empirical research</i>	72
<i>Inequalities within Schools</i>	76
<i>Social reproduction: theory and evidence</i>	77
<i>Equity traps</i>	80
<i>Equity audits</i>	83
<i>Structural inequalities within schools</i>	86
The Shift to Output Criteria: School Accountability Systems.....	87
<i>Overview</i>	88
<i>No Child Left Behind</i>	91
<i>Adequate yearly progress</i>	92
<i>Subgroups</i>	92
<i>Unit of accountability</i>	94
<i>Potential benefits and pitfalls</i>	95
<i>State Systems</i>	97
<i>Texas</i>	97
<i>California</i>	98
<i>Florida</i>	98
<i>Arizona</i>	99
<i>Promise and Limitations of Accountability Systems: Empirical Findings</i>	101
<i>Benefits for “bubble” students</i>	101
<i>Subgroup distortions</i>	102
<i>Reliability</i>	104
<i>NCLB and segregation</i>	106
<i>Adjusting Accountability Systems</i>	107
School Leaders in the Accountability Environment.....	113
<i>School Leaders and Educational Equity</i>	113
<i>Fostering new meanings about diversity</i>	114
<i>Promoting inclusive school practices: instructional leadership</i>	116
<i>Promoting inclusive school practices: molding school cultures</i> <i>that embrace and support diversity</i>	125
<i>Building communities</i>	126
<i>School Leaders and Accountability Systems</i>	128
<i>Responses of leaders to accountability incentives</i>	128
<i>Use of data</i>	132
<i>Leader behaviors and effective accountability practices</i>	133
CHAPTER 3. METHODS.....	136

TABLE OF CONTENTS – *Continued*

	Page
The Research Context.....	136
The Research Participants.....	141
Interviews.....	142
School and Participant Descriptions.....	143
<i>Group 1: Whitcomb Elementary</i>	143
<i>Group 2: Desert Springs Elementary</i>	146
<i>Group 3: Oakwood Elementary</i>	149
<i>Group 4: Arroyo Elementary</i>	156
Data Analysis.....	159
<i>The Equity Achievement Driver: A Tool for Measuring Educational Equity</i>	159
<i>Linguistic Inquiry and Word Count</i>	169
<i>Qualitative Analysis of Interviews</i>	169
CHAPTER 4. RESULTS.....	172
Introduction.....	172
Differences around Professions.....	179
Differences among Schools.....	186
The Core Category: Meeting Student Performance Challenges.....	190
<i>Using the Accountability Environment to Induce Accountable Behavior</i>	191
<i>Unit Level of Analysis</i>	198
<i>Individual student level</i>	199
<i>Grade level</i>	202
<i>Classroom level</i>	204
<i>Subgroup level</i>	206
<i>Utilizing Available Resources</i>	207
<i>Types of Student Performance Goals</i>	209
<i>Two perspectives on academic goals</i>	209
<i>Long-term goals</i>	213
<i>Summary</i>	214
Second Category: Investing in Teachers.....	214
<i>Recruiting / Hiring</i>	216
<i>Supporting / Retaining</i>	217
<i>Training / Developing</i>	220
<i>Qualities of a Good Teacher</i>	224
<i>Dealing with Resistors</i>	226
<i>Investing in Teachers and Meeting Student Performance Challenges</i>	228
Third Category: Creating a Productive Learning Environment.....	230
<i>High Expectations</i>	231
<i>Respect</i>	232

TABLE OF CONTENTS – *Continued*

	Page
<i>Professionalism</i>	233
<i>Safe, Inclusive, Caring Environments</i>	234
<i>Safe for Taking Risks</i>	236
<i>Community</i>	237
<i>Future Orientation</i>	238
Fourth Category: Achieving Educational Equity.....	239
<i>Educational Equity as a School Responsibility</i>	240
<i>Barriers to Equity</i>	241
<i>Overcoming Barriers to Equity</i>	244
Fifth Category: Testing and Data.....	248
<i>Collaborative Data Analysis</i>	250
<i>Types of Data</i>	251
<i>Teacher Judgment and Data</i>	252
<i>Leader Involvement with Data</i>	252
Sixth Category: Intervention for Struggling Students.....	253
<i>Initiation of Interventions</i>	255
<i>Types of Interventions</i>	256
<i>Time</i>	258
Seventh Category: Identity as a Leader.....	258
<i>Leader Beliefs and Dispositions</i>	259
<i>Background Experiences</i>	262
<i>Leader Actions</i>	264
Eighth Category: Dealing with the Accountability Environment.....	266
<i>Response to the Accountability Environment</i>	266
<i>Low Knowledge of Accountability System Elements</i>	269
<i>Perceived Accountability Environment Benefits</i>	271
<i>Neutral Areas</i>	273
<i>Perceived Accountability Environment Shortcomings</i>	273
<i>Earning / Retaining a Good Label</i>	279
Ninth Category: Coordinating Efforts.....	281
<i>Curriculum Calendars</i>	282
<i>Professional Learning Communities Focused on Student Data</i>	282
<i>Coordination between School and Home</i>	284
<i>Coordination between principals and instructional coaches</i>	284
<i>Leadership Team</i>	285
<i>School-wide Programs</i>	286
Summary.....	287
CHAPTER 5. EVALUATING SCHOOLS' EDUCATIONAL EQUITY.....	288

TABLE OF CONTENTS – *Continued*

	Page
Comparing Individual School Performance with the Equity Achievement Driver.....	288
<i>Schools that Gained in Rank</i>	302
<i>Schools that Declined in Rank</i>	306
<i>Schools that Held their Position</i>	307
Analyzing Groups of Schools with the Equity Achievement Driver.....	308
Use of a Correction Factor.....	310
 CHAPTER 6. SUMMARY AND DISCUSSION.....	 312
Statement of the Problem.....	312
Methodology.....	314
Results.....	315
<i>Differences in Participants that Clustered by Profession</i>	315
<i>Differences in Participants that Clustered by School</i>	316
<i>Meeting Student Performance Challenges</i>	317
<i>Investing in Teachers</i>	318
<i>Creating a Productive Learning Environment</i>	320
<i>Achieving Educational Equity</i>	321
<i>Testing and Data</i>	322
<i>Interventions for Struggling Students</i>	323
<i>Identity as a Leader</i>	323
<i>Dealing with the Accountability Environment</i>	324
<i>Coordinating Efforts</i>	325
Evaluating Schools' Educational Equity.....	326
Discussion.....	328
<i>A Conceptual Model</i>	330
<i>Implications for Practice</i>	346
<i>Implications for Research</i>	350
 APPENDIX: INTERVIEW GUIDE.....	 352
 REFERENCES.....	 361

LIST OF FIGURES

	Page
<i>Figure 1.</i> Conceptual framework.....	24
<i>Figure 2.</i> Reading performance of fifth graders on the Arizona Instrument to Measure Standards for elementary schools in four socioeconomic groups within the district in 2006.....	138
<i>Figure 3.</i> Math performance of fifth graders on the Arizona Instrument to Measure Standards for elementary schools in four socioeconomic groups within the district in 2006.....	139
<i>Figure 4.</i> Writing performance of fifth graders on the Arizona Instrument to Measure Standards for elementary schools in four socioeconomic groups within the district in 2006.....	140
<i>Figure 5.</i> Conceptual model showing interrelationships between categories in the study.....	331

LIST OF TABLES

	Page
Table 1, <i>Summary Data for Participating Schools</i>	160
Table 2, <i>Summary Descriptive Data for Principals and Instructional Coaches at Four Selected Elementary Schools</i>	161
Table 3, <i>AIMS Scores Disaggregated by Race for Tenth Graders in Three Subjects at Two High Schools</i>	167
Table 4, <i>Equity Achievement Drivers D_{ea} for High Schools D and E</i>	168
Table 5, <i>Linguistic Word Count Results for 64 Dimensions</i>	180
Table 6, <i>Reading Performance of Fifth Graders Ranked on the Basis of Raw Score by School</i>	289
Table 7, <i>Reading Performance of Fifth Graders Ranked on the Basis of Equity Achievement Driver by School</i>	296
Table 8, <i>Schools that Gained or Lost 10 or More Positions in Rank When the Equity Achievement Driver was Applied</i>	303
Table 9, <i>Performance of Fifth graders District-Wide and in Groups of Schools Stratified by Socioeconomic Status</i>	309

ABSTRACT

Principals and instructional coaches from four high-performing and equitable-performing elementary schools stratified by socioeconomic level were interviewed. The main concern of participants was meeting student performance challenges set by accountability systems and measured by mandated tests. This concern was manifested in a focus on the needs and educational progress of individual children. Other participant concerns included investing in teachers, analyzing test data collaboratively, intervening on behalf of struggling students, dealing with the current accountability environment, building productive learning environments, achieving educational equity, and maintaining identity as a leader. Participants demonstrated a rather low knowledge of accountability system mechanics and ambivalence about accountability system features; however, they fully embraced the spirit of “leaving no child left behind.” They also emphasized many elements of leadership that predate the current accountability environment, including instructional leadership, professional learning communities, and high expectations for students. Although categories were consistent among the schools, other factors co-vary with school size (performance unit of analysis) or school socioeconomic status (future vs. present orientation). A statistical measure for evaluating the educational equity of schools is introduced.

CHAPTER 1

INTRODUCTION AND PROBLEM STATEMENT

The challenge of ensuring an equitable education for all citizens in a society that is characterized by increasing racial, ethnic, linguistic and socioeconomic diversity has come into sharp focus at the beginning of the new millennium. Increasing diversity in society is reflected in the makeup of the national student population and is an important factor affecting demographic changes within regions, districts and individual schools. In such an environment, schools are being held increasingly accountable for educating students from diverse backgrounds in such a way that all groups of students meet required standards, and school leaders are the first to be held accountable.

The Requirement to Educate All Children

The new *accountability environment* that school leaders face as they seek to lead in educating diverse populations of students in the 21st century has the feature of requiring specific outputs mandated by increasingly involved federal and state governments. Prior to the imposition of federal and state accountability systems, a consensus had emerged among educators across the political spectrum that all students, regardless of background or personal characteristics, must be provided with the means to succeed in school (Riehl, 2000). The mandate to educate all children has been identified as a moral responsibility for educators (Fullan, 2003). Public education is seen as a common good that benefits all of society by preparing students to contribute to building society through meaningful careers and instilling in them the qualities necessary for citizenship in a diverse, democratic society. To fulfill these functions, education must

serve all children equitably. Singleton and Linton (2006) defined *educational equity* as “raising the achievement of all students while narrowing the gaps between the highest- and lowest-performing students; and eliminating the racial predictability and disproportionality of which student groups occupy the highest and lowest achievement categories” (p. 46). The failure to achieve educational equity is a fundamental failure of education itself. As one researcher has put it: “Having kids not educated in relation to race and class is a moral issue of enormous consequences” (Fullan, 2003, p. 65).

The familiar quote “all children can learn” has now evolved into to the federal mandate “all children must learn.” The No Child Left Behind Act (NCLB) of 2001 requires that all children meet minimum standards in reading and mathematics by the year 2014 (U.S. Department of Education [U.S. DOE], 2002). Student performance and progress in accordance with the act is monitored through state-mandated assessments. These assessments are high-stakes for both students and schools alike: Students must demonstrate mastery in order to graduate from high school in an increasing number of states, and schools must demonstrate that they are making *adequate yearly progress (AYP)* toward the goal of mastery by all students. In its assessment of adequate yearly progress, NCLB takes account of the performance of different racial, ethnic, socioeconomic, language, and disability groups within a school. Schools that repeatedly do not make AYP are subject to sanctions that include reorganization by the state and removal of the principal.

This continual rising bar of performance has been compared by one school board member (Hardy, 2003) to a room in which the water is rising, and “You can’t get out and

the room's half full. You know that by 2014 that room's going to be full of water,. . . . Eventually, the water's going to hit the ceiling, by definition" (p. 30). The ever-increasing pressure of NCLB (U.S. DOE, 2002) is a new reality that school leaders must take account of in their planning and practice.

Contextual Inequalities That May Hinder the Goal of Equitable Education

The achievement of educational equity has been an elusive goal. Its attainment may be hindered by several persistent inequalities in the nation's educational system that have been documented and tracked over several decades (Lee, 2004; McKenzie & Scheurich, 2004; McUsic, 1999; Orfield, 2001). Four types of inequalities relevant to the discussion of educational equity are (a) segregation or racial isolation, (b) school funding mechanisms, (c) achievement differences, and (d) inequalities within schools. These school inequalities mirror inequalities American society and culture such as the segregation of housing by race; the correlation of race with poverty; unequal access to higher education and high-paying jobs; and racism, which continues to be practiced and experienced in subtle and unsubtle ways. Schools must deal with many of the same problems embedded in the broader culture and are somehow expected to do better.

School inequalities are interrelated in ways that are only partly understood, and the ability of school leaders to control these inequalities varies. Inequalities within schools may be influenced by school leaders, while other inequalities, such as segregation and school funding, fall within the realm of public policy, legislation, and litigation. However, wherever the locus of control may exist, all types of schools and their leaders are affected to some degree by these inequalities, and under the new accountability

system rules, they are somehow required to overcome the hindrances these inequalities present within the 12-year limit imposed by NCLB (U.S. DOE, 2002).

Segregation and Resegregation

After the *Brown v. Board of Education of Topeka* (1954) decision, courts ordered systematic desegregation in all regions of the country. The desegregation trend continued into the mid-1980s, resulting in a significant decrease in the isolation of African American students (Orfield, 2001). The outcome of desegregation that was most relevant to this study is that as desegregation progressed, the achievement gap between African American and white students narrowed significantly for a time, as measured by standardized test scores (Lee, 2004). Since the stalling of the desegregation trend for African American students in the 1980s and a series of decisions by the U.S. Supreme Court authorizing a return to segregated neighborhood schools, there has been a general trend toward resegregation for these students (Orfield, 2001). Coinciding with the trend toward resegregation, there has been a flattening of the percentage of African American and Latino students achieving basic mathematics proficiency and an increase in achievement gaps as measured by the National Assessment of Educational Progress, or NAEP (Lee, 2004). That trend suggests that further resegregation may be accompanied by a further widening of achievement gaps.

School Funding Mechanisms

Laws at the local, state, and federal level have set up and ensured unequal funding of schools through mechanisms of unfunded federal mandates, local control of schools, and subsidization of private and parochial schools (McUsic, 1999). To correct

the inequalities inherent in the funding system, states make supplemental appropriations intended to equalize resources among districts. However, the promised equalization is rarely achieved, and in some cases the state interventions aggravate the inequalities.

Litigation has become the primary remedy for the unequal funding of schools (McUsic, 1999). Most legal challenges to school finance systems have relied on the education clause or the equal protection clause in state constitutions to show that unequal education is a constitutional violation. The original school finance lawsuits were desegregation cases in which courts intervened in school systems in all regions of the U.S. to limit freedom of choice, prescribe transfers of students within and across district boundaries, and order funding for desegregation projects. These measures resulted in a measure of funding equalization.

School finance litigation became an independent movement after the Supreme Court restricted desegregation orders to apply only within district boundaries (*Milliken v. Bradley*, 1974 & 1977). School finance litigation has evolved through several phases; the most recent cases, such as *Rose v. Council for Better Education* (1989), were built on the criterion of *adequacy* (Roellke, Green, & Zielewski, 2004). This approach began with the question, “What constitutes an adequate education?” Such litigation seeks to measure the educational services being delivered and show that they are insufficient to satisfy constitutional requirements. Litigants then seek to restructure the educational delivery system to bring the achievement of all students and especially the disadvantaged up to an acceptable, adequate standard.

Differences in Performance

In many schools and districts, large numbers of students are not meeting minimum standards necessary for graduation, and differences in achievement that were identified decades ago between different racial, ethnic and socioeconomic groups have stubbornly persisted. These performance differences have been referred to as achievement gaps, and often, monolithically, as *the achievement gap*. *No Child Left Behind* (U.S. DOE, 2002) focused attention on achievement differences and required their elimination by 2014 (U.S. Department of Education, 2002). Since the adoption of the NCLB, however, urban school districts have made only modest progress in reading and math at the fourth and eighth grade levels, and achievement gaps may be growing, according to the National Assessment for Academic Progress (Sanchez, 2005; U.S. Department of Education, 2005).

Differences in performance are evaluated from widely different educational and political perspectives. Some theorists emphasize race as the key factor; others point to differences based on socioeconomic status. The high correlation between these two variables contributes to the ongoing argument. Class difference theorists such as Richard Rothstein (2004) argued that socioeconomic or class difference regardless of race was the major contributor to academic performance differences between children from middle class families and children from poor and working class families; and that equity could not be achieved until class differences were reduced. Conservative theorist Chester Finn (2006) countered with examples of hi-poverty schools that beat the odds and attained

high performance levels. For other scholars, latent institutional racism trumped class as the source of achievement gaps (Singleton & Linton, 2006).

Achievement gaps have been framed as a multidimensional problem involving issues of equality, adequacy, and reciprocity (Lee, 2004). Such an approach allows for an analysis and evaluation of the extent of inequity and injustice underlying the gaps. In contrast, critical race theorists, while acknowledging performance differences, criticized the formulation of and emphasis on the achievement gap as a subversion of the goals of the Civil Rights Movement.

Inequalities within Schools

Inequalities of instruction, access, and participation between White students and students of color within schools and districts can also hinder equity in schools. Students of color are more likely than White students to be educated by less experienced teachers; to be over-represented in special education programs and lower level courses; to be under-represented in programs for the gifted and talented; to drop out of school; to be educated by teachers who are not certain they can learn or are actively negative in their attitude toward these students; and to be suspended or expelled (McKenzie & Scheurich, 2004). Compounding these inequalities is the finding that issues of equitable practice are rarely discussed in school buildings (Riehl, 2000). Of all the factors affecting equity in schools, within-school inequity may be the factor most susceptible to leverage by the principal and other school leaders.

NCLB and the Shift of Accountability

A concern for educational inputs has historically driven movements for school equity such as desegregation and school finance reform. An assumption by the proponents of these movements has been that as inputs in the form of educational opportunities and resources were equalized, equitable outputs in the form of achievement gains and narrowing of achievement differences would result (McUsic, 1999). This type of result was observed in the case of African American students relative to their White peers during the era of desegregation (Lee, 2004).

A recent development has been a shift away from a strict focus inputs as the sole measure of equity towards a more nuanced consideration of both inputs and outputs. The concept of adequacy has been important in this shift. From an adequacy perspective, a specific desired outcome is first defined; then, restructuring and inputs required to achieve the output are considered and pursued (Alexander, 2004). Coincidental with the shift in focus from inputs to both inputs and outputs was the standards movement. Since the 1980s, state legislatures, educator organizations, and boards of education have designed specific academic standards and measured their attainment by students. The standards movement has proved persistent and forms the basis of current accountability systems. This persistence may be based on the measurability of standards and their applicability across states. Standards are potentially valuable tools in defining what is meant by an adequate, constitutionally-guaranteed level of education (McUsic, 1999).

With the advent of NCLB (U.S. DOE, 2002) came a further shift toward emphasis on educational outputs and a de-emphasis on inputs. An explicit goal and requirement of

NCLB was the closing of achievement gaps between disadvantaged students and students of color and their peers. In this respect, NCLB aligned with school finance reform and the desegregation efforts, and it elevated the goal of educational equity to the level of national policy. However, NCLB diverged from the other movements as to how the goal would be achieved and how accountability was assigned. School finance reform and equity movements insisted on inputs necessary to achieve specified standards, whereas NCLB demanded uniform outputs from differently advantaged schools (Fusarelli, 2004). Desegregation and school finance rulings generally held states accountable for providing adequate education, yet NCLB placed accountability for student performance primarily on districts and individual schools and largely decoupled educational delivery from funding. A researcher of the relationship between achievement gaps and equity has summarized the shift embodied in NCLB: “The focus of national concerns about racial equity has shifted from school inputs to student outcomes and from redistribution of social relations and opportunities to redistribution of academic achievement” (Lee, 2004, p. 52). To the extent that NCLB assigns accountability without supplying adequate funding, it represents an additional and enormous unfunded federal mandate. Critics responded to this mandate by asking how the redistribution of achievement could be attained without a redistribution of opportunities and resources (Fusarelli, 2004).

The Challenge for School Leaders

Research is needed on how the accountability environment has affected school leaders’ practices and attitudes related to educational equity. After all, educational equity is the espoused cornerstone of the accountability movement and NCLB (U.S. DOE,

2002). Limited research on leadership in diverse schools has been conducted since the NCLB legislation. Evidence from North Carolina suggests that a carefully designed accountability system has a powerful influence on principal behavior (Ladd & Zelli, 2002). More research is needed on how principals and other school leaders are responding in general to the new environment: how they are viewing educational equity, focusing their energies, using data, and adapting their policies and practices.

This study focuses on the particular role of school leaders in promoting educational equity in the current environment of accountability. A conceptual framework for the study is shown in Figure 1. The framework begins with educational leaders' beliefs about issues of educational equity. Krug (1990) documented that principals' beliefs are a powerful factor influencing leadership effectiveness. If a particular belief system enables school leaders to promote educational equity, it is important to discover what that belief system is.

The model proceeds to conjecture that leader beliefs are mediated through several factors that influence the leaders' present actions and practices, and that these actions and practices in turn affect the equity outcomes of students. The three factors or conditions in the framework that mediate leader beliefs are (a) previous practices related to achieving educational equity, (b) requirements of accountability systems, and (c) demographic profile of the leaders' schools. Each of these will now be described.

Previous Practices Related to Achieving Educational Equity

Educators' concern for educational equity predated the current accountability systems. Many principals and teachers view themselves as advocates for children and

especially for those children who may be disadvantaged because of race, ethnicity, class, language background, or disability. A review conducted before the NCLB (U.S. DOE, 2002) legislation indicated that principals created inclusive schools for diverse students in three broad ways: (a) by fostering new meanings about diversity, (b) by promoting inclusive school practices, and (c) by building connections between schools and communities (Riehl, 2000). These categories encompassed a wide repertoire of principal beliefs, dispositions and actions. For example, fostering new meanings about diversity implied a willingness to openly discuss issues of diversity that tended to be ignored in schools under a system of norms referred to as *pathologies of silence* by Carolyn Shields (2004) and *equity traps* by McKenzie and Scheurich (2004). An articulation of these issues by principals in an environment of open, democratic discussion described by Shields as *dialogic leadership* could be transformative. Similarly, Singleton and Linton (2006) called for “courageous conversations” (p. 1) about race as necessary to disrupt deeply embedded beliefs and understandings that impede educational equity.

The second area of leadership for equity identified by Riehl (2000)—promoting inclusive school practices—encompassed two types of activity: the improvement of instruction and the molding of school cultures that embrace and support diversity. Instructional leadership has been studied within effective schools research (Fullan, 2003; Hallinger & Heck, 1998; Liethwood & Jantzi, 1990; Riehl, 2000; Rosenholtz, 1985). Researchers have found that principals of effective schools are intensely focused on student learning, and that they recruit outstanding teachers, buffer teachers organizationally to keep the focus on increased student achievement, provide professional

development for teachers, facilitate a collegial environment, and provide specific, concrete goals as benchmarks (Rosenholtz, 1985). Implicit in the inclusion of instructional leadership in the discussion of diversity is the idea that good teaching for diverse student bodies requires good teaching, period.

The molding of school cultures to embrace and support diversity requires a focused care for students (Riehl, 2000). It is focused because it takes account of the specific cultures, knowledge, and needs that children bring with them to school. It is caring because it advocates for children as individuals, holds high expectations for all students, seeks to utilize culturally sensitive instructional methods, and explicitly confronts issues of bias and exclusion. Leaders molding inclusive school cultures may actively promote curricular approaches such as culturally relevant pedagogy (Ladson-Billings, 1994). They also are sensitive to structures of power, control, and privilege within schools and strive to create programmatic equity, which refers to equality of access for all children to resources, programs, and experiences (Skrla, Scheurich, Garcia, & Nolly, 2004).

Building communities is the third area of principal leadership in diverse schools identified by Riehl (2000). *Professional learning communities* within schools can facilitate a collegial environment focused on instructional practice (DuFour, DuFour, Eaker & Karhanek, 2004). Creating connections between schools and the larger community involves mobilizing schools within overall processes of community development and coordinating services for children which can often be fragmented and

overlapping. Some of these services include Title I, special education, drug prevention, counseling, and language acquisition programs.

Requirements of Accountability Systems

The second factor in the conceptual framework of this study that mediates between school leaders' beliefs and their practice is the current accountability environment (see Figure 1). The preceding discussion has outlined this environment: Increasingly and dramatically with No Child Left Behind (U.S. DOE, 2002), educational equity was measured in terms of outputs in the form of performance relative to academic standards. Accountability was shifted from states to districts and individual schools, particularly to the school principal and other school level leaders. The national system under NCLB set multiple targets, and the achievement of adequate yearly progress by a school was dependent on meeting every target. States created their own accountability systems, which might resemble or overlap with the national system in many respects, but which also contained their own unique requirements. Thus, schools and their leaders simultaneously faced demands imposed by two accountability systems. Within the conceptual framework, the accountability environment could act in two directions, influencing both leader beliefs and actions. This is shown by arrows in Figure 1.

Priorities and incentives associated with multilayered accountability systems affected schools and their leaders in complex ways. Some incentives were intended, such as raising student performance in reading. Other incentives associated with accountability systems were unintended and may be counterproductive or perverse. Possible negative incentives associated with NCLB (U.S. DOE, 2002) were a narrowing

of the curriculum, an emphasis on test preparation at the expense of authentic learning experiences, a lowering of standards by states in order to facilitate the attainment of adequate progress, a neglect of subjects not tested by the accountability systems, a de-emphasis of cultural competency and multicultural education in schools, and a reduction of diversity in schools (Fusarelli, 2004). Biases in accountability systems increased the probability that likely more diverse schools and schools of different sizes would be singled out for sanctions (Kane & Staiger, 2002b). Understanding how school leaders perceived and responded to the intended and unintended incentives of accountability systems was one of the objectives of this study.

A striking feature of the new accountability environment was its emphasis on performance data. NCLB (U.S. DOE, 2002) required that children be tested yearly and that the results be broken out by poverty, race, ethnicity, disability and English proficiency (U.S. Department of Education, 2002). Although in the past school leaders may have used achievement data to inform their decision making, they lacked access to the kind of output data that are now available. By contrast, in the new environment of accountability, school leaders were awash in data and were continually confronted with the implications of that data. Although these data were widely available, they could be difficult to interpret. The sheer amount of data, the number of available comparisons among subgroups, the varying level of representation of subgroups, and the need to look at trends over time, all made the breakout analysis a daunting task. There was a need for new tools to use in evaluating student performance and educational equity. This study

introduces such a tool and seeks to discover how school leaders are coping with the challenge of data analysis.

Currently, little is known about how school leaders were responding to the new accountability environment as it specifically affected issues of educational equity. Several possible responses could be conjectured. Principals and other school leaders might choose to continue with practices in use before the new accountability pressures, perhaps out of a strong belief that these practices were the best vehicles in the long term for student achievement and equity. Alternately, leaders might significantly change practice in direct response to perceived incentives or threats within the accountability systems while simultaneously abandoning previous practices that were not perceived to have the same urgency as the demands of the accountability systems. A third possibility is that leaders might retain past practices to the degree possible while additionally using accountability data to diagnose and inform practice (J. Cheslock, personal communication, December 19, 2005). A key question was whether school leaders allowed accountability systems to determine their agenda or whether they integrated the new requirements as an important but not exclusive component of their agenda.

The Influence of School Profile

The conceptual framework suggests that the actions of school leaders may be influenced by the demographics of the school in question and the school's current performance status within the accountability systems (see Figure 1). For example, schools with high socioeconomic profiles had less student diversity on average and tended to experience higher student performance. The concerns of leaders at such

schools might vary significantly from those of leaders with diverse socioeconomic and racial student bodies or schools. In the latter case, sanctions and biases embedded in accountability systems might loom as more immediate threats and incentives to action. Moreover, leaders from schools that were poor and racially isolated had a different set of challenges. In this study, schools from a stratified sample of schools were sampled to investigate differences in leader approaches to educational equity.

Research Questions

Four elementary schools at different socioeconomic levels were identified for the study. From each group, a school was selected that had both high performance and high equity in comparison to its socioeconomic peers. In the research questions that follow, the phrase “high performance and equity” is used to describe schools that met these criteria. The purpose of selecting high-performing, high-equity schools was to limit the scope of the research and to discover the response of leaders in those high-performing schools to the needs of diverse students within the current accountability environment.

1. *What are the beliefs of leaders in schools with high performance and equity about educational equity?* Leader beliefs about their schools’ ability to achieve high levels of performance were investigated under this research question. The investigation included leaders’ perceptions of diversity, educational equity, obstacles to educational equity, optimism or pessimism about their schools’ prospects, and beliefs about the capacity of schools to overcome inequalities related to racial and socioeconomic isolation, funding disparities, poverty, and racism.

2. *Given the new accountability environment, what strategies have leaders in schools with high performance and equity used to achieve educational equity?* This question sought to discover if school leaders have changed their way of doing things with regard to educational equity in the new accountability environment. The question assumed that while accountability requirements and educational equity might be related and overlapping, they were not identical. Practices focused toward meeting student performance challenges might reduce achievement differences and thus create greater educational equity. Instructional leadership is an example of a set of practices that had the double effect of improving student performance attaining more equitable outcomes (Riehl, 2000; Rosenholtz, 1985). In addition to instructional leadership, the establishment of professional learning communities effectively extended the agenda of the instructional leader into conversations and situations where the school leader was not present. Therefore, school leaders facing demands of the institutional demands accountability environment and the moral demands of equity might turn to these two models as strategies.

However, accountability requirements might also replace or crowd out practices related to educational equity that were described by Riehl (2000) prior to No Child Left Behind. Those practices focused on fostering new meanings about diversity, promoting inclusive school practices, and building connections between schools and communities.

Such practices might not be perceived as relating directly to the student performance requirements of accountability systems. School leaders might de-emphasize or abandon these areas as the focus was shifted to urgent demands such as test preparation. Such abandonment might prove counterproductive for students from diverse backgrounds and to achievement in the long run (Fusarelli, 2004).

3. *How have leaders in schools with high performance and equity responded to the requirements of the accountability environment?* This research question encompassed several component questions: (a) How well do school leaders understand the requirements and incentives of accountability systems?, (b) Are leaders' perceptions of success congruent with success as defined by accountability systems?, (c) Do leaders' perceptions and attitudes toward accountability systems affect their response to those systems?, (d) What strategies are leaders using to meet the requirements of accountability system?, and (e) Do leaders believe they can succeed in meeting accountability requirements?

School leaders had to meet requirements of both national and state accountability systems. One of the functions of an accountability system was to create incentives, and the effectiveness of a system depended on how well incentives were understood. The national system under No Child Left Behind (U.S. DOE, 2002) defined several criteria that must be met in order for a school to make Adequate Yearly Progress. State accountability

systems varied in their incentives as well as their methods of assessing performance. As discussed in Chapter 2, Arizona's accountability system utilized three different models of assessment, while North Carolina's system focused strongly on improving the achievement of low-performing students. School leaders might be expected to respond more decisively to the clearer NCLB and North Carolina systems than to the more complicated Arizona system.

Leaders' perceptions of success might differ from success as defined by accountability systems. In addition to meeting achievement standards, leaders might seek to foster civic values, learner qualities, diversity appreciation, arts appreciation, community partnership, and other school characteristics that were not directly measured by accountability systems. This study explores these leadership perceptions.

In addition to the perceptions and attitudes of school leaders, their practices were of primary interest. Thus, the study also investigated the actual practices and strategies leaders implemented in the accountability environment. The results of a North Carolina study indicated that school leaders who held varying levels of personal support for an accountability system acted similarly to implement the goals of the system (Ladd & Zelli, 2002). These results were interpreted as evidence of the strength of the accountability system's incentives.

The specific strategies school leaders used to meet the requirements of accountability systems were also investigated. It has been suggested above that leaders might utilize models of instructional leadership and professional learning communities as efficient strategies for improving achievement outcomes (Riehl, 2000). In addition, No Child Left Behind (U.S. DOE, 2002) contained a built-in incentive for schools to shift resources and attention to students who are on the margins of passing. A study of the Texas system, which contained the same incentive, indicated that the incentive had its intended effect: Scores increased for students at the margins of passing and for low-achieving students in general (Reback, 2005). Although evidence existed for targeted attention to underachieving students regardless of subgroup designation, there was little evidence of attention targeted toward racial, subgroups monitored by No Child Left Behind and some state systems, even though subgroup performance determined a school's rating and attainment of adequate yearly progress (Kane & Staiger, 2002a).

4. *How does the socioeconomic and racial/ethnic profile of a school with high performance and equity influence the actions of school leaders to create schools with educational equity?* Because accountability systems impose immediate consequences on schools that are falling below standards, and schools that fall below standards are more likely to be poor and racially isolated, it may be expected that leaders in those schools

would respond more vigorously and urgently to the accountability systems. Leaders in schools with high achievement, an abundance of resources, and few subgroups might be less likely to feel pressure to make systemic changes in the initial years of the accountability systems.

However, leaders in well-to-do schools might also feel intense pressure to retain their high rating. The use of a stratified sample of schools with different socioeconomic profiles allowed the investigation of this research question.

5. *What are the challenges that exist in accurately measuring a school's educational equity?* In this study, a measure called the *equity achievement driver* is proposed as a measure of educational equity. In order to be effective, such a measure must reliably evaluate educational equity, provide information that aggregate data do not, be easily calculated and interpreted, and invite further analysis of the data. This research question explored the advantages and problems associated with this measure.

It is important to note that although the conceptual framework given in Figure 1 and the research questions in this section formed the basis of inquiry for the study, they were only a starting point. The study utilized a grounded theory approach, in which categories emerging from the data were described and linked theoretically (Glaser & Strauss, 1967; Glaser, 1978, 1992, 1998). In this type of research approach, the emergent categories might differ from the proposed conceptual framework and might fall outside the research questions.

Description of the Research

Four elementary schools at different socioeconomic levels were identified for the study. Selected schools outperformed their socioeconomic peers and had small achievement gaps. Performance criteria were the aggregate scores and the equity of performance across racial groups of fifth graders on the most recent state-mandated assessment. The principals of selected schools had tenure of at least three years at the school.

Principals and instructional coaches from each of the selected schools were interviewed. The study's conceptual framework and research questions were the basis for interview questions. Specific question topics included instructional leadership, inclusive cultures, culturally relevant curriculum, professional learning communities, connections between school and communities, use of data, and the effect of the accountability environment on leader practice (see Appendix). Emergent insights, categories, and patterns were coded from the school leaders' responses. Responses that varied by school and profession were noted.

Professional Significance of the Study

The study extends the knowledge base of how school leaders address equity needs of students within the current environment of national and state accountability systems. Empirical studies on principal responses to this environment have focused on the potency of the accountability systems themselves (Ladd & Zelli, 2002), the role of data in principal practice (West & Macharia, 2002), and the adherence of principals to best practices (Englert, Fries Goodwin, Martin-Glenn, & Michael, 2004). These studies

utilized survey information from a large number of principals. None of the recent studies focused on beliefs and practices relating to educational equity in the accountability environment. The current study focused on educational equity in the accountability environment in order to address this neglected area of research. It also analyzed the responses of a small number of school leaders more deeply. Moreover, the current study investigated how adaptation to the requirements of accountability environment affected or displaced other principal practices that relate to equity in schools.

In the domain of practice, this study introduced a new measure intended assist schools and districts to monitor their progress in both equity and achievement. Also, by synthesizing research on contextual variables affecting equity and accountability systems, the study sought to assist school leaders in negotiating the multiple and sometimes competing demands of equity and accountability.

Limitations of the Study

The study was limited first by its focus on a particular level of schooling, e.g., elementary schools. Findings may not generalize to secondary school levels. Second, the sample size was small, and the results were descriptive; therefore, it was not possible to determine the size effects of variables and categories. However, description also has advantages in establishing categories in a way that larger quantitative studies do not. A third limitation is that the study was cross-sectional and not longitudinal. Fourth, only schools that were relatively successful in student performance and educational equity were included in the study; the beliefs and practices of leaders in lower-performing schools were not investigated.

Definitions

For the purposes of this study, the following definitions are offered:

Accountability Environment. The current context of public schools, which is characterized by accountability of schools and districts for the achievement of national and state goals by students.

Accountability System. A set of legislated educational mandates from the federal government or states. Accountability systems confer various rewards and sanctions on schools depending on whether schools meet mandated performance requirements. An important feature of accountability systems is the requirement to report student performance data disaggregated by grade level, race, poverty, language background, and disability. No Child Left Behind (U.S. DOE, 2002) and Arizona Learns (Arizona Department of Education [ADE], 2005) are accountability systems.

Achievement Gap. Performance differences between racial, socioeconomic, language or disability subgroups of students on a national, state, or local level.

Adequacy. Legal concept in school finance reform law; it seeks to set a minimum standard for educational services that meets state constitutional requirements.

Adequate Yearly Progress (AYP). Standard for school success under No Child Left Behind (U.S. DOE, 2002) based on progress in mathematics and reading for a school overall and for racial, socioeconomic, language and ability subgroups within a school. Each year a school is deemed to either make or not make AYP.

Arizona Instrument to Measure Standards (AIMS). State mandated assessment for Grades 3-8 and 10-12 that tests students in Reading, Writing, and Math. Performance

is reported as *Exceeds*, *Meets*, *Approaches*, and *Falls Far Below* the standard. Both *Meets* and *Exceeds* are considered to be passing.

Arizona Learns. The accountability system for Arizona (ADE, 2005) mandated under the federal No Child Left Behind (U.S. DOE, 2002) legislation. It credits points to schools based on various achievement and non-academic measures and on the basis of these points assigns each school a label: *Excelling*, *Highly Performing*, *Performing Plus*, *Performing*, *Underperforming*, and *Failing*.

Category. A conceptual name given to a pattern of similar incidents in grounded theory analysis.

Constant Comparison. A procedure of grounded theory in which the researcher compares incidents within a set of data, such as interview notes, and between sets of data, to generate theoretical categories that conceptualize behavior. Initially, one incident is compared to another; later, incidents are compared with theoretical categories.

Core Category. A category generated in grounded theory analysis that accounts for most of the variation in a pattern of behavior.

DIBELS. The Dynamic Indicators of Basic Early Literacy Skills (Gordinier & Foster, 2004) are a set of standardized, individually administered measures of early literacy development. This is a one-minute fluency test used to regularly monitor the development of pre-reading and early reading skills.

Educational Equity. "Raising the achievement of all students while narrowing the gaps between the highest- and lowest-performing students; and eliminating the racial

predictability and disproportionality of which student groups occupy the highest and lowest achievement categories” (Singleton & Linton, 2006, p. 46).

Equity Achievement Driver (D_{ea}). A statistical measure formulated to evaluate the educational equity of a school or district. It is described in Chapter 3.

Grounded Theory. A methodology for qualitative research that seeks to discover and generate theory from data rather than to impose existing theory on data. Procedures of grounded theory include constant comparison, coding, memoing, sorting, and writing.

Memoing. A procedure of grounded theory in which the researcher notes theoretical ideas or hypotheses that emerge during the coding of data. Memoing can result in new categories, a core category, or descriptions of relationships between categories.

Mobility. A measure of how many students are transferring in and out of a school. Higher mobility tends to correlate with lower achievement. In the district under study, mobility is calculated by the following formula: $100 * (\text{Entries after First Day} + \text{Reentries} + \text{Withdrawals}) / (\text{First Day Enrollment} + \text{Entries after First Day})$.

No Child Left Behind (U.S. DOE, 2002). Federal legislation requiring that all children meet minimum standards in reading and mathematics by the year 2014. It requires incremental progress toward this goal, known as Adequate Yearly Progress (AYP), and requires states to set standards for AYP and assess students yearly.

Open Coding. Initial coding process of grounded theory in which the researcher generates conceptual categories from the comparison of incidents in data. Open coding ceases when it yields a core category.

Opening Minds through the Arts (OMA). A nationally recognized research-based program that integrates the arts into teaching reading, writing, math and science in public elementary and middle schools (Preston, 2003).

Professional Learning Community (PLC). A group of educators whose common interest is to work collaboratively to access the evolving educational knowledge base with the goal of improving student performance (DuFour & Eaker, 1998).

Selective Coding. Secondary coding process of grounded theory in which the researcher compares incidents from the data and categories generated from the data to the core category.

Stability. A measure of how much of the school year a student spends at a single school. It is negatively correlated to mobility.

Theoretical Sampling. A process in grounded theory analysis in which the researcher returns to the field or the data to add properties and diversity to the concepts and hypotheses that have been generated.

Weighted Combination of Achievement Measures. A combined measure of all the assessment measures used by the district under study, calculated by the district's accountability and research department. It takes account of both current standing and growth of achievement for a school. It is reported as a whole number z-score ranging from 1 (*poor*) to 6 (*excellent*).

The remainder of the study is organized as follows: Chapter 2 reviews the literature; Chapter 3 describes in detail the methodology for the study; Chapters 4 and 5 present the results; and Chapter 6 summarizes and discusses the findings.

CHAPTER 2

REVIEW OF THE LITERATURE

This literature review brings together theoretical and historical research from several fields that affect schools and school leaders in the new environment of accountability. Those fields include educational policy, law, sociology, economics, effective schools research, and educational leadership. The summary and synthesis of this varied literature set a broad context for the study and provide a perspective on the position of school leaders within that context.

Search Process

The following review resulted from a search that had four components: (a) a survey of a major journal of educational administration, (b) a key word search, (c) an author search, and (d) a search of the reference lists in the sources that were generated through the first three components. These components were completed in an overlapping rather than a sequential manner.

For the first component, the journal *Educational Administration Quarterly* was surveyed over the past several years for articles related to leadership in diverse schools. This search yielded several sources and references. For the second component, a key word search through the Educational Resources Information Center (ERIC) was undertaken involving the following key words singly and in combination: accountability, achievement, black achievement, data, desegregation, diversity, effective schools, equity, incentives, leadership, minority achievement, principal, and response. At the suggestion of members of the dissertation committee, a decision was made to expand the literature

review to include contextual factors that affect principals as they seek to lead in diverse schools. School finance and accountability system incentives were recommended, with a key article by Molly McUsic (1999) for the former topic, and articles by Thomas Kane and Douglas Staiger (2002 a, b) for the latter topic. A Google Scholar search was undertaken for these authors; this yielded additional articles by them and additional sources by others who had cited the authors.

Investigation of school finance and accountability system incentives revealed that these variables affect schools with diverse student bodies by creating, sustaining, or ignoring inequalities in schools. This led to the exploration of three additional inequalities: segregation, inequalities of student performance, and inequalities within schools. These factors are highly interrelated. Segregation was included because it directly related to student diversity, a main theme of the study. Student performance is the primary indicator for accountability goals and educational equity in the current educational context. Moreover, discussion of performance differences or gaps is prominent the literature on achievement; hence, its inclusion. Inequalities within schools comprise a variety of conditions and practices that affect educational equity and which are probably the most susceptible to influence by school leadership practice. Thus, five different but interrelated contextual variables were identified as being relevant to the study. Word searches and Google Scholar searches similar to those described above were made in connection with these variables.

The fourth component of the search involved the examination of the reference lists of the sources recovered. Titles and abstracts were used to identify potential sources.

Preference was given to sources from the past 10 years. Exceptions to this preference were made in some sources involving effective schools and desegregation, which are fields in which seminal research was completed more than 10 years ago. In addition, preference was given to refereed journal sources and to primary sources. Sources identified as the most useful by the process above were retrieved.

Due to the large body of work on school leadership, a limitation was needed for this area of the literature. A review by Riehl (2000) about the principal's role in creating inclusive schools for diverse students provided a framework for selection of topics and many references. The review covers literature on those topics up to the time just predating *No Child Left Behind* (U.S. DOE, 2002). For example, instructional leadership is included in the Riehl review and this chapter while the concept of leadership frames is not. This is not to say that other leadership domains do not benefit principals and students of diverse schools. It is simply necessary to limit the literature to domains most directly connected to educational equity.

Inequalities that Hinder Educational Equity

This study focused on how school leaders think about and respond to the equity demands of their diverse student bodies in the current environment of accountability. It was assumed that leaders' beliefs and behaviors could affect educational equity and student learning. The study did not attempt to prove this assumption but presented findings of studies by others on principal effects on student learning. In addition to leadership, accountability systems themselves were intended to foster equitable student

outcomes; thus, accountability systems represented contextual factors in school equity as well and are examined in this review.

In addition to leadership and accountability systems, four other contextual variables have been associated with school equity: segregation, school funding, inequalities of achievement, and inequalities within schools. There is variation in how these categories were framed; some authors considered within-school inequalities as an aspect of desegregation (Bates, 1990). In addition, the variables were drawn from broadly varying areas of the literature and each of them stood as a major area of research, sometimes with journals devoted exclusively to their study. Although there was a danger of casting too large a net in considering the effects of these variables, a danger also existed in ignoring them due to their potential effects, interrelationships, and dynamic nature. Their inclusion enabled a contextual synthesis and a consideration of the *big picture* with which school leaders were confronted.

Segregation

In this review, segregation is considered to be synonymous with racial isolation whether it occurs through legal means or indirectly. Court-mandated desegregation in education gathered strength for over 30 years and then declined for 15 years (Orfield, 2001). Studies on the effects of desegregation indicated a beneficial effect on the achievement of African American and Latino students when the desegregation occurred very early in their schooling (Crain & Mahard, 1981). Of potentially greater significance were the long term effects of desegregation in breaking cycles of isolation for these youth and thereby improving their life chances (Wells & Crain, 1994). Although mandated

desegregation declined and there was gradual trend toward resegregation, cases aimed at removing the vestiges of past segregation continued to be brought (Weiler, 1998). In addition, voluntary plans such as controlled choice were in place in many districts and might hold promise for the future of desegregation (Willie, Edwards, & Alves, 2002).

Desegregation and resegregation trends. Judicially-enforced desegregation in all regions of the country beginning after the *Brown v. Board of Education of Topeka* (1954) decision and continuing into the mid-1980s resulted in a significant but temporary decrease in the racial isolation of African American students, followed by a gradual increase in segregation through the present. Orfield (2001) found that the isolation of African American students decreased from the late 1960s to a minimum in late 1980s but increased steadily thereafter. He measured isolation in two ways with consistent results: the percentage of African American students attending predominantly non-White schools, and the percentage of White students in schools attended by the typical African American student. Orfield also found that desegregation and resegregation trends have not affected all regions of the country equally. Desegregation had its greatest effect on the South, which became the most integrated region. Since 1990, all regions have undergone resegregation; the area of greatest isolation was the Northeast, where over half of African American students attended schools that are between 90 and 100% non-White.

The rising and falling fortunes of mandated desegregation have been driven by changing policies on the part of government branches. During the most active period of between 1964 and 1968, desegregation was national policy pushed by the executive branch as well as the judicial branch (Orfield, 2001). Executive support was reduced

after 1968; however, the courts continued to actively require desegregation. The first major break in judicial support came in the *Milliken v. Bradley* (1974 & 1977) case, in which the Supreme Court exempted suburban districts from participation in desegregation plans. Further weakening of desegregation came in three Supreme Court decisions in the 1990s. The *Board of Education of Oklahoma City v. Dowell* (1991) ruling released districts that had implemented past orders from enforced busing and allowed them to resume assigning students to segregated neighborhood schools. In this ruling, the Court also specified that school districts were not responsible for remedying local conditions such as segregated housing patterns (Weiler, 1998). In *Freeman v. Pitts* (1992), the Court made it easier for districts to be released from court supervision and achieve unitary status through the incremental rather than complete achievement of desegregation benchmarks. And in *Missouri v. Jenkins* (1995), the Court ruled that low achievement scores of target population students were not sufficient grounds to require the continuation of a desegregation plan. However, the state was still required to remedy the vestiges of past discrimination.

The history of segregation in policy and practice for was quite different for Latinos, now the largest minority population in the country (Orfield, 2001). The right of Latinos to desegregation was not recognized by the Supreme Court until 1973, long after the *Brown* decision. However, the ruling was never significantly enforced, and actual desegregation occurred only in Colorado. As an alternative to desegregation, the policy of the Nixon administration was to enforce bilingual education rights for these students. It was therefore not surprising that Latino students never experienced the decrease in

isolation that their African American counterparts experienced. Using measures of the percentage of Latino students in predominantly White schools and the percentage of White students in schools with a typical Latino student, Orfield found that the isolation of Latino students, which was somewhat less than that of African American students in the late 1960s, increased steadily from that point forward. The level of Latino isolation in 2001 was equal to that of African American students in the late 1960s. Thus, Latino students began the desegregation era somewhat less isolated than African American students; however, because they did not experience the effects of desegregation laws and policies, they became more segregated than African American students.

An associated consequence of segregation is concentration of poverty in segregated schools. In 1999, a typical African American student attended a school in which the percentage of families in poverty was 39%; a typical Latino student attended a school in which the percentage of families in poverty was 44%; and a typical White student attended a school in which the percentage of families in poverty was only 19% (Orfield, 2001). Predominantly African American and Latino schools were much more likely than predominantly White schools to experience concentrations of poverty as determined by the percentage of students receiving free and reduced lunch. The negative consequences of concentrated poverty included lower performing schools with less competition, larger achievement gaps, fewer qualified teachers, fewer advance placement courses, and higher dropout rates. Some researchers believed that concentration of poverty was the primary problem desegregation was intended to alleviate (Crain & Mahard, 1981; Willie et al., 2002).

The effect of desegregation on student outcomes. Orfield (2001) pointed out that desegregation was not intended primarily as an “educational treatment” (p. 9) but rather part of a movement for racial and economic justice. However, proponents of desegregation also expected that the *Brown* decision would serve as an equalizer that would guarantee equality not only of access to educational opportunity but also of outcomes (Love, 2004). As a result, the relationship of desegregation to student outcomes was the topic of several studies since the 1970s (Coleman, 1968; Crain & Mahard, 1981; Stevens & Dial, 1993). Early studies focused on the effects on student achievement. More recently, studies have appeared on the long-term effects of desegregation (Wells & Crain, 1994). Also, an association between desegregation and reduction of achievement differences between racial groups has been noted (Lee, 2004).

The first major study on the effects of segregation and desegregation was the Coleman Report, which concluded that African American students in predominantly Anglo schools scored higher on achievement tests (Coleman et al., 1966). These researchers found that an integrated environment was more important for student success than other factors such as Title One resources. Coleman (1968) hypothesized that the positive effect of desegregation might be due to a change of attitude in African American students, who gained a greater sense of control and destiny when they realized that they could do some things better than their White counterparts. Other researchers attributed the positive effects of segregation at least in part to the mixing of students from different socioeconomic backgrounds (Crain & Mahard, 1981; Orfield, 2001; Willie et al., 2002).

Findings of studies contemporary with the Coleman (1966) report indicated the importance of desegregation in the elementary years: African American students moving from segregated elementary schools to integrated secondary schools did not make gains (Crain & Mahard, 1981). In addition, certain theories about the mechanism of desegregation effects were supported while others were refuted. The *teacher expectation theory* proposed that higher teacher expectations geared to the middle class average of integrated schools raised African American student achievement. The findings of Crain and Mahard supported this theory. The *lateral transmission theory*, which proposed that African American students benefited from exposure to values and behaviors of White students, was not supported by the evidence. This finding may answer more recent critics like Love (2004) who questioned the argument that African American children must be in the classroom with White children in order to learn.

In a seminal study, Crain and Mahard (1981) conducted a meta-analysis of the 93 existing studies on desegregation and student achievement. They found that methodological problems were the cause of ambiguity in the research results at that time. Studies using the most reliable design methods yielded consistent findings; these studies either randomly assigned students or utilized segregated African American students with similar backgrounds as a control group. They found that (a) desegregation in the early primary grades had a significant positive effect on African American student achievement that was sustained through later years, (b) desegregation beginning at Grade 6 or later had no effect on achievement, and (c) the duration of desegregation had no effect on improvement of achievement. The average improvement for students in

desegregated schools was one-third standard deviation on a norm-referenced Comprehensive Test of Basic Skills; this improvement was equivalent to a one year academic gain. The researchers also measured significant gains in I.Q. and achievement test scores in all subject areas in association with early childhood desegregation.

An additional finding of the study was that there was an optimal range of student body composition for maximizing African American student achievement (Crain & Mahard, 1981). This range was between 19 and 29% in Southern schools and between 9 and 19% in Northern schools. In overwhelmingly White schools where their representation was less than 10%, African American student achievement declined. Crain and Mahard attributed this finding to racial isolation and in support for the concept of a *critical mass* of non-White students in integrated schools.

In interpreting their findings, Crain and Mahard (1981) hypothesized that desegregation created a burst of achievement growth lasting through the early elementary grades. Since this analysis, other researchers have made similar findings regarding the benefits of desegregation for achievement of non-White students. In a longitudinal study spanning fourth through eighth grade, Stevens and Dial (1993) compared the performance of different racial groups in different types of schools. They found that African American and Latino students in multi-ethnic schools performed better than their counterparts in racially isolated schools. Willie et al. (2002) found that African American students performed the best when they were enrolled in racially mixed schools that were not poverty-concentrated. They also found that schools rated *highly performing* in the Florida accountability system were more likely to be racially and

socioeconomically diverse than schools that were underperforming. These findings led them to conclude that racially and socioeconomically diverse schools were more effective learning communities than poverty-concentrated and racially isolated schools and that achieving diversity should be the goal of desegregation. In findings similar to those of Crain and Mahard, both of these studies indicated a negative effect on non-White students' achievement when they were racially isolated in either hi-poverty non-White settings or nearly all-White settings.

Another area of research focused on the long-term effects of desegregation on students. This research became possible when school-age children in the desegregation era came of age. This research took the premise that desegregation was not intended primarily to increase achievement scores. Rather, it was intended to improve the life chances of students by breaking cycles of isolation and enabling access to high-status institutions and social networks (Wells & Crain, 1994). Wells and Crain drew from a theoretical framework consisting of *perpetuation theory* and *network analysis*.

Perpetuation Theory proposed that segregation perpetuates itself through people's lives and institutions when individuals have not experienced desegregated settings early in life (Braddock, 1980). For example, Black and Latino students might overestimate the racial obstacles they would encounter in integrated settings or underestimate their skills at coping in such settings, and therefore hold themselves aloof. Network analysis was an extension of this theory (Wells & Crain, 1994). It hypothesized that African Americans and Latinos lacked access to informal networks that provided access to integrated institutions and employment.

Wells and Crain (1994) tested these ideas by using national databases to investigate the effects of desegregation on high school students' occupational aspirations, college selection, and occupational attainment. They found that desegregated African American students had higher aspirations than segregated African American students and that these aspirations were more realistically related to the desegregated students' educational attainment than those of segregated students. In regard to college selection, they found that African American graduates of desegregated schools were more likely than graduates of segregated schools to attend desegregated colleges. Finally, they found that students of desegregation were more likely to have desegregated social networks, desegregated employment, and jobs in the professional private sector than segregated students. Wells and Crain concluded their findings were consistent with their theory that a major function of desegregation was to break perpetuating cycles of isolation. There is additional evidence of the long-term benefits of desegregation in the research. Orfield and Whitley (1999) showed that most African American and Latino students admitted to elite law schools came from integrated school backgrounds. Camburn (1990) found that students from integrated schools were more successful in college than students from segregated backgrounds with the same test scores.

Circumstantial evidence for the benefits of desegregation came from trends in performance differences, known as the *achievement gap*. The desegregation era coincided with a significant narrowing of the test score gap between African American and White teenagers through the 1970s and 1980s as measured by the National Assessment of Educational Progress (NAEP) in mathematics (Lee, 2004; Orfield, 2001).

However, achievement differences between these groups leveled off or expanded during the period of resegregation in the 1990s.

The future of desegregation. Although empirical evidence consistent with theory indicated that desegregation benefitted students of color in terms of short-term achievement and long-term life chances, desegregation was rolled back as a practice and policy over the past 20 years and a trend toward resegregation was underway. Resegregation was a gradual phenomenon that does not seem to be the result of public opinion. Survey results indicated the public strongly supported diversity and integration in schools, in spite of simultaneous support for neighborhood schools and school choice, which might be incompatible with diversity (Orfield, 2001). Rather, a confluence of law, policy, and demography appeared to be driving resegregation. Court reversals were mentioned earlier. Many large districts were granted unitary status by courts that considered desegregation to be a temporary remedy. In the policy arena, executive branch of the federal government pushed for the termination of desegregation since 1970. In the demographic domain, a major shift toward a metropolitan society dominated by suburbs was ongoing, and suburbs were becoming less White and racially isolated. In this situation, the demographics of neighborhood schools were extremely unstable: Integrated schools on the boundaries of segregated communities quickly became racially isolated within just a few years as racial changes in schools outpaced changes in the neighborhoods themselves (Orfield, 2001). A large wave of immigrants accelerated this rapid racial transformation.

One type of desegregation plan that was fairly widely implemented and which held promise for achieving racial and socioeconomic diversity was *controlled choice* (Willie et al., 2002). Designed to address the problem of segregated schools resulting from segregated neighborhoods, controlled choice programs abolished neighborhood schools. In their place, the plans divided large metropolitan districts or counties into zones with roughly equivalent demographics. Parents selected among available schools in their zone, with entry points at the beginning of elementary, middle, and high school. Selection criteria included zone, seat availability, family proximity to the school, and sibling preference. Willie et al. reported several advantages of the controlled choice plans they studied: (a) Resegregation was prevented even when demographic changes occurred; (b) mandatory assignments were rare, and 90% of families received their top school choice; (c) no court order was required; (d) plans were compatible with unitary status; (e) effects of controlled choice were comprehensive and immediate rather than piecemeal as in many court rulings; and (e) there was a strong incentive for underperforming schools within the controlled choice systems to improve in order to attract students.

Controlled choice held promise for districts with requisite student economic and racial diversity or for areas in which smaller districts voluntarily banded together into service delivery zones (Willie et al., 2002). However, most segregation was between districts of differing racial composition rather than within districts (Orfield, 2001). In addition, a number of large urban districts were one-sixth or less White, a situation that made meaningful desegregation within these districts impossible (Weiler, 1998). For

such areas, court-ordered desegregation would seem to be the only available remedy. A recent example was the *Sheff v. O'Neill* (1996) case, in which the Connecticut Supreme Court declared that racial isolation in Hartford violated the state constitution. The state legislature subsequently authorized inter-district transfers and the enhancement of urban schools with magnet and charter schools. The ruling indicated that a judicial option remained a viable remedy in cases of extreme segregation.

This section on segregation concludes with an example of how contextual variables of this study, e.g., segregation and accountability systems, can interact. Debray (2005) documented a conflict that arose between a federal desegregation order and requirements of the No Child Left Behind Act of 2001 (U.S. DOE, 2002). The conflict illustrated the collision of two federal mandates and philosophies regarding choice and diversity. District officials in Pinellas County, Florida sought protection from the section of NCLB that allowed parents to remove their students from schools that did not make adequate yearly progress in two consecutive years. The officials claimed this section disrupted a court ordered controlled choice program already in place. The judge ruled that the original controlled choice plan was to be protected from changes through 2008.

School Funding

The unequal funding of schools as a factor in educational equity is now considered. Inequalities of funding spring from historical structures and mirror other societal inequalities (McUsic, 1999). Although equity in schools was sought after for several decades, it was a concept that was difficult to define. Through a four-decade long process of litigation, an evolution in school finance occurred centered around the notion

of what constitutes an adequate education (Roellke et al., 2004). In addition, progress was made in more closely linking educational resource inputs with student outcomes and in estimating the costs of adequacy (Baker, 2005).

How funding inequalities are structured. An assumption generally shared among educators and policy makers is that students from diverse backgrounds in different schools and regions should experience an equality of educational delivery and demonstrate comparable mastery of knowledge and skills. That is to say, students should experience equitable inputs and produce equitable outputs. McUsic (1999) found that in the decades preceding No Child Left Behind (U.S. DOE, 2002), equity was measured primarily through *inputs* to schools in the form of funding. However, laws at the local, state and federal levels, have set up and perpetuated the *unequal* funding of schools, and litigation over unequal funding has emerged as the primary remedy.

Local, state, and federal laws created funding inequalities through three mechanisms: (a) unfunded federal mandates, (b) local control of schools, and (c) subsidization of private and parochial schools (McUsic, 1999). First, federal law created inequality by imposing unfunded mandates for students with special needs who were not evenly distributed in the population, but instead were concentrated in schools with high numbers of economically disadvantaged students. These mandates required schools to provide expensive services to students with special needs before attending to general education needs or risk losing federal funds. Compensatory funding received through Title I and other programs were not sufficient to offset expenditures on the special needs

students. In addition, while the federal government contributed to inequalities through unfunded mandates, it was under no obligation to equalize funding among states.

Second, the system of local control of schools established in the 19th century ensured inequality of funding due to differences in wealth between school districts (McUsic, 1999). The mechanism of inequality began with property taxes based on land values that provided more than half of school funding nationwide. Because no state drew boundaries to equalize property values within districts and most neighborhoods were segregated by class and race, revenue and per-pupil funding varied greatly from one district to the next. Wealthy neighborhoods could finance their schools with a low property tax rate, while poor districts had difficulty financing their schools even at a high property tax rate. In order to correct inequalities inherent in the funding system, states made supplemental appropriations intended to equalize resources among districts. Supplemental state appropriations accounted for nearly 50% of school expenditures in the 1990s. However, the promised equalization was rarely achieved, and in some cases the state interventions aggravated inequalities.

A third source of inequality of funding came from state subsidization of private and parochial schools (McUsic, 1999). State governments accredited and subsidized those schools through tax credits and exemptions, transportation, and material support. Private and parochial schools that served less than ten percent of the school age population drew from relatively well-to-do families who could afford to pay the tuition. State resources given to private schools were diverted away from public schools.

Three waves of litigation. Beginning in the 1960s, school finance litigants attempted to force changes in how schools were financed by claiming that inequalities of educational resources resulted in inequalities of educational outcomes. Three waves of litigation have been identified, and the third wave is currently in progress. The first wave spanned the years from 1969 to 1973 and consisted of cases brought on the basis of the federal constitution's equal protection clause (Roellke et al., 2004). These were desegregation cases built on the concept of *fiscal neutrality*, which specified that the quality of education delivered to children of a state may not be a function of personal wealth. Through these cases, litigants sought to limit freedom of choice, prescribe transfers of students within and across district boundaries, and order funding for desegregation projects. The first wave abruptly ended with the *San Antonio v. Rodriguez* (1973) case, in which the U.S. Supreme Court ruled that there was no federal constitutional protection from inequalities arising from differences in property wealth.

In the second wave, which occurred between 1973 and 1989, plaintiffs used both state education clauses and state equal protection clauses to challenge school finance systems (Roellke et al., 2004). Roellke and colleagues attributed the mixed results this approach has had to several factors: (a) courts have been unwilling apply strict scrutiny to school finance inequalities, (b) courts have accepted local control as a reasonable defense for inequalities, and (c) plaintiffs failed to establish a strong causative relationship between funding and the education provided. Despite these problems, an economic study indicated court-ordered finance reform during the period from 1971 and 1996 reduced within-state inequality 19 to 34% by raising spending in the poorest districts while

leaving spending in the wealthiest districts unchanged (Murray, Evans, & Schwab, 1998). The additional spending came from higher state taxes. These findings indicated a measure of effectiveness of second-wave school finance litigation.

McUsic (1999) described the difficulty of the first two waves of school finance litigation in defining precisely what is to be equalized. Equalization efforts could focus on funding, facilities, student outcomes, or other variables. Dollar equality as represented by per pupil funding was questioned by those on both sides of the debate: by those who contend there is no proof that greater funding results in academic gain, and by those who argue that equal funding where needs greatly differ is insufficient and unjust. In response to this seemingly intractable argument, a third wave of school finance litigation cases has relied less on the criterion of equity and more on the criterion of adequacy.

The adequacy approach to school finance began with the question, “What constitutes an adequate education?” Such litigation then sought to measure the educational services being delivered and show that they were insufficient to satisfy constitutional requirements. McUsic (1999) wrote, “This approach requires lawyers to frame the lawsuit around educational theory, rather than legal theory—on what works, rather than what is equal” (p. 115). Relying on the expertise of educators, adequacy litigation sought to restructure the educational delivery system to bring the performance of all students and especially disadvantaged students up to an acceptable, adequate standard. As defined by Rebell (1994), an adequate education had several components: (a) It prepared students for life in a competitive society, (b) it took account of the

educational needs of students in present society, (c) it prepared student for higher education, and (d) it ensured availability of essential resources in order to meet standards.

One reason that third wave cases succeeded where the first two waves failed is that adequacy claims did not violate conceptions of the *separation of powers* (Roellke et al., 2004). Courts set the standards for educational adequacy, often with significant input from educators, and then left implementation to state legislatures. Another reason adequacy claims succeeded is that they were less likely than other approaches to conflict with concerns about local control. For example, a minimum adequacy standard didn't penalize wealthy districts for higher academic success. A third strength of the adequacy approach was that it provided specific standards for measuring the adequacy of a finance system.

Some limitations of the third wave have been identified (Roellke et al., 2004). First, the determination of adequacy was highly complex, because it involved evaluation of both school inputs and school outputs. Second, rulings might limit the autonomy of local districts. And third, the capacity of districts to correctly use resources wasn't fully addressed. Despite the relative success of adequacy litigation, translating courtroom success to classroom implementation has proven more difficult. McUsic (1999) described a process in which courts that found school finance systems unconstitutional typically instructed state legislatures to craft a remedy. However, transferring funds to poor districts and schools was not a political action legislatures were inclined to. The dynamics of the political process ensured that legislative school reform focused on equity was always a response to real or threatened litigation. Even so, the various types of state

equalization plans put forth by legislatures either failed to achieve equalization or had the effect of decreasing the overall funding for schools as funds were diverted to other state needs.

Theoretical considerations. Equity in school finance can be considered from the perspectives of *horizontal equity* and *vertical equity*. Horizontal equity is concerned with the equal allocation of resources across different groups of students. It calls for the equal treatment of equals. Vertical equity modifies this approach by recognizing that students bring certain education needs to the classroom that require additional resources (Berne & Steifel, 1984; Rodriguez, 2004). It calls for the unequal treatment of unequals. The goal of vertical equity is to make schools responsive to the varying needs of students.

Rodriguez (2004) provided criteria for effective vertical equity standards; they must (a) address schooling in its complexity, (b) support school responsiveness to both student and staff needs, and (c) assess the impact of resources on identified needs. Two practices that utilize vertical equity are pupil weighting and targeted or categorical funding. The former is illustrated by additional funding weight given to categories of students such as special education students or English language learners, and the latter is exemplified by dropout prevention and school-to-work programs, among others. The goal of vertical equity theorists is to establish a tight coupling between resource inputs and the schooling experiences of individual students, and to monitor that coupling.

Vertical equity recognizes that varying resources may be required in order for schools to be responsive to diverse student needs. One theorist was careful to distinguish this recognition from *deficit model thinking*, which places the burden of failure on racial,

ethnic and class backgrounds of students, rather than on institutional biases, practices, and lack of responsiveness (Rodriguez, 2004). Rodriguez argued that school finance scholarship was susceptible to deficit model thinking when it focused on educating for the norm rather than for the success of a diverse population of students.

In addition to vertical equity, another concept in school finance with theoretical considerations at the school level is *adequacy*. In the preceding section on litigation, the success of plaintiffs arguing for adequacy was due in part on their ability to link resource inputs with student outcomes. Alexander (2004) provided a theoretical framework for adequacy. The model identified how inputs were transformed into an educated adult along a production continuum that moved from inputs, through schools process, and terminated with outputs. Alexander argued that the adequacy of the whole system depended on adequacy of each part of the continuum. In her theoretical framework, adequacy of educational inputs included concepts such as horizontal equity, vertical equity, and equity of access. Priority was placed on distinguishing between *basic* and *luxury* inputs. Time was also considered as an input.

An important contribution of this analysis was the identification of school processes as a necessary part of the framework that was neglected in models that considered only inputs and outputs (Alexander, 2004). Processes made sense of how inputs were used and transformed into outputs. Factors that influenced the adequacy of school processes included effective schools practices, intervention priorities for students at different performance levels, transformative processes, and capacity.

The adequacy of outputs rested on the concept of *results neutrality*, which specified that differences students bring with them to the classroom should not be associated with their exit outcomes under an adequate funding system (Alexander, 2004). Like vertical equity, results neutrality recognized differences in student needs and concluded that it was incumbent on schools to address these varying needs equitably. Thus, schooling was seen as an equalizer. Alexander predicted that results neutrality would increasingly become the standard in school finance litigation.

The cost of adequacy. Baker (2005) used empirical data to validate theoretical assumptions about the cost of an adequate education. He began by formulating six theoretical assumptions from an economic perspective:

1. Cost varies by the desired outcomes, e.g., there are diminishing marginal returns as more resources are invested to raise desired outcomes.
2. Cost varies by district size, e.g., there are economies of scale that are realized for districts with 2,000 to 6,000 students.
3. There are additional costs associated with achieving desired outcomes for different student groups.
4. Variations in prices across regions influence the cost of education in different districts.
5. The realities of assumptions 2, 3, and 4 interact in a multiplicative way.
6. Marginal costs of achieving outcomes increase as performance standards increase and decrease as performance standards decrease.

Baker (2005) tested these assumptions using several empirical studies in the literature that included the measurement of student outcomes. Specific outcome measures included Scholastic Aptitude Test (SAT) performance, state accountability test performance at different grade levels, and percentage of students completing advanced coursework. Studies employed one of two approaches for calculating the cost of adequacy: (a) a top-down *education cost function*, which began with district spending and estimated its effect on student outcomes using appropriate statistical controls; and (b) a bottom-up *resource cost model*, which began by identifying the inputs necessary for an adequate education and proceeded to estimate the total costs. In an ideal situation with no distortions, estimates gained from the two models for a given finance system would be identical. Baker found a high degree of support in the empirical studies for his six assumptions, although there were some differences in magnitude depending on the model used. Specifically, he calculated the extra weighting factor for a poor child to be between 35% and 100% depending on district size and the model used; this student would require up to double the funding of a student who wasn't poor to be adequately educated. However, most districts provide less than a 35% additional weighting. A limited English proficient student's additional weighting requirement was also calculated at 100%, significantly higher than the 20% that is often recommended.

Baker's (2005) findings confirmed that educational outputs don't come for free. They also signaled a theoretical and empirical precision that was previously lacking in funding formulas and policies. This precision should aid legislatures, districts and courts in the determination of the cost of an adequate education.

Inequalities of Performance

Differences in performance between different racial, ethnic, and socioeconomic groups constituted a third variable related to equity in schools. In the framework of No Child Left Behind (U.S. DOE, 2002), this variable is implied to be equivalent with equity, but this has not always been the case. During the Civil Rights era, the focus of equity was on educational opportunity rather than student performance (Love, 2004; Wells & Crain, 1994). However, improved student outcomes have always been one of the goals associated with educational equity, and these outcomes include performance. Thus, student performance can be considered as one variable among other contextual variables that determine educational equity, including opportunity, segregation, instruction, school funding, and other inequalities or gaps that children experience. Consideration of the *achievement gap* in isolation of these other variables can be misleading and counterproductive. This section reviews several theoretical approaches to performance differences and then summarizes empirical research that demonstrates the inequity of current performance differences.

Theoretical literature. Two different theoretical perspectives on the causes of achievement differences between children of different races and ethnicities are offered by *class difference theory* and *critical race theory*. Class difference theory attributed achievement differences to socioeconomic or class differences (Rothstein, 2004). In this approach, race was a proxy for social class, because the two variables were highly correlated. Even when income levels of different families were similar, income could mask class differences in form of varying family assets, savings, and educational

histories. Rothstein argued for this perspective by identifying three gaps that existed between middle class children and children of working-class and poor families: (a) a conversation gap, (b) a role model gap, and (c) a health and housing gap.

The *conversation gap* referred to differences in the way middle class and working class parents read to their children and conversed with them. The differences were in the quality and quantity of interactions. Middle-class parents were more likely than working-class parents to use encouragements rather than reprimands, use questioning strategies and indirect means of communication, negotiate with their children, include children in adult conversations, provide effective assistance with homework, instill a sense of entitlement and self-confidence in children, and expose children to organized activities outside of school that reinforce those patterns (Rothstein, 2004). In addition, middle class parents spoke roughly 50% more words to their children than working-class parents, resulting in wide differences in acquired vocabulary by age three.

The *role model gap* referred to the greater exposure of middle class children to social networks and professional acquaintances in comparison to their working-class counterparts (Rothstein, 2004). This notion was discussed earlier in connection with *perpetuation theory* and *network analysis* (Wells & Crain, 1994). Rothstein argued that differences of exposure combined with the evidence that academic effort was not equitably rewarded in the labor market could result in African American children investing less in school.

The *health and housing gap* referred to the lower overall health and higher mobility of working-class children compared to middle-class children (Rothstein, 2004).

Health issues such as vision problems, oral hygiene, nutrition problems, pediatric care, asthma, and exposure to diesel and hi-sulfur fumes disproportionately affected the capacity of lower income children to learn and achieve. In addition, the tendency of children from working class families to move from school to school was an indicator for low achievement.

It is notable that in describing these gaps, Rothstein (2004) characterized the conditions lower income children confront as *deficits*. Although it is clear that he did not impute blame to children through this terminology, it is also true that the term has been used and interpreted in ways that denote inferiority. On account of this history, some researchers, including critical race theorists, took issue with class differences theory.

In light of these multiple gaps, Rothstein (2004) argued that improvement of instruction in schools was a necessary but insufficient strategy for closing achievement gaps. Social and economic reform aimed at narrowing social and economic differences between children were also required. Those reforms must target the class characteristics that produce differences in achievement. He called for a redefinition of school to encompass a greater part of early childhood, through high quality after-school and summer programs that went beyond remediation. He wrote,

The advantage that middle-class children gain after school and in summer comes from the self-confidence they acquire and the awareness of the world outside that they develop through organized athletics, dance, drama, museum visits, recreational reading, and other activities that develop inquisitiveness, creativity, self-discipline, and organizational skills. (p. 21)

In addition, Rothstein advocated for the establishment of greater healthcare services and clinics associated with schools.

Whereas class difference theory sought to explain the existence of achievement gaps, critical race theory challenged the concept of the achievement gap as an accusation against children of color and as a tool of dominance (Love, 2004). This perspective considered seemingly just concepts in society as camouflages for the self-interest of dominant groups and insisted that the stories and voices of subordinated peoples be included in discussions pertaining to them. Love argued that current discussions of the achievement gap combined with the myth of meritocracy provide a justification for the continued domination of Whites and subordination of African Americans.

Love (2004) began with the premise that the achievement gap could be considered in isolation from other related gaps such as the resource gap, the expectations gap, the teacher efficacy gap, the commitment gap, the funding gap, and the school responsiveness gap. In addition, Love argued that consideration of achievement differences must take account of historical roots of African American schooling. She conceptualized this history, which included the denial of education to and bondage of African Americans, the proclamation of their inferiority, and the maintenance of this belief through pseudo-sciences, as a *majoritarian story*. She then framed the present preoccupation with the achievement gap as an extension of that majoritarian story of African American inferiority.

She proceeded to describe the *myth of meritocracy* that is built on the beliefs that the educational and employment system is fair, that those who succeed do so on the basis

of individual effort, and those who do not succeed also get what they deserve (Love, 2004). Inherent in meritocracy are concepts of neutrality, color-blindness, objective standards, equal opportunity, fair evaluation, neutral reporting, and equitable allocation of merit; yet, each of these concepts is susceptible to critique. The notion of an impartial system of merit combined with a non-contextual fixation on the achievement gap creates a harmful distortion that must be countered. Further, the proclamation by the Supreme Court in *Brown v. Board of Education of Topeka* (1954) that separate educational facilities are inherently unequal inadvertently left the ruling open to be used in support of an ideology of White supremacy.

Love (2004) listed requirements for equity in schools: teachers who believe children can learn and believe in their own efficacy to teach them, adequate funding, and refutation of the majoritarian story of inferiority. A comparison of the class differences and critical race theory perspectives on educational equity indicates areas of agreement, areas of disagreement, and some issues of language. There is agreement about the need for good teaching and appropriate funding. There is disagreement over the sufficiency of good teaching and the roles of race and socioeconomic factors in student performance differences. In regard to language, critical race theory rejects the concept of students possessing *deficits* and questions the current conceptualization of the *achievement gap*. It will be important for scholars to closely examine these terms, which are laden with historical and emotional content, and carefully sort out their disagreements and areas of common ground beyond the constraints of current language.

Equity is the final topic in this theoretical section. Equity is viewed as a condition of social justice. The American Heritage Dictionary defines it as “the state, ideal, or quality of being just, impartial, and fair” (DeVenne, 1982, p. 462). Student achievement or performance is central to current concepts of educational equity. Singleton and Linton (2006) defined educational equity as “raising the achievement of all students while narrowing the gaps between the highest- and lowest-performing students; and eliminating the racial predictability and disproportionality of which student groups occupy the highest and lowest achievement categories.” (p. 46) Lee (2004) framed educational equity as a multi-dimensional concept with three components: (a) equality, (b) adequacy, and (c) reciprocity. Equality referred to allocation of resources and opportunities to diverse students groups and to their comparative outcomes; adequacy was concerned with achievement performance relative to academic standards; and reciprocity was concerned with the opportunity of diverse student populations to learn and benefit from each other in integrated educational settings. Lee argued that to the extent that these three conditions were met, educational settings could be said to be equitable; and to the extent that they were not met, there was a violation of equity. Systematic relationships between educationally irrelevant variables such as race and achievement were indicators of inequity in a system. From this multidimensional perspective, an identical achievement gap or difference in two different settings could disguise different levels of inequity.

Lee’s (2004) definition of reciprocity as the opportunity for minority student to learn in a racially integrated school and benefit from the achievement of white students is problematic; he did not cite any studies in support of this idea, and earlier studies have

not supported the *lateral transmission theory*, which proposed that African American students benefit from exposure to values and behaviors of White students (Crain & Mahard, 1981). The benefits of integration appear to originate from factors other than that exposure. It is true that Lee framed reciprocity as a two-way process, since within-group variability of performance is greater than variability between groups. This means that there will always be non-White students who achieve higher than White students in any given school, and presumably the White students could benefit from the achievement of their non-White peers. Even without empirical backing for the concept as it is framed, the idea of reciprocity could have value when considered from the perspective of *network analysis*, a theory that does have empirical support (Wells & Crain, 1994). In that analysis, African Americans and Latinos in integrated schools gain access to informal networks that enable access to integrated institutions and employment. This enables a reformulation of Lee's concept of reciprocity: integrated settings provide students of color exposure to educational benefits they would not experience in isolated settings. The next section summarizes Lee's findings with regard to the three criteria.

Empirical research. A study of racial performance differences by Lee (2004) integrated empirical evidence on achievement differences with data from other contextual variables, e.g., segregation and school funding. It also met the requirement set by Love (2004) and others for considering achievement differences in the context of other inequalities. Lee monitored trends in the *National Assessment of Educational Progress* (NAEP) and long-term NAEP mathematics assessments for White, African American and Latino students at different age levels. He found that between 1978 and 1990 the gap

between African American and White achievement decreased by 0.2 to 0.4 standard deviations. From 1990 onwards, that gap widened across three age levels to between 0.7 and 1.1 standard deviations, returning to the 1980 level. During the period when the gap narrowed, African American students made greater gains than their White counterparts. Latino students, who did not experience significant desegregation as a whole, did not experience large changes in an achievement gap relative to White students during the same period. More recently, results of an NAEP Trial Urban District Assessment in mathematics indicated that only modest progress was made by fourth and eighth graders in major urban centers during the period between 2003 and 2005, and that achievement gaps may be widening (Sanchez, 2005; U.S. Department of Education, 2005).

In order to assess the equity of the observed achievement differences, Lee (2004) used several databases to measure variables relating to the opportunity, adequacy, and reciprocity experienced by African American, Latino, and White students. In the area of equality of educational opportunity, the variables he measured were exposure to teacher quality, exposure to advanced courses, per pupil school expenditures, poverty, single parent household status, and parent educational attainment. He found the following:

1. Math teachers in predominantly African American or Latino high schools were less likely to be certified in math than teachers in predominantly White high schools.
2. The enrollment of African American and Latino students in advanced courses increased from 1978 to 1990 and then remained flat at a level lower than for White students.

3. Per pupil expenditures in predominantly Black and Hispanic school districts were significantly lower than in predominantly White school districts in 2000.

In all of the variables related to home conditions, performance gaps narrowed between students of color and White students in the period between 1970 and 2000. However, performance gaps in these variables remained very high, at greater than a two to one ratio for African American single household status, African American and Latino poverty, and Latino parental college attainment. These findings led Lee (2004) to conclude that gaps exist for African American and Latino students in many areas of educational opportunity related to school and home, and that by the criterion of equality of opportunity, achievement gaps are inequitable.

In the area of adequacy, Lee (2004) assessed student performance against an adequate standard. He selected three standards: (a) an NAEP math score of 300 for 17 year olds; this is a score considered necessary to secure middle class employment; (b) the percentage of African American and Latino high school seniors scoring at *basic* on the national NAEP; and (c) dropout rates. Progress in all of these measures was observed from 1978 through 1990, but flattened in the 1990s. In 1999, only 27% of African American and 38% of Latino 17 year olds achieved at or above Level 300, and the Latino dropout rate stood at 28%. Lee concluded that these students were not experiencing standards of adequacy and that their progress toward achieving adequacy stalled in the 1990s.

In the area of reciprocity, Lee's (2004) findings confirmed declining performance differences between African American and White students during the period of maximum desegregation, followed by steady or increased differences associated with a resegregation trend during the 1990s. Latino students experienced steadily increasing isolation during the last three decades.

On the basis of these collected findings, Lee (2004) concluded the existing achievement gaps for African American and Latino students were inequitable and unjust from the perspectives of opportunity, adequacy, and reciprocity. That is, the gaps are inequitable in multiple ways. Continuing gaps in educational funding and the trend toward resegregation were factors that exacerbated the inequity.

This section concludes with empirical findings that shed light on the relative importance of race and socioeconomic status in student achievement. Willie and colleagues (2002) correlated achievement on the Florida Comprehensive Assessment Test with these two variables using different Florida districts as the unit of analysis. They used the proportion of students qualifying for free and reduced lunch in a district to measure socioeconomic status and the proportion of non-White students enrolled in a district to measure racial composition. In a separate partial regression analysis, they looked at one variable while holding the other constant. They found that

1. Both SES and non-White race had significant and negative associations with achievement.
2. However, the contribution of SES was greater.

3. The association between achievement and race was weakened more when SES was held constant than the association between achievement and SES was weakened when race was held constant. This finding indicated that race as a factor in student performance was highly influenced by the socioeconomic balance that existed in a school.

In another study, Willie et al. (2002) made two additional findings that illustrate the interaction of race, racial isolation, and SES. First, they found that the amount of performance variation attributable to race was greater in segregated schools than in racially balanced schools; and second, they found that high-achieving schools were more likely to be racially balanced than socioeconomically balanced. These findings suggested that racial diversity of a student body was strongly related to high achievement and that socioeconomic homogeneity in the form of concentrated poverty was strongly related to low achievement. The researchers concluded that the best environment for African American and Latino students were those that were both racially and socioeconomically diverse.

Inequalities within Schools

Research has identified other inequalities pertaining directly to schools that serve low income students and students of color. Sometimes these inequalities are associated and discussed with the contextual variables already reviewed: segregation, school funding differences, or achievement differences. The terms *within-school equity* and *within-school integration* have been used to refer to the elimination of all vestiges of legal segregation from policies and practices within a given school (Willis, 1994). In

some court cases, plaintiffs have argued that achievement gaps are vestiges of legalized segregation (Weiler, 1998). A similar analysis could be applied to inequalities at the district level. This section begins with a discussion of theory and proceeds to examine various school inequalities associated with teaching, programs, and achievement. Four topics will be reviewed: (a) social reproduction, (b) equity traps, (c) equity audits, and (d) within-school inequalities. The analysis pertains to inequalities within individual schools and inequalities between schools within a district.

Social reproduction: theory and evidence. Whereas Rothstein (2004) described the class differences children bring with them to school, Carnoy and Levin (1985) theorized on the role of schools in perpetuating and reproducing those differences in service of the labor market. They proposed that educational processes differed according to the class origins of children, with the result that children were socialized for the types of employment their parents held. Four classroom attributes they used to describe this socialization were (a) external versus internal standards of authority, (b) future versus present orientation, (c) verbal self-presentation skills, and (d) emphasis on cognitive skills and achievement. Social reproduction can be considered an extension of class difference theory to consider the role of schooling in maintaining class differences.

The first attribute, external versus internal standards of authority, was illustrated in different types of jobs (Carnoy & Levin, 1985). Workers at lower level jobs function in an environment of rules, regulations and external supervision. In contrast, high-end employees require self-direction and internal norms congruent with organizational goals. For the second attribute, orientation toward the present also aligns with lower level jobs,

in which attention to work routines and procedures are paramount. By contrast, workers in managerial jobs must always consider the future consequences of their decisions and actions. The third attribute, verbal self-presentation, manifests in the types of verbal responses required. Lower level workers are required to provide brief, factual responses; while managerial employees must be able to engage in verbal problem solving and analysis. Fourth, managerial jobs require stronger cognitive skills than lower-end jobs.

Carnoy and Levin (1985) predicted that school socialization for the two types of employment would mirror the contrasting attributes of those types. Constant teacher supervision, externally imposed standards, an orientation to the present, emphasis on brevity, and de-emphasis of cognition and achievement would socialize students for lower-end jobs. Internalization of norms, self-direction, orientation to the future, extensive practice with verbal presentations, and emphasis on cognition and achievement were elements that would support the socialization of future managers. The researchers conducted an ethnographic study that tested these hypotheses in two first grade classes, one in an upper-middle-class neighborhood and one in a lower middle class neighborhood. There was a 40% difference in per pupil funding between the two schools. The researchers coded teacher-student interactions for message and strategy as they pertained to the four attributes.

They found confirmation for their hypotheses in all four attributes (Carnoy & Levin, 1985). For example, internal standards accounted for 52% of the interactions in the wealthier school compared to 17% in poorer school. Children in the wealthier school were encouraged to use their time wisely, take responsibility, and know the consequences

of their actions, whereas children in the poorer school were given more commands, direct praise, and blame. Children in the upper middle class school were eight times more likely to have their attention focused on the future, whether it was tomorrow's learning, what one needed to know to succeed in the next grade, or the prospect of college and employment. Although both classes had verbal presentations in the form of show and tell, the quality and quantity of presentations differed, with wealthier children receiving detailed feedback and questions about their presentations. In the area of achievement and cognition, the focus for the poorer children was on completion of assignments, play opportunities, and quickly closed teacher-student interactions. For the wealthier children, value was placed on the quality of work, mastery of skills, habits of mind, and sustained teacher-student interactions.

These findings lend support to the hypotheses that schooling differs for children of different classes in different schools and that it can act to reproduce the class differences children bring to school. Additional findings of the study regarding the students' parents reinforced the observed patterns (Carnoy & Levin, 1985). Parents in the upper-middle-class school were perceived by the researchers and school personnel as powerful, successful, optimistic, demanding, and willing to challenge school policies. Parents in the poorer school were more timid and fatalistic.

The role of schools in perpetuating class differences rather than intervening to alter them demonstrates that the issue of class in schools is not limited to student performance. It also leads to the question of how this situation might be changed or disrupted. Promoting effective teaching strategies might be one solution; however, it can

be argued that the contrasting teaching styles just described are perfectly effective for producing different kinds of workers. Is instruction targeted toward producing managers to be favored, then? The creation of socioeconomically diverse schools may have potential for disrupting social reproduction by placing children from different social classes in the classroom together. *Teacher expectation theory*, cited above in the review of segregation, may provide a mechanism by which children in those diverse settings may benefit (Crain & Mahard, 1981). Further research on the role of schools in the maintenance of class differences is needed.

Equity traps. The attitudes and behavior of school personnel toward students of color and low socioeconomic backgrounds can be sources of inequality that have a great impact on these students' educational outcomes. McKenzie and Scheurich (2004) described *equity traps* as conscious and unconscious thinking patterns that trap educators by preventing them from believing their students of color can be successful learners. These traps are perpetuated through "dysconsciousness" (p. 603), an uncritical habit of mind that justifies inequity and prevents the belief that all students can achieve. This pattern of belief sets in motion a vicious cycle of lowered expectations by teachers and negative self-view by students. What all equity traps have in common is that they effectively permit school personnel to disavow any responsibility for observed achievement inequalities. The authors' premise was that a substantial portion of that inequity is caused by the attitudes, beliefs, and assumptions of teachers and administrators.

McKenzie and Scheurich (2004) conducted a qualitative study in which they interacted with eight experienced White teachers in a predominantly African American, Latino, and low income elementary school. Interactions consisted of interviews and journals kept by the teachers. The researchers identified several themes in the teachers' comments about their students, and from these themes they distilled four equity traps: (a) the deficit view, (b) racial erasure, (c) employment and avoidance of the gaze, and (d) paralogical beliefs and behaviors.

Deficit thinking attributes failure in school internal deficits or deficiencies within students that manifest in limited intellectual abilities, linguistic shortcomings, lack of motivation to learn and immoral behavior (McKenzie & Scheurich, 2004). Such deficits are often viewed as generational. Holders of the deficit view often espouse sympathy for children perceived to be afflicted with those deficits. The authors believed the appropriate response to the deficit view was to reframe thinking about students, their families, and communities from a deficits-based model to an assets-based model, e.g., to recognize the resources and assets that students bring with them to school. They also recommended specific actions for achieving this shift. The first action was for educators to meet and know students and families on a personal level, through neighborhood walks and conferences held in students' homes. They pointed to research linking home visits to increased student achievement. A second action was to gather oral histories from people in their communities through student interviews of families and neighbors which could then be presented at school. A third action was the three-way conference. In this format, the teacher, student, and family member met twice a year to establish goals, and the

student took a leadership role by presenting his own portfolio. The authors found that when such actions were practiced, educators were impressed with the genuine interest families and neighbors had in children's academic success, and this experience went a long way toward altering educators' deficit thinking.

The second equity trap, *racial erasure*, is equivalent to the color-blind view (McKenzie & Scheurich, 2004). Underlying this view is an unwillingness to address race and the issues it presents. The researchers observed that teachers who practiced racial erasure were able to deny any possibility that they discriminated against students on the basis of race. As with the deficit view, the authors recommended several actions for addressing racial erasure. They include study groups on racial and cultural topics in education and equity audits discussed below.

The third equity trap was *avoidance and employment of the gaze* (McKenzie & Scheurich, 2004). The gaze was a concept formulated by the philosopher Foucault (as cited in McKenzie & Scheurich, 2004). It is defined as "surveillance for the purpose of controlling behavior" (p. 619). Avoidance of the gaze is exemplified by teachers who endeavor to avoid scrutiny by administration and parents in more affluent districts by working in less privileged schools where their mere presence is appreciated. Employment of the gaze is manifested in the norming and sanctioning of the behavior of other teachers who might speak against deficit thinking and other inequities. When the gaze is employed on these would-be rebels in the form of peer pressure, they are effectively silenced. To counter this equity trap, the authors recommended care in the hiring process through stakeholder representation at interviews and interview questions

that serve as checks on applicants' attitudes toward diversity. A second recommendation was to establish norms for democratic discourse and dialogue that are necessary to professional learning communities and which respect and uphold points of view that advocate social justice. A third recommendation was to create schools so thoroughly collaborative that no one can hide destructive or deficit beliefs and behaviors. Frequent classroom visits by administrators and facilitation of parent visits could reverse the destructive effects of the gaze through a different kind of gaze that fostered accountability, growth, and transparency.

McKenzie and Scheurich (2004) described the fourth equity trap, *paralogical beliefs and behaviors*, as a habit of mind that draws conclusions from premises that don't warrant the conclusion, and as self-deception or self-rationalization. Within this framework, educators commit the double offense of treating students unfairly and then blaming the students for the educator's behavior. Recommendations for responding to this trap included having teachers visit the rooms of successful teachers. When teachers observe students succeeding in another classroom, their deficit ideas are deconstructed. Another recommendation was for the provision of instructional coaches to model effective practices, mentor teachers, and assist them in reflecting on practice. A third recommendation was to develop a critical mass of advocates for equity among the faculty, starting with a small group of teachers.

Equity audits. A systematic approach for identifying inequalities in schools through the use of equity audits was set forth by Skrla, Scheurich, Garcia, and Nolly (2004). Borrowed from audits already in existence which focused on civil rights,

curriculum, and school reform, the reconceptualized equity audits are a more limited, focused and public means of transmitting data that indicate the level of equity in specific areas of schooling and suggest directions for change. The researchers noted that equity audits can overcome a tendency of school personnel to avoid discussions of race and socioeconomic status through a formal procedure that breaks the silence and promotes insight and response.

Equity audits included three dimensions: (a) teacher quality equity, (b) programmatic equity, and (c) achievement equity (Skrla et al., 2004). Achievement equity (AE) was conceptualized as the sum of teacher quality equity (TQE) and programmatic equity (PE): $TQE + PE = AE$. Teacher quality equity referred to the distribution of high quality teachers among schools within a district and to distribution of access to such teachers by students within a school. The researchers noted that students of color and low income students are often taught by less experienced teachers. In the conceptual framework, teacher quality equity was broken down into components of teacher experience, teacher mobility, and proper teacher subject placement. These components could be applied to a district audit in order to compare, for example, differences in teacher quality between high SES schools and low SES schools. They could also be applied within a single school audit to examine differences between teachers of gifted and advanced placement classes to those of lower tracks.

Programmatic equity, the second dimension of the model, addressed the under-representation of low income and students of color in school programs such as advanced placement and their over-representation in others such as school discipline and special

education (Skrla et al., 2004). The authors proposed that representation in these programs be disaggregated and examined as part of the audit process. They argued these inequitable patterns had systemic solutions despite the common claim of educators not to have control over them. The third variable, achievement equity, was considered to be the sum of teacher quality equity and programmatic equity. However, it also had its own measurable indicators: achievement test results, access to advanced courses, dropout rates, high school graduation tracks, and participation and results in college entrance exams (SAT, ACT, and advanced placement). Because the authors considered state achievement tests as a necessary but insufficient baseline for assessing achievement equity, they included the additional indicators.

After providing the conceptual framework, Skrla et al. (2004) suggested seven steps for conducting equity audits:

1. Gather a committee of relevant stakeholders.
2. Present the data to the committee and have members graph it in colorful ways.
3. Discuss the meaning of the data, utilizing experts where possible, under the guidance of a facilitator.
4. Discuss potential solutions, giving consideration to the strengths, weaknesses, and costs of each.
5. Implement solution(s).
6. Monitor results and report to the school community.

7. Celebrate if successful; if not successful, return to step three and repeat the process.

The authors warned that school leaders must be prepared for the fearful or negative reactions of some, whether parents, educators, or even leaders, to the open discussion of inequities.

Structural inequalities within schools. Noguera (2001) observed extreme disparities in academic and behavioral outcomes between White, Latino, and African American students in an integrated high school. In a qualitative study, he found that these disparities were not due to a conspiracy to deny educational opportunity to students of color. Rather, disparities were variously ignored, rationalized, or tolerated. He concluded that two virtual schools inhabited a single campus; one an elite college preparatory program, and the other a low-achieving, inner-city school for students of color. This prompted an analysis of the systemic equity and opportunity structure within the school.

Underlying the disparities were two apparently neutral elements of school structure and culture (Noguera, 2001). First, assignment and sorting of students into different academic tracks upon enrollment in ninth grade mathematics determined the “course trajectory” (p. 34) of a student’s entire high school career. Second, every club and team at the school was segregated racially. The fact that this segregation was voluntary gave school personnel a reason not to feel responsible. Noguera concluded that the subtle structural mechanisms through which racial and socioeconomic inequality is

reproduced must be addressed before the achievement gap within schools can be narrowed.

However, addressing these mechanisms inevitably involves political conflict, because it includes the reallocation of resources or the restructuring of programs (Noguera, 2001). He observed that parents of high achieving students fiercely resisted such measures as the reduction of tracking or broadening access to advanced placement courses. A powerful belief that academic excellence and educational equity are at odds seemed to fuel this resistance. Noguera distinguished between academic excellence, which is compatible with equity, and academic privilege, which is not. He concluded that political strategies, rather than educational strategies alone, were necessary to respond to inequalities within schools. Such strategies included Black and Latino parents organizing in order to advocate for their children.

The Shift to Output Criteria: School Accountability Systems

Accountability systems operate within the context of the inequalities for schools that have been described and may be viewed as another contextual variable that school leaders must take account of. The current accountability system is two-tiered, originating at the national and state levels. A stated goal of the national No Child Left Behind Act (U.S. DOE, 2002) is the elimination of one type of inequality, performance gaps; however, it is silent on issues of resegregation, school funding differences, and inequalities within schools and districts. State accountability systems come in several varieties. Together, the two systems place demands and incentives on schools and also threaten sanctions. This section describes the features, incentives, promises, and

problems of accountability systems with special attention to their influence on school leaders and educational goals. The section begins with an overview, then takes up NCLB, followed by state systems, and concludes with an examination of reliability problems of the systems and possible remedies.

Overview

Accountability systems can be confused with testing and assessment systems. Although they contain those elements, accountability systems are the ways in which test data and other measures are used to assess the performance of students, teachers, schools, and districts (Sirotnik & Kimball, 1999). The purposes of accountability systems have been identified as (a) the measurement of outcomes; (b) the improvement of instruction; (c) the creation of incentives; and (d) the imposition of interventions, rewards, or punishments (Hanushek & Raymond, 2002). The goals of the national accountability system are the improvement of student achievement as measured by standardized tests and the elimination of achievement gaps (U.S. DOE, 2002). Because each state implements NCLB and its own accountability system differently, the implementation of national policy is decentralized (Fusarelli, 2004). This has prompted discussion by scholars on how to effectively implement accountability systems.

Izume and Evers (2002) provided five questions regarding the characteristics, quality, and administration of assessments:

1. What are the schools accountable for?
2. For whom are schools accountable?
3. Is progress measured absolutely or relatively?

4. Is it fair?
5. Is it informative?

The third question relates to how performance is evaluated, and this varies from among the different accountability systems. Models based on performance status, improvement, and student gain are in use, and each has distinct assumptions and outcomes within an accountability system.

Seven general characteristics for effective accountability systems identified in the research literature were summarized by Englert et al. (2004): (a) high expectations for all students; (b) high quality assessments aligned with standards; (c) alignment of resources, support, and assistance for improvement; (d) sanctions and rewards linked to results; (e) multiple measures; (f) diagnostic uses for data; and (g) readily understandable to the public. It should be noted that alternative criteria developed by Sirotnik and Kimball (1999) overlap somewhat but are quite different in aggregate. Their criteria focused on many elements outside standardized testing, including community-based indicators, the monitoring of equitable inputs, flexibility, professional development for teachers, classroom-based assessment, and adequate funding. As desirable as these elements may be, they are not emphasized in current accountability systems. Therefore, the characteristics specified by Englert et al. will be the focus of this review.

The first characteristic, high expectations for all students, is a familiar element of effective schools research (Englert et al., 2004). One way this expectation was expressed in NCLB (U.S. DOE, 2002) and many state accountability systems was in the requirement to test all students, including subgroups that have been exempted in the past.

The second characteristic, high quality assessments aligned with standards, has been difficult to achieve in some states. Alignment to standards was poor in California in the early 2000s when the norm-referenced Stanford Nine test was used as the sole assessment. Although most states have now achieved good alignment, many issues of quality inherent in the tests themselves have emerged, including statistical bias associated with various models, statistical noise, ineffective incentives, perverse incentives, effect of school diversity on results, and effect of school size on results. Each of these issues is discussed in detail later in this review. Their emergence suggests the addition of three questions about accountability systems in addition to the five provided above:

6. Are assessment results statistically reliable?
7. Does the system provide effective incentives?
8. Does the system conform to educational research?

Alignment of resources, support and assistance for improvement was the third characteristic of effective accountability systems (Englert et al., 2004). This characteristic highlighted the importance of adequate funding resources targeted toward professional development for teachers and administrators, especially those serving diverse and low-income populations. The fourth characteristic, sanctions and rewards linked to results, is a subject of varying practice and opinion. Some states simply report accountability results while others give tangible rewards in the form of monetary rewards for schools and teachers on the one hand, and sanctions in the form of closures, reorganization, and the withholding of funds on the other hand.

The fifth characteristic was multiple measures (Englert et al., 2004). In addition to test scores, other types of student outcomes such as attendance and dropout rate are included in many accountability systems. Other types of data newly available to schools such as equity audits can assist in school quality evaluations (Skrla et al., 2004). The possible diluting effects of multiple measures in accountability systems is discussed later in the review.

The sixth characteristic of effective accountability systems was the use of data by district and school personnel for diagnostic purposes and changed practice (Englert et al., 2004). Effective use of data requires proper presentation, interpretation at different levels of the school system, and the capacity, leadership and will to respond appropriately. The seventh characteristic of effective accountability systems was the requirement that they be informative to parents and the community. This is a requirement of NCLB, which mandates data-rich report cards for every school (U.S. DOE, 2002). Public reports should be user-friendly and include information about the elements and weighting of the system. However, the complicated nature of some accountability systems can make user-friendliness and transparency difficult.

No Child Left Behind

Characteristics of the national accountability system are now discussed. In its fact sheet, NCLB (U.S. DOE, 2002) stated that its goal was to “close the achievement gap between disadvantaged and minority students and their peers” (p. 1). It proposed to reach this goal by on the basis of the principles of accountability, flexibility, parental choice, and effective teaching. The law set out standards for adequate yearly progress and

required states to implement accountability systems to fully meet standards within twelve years. The accountability systems must include the testing of all students in Grades three through eight, the reporting of test disaggregated data, and rewards and sanctions. Also, districts must certify that all teachers in core subject areas were highly qualified, that is, teaching in an area of expertise in which they were fully certified.

Adequate yearly progress. Schools' adequate yearly progress or AYP in NCLB (U.S. DOE, 2002) is a yes or no proposition that is determined by three requirements: (a) achievement targets called *annual measurement objectives* (AMO's), which are achieved on criterion referenced tests in reading and mathematics; (b) a minimum requirement of 95% of students taking the test; and (c) a third measure, which is often attendance in elementary and middle school and graduation rate in high school (R. Hahn, personal communication, December 20, 2005). These requirements must be satisfied by each subgroup at each grade level each year. Failure to achieve only one measure in one grade level results in a school not making AYP. Three factors built into the system influence and mitigate the strict requirements of AYP: (a) subgroup size, (b) confidence intervals, and (c) the *safe harbor* provision.

Subgroups. The determination of subgroups within NCLB affects schools powerfully. Subgroups specified by NCLB (U.S. DOE, 2002) are race, ethnicity, poverty status, disability and limited English proficiency. Students may be counted as belonging to multiple subgroups, and states may determine the threshold size for subgroups. Currently, that threshold is 40 students in Arizona. Although all students' scores and participation are counted in aggregate, for purposes of determining AYP, their scores and

participation are only counted as part of subgroups that exceed 40. The second factor influencing AYP attainment is the statistical confidence interval at the .01 level of statistical significance around the AMO targets. These intervals create a built-in margin of error. The third factor, the safe harbor provision, applies to schools containing subgroups who have missed AMO's but have achieved a 10% improvement plus either a 94% attendance rate or a one percent improvement in attendance for any given year. For example, many schools in Arizona achieved AYP in 2004 through the safe harbor provision after the mathematics portion of the state-mandated assessment was re-normed (R. Hahn, personal communication, December 20, 2005). Re-norming lowered the cut score for mastery, allowing more students to pass the test.

The AMO targets increase so as to reach 100% mastery for each subgroup by 2014. Initial AMO levels in Arizona were set at the 20% level, the average aggregate score of the school in the twentieth percentile within the state. States may determine the manner in which AMOs increase. For example, Arizona uses a plateau model in which AMOs rise every three years for the first nine years of the period covered by the legislation and then rise every year for the final four years, thus effectively postponing most of the required growth (Arizona Department of Education, 2005). Because AMOs are absolute standards that apply to all subgroups, accelerating gains are required of low scoring subgroups in order for schools to achieve AYP. When schools miss AYP due to a single reason measure, such as the failure of eighth grade English language learners to meet an AMO in reading, the system possesses a high degree of discrimination that may

assist in making a diagnostic corrections; however, as targets escalate with the expected outcome that AYP is missed for multiple measures, discrimination is reduced.

Unit of accountability. NCLB (U.S. DOE, 2002) effectively shifts accountability from states to schools and districts. A telling definition in the NCLB act that illustrates the shift of accountability is that of the *failing school*; that term refers to a school that consistently does not make adequate yearly progress towards state mandated goals, including achievement targets. The legislation provides choice for parents whose children are “trapped” (p. 2) within such schools. Districts must give students attending schools identified for improvement an opportunity to attend another school in the district, provide transportation, and allocate at least five percent of its Title I funds for these purposes. Further, the act requires districts to use up to 20% of Title I funds to provide school choice and support services for students from failing schools, including their enrollment in private schools.

In determining whether a school is failing, NCLB (U.S. DOE, 2002) did not consider conditions such as racial imbalances, concentrations of poverty, or school funding disparities. On the contrary, all schools regardless of context were expected to meet the escalating targets. Although NCLB authorized additional federal funds for states, mainly in the form of competitive grants, and also provided for flexibility in the allocation of Title I and other federal funds by districts, it did not provide funding based on adequacy model projections (see Baker, 2005). Thus, NCLB ignored many input variables that are associated with inequalities of performance. One superintendent criticized NCLB for misplacing accountability in this way: “Schools labeled as ‘failing’

will not receive their label because they have failed. Rather, schools will be branded because they are in poor or diverse neighborhoods, because they are small and rural, because they are underfunded, and because the AYP system cannot tell the difference between a learning gain and random noise” (Mathis, 2003, p. 6).

Potential benefits and pitfalls. Potential positive and negative effects of NCLB on educational equity in schools were listed by Fusarelli (2004). On the positive side, the legislation required attention to narrowing performance gaps, the focused use of resources, and improved systems of data gathering and reporting. By requiring the use of data disaggregated by subgroup, NCLB shone a light on subgroup performance and prevented school districts from hiding low subgroup performance within aggregated averages.

Possible negative effects of NCLB on educational equity included the all-or-nothing character of AYP, an over-emphasis on test preparation at the expense of authentic learning experiences, a narrowing of the curriculum so that non-tested subjects become marginalized, and a decrease in creativity and flexibility of teachers (Fusarelli, 2004). The legislation held educators responsible for achievement gaps that have persistent complex socioeconomic and societal origins, yet it ignored funding, racial, and socioeconomic differences between schools. NCLB contained no language pertaining to multicultural or culturally relevant approaches and methods. Although it didn't prohibit those approaches, they could become a casualty in an accountability environment where test scores determined the educational agenda. In the same way, students who learned differently due to cultural and linguistic differences could be at risk.

Fusarelli (2004) argued that by delegating nearly all accountability responsibilities to states and school districts, NCLB created rather than solved problems in educational equity. Given the substantial inequalities that exist between and within states, the use of a decentralized system to achieve equity for all students was unlikely to succeed. Not all states had well-developed curricula, standards, and assessments. A perverse incentive existed for states to lower standards in order to make AYP easier to achieve. In addition, underperforming and failing labels were more likely in poor schools with diverse student populations. Educators might become demoralized. These are some of the potential unintended consequences of a decentralized system.

Arguing from a structural perspective, Fusarelli (2004) claimed that if closing achievement gaps was a goal of national importance, then appropriate federal centralization and support must be provided. Furthermore, federal intervention must take account of complex causes of achievement gaps, including ethnic and class bias, interest group pressure, structural and societal inequalities, and problems within families and communities. When it is considered in light of the seven characteristics for effective accountability systems described above (Englert et al., 2004), NCLB (U.S. DOE, 2002) seems strongest in the areas of high expectations for all students, multiple measures, and reporting of results in ways readily understandable to the public. Due to its decentralized application, areas in which NCLB appears weaker include high quality assessments aligned with standards, alignment of resources and support, and diagnostic use of data.

State Systems

State systems represent a second layer of accountability for schools and educators. Some state accountability systems predated and foreshadowed NCLB (U.S. DOE, 2002); the current federal legislation requires that all states implement a system. Each state does this in its own way, as is illustrated in the following examples.

Texas. Texas used a system of absolute performance standards to determine adequate yearly progress in the same manner as NCLB (U.S. DOE, 2002). Each school received a rating within a five-level rubric. The highest rating, *exemplary*, required that 90% of students in a school and in each ethnic and socioeconomic subgroup pass on each section of the Texas Assessment of Academic Skills (TAAS; Izume & Evers, 2002). Lower ratings were characterized by lower thresholds of performance. The TAAS was a minimum proficiency test that was criticized for not being sufficiently rigorous. When standards were revised and more difficult questions were added to the test in 2001, the cut-scores for passing were also adjusted downwards, creating ambiguity regarding student performance and undercutting the stated purpose of raising standards.

In addition to its AYP component, the Texas accountability system included a separate feature that monitored the growth of schools using a statistical index called *comparable improvement* (Izume & Evers, 2002). This measure compared schools with similar ethnic, socioeconomic, and language characteristics. Various rewards and sanctions were allocated based on the combined measures from the AYP component and the growth component of the system.

California. In contrast to Texas, California eschewed absolute standards and used relative growth in its accountability system (Izume & Evers, 2002). An *academic performance index* (API) based on state assessment scores was calculated at each grade level for each subgroup and compared from one year to the next. Annual growth targets for schools were determined using a similar schools index similar to the comparable schools measure in Texas. In order to meet a target, each subgroup within a school must achieve 80% of the school's growth target. Monetary rewards and bonuses up to \$50,000 per teacher were allocated to schools that meet targets. Schools that did not meet targets were eligible but not required to apply for intervention assistance, and were subject to various sanctions after multiple years of underperformance. Under this system, some of California's worst performing schools received no interventions.

Florida. Florida's accountability system combined a performance level component with a measure of student gain. Gain or value-added analysis differs from growth models by tracking individual students' progress from one year to the next. Growth models, by contrast, focus on year-to-year changes at each grade level. Florida specified a five-level rubric for scores on the Florida Comprehensive Assessment Test (FCAT; Izume & Evers, 2002). Schools received an annual report card based on three factors: the percentage of students receiving a three or better on the test, a measure of individual student gain, and a measure of the improvement of students in the lowest quartile. Through this third measure, the system ensured that high scoring students did not carry a school. In addition, unlike California and Texas, ethnic and socioeconomic subgroup scores did not play a major role in determining a school's status. Although it

did not set subgroup targets, Florida set a minimum achievement level under which subgroup achievement may not fall.

Other features of the Florida accountability system were merit pay, hi-stakes accountability, and exit vouchers (Izume & Evers, 2002). Teacher salaries were partially determined on the basis of student achievement, with bonuses of up to 5%. High stakes were operative at every point in the system: Student promotion to the next grade was dependent on FCAT performance. Parents of students in underperforming schools had the right to vouchers that enabled them to enroll their student in private schools. Although they have been rarely used to date, vouchers were intended as an incentive for low performing schools to raise their rating.

Arizona. Arizona's accountability system combined elements of performance level models, growth models, and value added models. Schools accumulated points in each area. Data was disaggregated by grade level and subject only; race/ethnicity and SES subgroup performance was reported but did not figure into the point system. Absolute or status points were derived from the percentage of students at each grade level that pass the Arizona Instrument to Measure Standards (AIMS; Arizona Department of Education, 2005). Improvement or growth points utilized a three year interval in which the first year was the baseline year and the second and third year created a pooled average. The improvement points took account of students who moved into the passing category as well as lower-performing students who moved closer to passing. Performance level points and growth points were added using a 70-30 weighting that gave the best advantage to a school. Value-added or gain points were used only at the

elementary level; they were calculated as a pooled three-year average of the ratio of students achieving *one year's growth* on assessments divided by the number of students taking the assessments. Thus, the value added component was an aggregate measure of individual student gain. Additional points were earned based on NCLB (U.S. DOE, 2002) adequate yearly progress, high-school dropout and graduation rates, and the percentage of students that *exceeded* standards on the AIMS test. The Arizona system did not provide monetary rewards; progressive interventions and sanctions were imposed on underperforming schools.

Distinctive features of Arizona's system which may be seen as advantages in comparison to NCLB included the following:

1. Because points were accumulated additively, no single measure resulted in an underperforming rating.
2. Incentives were built into the system for improvement at all levels of the achievement continuum and not only at the cutoff between passing and not passing.
3. As mentioned above, the system combined elements from three prevalent models: absolute, growth, and value added.

However, these same features may also be disadvantages. The complexity of the system and number of measures may make the intended incentives difficult for educators to understand, lessening their effectiveness. Also, the pooling methods and range tables used to determine points resulted in considerable data loss and seemed more suited to a tax table than to educational considerations. Moreover, although the value-added

component of the system purported to use individual students as a control for measuring gain, the measure was actually aggregate in nature. It combined the number of students achieving one year's growth on both the math test with those achieving one year's growth on the reading test. Because these two groups overlapped but were not equivalent, the measure gave an inflated estimate of student gain.

Promise and Limitations of Accountability Systems: Empirical Findings

In this section, empirical findings on the promise and limitations of accountability systems are discussed. One study indicated that incentives of accountability systems influenced the achievement of students at the margins of passing (Reback, 2005). Other studies indicated that issues of statistical reliability made the attainment of accountability goals difficult. These issues centered on subgroup size, statistical noise, and perverse incentives (Kane & Staiger, 2002a, 2002b).

Benefits for "bubble" students. One incentive of NCLB (U.S. DOE, 2002) and state systems based on AYP was for students on the margins of passing, or *bubble students*, to improve performance. Schools would be predicted to shift resources and assistance to students at the margin. Using data from Texas in the 1990s, Reback (2005) found evidence in support of this incentive. He observed higher than expected scores from marginal students when their test score was important to the school's rating, higher scores in general of all low achieving students, and lower scores than expected from high achieving students. He hypothesized that the effects of the incentives were due to shifts of resources mediated through broad instructional changes rather than efforts targeted at specific students.

Subgroup distortions. The designation of the minimum number of students required to constitute an official subgroup has an impact on a school's chances of achieving adequate yearly progress under NCLB (U.S. DOE, 2002) and many state accountability systems (Kane & Staiger, 2002a). That is due to the statistical property that achievement variation within a group is much greater than variation between groups. For example, in a North Carolina elementary assessment, there was considerable overlap in the distribution of statewide scores of African American students with scores of the total population, with 30% of African American students scoring above the statewide mean. However, at the level of subgroups within schools, the overlap of distributions was much reduced, so that only two percent of African American students attended schools where the mean performance of African American students exceeded the statewide mean. This statistical property makes it highly likely that schools with African American, Latino, special education, or English language learner subgroups will fail under NCLB.

Currently, many states set the minimum number of students for a subgroup at 30 or 40, which is an important threshold for statistical validity (R. Hahn, personal communication, December 20, 2005). In Arizona, 60% of public schools had Latino or African American subgroups in 2001 (Kane & Staiger, 2002a). However, the numbers 30 or 40 are arbitrary from an educational point of view. Under NCLB and state systems with similar performance level standards, each additional subgroup places a school at a statistical disadvantage, since it represents an additional roll of the dice and a chance to fail. Schools that are more integrated or racially balanced are, all other factors being

equal, more likely to experience a higher failing rate under NCLB. Thus, the national accountability system and some state systems have a built-in perverse incentive for segregation. There are additional perverse incentives to reclassify students in order to avoid subgroup status and to encourage specific students not to be present on the testing day. The problem of subgroups is exacerbated when rewards and sanctions are considered. Because children of color are more likely to attend diverse schools than white students, subgroup rules put these students at a statistical disadvantage in the competition for monetary rewards in states like California.

Kane and Staiger (2002a) investigated whether despite these problems, subgroup rules might have a beneficial impact on achievement for students of color. They reasoned that accountability systems with subgroup components should create incentives for improved achievement just above the numerical threshold where subgroups are counted. They compared student performance between Texas schools with Latino populations just above and just below the threshold and found no difference in performance between the two groups of schools. They obtained similar findings for California. Kane and Staiger concluded that subgroup rules were not having their intended effect of raising the achievement of targeted subgroups and argued against the use of subgroup targets in accountability systems altogether. In 2002, NCLB (U.S. DOE, 2002) and eleven state systems sanctioned subgroup underperformance, including Texas and California. States that did not use subgroup targets in their accountability systems included North Carolina and Arizona, which simply reported subgroup results. Florida

took a middle approach; although it did not set subgroup targets, it set a minimum achievement level under which subgroup achievement could not fall.

One solution to the disadvantages faced by schools with many subgroups or racially imbalanced schools with high populations of students of color would be a system that sorted and compared schools with similar student profiles. Components of the Texas and California accountability plans utilized such an approach. However, this approach was inconsistent with the premise and structure of NCLB, which required that all groups meet absolute performance standards.

Reliability. Variability in test scores from one year to the next can be divided into persistent and non-persistent sources (Kane & Staiger, 2002b). The persistent variation, also called *signal*, is associated with educational factors such as teacher quality, instructional practice, and student learning. The non-persistent variation consists of sampling variation and one-time systemic shocks; together these factors are designated as *noise*, since they do not reflect actual educational inputs or outputs. Kane and Staiger found that the level of noise in accountability systems was unexpectedly large and could lead to misinterpretations of the data and the misallocation of sanctions and rewards. They also found that the level of noise depended on the type of accountability system being used and on the size of schools.

Noise levels in data from North Carolina in the 1990s were found to be the least for data reported in absolute or performance level terms (Kane & Staiger, 2002b). NCLB (U.S. DOE, 2002) and Texas utilized the performance level model. Because of the all-or-nothing nature of these systems, many states chose systems based on two other models:

growth models based improvements from one year to the next for grade levels and subgroups; and value-added or gain models that track the progress of individual students. California used the growth model in an effort to solve the problem of low subgroup performance discussed above. As was mentioned earlier, Arizona combined elements of all three models. Kane and Staiger found that noise levels for all three models increased dramatically as school size decreased. The noise level associated with the value-added model was much greater than in the performance level model, exceeding it by 50% in the quintile of schools with the smallest enrollment. Noise was even greater for the growth model, exceeding the noise level for the performance model by 80% for the smallest quintile of schools.

The consequence of noise is that data is more volatile under growth and value-added models and for small schools (Kane & Staiger, 2002b). A study in Colorado detected volatility so high that correlations among scores during a four year period were near zero (Linn & Haug, 2002). Volatility resulted in both rewards and sanctions being meted out and conclusions drawn based on unreliable results. In a study conducted in Chile, Chay, McEwan, and Urquiola (2005) demonstrated that noise led to a gross overestimation of the impact of a school intervention program. On the other hand, large schools that were less prone to volatility were less likely to experience either rewards or sanctions and might not be responsive to accountability incentives.

Kane and Staiger (2002b) offered recommendations to alleviate problems associated with noise. First, they proposed that test results be averaged over several years to reduce the level of noise and yield more reliable results. Although NCLB (U.S. DOE,

2002) provided for weighted averages over three years, there was an even greater need for state systems using gain or growth models to utilize such averaging. Sophisticated moving averages derived from Bayesian modeling have been developed. One such model removed a bias in value-added models (Li & Tobias, 2003). A disadvantage of statistical modeling is that it reduces transparency of reporting to the public; however, this loss of transparency may be an acceptable trade-off for increased precision (Linn & Haug, 2002). The second recommendation of Kane and Staiger was to sort schools into separate classes based on size and establish thresholds specific to those size classes for growth and gain targets. Where monetary rewards were given, they could be spread out in smaller amounts to a larger number of schools in order to further minimize the effects of volatility.

NCLB and segregation. The interaction of accountability systems with segregation, a source of inequality discussed earlier in this chapter, is now considered. As noted earlier, NCLB (U.S. DOE, 2002) did not deal with the issue of segregation; however, the converse was not true. Stiefel, Schwartz, and Chellman (2007) demonstrated that segregation had a profound effect on the provisions of NCLB and questioned the premise that accountability at the school and district level reduced gaps within states and nationwide. They found that in New York State, elementary and middle school students were segregated to such an extent that over one half the schools were too homogenous to be subject to ethnic/racial subgroup accountability under NCLB; that is, their only accountability under the law was for aggregate performance. NCLB included a provision that if a school had too few students in a subgroup to be held accountable for

the group, the district would be held accountable through the pooling of subgroup members at the district level. This provision resulted in the capture of additional schools that otherwise would not be accountable for subgroups. However, inter-district segregation in New York was so pervasive that even under the provision, 25% of elementary schools and 36% of middle schools remained unaccountable for subgroups; and these unaccountable schools were the most segregated schools.

Compounding this situation was the finding that achievement gaps in New York were greater across segregated schools than within partially integrated schools (Stiefel et al., 2007). That is to say, schools held accountable under subgroup rules were not the schools with lowest test scores for children of color. Rather, non-accountable segregated schools had the lowest performance for those students. On a fourth grade Language Arts test, there was a 32 point gap between White students and children of color in segregated schools, while the gap in partially integrated schools was 20 points. As a result, NCLB (U.S. DOE, 2002) did not leverage the problem where it was greatest. Rather, the authors found that schools held accountable were most likely to be large, urban, and somewhat integrated. They suggested that the accountability unit may need to be raised from the district to the state level in order to overcome the distorting and damaging effects of segregation, which NCLB ignored.

Adjusting Accountability Systems

This section concludes with a discussion of how the advantages of accountability systems could be maximized and the problems minimized from an educational perspective. First, it is useful to evaluate the three most common accountability models.

Performance level models succeed in setting an egalitarian standard and seem to create an incentive for helping students at the margin of passing. In performance level systems such as NCLB (U.S. DOE, 2002), the unit of analysis is the school, which must meet all performance requirements each year in order to be certified as making adequate yearly progress. Performance level systems are also less susceptible to statistical noise than other accountability models. However, performance level systems are weak in providing the means to achieve success. They do not take account of either of existing and persistent differences in student achievement or of other systemic inequalities. They place schools with diverse student populations and concentrations of poverty at a statistical disadvantage. Due to the high profile of NCLB, performance level systems have undergone more scrutiny and been more influenced by the political process than other models.

In educational terms, the performance level model implies what Ronald Edmonds claimed in 1979: educators have it within their power under present circumstances to educate all children to minimum but sufficient standards. Edmonds believed that the primary requirement for educators was the will to do so. He also viewed the debate about educating diverse and poor students as centered on two opposing premises: (a) family background determines school performance, and (b) school response to family background determines school performance. This review indicates a more complex picture: that background and school response are interactive, and that other intervening variables such as racial isolation and inequalities of funding also influence student performance.

In addition to implying that educators can raise achievement and eliminate gaps, performance level systems hold educators responsible for any failure to meet this goal. NCLB (U.S. DOE, 2002) provided a 12-year period of time for achievement of the goal, and maintenance was expected thereafter. However, NCLB contained inconsistencies in the form of mitigating factors as hedges against its absolute requirements. Subgroup size thresholds, safe harbor, and confidence intervals may be welcome allowances that enable a given school to make AYP in a particular year, yet these are allowances more rooted in statistics and politics than in educational theory. Moreover, the decentralized way in which NCLB was implemented allowed states to create their own assessments, re-norm them at will, and postpone the required gains until the end of the specified compliance period. If NCLB was a high-jump bar for educators to clear and a stick to beat them up if they don't, then the flexibility built into the bar allowed it to sometimes stay in place when grazed and may also lessen the sting of the blow.

Growth model such as California's focused on relative or incremental progress. The unit of analysis for this model was the classroom teacher, since she or he was the constant in a system in which the performance of successive groups of students was monitored. This system could be compared to observing a waterfall and focusing on a single location; was the flow greater or less at that location in any given year? In educational terms, growth models lead naturally to a focus on increased teacher efficacy and capacity over time. This implies ongoing professional development. It is not surprising that California's system included monetary rewards for teachers.

As noted however, growth models are plagued by statistical noise and high volatility in smaller schools. Because of their relative nature, they are also susceptible to having overly modest improvement targets that fail to narrow achievement gaps. This is the opposite problem of performance level systems that require eventual 100% mastery. The problems of growth models can be remedied through the averaging of test results over several years, sorting of schools into size classes, realignment of rewards in order to avoid misallocation, and the political will to maintain high expectations. One shortcoming of the growth model that is more difficult to remedy is that it is not student-centered.

The value-added or gain model may be the preferred model from an educational perspective. It takes the individual student as its unit of analysis. It asks the question: “On the continuum of mastery of standards, how far has this individual student progressed?” This approach is analogous to observing a waterfall not at a particular location, but rather following an individual water molecule on its path through the fall. Gain models are consistent with current educational theories and practices such as differentiated instruction and those described in *The Skillful Teacher*, a popular handbook of instructional practice (Saphier & Gower, 1997). It can also be argued that student gain is more susceptible to influence by principals. Current data systems allow monitoring of individual student progress disaggregated by specific standards and competencies. Student progress could be the focus of the evaluation process principals conduct with teachers. Student progress is free from cohort effects and may be easier to measure than teacher effectiveness.

However, the measurement of student gain on a district wide basis requires investment in sophisticated data systems (Kane & Staiger, 2002b). It is also not without bias and is nearly as susceptible to noise and small-school volatility as growth model systems. These problems can be remedied through some of the same measures recommended for growth models plus the use of statistical modeling to remove bias.

The use of subgroup targets is problematic, whether they are absolute as in NCLB (U.S. DOE, 2002) and Texas, or relative as in California. Subgroup targets take specific racial, ethnic, socioeconomic, language, gender, and disability groups as their unit of analysis. On the one hand, states, districts, and schools have an interest in monitoring the progress of these subgroups. Furthermore, the capacity for monitoring is now largely in place. On the other hand, the determination of subgroup size is arbitrary from an educational perspective; subgroups have a high degree of overlap; the statistical properties associated with subgroup monitoring create several perverse incentives and disadvantage students in diverse and poor schools; and there is evidence that subgroup targets are not effective in raising the achievement of targeted groups (Kane & Staiger, 2002a). A possible explanation for this latter finding is that educational literature and theory is not focused on the narrowly defined goal of raising subgroup achievement. To be sure, there is research on differentiated instruction for learners at different levels and culturally relevant teaching for students of different backgrounds (Echevarria, Vogt, & Short, 2004; Saphier & Gower, 1997); however, these strategies may fit more closely with an individual student-centered gain model than with a group-centered gain model.

Based on this analysis, it would seem advisable to require the reporting of subgroup performance as part of all accountability systems but to abandon subgroup performance targets. Although subgroup targets were required by NCLB (U.S. DOE, 2002), states had discretion to exclude them from their accountability systems. A system like Florida's was a compromise approach that may have value; it penalized schools when subgroup performance dropped below a floor level but did not set improvement targets. Simple reporting does not require size thresholds which are statistically necessary but educationally irrelevant. Reporting of disaggregated data for achievement measures and other measures of equity can prompt reflection at the district and school levels and serve as the basis for an equity audit (Skrla et al., 2004). In addition, new measures that take account of subgroup performance, such as the equity achievement driver introduced in this study, can assist in the process of diagnosis and change.

Although the state of Arizona did not include subgroup targets, it did combine elements of status, growth, and gain models in its accountability system. In so doing, *Arizona Learns* (ADE, 2005) met the effective accountability system criterion of multiple measures (Englert et al., 2004). However useful these measures may be in assessing school progress, they are of questionable value in motivating change. Given the task complexity associated with any single accountability model, the combination of three models increases the complexity. In addition, the Arizona system reduces, weights, and hedges data at various levels. It appears that Arizona Learns may not effectively provide school leaders with tools to focus the energies of the staff and school community toward

improvement. Responses of school leaders to Arizona Learns are investigated in this study.

School Leaders in the Accountability Environment

This review has focused thus far on describing the current context of school accountability. It now proceeds to locate school leaders within that context. This section begins with a discussion of the relationship between principals and equity in schools. The discussion includes three themes: fostering new meanings about diversity, promoting inclusive school practices, and building communities. Much of the literature reviewed for this discussion necessarily predates NCLB (U.S. DOE, 2002). Next, the response of principals to accountability systems is considered. Findings indicated that accountability systems were affecting principal behaviors and that principals sought more support and resources to implement data-driven strategies to meet the requirements of accountability systems.

School Leaders and Educational Equity

The task of organizing and administering schools for diverse students was addressed in a review of principals and inclusive schools by Riehl (2000). She argued that although administrators tend to be oriented toward the status quo and do not readily admit publicly to problems of race, class and gender within their schools, they must promote change in order to meet the needs of their diverse students. Riehl identified three change tasks for leaders of these schools: (a) fostering new meanings about diversity, (b) promoting inclusive practices within schools, and (c) building communities. These tasks provide the framework for the following discussion of principals and equity.

Fostering new meanings about diversity. Riehl (2000) hypothesized that for inclusive structures and practices to be lasting, they must be accompanied by new understandings and values. In his theory of *organizational sense-making*, Weick (1995) proposed that lasting change results not primarily from technical changes but from the collective construction of new understandings by persons inside and outside schools. Principals play a key role in organizational sense-making through their ability to define situations and their meanings (Smircich & Morgan, 1982). Literature in the critical tradition indicated that an atmosphere of democratic discourse enables principals and school staff members to co-create policy based on new understandings (Corson, 1995). Democratic discourse begins with questions such as, “Who benefits from what goes on here?” and is characterized by free exploration, honest expression, and non-manipulative discussion (Riehl, 2000). Democratic discourse may be a rare occurrence in schools, hindered by hierarchical relationships and a lack of attention to issues below the surface.

Democratic discourse was a component of the leadership framework promoted by Carolyn Shields (2004). Drawing on scholarship about leadership for social justice and empirical research, she described leadership as a primarily moral activity in which the goals were the enhancement of equity, social justice and the quality of life for students. In this framework, educational equity is one measure of social justice. Permitting or expecting inequitable outcomes for poor students and students of color was symptomatic of “pathologies of silence” (p. 117) on the part of educators. Pathologies of silence consisted of ways in which school personnel responded to the differences of race, ethnicity and class among students. These differences were pathologized in two ways: by

being ignored or by being treated as deficits (see the discussion of equity traps earlier in this review; McKenzie & Scheurich, 2004). Shields argued that pathologies of silence are overcome by a moral dialogue initiated by leaders, e.g., *dialogic leadership*.

Shields' (2004) leadership framework consisted of three components: (a) transformative leadership, (b) the centrality of relationships, and (c) the facilitation of moral dialogue. The first component, transformative leadership, was distinguished from transformational leadership, which was concerned with the alignment of collective interests through institutional and organizational arrangements (Liethwood & Jantzi, 1990). In contrast, transformative leaders were concerned with rooting morality in the social context and creating the value ends of equity, social justice and quality of life. The second component of Shields' framework, centrality of relationships, focused on relationships as the basis of knowledge, curriculum, and learning. In curriculum, relationships among things rather than isolated facts became the basis for understanding. In relationships among people in schools, the theme of caring is central. Shields cited Noddings on the *pedagogy of care* centered on relationships and ideas. The third component of Shields' leadership framework was moral dialogue. Consistent with the concept of democratic discourse discussed by Riehl (2000), moral dialogue was initiated by leaders with the goal of discovery and new understandings.

Several indicators would be anticipated in a school where this leadership framework is practiced (Shields, 2004). Differences would be openly discussed and questions of inclusion explored. Classroom discussion of differences would not be suppressed. Students would not be singled out and embarrassed with regard to

differences; however, discussions about color, wealth, poverty, and power would occur. Middle class experience would not be the sole standard of discourse, and unjust situations would be spoken out against and corrected. Actions and arrangements would be evaluated as to whether they were just, democratic, caring and optimistic. Frank examinations of school practice such as equity audits would be a part of school life (Skrla et al., 2004).

Promoting inclusive school practices: instructional leadership. The second task for principals in diverse schools was promoting inclusive practices within their schools, and it was divided into two subtasks: (a) improving instruction so as to enable diverse students to succeed, e.g., instructional leadership, and (b) molding school cultures that embrace and support diversity (Riehl, 2000). Instructional leadership has been cited as an element of effective schools (see Edmonds, 1979). Although the literature on instructional leadership has been criticized for failing to explore issues of student diversity explicitly, some studies built theoretical frameworks about leadership and instruction in diverse environments.

For example, Rosenholtz (1985) collected and analyzed case studies conducted in poor, urban elementary schools. Empirical findings from these studies consistently indicated that, other factors being equal, there was a high degree of variation in achievement effectiveness among these schools. In addition, some schools demonstrated turnaround characteristics after new leadership was installed. She identified the central issue for these schools as the recruitment, retention, and development of high quality teachers.

Rosenholtz (1985) proposed a theory of exchange between inducements and contributions to explain effective urban schools. She argued that the balance between inducements or rewards teachers experienced and the contributions they made determined the decision to teach. The primary reward was the achievement of their students. In poor, urban schools, these rewards were more difficult to achieve and were often outweighed by frustrations. In effective urban schools, principals were focused on student achievement and conveyed a certainty that teachers could improve student performance and that students were capable of learning. This resulted in greater rewards for teachers, lower teacher turnover, and a higher proportion of experienced teachers in the school.

Principals conveyed certainty about achievement goals by orienting a teaching staff around organizational goals (Rosenholtz, 1985). Ineffective urban schools had loosely coupled environments characterized by ambiguous goals, unclear technology, and fluid participation. Goal ambiguity occurred when professed goals for student achievement were contradicted by actual practice. Unclear technology referred to teacher uncertainty about practice, which was often exacerbated by a culture of isolation. Fluid participation was seen in high turnover rates and withholding of services through absenteeism.

In contrast, effective urban schools had practices tightly coupled to organizational goals (Rosenholtz, 1985). Official goals about student skill acquisition were clearly defined, and there was agreement between principals and teachers about goal importance and means of implementation. Principals employed several strategies to achieve

adherence to organizational goals: recruiting like-minded teachers, inducting them into school norms, buffering them to ensure that their efforts are directed toward raising student achievement, providing professional development opportunities to further goals, and monitoring instruction and teacher development.

In the area of teacher recruitment, Rosenholtz (1985) found that principals of effective urban schools were directly involved in the process, screened applicants carefully, and used interviews to articulate school goals and expectations. Indeed, most principals in the effective schools she analyzed were involved in hand-picking a majority of the staff. This is an illustration of the organizational priority proposed by Collins (2001) of “getting the right people on the bus” (p. 41). Selection of teachers was followed by careful monitoring to determine where they fit best, e.g., finding the right seat on the bus; and sanctioning or removing teachers who threatened the organizational norm of contributions and inducements. The success of a school in recruiting and retaining teachers created high visibility for the school; this maintained a synergistic cycle of teacher quality.

Another difference between effective and ineffective urban schools was induction into teaching and socialization to school norms (Rosenholtz, 1985). Ineffective schools were characterized by isolation from professional knowledge and a norm of autonomy. By contrast, effective schools were characterized by collaborative planning, collegial work, intellectual sharing, and norms against malefactors. In collegial versus isolated settings, teachers had more contacts based on their work, more task-related discussion, and a greater belief in the availability of professional knowledge from colleagues. They

were also absent less often and asked for assistance more often. Staff stability and time were required to establish collegial norms. Rosenholtz suggested that principal intervention was important in establishing collegial norms through communication patterns and structures of shared decision making. Shared decision making focused on instructional tasks promoted collegiality by increasing ownership and coordination of the curriculum.

The certainty that instruction increased student learning implied a commitment to protect instruction (Rosenholtz, 1985). Principals of effective urban schools buffered their staff from non-instructional intrusions on teachers by attending to material requirements and the organization of instructional programs, providing clerical assistance for paperwork, mobilizing outside resources to assist teachers with non-teaching tasks, formalizing rules and procedures, and making discipline policies explicit and shared.

The final area in which principals of effective urban schools created certainty was in their supervision of instruction (Rosenholtz, 1985). Principals gave clear criteria for evaluation, monitored instruction frequently, gave feedback, and provided assistance for struggling teachers. Careful monitoring of instruction had symbolic significance for organizational goals. Rosenholtz found that a high degree of variation in achievement between schools could be explained by the combined effects of two components: frequent feedback based on observation, and teacher certainty. Certainty was enhanced through opportunities to meet with colleagues to formulate instructional plans, evaluate student work, and analyze achievement data.

Rosenholtz's (1985) study had two limitations: its findings in urban elementary schools focused on basic skill acquisition may not generalize to different levels of schooling or schools in other settings, and its dependence on case studies did not allow for a full assessment of the strength of the variables identified. In other words, principal effects on student outcomes were inferred by association rather than measured. Since principals are being held accountable for achievement gains in the new accountability environment, it is relevant to understand the nature and extent of their influence on achievement. Evidence from quantitative studies indicated that principals have a small but significant effect on achievement and may influence achievement through indirect means. Although studies were not focused solely on principals of diverse schools, they did investigate the effects of family class background.

Hallinger and Heck (1998) reviewed quantitative studies of the effects of principal beliefs and behavior on school outcomes, usually in the form of student performance. They identified three conceptual approaches. The earliest and most common approach was the *direct effects model*, in which the principal's actions were assumed to influence school outcomes independently of other school variables. The second approach was the *mediated effects model*, in which it was assumed that principals influenced outcomes indirectly through organizational characteristics and people. The third approach was the *reciprocal effects model*, in which principals were assumed to interact with school variables and change their thinking and behavior over time and thereby affect outcomes.

The direct effects model produced no consistent evidence of leadership effects on student outcomes (Hallinger & Heck, 1998). Results of studies based on the model were either mixed or negative. The researchers concluded that although this model tested an empirical relationship, the findings revealed little about how leadership operated. Using a more sophisticated meta-analysis, Witziers, Bosker, and Kruger (2003) detected significant but small direct effects of four principal behaviors on student achievement: (a) defining and communicating a mission, (b) visibility, (c) monitoring, and (d) supervising and evaluating. However, less than 1% of variation in student achievement was associated with differences in these areas.

In contrast, studies based on the mediated-effects model produced consistent evidence of small leadership effects on outcomes (Hallinger & Heck, 1998). Statistical path analysis allowed for the calculation of the strength of indirect effects. Using this approach, Hallinger and Heck were able to detect significant effects in an earlier direct-effects study that had found no effects. To describe how effects are mediated, the authors identified three domains around which to organize significant effects: (a) purposes and goals, (b) structure and social networks, and (c) people. They also distinguished between instructional leadership and transformational leadership orientations. Instructional leadership was focused on tight coupling between achievement goals and practice. Transformational leadership looked beyond curriculum and instruction and sought to increase an organization's capacity to innovate. Hallinger and Heck found that the domain of *purposes and goals* was the strongest for principal influence. In this domain, principal behaviors in the following areas were associated with improved outcomes:

selection and motivation of teachers, framing of instructional goals, framing of transformational goals such as professional development, and achieving staff consensus on school goals.

In the domain of *structure and social networks*, involvement by principals in supporting teachers, creating challenging work, framing mission and vision, and promoting collaborative decision making and parental involvement were associated with improved student outcomes (Hallinger & Heck, 1998). In the domain of *people*, principals positively influenced outcomes by fostering group goals, modeling desired behavior, stimulating staff members intellectually, and providing individual support and recognition. The congruence of Hallinger and Heck's findings with those of Rosenholtz (1985) obtained from case studies is notable.

Antecedent effects of socioeconomic status were noted in the study (Hallinger & Heck, 1998). Principals had indirect effects on reading and math performance in low SES and predominantly African American schools but not in high SES or predominantly White schools. Principals in low SES were more involved in promoting basic skill acquisition. In addition, principals in effective low SES schools acted as buffers to parent involvement, while principals in successful high SES schools selectively encouraged various types of parent involvement.

Although the results of Hallinger and Heck's (1998) study seemed to suggest indirect-effects model studies had greater power to detect leadership effects than direct model studies, indirect-effects studies were rare and have yet to be evaluated through meta-analysis techniques (Witziers, Bosker, & Kruger, 2003). Moreover, reciprocal

effects studies that required more sophisticated methodologies and longitudinal data have not yet been attempted. These areas require further research.

In a follow-up study, Leithwood and Jantzi (2000) again evaluated principal effects on student outcomes quantitatively. As their dependent variable, they designated student engagement with school rather than student performance. They argued that student engagement, while related to performance, was actually an earlier indicator of student success that extended knowledge beyond academic achievement. They described student engagement with school as a combination of behavioral and psychological components. The behavioral component took account of participation in school activities, and the psychological component was focused on how students identified with the school. The authors formulated a “Student Engagement and Family Culture Survey” (p. 6) to measure student engagement with school. They carried over the domains of Hallinger and Heck (1998) and added several more, for a total of eight domains. These domains included purposes and goals; school planning, organizational culture, structure and organization, information collection and decision making; classroom conditions, family educational culture; and the dependent variable, student engagement with school. The authors defined family educational culture as the norms and values held by families about intellectual work and school and specified this variable in lieu of socioeconomic status. In addition to principal influences, the authors investigated the influences of teacher leadership on outcomes.

The researchers found that both principal and teacher leadership had significant but weak relationships with one of the domains, classroom conditions (Leithwood &

Jantzi, 2000). Principal leadership had a stronger relationship than teacher leadership with school conditions and with student identification with school. However, it was a non-school factor, family educational culture, which related most strongly to student engagement and identification with school. Interpreting their findings, the authors commented on the seemingly modest effects of principal leadership. They concluded that even a modest contribution toward student engagement might still be important. Moreover, they found that principal leadership affected school conditions, which in turn was found to affect student engagement. This indicated an additional indirect effect of principal leadership on student engagement.

This section on instructional leadership concludes with a brief discussion of culturally relevant teaching. Researchers in this area take the premise that culturally diverse students pose opportunities instead of problems for teachers, and they assert that teachers promote learning among diverse students when they honor students' different ways of knowing and sources of knowledge (Ladson-Billings, 1994). Some practices of teachers who embraced this perspective included (a) allowing students to speak and write in their own way, (b) using compatible communication styles with students, (c) expressing cultural solidarity with students, (d) sharing power with students, (e) caring for the whole child, and (f) maintaining high expectations for all students (Riehl, 2000). Although the role of administrators has rarely been studied in the literature on culturally relevant teaching, school leaders have promoted and modeled culturally relevant strategies in their interactions with teachers and parents (Riehl, 2000). Shields (2004) envisioned this type of leadership behavior in her concept of dialogic leadership.

Promoting inclusive school practices: molding school cultures that embrace and support diversity. In addition to instructional leadership, molding school cultures was another domain for principals seeking to promote inclusive practices (Riehl, 2000). Culture has been defined as the norms, values, and understandings that are manifested through structures, activities and interactions within a school (Bolman & Deal, 1997). By this definition, when principals engage in the first task of fostering new meanings about diversity, they are already molding inclusive cultures. Although promotion of inclusive cultures by principals has not been studied as much as culturally responsive teaching, and empirical evidence is limited, both literatures agree on a core set of administrative practices, which are quoted here from Riehl:

1. Adopting a personalizing strategy and treating children as individuals rather than as representatives of a social group
2. Appreciating the cultural knowledge that students bring to school and using it to help teachers and principals learn students' cultures
3. Embracing interethnic conflict when it occurs and using it as an opportunity for making positive changes
4. Creating a caring environment and a high level of cooperation among students, teachers, and families
5. Holding high expectations for all students
6. Focusing on academic achievement and providing appropriate supports
7. Reconfiguring school structures through looping and de-tracking to ensure equal and effective access to instruction

8. Encouraging teachers to examine their practices for possible race, class, or gender biases
9. Taking strong steps to work with parents, meeting parents in their homes and work sites, establishing linguistic equity by providing translators whenever needed, and developing parent competencies in leadership and other areas
10. Taking an advocacy approach regarding various forms of discrimination or inequity
11. Maintaining an environment of critique and deconstructing the regularities of practice that serve to dis-empower some persons and groups. (p. 65)

Riehl identified the latter two practices as transformative, because they explicitly addressed social justice.

Building communities. The third task of administrators in diverse schools was to build communities, both within the school and between the school and the outer community. A body of literature developed around concept of *professional learning communities* (DuFour & Eaker, 1998). DuFour et al. (2004) defined three goals for a professional learning community: (a) to identify what it is the students are to learn, (b) to know when they've learned it, and (c) to know how to respond when they do not learn. Thus, this highly instructional model was concerned with coordination of instruction and curriculum, assessment, and intervention. The professional learning community model may be especially relevant for urban schools, where resources for other types of school reform are limited. Administrators assisted in the development of professional learning

communities by helping individual teachers to grow and develop and by creating and sustaining networks of conversation around issues of teaching and learning (Riehl, 2000).

In addition to building up professional communities within the school, administrators of diverse schools had the task of building connections between schools and communities that they serve. Riehl (2000) proposed that schools were embedded within the broader context of community organizations and that they functioned not only to improve the lives of children but also to improve the social fabric of neighborhoods and communities. The task of community building had two parts: (a) mobilizing schools within processes of community development, and (b) working with other organizations to deliver coordinated services to children. In the first part, because schools were embedded within communities, they had the potential to either become targets of activism or positive community partners. Partnership was preferred, since a defensive posture drained energy from the school and its leadership as attempts were made to buffer against outside bids for control.

For the second part of community-building, coordination of services, Riehl (2000) pointed out that students from diverse backgrounds were more in need of these services than their counterparts in less diverse schools. Two problems that could arise were the failure to provide services where they were needed and fragmentation of service delivery. Schools were a logical hub for coordinated services, and principals had an important role in bringing separate organizations together in coordinated efforts. Although empirical evidence on the effectiveness of service coordination still needs to be collected, Riehl suggested that principals committed to service coordination must focus on outcomes and

implementation issues and that they must articulate coordinated services as part of the school mission and accountability system.

School Leaders and Accountability Systems

Research on school leaders' responses to accountability systems is limited. The three empirical studies reviewed here each considered the behaviors and attitudes of principals from different perspectives: as a response to the incentives of a state accountability system (Ladd & Zelli, 2002), as a general response to the importance of data monitoring (West & Macharia, 2002), and in reference to accountability practices identified as effective in the research literature (Englert et al., 2004). Findings indicated that carefully constructed accountability systems had a powerful influence on the behavior of principals; that principals became increasingly involved in monitoring and interpreting data; and that principals in high-performing or improving schools were more likely to use elements of effective accountability systems.

Responses of leaders to accountability incentives. The effects of a well-defined state accountability system on principal behavior were investigated by Ladd and Zelli (2002). The *ABC's* system of North Carolina was notable for its value-added component and for strong observed gains in student achievement. This system was designed to create incentives to increase student performance in basic reading, writing, and mathematics skills and to focus attention and resources on low-performing students. In support of these incentives, the program gave principals added discretion on budgetary transfers, class size limits, and choice of textbooks. The program also provided

assistance to low-performing schools and reward bonuses to teachers of schools that met expected targets.

Ladd and Zelli (2002) framed school leaders' roles in North Carolina's system within a theoretical framework known as the *principal agent model*. In this model, a *principal* (the state policy maker) hired agents (school principals) to perform work to raise student achievement. The model assumed that agents might not share policy makers' values and wouldn't work to promote their goals unless given appropriate incentives. Because policy makers couldn't observe school principals directly, they implemented rewards and sanctions linked to the desired outcomes. The model was subject to two constraints: first, the agents' jobs must be sufficiently attractive to induce them to participate; and second, incentives must be compatible with desired outcomes. Factors that complicated the model included the fact that some outcomes were easier than others to achieve; that agents didn't have full control over factors that affect outcomes, such as school funding and racial isolation; and that school principals coexisted with other agents within a hierarchical structure.

With this framework, Ladd and Zelli (2002) conducted structured interviews with elementary principals focused on several areas of the *ABC's* accountability program and followed up with these principals two years later. Of greatest interest to the present study are the reported effects of the system on principal actions. The researchers found that high percentages of principals focused their energy and attention to instructional leadership activities consistent with the desired outcomes of the program. These included holding grade level meetings with teachers to prepare for state assessments; using state

assessment scores as a diagnostic tool to help teachers improve instruction; spending more time with teachers in classroom observations; creating extracurricular programs focused on math and reading; and increasing the integration of math and reading across the curriculum. In addition, the level of participation by principals in many of these activities increased significantly in the two year interval over which the study was conducted. This led the authors to conclude that the *ABC's* program' incentives were effective in influencing the behavior of school leaders toward desired program outcomes.

Other findings reinforced this conclusion. In the initial round of interviews, principals who personally supported the *ABC's* program reported stronger behavioral responses than principals who did not support the program (Ladd & Zelli, 2002). However, in the follow-up interviews two years later, behavior differences between the two groups had vanished. Apparently, the program's incentives were powerful regardless of the personal beliefs of principals. In addition, principals reported shifting efforts and resources to low-performing students. This finding may be a result of the value-added component of the accountability program, which monitored the percentage of students in a school achieving one year's growth.

Other effects of North Carolina's program were not unambiguously positive. Principals reported an increasing tendency to redirect funds to math and reading from other subject areas and to focus resources on test-taking skills (Ladd & Zelli, 2002). These actions could be interpreted as narrowing the curriculum and neglecting some important school goals and subjects in favor of those that are rewarded. In addition, there were differential response patterns between principals of high and low SES schools and

between principals of high and low-performing schools. Principals of low SES and low-performing schools were the least supportive overall of the ABC's program. The authors conjectured that this finding was due to the greater likelihood that low SES and low-performing schools failed to make their expected achievement growth targets. Moreover, principals of low SES schools perceived they had less power to remove poor teachers than principals of high SES schools. The authors reasoned that low-quality teachers removed from high performing and high SES schools may end up concentrated in low performing and low SES schools. Supporting this hypothesis was the finding that principals from low-performing and low SES schools complained of the difficulty of recruiting and retaining high quality teachers. Due to this frustrating dynamic and the awarding of bonuses to employees of high-achieving schools, the accountability system may have carried a built-in perverse incentive for high quality teachers and principals to leave schools with low-performing students, with the result that these students were disadvantaged relative to their peers.

In view of the powerful effects of *ABC's* incentives on principal behavior and their possible counterproductive outcomes, Ladd and Zelli (2002) suggested that these incentives should be used cautiously. They recommended that incentives in the form of bonuses be kept small, since they are based on imperfect measures. Additional support for this recommendation derives from the previously discussed susceptibility of value-added systems to statistical noise. Another recommendation, not mentioned by the authors, would be to build incentives into the program for high-quality teachers and administrators to work at low-performing and low-SES schools.

Use of data. As this review has demonstrated, data tracking is a significant activity within accountability systems. No Child Left Behind (U.S. DOE, 2002) and state systems required the reporting of achievement data that were disaggregated in several ways. Evaluation of data and schools can follow several models: performance levels, growth targets, student gain, or combinations of these. West and Macharia (2002) investigated attitudes of principals and teachers in Tennessee about working with data. Through interviews consisting of open-ended questions, the researchers found widespread agreement that principals needed to be skillful translators and interpreters of data for their staffs and the public. Another finding was that principals required knowledge and skills to use data to improve teaching and learning. The study was conducted shortly after the passage of NCLB. As implementation of NCLB and state accountability systems have continued, further research is needed on how principal attitudes toward data have evolved.

Principals reported that as translators and interpreters of data, they must eliminate jargon and proceed in a discreet and sensitive manner, taking different approaches for school-wide and classroom level data (West & Macharia, 2002). Differences were noted in principal approaches to data presentation: some principals took primary responsibility for presenting data to staff, while others involved teachers in grappling with data to build teacher ownership. Principals placed importance on celebrating positive data and structuring constructive approaches to less positive data.

Using data to improve teaching and learning was an important priority identified by principals (West & Macharia, 2002). The primary way principals accomplished this

was by forming and promoting small teams of teachers that interpreted disaggregated data and diagnosed instructional needs. Collegial relationships within these teams assisted the process. Other data-related leadership processes for the improvement of instruction included curriculum mapping, in which standards and instruction were aligned; the promotion of differentiated instruction tailored to individual student needs; and the use of data to determine school improvement, funding priorities, and staff development priorities.

Leader behaviors and effective accountability practices. In a cross-state survey of principals using closed and open-ended questions, Englert and colleagues (2004) sought to align principal responses in hi-performing and low-performing schools with areas of effective accountability practice identified in the research literature. They rated schools based on a composite of principal perceptions of overall school proficiency, improvement on state assessments, achievement differences between subgroups, and other achievement measures. Schools above the composite mean were designated *high-proficient* and those below the mean were designated *low-proficient*. Their seven research based practices for effective accountability systems have been listed earlier in this review. Consistent with the findings of West and Macharia (2002), principals in this study reported using data to a large extent and encouraging their staff to do so to improve instruction. Principals also expressed a desire to implement data-driven school improvement more effectively and cited a lack of time and resources as constraints. In addition, the authors found that principals of high-proficient schools were more likely to report the use of effective accountability practices than principals of low-proficient schools.

Among the seven accountability measures, principals responded most strongly to survey items related to holding high expectations for all students (Englert et al., 2004). This category was followed in order of importance by data usage for diagnostic purposes; high-quality assessments aligned with standards; and alignment of resources, support and assistance for improvement. It is important to note that some of these categories were more within the control of principals than others; for example, some standardized assessments were a state responsibility.

Englert et al. (2004) found differences between responses of principals in high- and low-proficient schools. Principals of high-proficient schools had higher response rates in the areas of high expectations for all students, alignment of resources for improvement, data usage, and high-quality assessments. In contrast, low-proficient schools had higher percentages of low SES and students of color. The possible confounding effects of these antecedent variables made it impossible to determine whether the observed differences in principal attitudes were related to actual differences in performance. However, principals of low SES schools who reported their schools were improving were also more likely to report the use of effective accountability practices.

The three studies on principal responses to accountability systems reviewed here consistently indicated that accountability systems had an important and increasing influence on the behaviors of principals. Although each of the studies took a particular perspective and obtained findings specific to its research objective, the findings overlapped significantly and were consistent with each other. However, none of the

studies focused specifically on issues of educational equity as they relate to principals in the accountability environment. It was the goal of the present study to examine the relationships between principal actions and beliefs, accountability systems, and the demands of educational equity.

CHAPTER 3

METHODS

This chapter describes the methods used to carry out the study. Schools were grouped into four socioeconomic strata based on free and reduced lunch percentage. Within each stratified group, a school was selected that demonstrated high student performance and equity of performance among its socioeconomic peers. The researcher interviewed principals and instructional coaches from the selected schools and analyzed the interview transcripts using two methods: a quantitative word count and a qualitative grounded theory approach. The researcher also evaluated published performance data of elementary schools in the district for educational equity using the equity achievement driver.

The Research Context

The study was conducted in a large urban district that served over 60,000 students in Arizona. The aggregate socioeconomic status of students in the district was characterized as follows: 24% of students were in poverty and 56% received free and reduced lunch. The racial and ethnic composition of the district was 7% African American, 3% Asian American, 53% Latino, 4% Native American, and 33% White. Desegregation orders dating from the 1970s affected 36 of the district's 117 schools. The district was in the process of petitioning the courts for unitary status, e.g., to be released from the desegregation orders.

Funding of Arizona public schools ranked 41st in the United States in fiscal year 2005 at a rate of \$7,538 per pupil compared to a national average of \$8,061 per pupil (Zhou &

Gaviola, 2007). The district in which this study was conducted made expenditures of \$7,140 per pupil for the same year. Therefore, per pupil funding in this district was below the statewide average in a state that was in the lower tenth of the national distribution of funding.

The district as an entity was in improvement plan status under No Child Left Behind (U.S. DOE, 2002), due to the failure of its special education students and English Language Learners to meet performance testing objectives. The district's 81 elementary schools in 2005-2006 earned school ratings under Arizona Learns (ADE, 2005) in the following numbers: excelling (7), highly performing (10), performing plus (35), and performing (29). None of the schools was rated as underperforming or failing.

In this study, elementary schools in the district were separated into four socioeconomic groups. The variable used to identify socioeconomic levels was the percentage of students in a school receiving free or reduced lunch. Among elementary schools in the district, Group 1 consisted of five schools for which the free and reduced lunch percentage was 25% or less, Group 2 consisted of 13 schools for which the percentage was between 25% and 50%, Group 3 consisted 24 schools for which the percentage was between 50% and 75%, and Group 4 consists of 29 schools for which the percentage was between 75% and 100%. The terminal year for most district elementary schools is fifth grade; thus, performance of fifth graders on the state-mandated test were taken as a summative measure of performance for the school as a whole.

Figure 2 shows the performance on the Arizona Instrument to Measure Standards (AIMS) in reading of fifth graders for the 2005-2006 school year in each of the four

socioeconomic groups. It can be seen that the average score was highest for the Group 1, the highest socioeconomic group, and decreased for each successive group. Although the

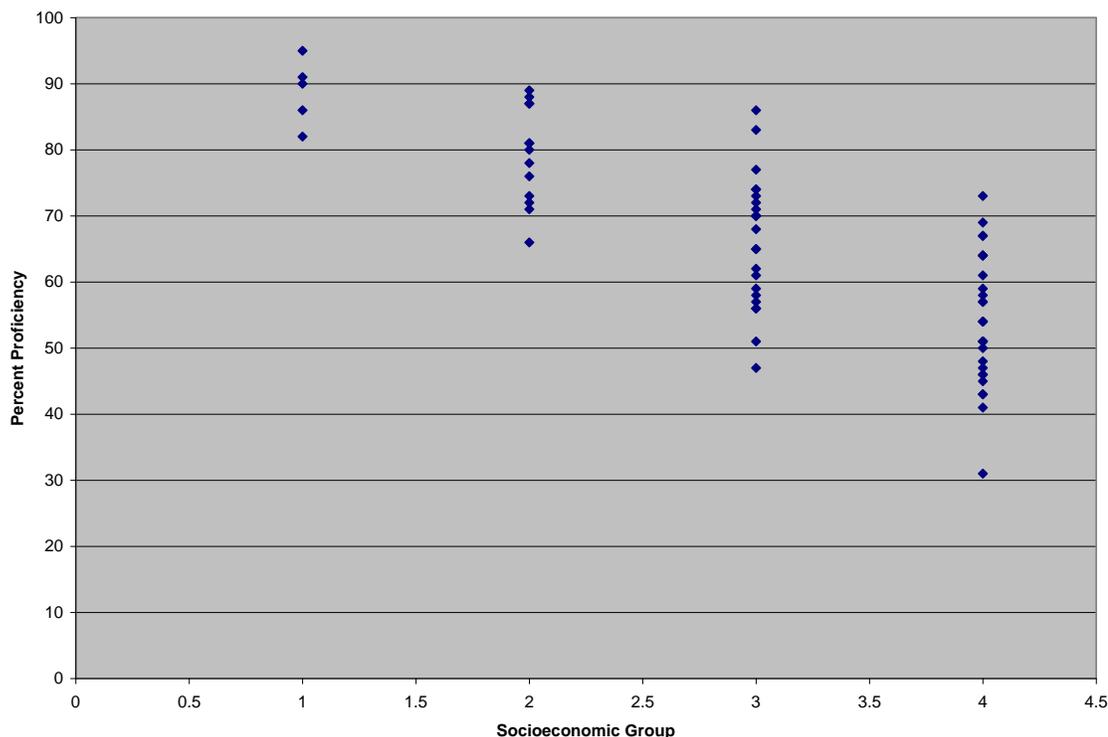


Figure 2. Reading performance of fifth graders on the Arizona Instrument to Measure Standards for elementary schools in four socioeconomic groups within the district in 2006. Points represent the percentage of students meeting or exceeding proficiency standards.

correlation of achievement and socioeconomic status was consistent with other studies (Willie et al., 2002), what was of interest for this study was the large amount of variation of performance among schools within specific socioeconomic groups of schools. This variation created significant overlap between the groups, such that some schools in Group

3, the second poorest group of schools, had reading performance that exceeded the performance of some schools in Group 1, the highest SES group. In Figures 3 and 4, which show math and writing performance respectively for fifth graders in 2005-2006, the same pattern is observed with even greater variation, to the extent that there was overlap between the performance of schools in the richest and poorest groups of schools.

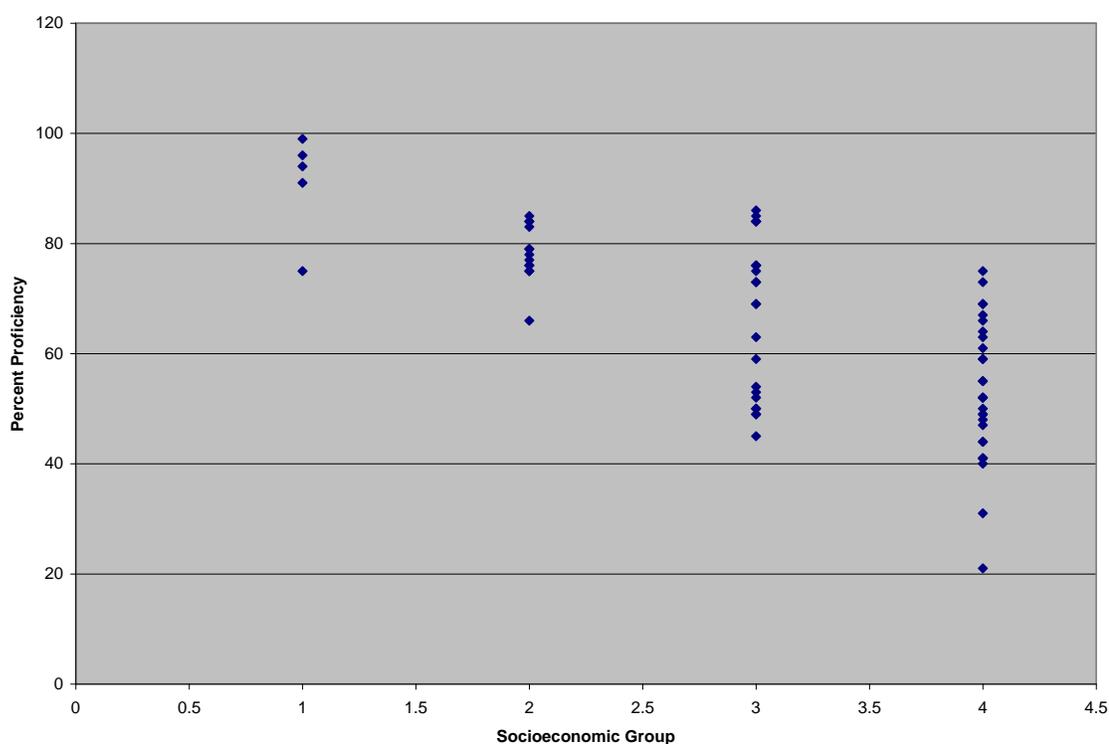


Figure 3. Math performance of fifth graders on the Arizona Instrument to Measure Standards for elementary schools in four socioeconomic groups within the district in 2006. Points represent the percentage of students meeting or exceeding proficiency standards.

The large amount of variation in performance among schools in each subgroup suggested that a school's socioeconomic profile influenced but did not determine student achievement and that performance may have been leveraged by other factors that counteracted the influence of socioeconomic status. In this study, the factor that was investigated as influencing overall performance and educational equity was leadership.

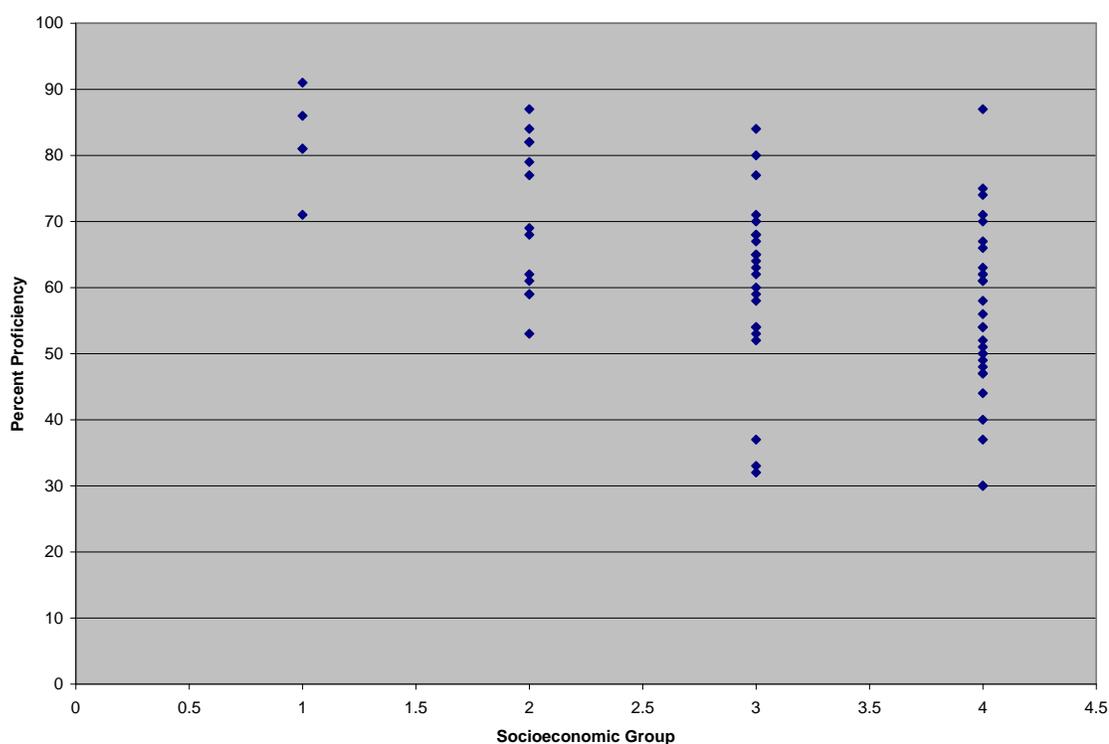


Figure 4. Writing performance of fifth graders on the Arizona Instrument to Measure Standards for elementary schools in four socioeconomic groups within the district in 2006. Points represent the percentage of students meeting or exceeding proficiency standards.

The Research Participants

Elementary schools in each of four socioeconomic subgroups were ranked in two ways: (a) performance of fifth grade students on the 2005-2006 Arizona Instrument to Measure Standards (AIMS) in the subject areas of Reading, Writing, and Math and (b) a measure of educational equity called the *equity achievement driver*, which is described below. From each subgroup, a school was selected that ranked highly in both of these measures according to the following selection criteria. Each selected school had scores on the AIMS that were in the top third for their socioeconomic level for all three subtests. The schools also demonstrated educational equity by having small achievement gaps on the AIMS. When the equity achievement driver was applied to the disaggregated test data, each selected school lost no more than one position of rank in each of the subtests, and on several of the subtests the selected schools improved their rank.

A detailed description of each school follows that includes the following elements: student demographics, educational programs, national and state accountability system status, and leadership organization. Additional information about each school was gathered during the interview process. Participants in the study were the principals and instructional coaches at the selected schools. Principals were chosen because they are the primary school leaders whose attitudes and actions are assumed to influence school culture, vision, and direction. Instructional coaches embody a new type of *teacher-leader* deeply involved in instructional leadership issues. As such, they can provide valuable insight, evidence, and triangulation to the responses of principals. The district under study took an initiative in 2004 to provide each of its schools with an

instructional coach. A detailed description of each participant follows. The role of the researcher was that of an interviewer.

Interviews

Interviews with principals and instructional coaches were structured according to the model described by Seidman (2006) and also used elements of the focused interview described by Merton, Fiske, and Kendall (1990). Seidman's procedure consisted of three interview phases: (a) a life history of the participant as it related to the topics under consideration, (b) an in-depth discussion of the participant's involvement with the topics, and (c) a reflection. In this study, the first phase focused on leaders' beliefs about educational equity; the second phase focused on leaders' beliefs about and responses to the new accountability environment; and in the third phase, participants were asked to reflect on the interrelationships between the accountability environment and educational equity.

Focused interviewing (Merton et al., 1990) seeks to assess the experience of participants rather than to monitor their opinions. It is appropriate when participants have been involved in a particular situation that the researcher has investigated and developed tentative hypotheses about. The researcher develops an interview guide and conducts the interview with the goal of evaluating the subjective experiences of the participants regarding the situation. Although it is more structured than exploratory interviewing, focused interviewing is open to departures from the interview guide and to new hypotheses that emerge from participant responses. It also avoids forcing topics on

participants that are not relevant to participants' experience. The goal is to uncover diverse responses that may have been unanticipated by the researcher.

Interview questions were vetted by a group of scholars consisting of a professor of psychology and two doctoral candidates in urban superintendence. Suggestions of the group led to revisions in the interview guide. Prior to interviews, the protocol was piloted with a central office administrator from the district under study.

School and Participant Descriptions

Schools in the study were given fictitious names for purposes of anonymity. The participating principals and instructional coaches were identified throughout with their school's fictitious names. The selected schools and their participants are now described.

Group 1: Whitcomb Elementary

The selected school from Group 1, the highest SES subgroup, was Whitcomb Elementary. Whitcomb had an enrollment of 359 students and an attendance rate of 96% in 2005-2006. On the 100th day of that school year, Whitcomb's demographics included the following features: 65.5% White students, 20.6% Latino, 6.4% African American, 6.4% Asian American, 1.1% Native American, and 18.4% of students receiving free or reduced lunch. The student to teacher ratio was 22.4/1, compared to 20.7/1 for the district as a whole. *Mobility* is a measure of how many students transfer in and out of a school; high values of this measure tend to correlate with lower performance. *Stability* is a measure of how much of the school year a student spends at a school and is negatively correlated with mobility. Whitcomb's mobility was 23% in

2005-2006, compared to 37% for the district; and its stability was 90%, compared to 86% for the district. Whitcomb therefore had mobility and stability rates that favored higher performance.

For the second consecutive year, Whitcomb earned the highest possible rating under the Arizona Learns (ADE, 2005) accountability system in 2005-2006: *excelling*. In 2004 and 2003, the school had earned the next highest rating of *highly performing*. Whitcomb made Adequate Yearly Progress under NCLB (U.S. DOE, 2002) in 2005-2006. In a *weighted combination of achievement measures* calculated by the district's accountability department, Whitcomb earned a z-score of four on a scale from one to six when current status and growth were weighted equally for various achievement measures. By comparison, 40% of the elementary schools in the district earned z-scores of four or above. One of the reasons for this somewhat average z-score was that Whitcomb's high scores did not leave it much room to improve, and gains at the upper levels became increasingly difficult.

The AIMS scores of Whitcomb's fifth graders in 2005-2006 were among the highest in the district. State-mandated benchmarks were met by 96% of those students in math, 90% in reading, and 86% in writing. These figures compared with 64%, 61%, and 64% respectively for the district as a whole. Among its peer schools in Group One, which had free and reduced lunch populations of less than 25%, Whitcomb's performance was in the upper third of the score range for Group 1 in all three subject areas. In addition, when AIMS scores of Whitcomb's fifth graders were disaggregated by race/ethnicity and evaluated for educational equity using the equity achievement

driver, Whitcomb retained its position among its peers. The school accomplished this while having the highest free/reduced lunch percentage among Group One schools and the highest percentage of students of color. Although three other schools in this elite group of five had impressive achievements and could have been selected, the factors cited here gave Whitcomb the edge.

Entering Whitcomb, a visitor had an impression of a small school in an ordinary building that was very clean and well-kept. Colorful student work adorned the hallway walls. One was immediately greeted by friendly adults and children; however, it was clear that the adults are on the lookout and very protective of the children. Many parents were in the building, and they communicated comfortably with teachers and staff.

Whitcomb's principal was White and had been in her position for seven years. The instructional coach was also White and had served in that capacity for two years, since the instructional coach program was initiated by the district. Prior to that, she worked at the school for 16 years as a teacher. Because of the small school size, there was no assistant principal or other administrator at the school. On the annual school quality survey administered by the district, the principal was rated an effective leader by 100% of the staff participating in the survey, making her the only principal in this study to receive unanimous endorsement.

Whitcomb's principal made a personable, serious impression, but she also displayed a sense of humor. Her interview responses were detailed, thoughtful and measured. Themes that she repeatedly returned to in regard to her work and role were the motivation to do what's right, to make a difference, and to focus on each child as an

individual. She commented often that she didn't ask any employee to do something she herself was not willing to do, including teaching classes and cleaning bathrooms. More will be said about Whitcomb's principal in Chapter 4.

Like her principal, Whitcomb's instructional coach held high expectations of herself as a leader and had a singular focus on children as unique individuals. She repeatedly expressed her admiration and support for the teachers at Whitcomb. This coach's responses focused more on technical areas of testing than her principal's responses, and she was somewhat more skeptical about testing and the accountability systems that mandated it. More will be reported about Whitcomb's instructional coach in Chapter 4.

Group 2: Desert Springs Elementary

Desert Springs Elementary was a somewhat larger school; it had an enrollment of 501 students and an attendance rate of 94.4% in 2005-2006. The 100th day demographics for the 2005-2006 school year at Desert Springs were 55.7% Latino students, 34.7% White, 4.4% African American, 4.0% Asian American, 1.2% Native American, and 33% of students receiving free or reduced lunch. The student to teacher ratio was 24.5/1. Mobility was 19%, and stability was 93%; both of these measures exceeded district averages and favored higher student performance.

Desert Springs earned an *excelling* rating under Arizona Learns (ADE, 2005) in 2005-2006. It was one of three schools in this socioeconomic group and one of seven schools district wide to earn that rating. The previous year its rating was *highly performing*, and the year before that it was *performing*. Thus, Desert Springs' rise was

dramatic. The school achieved Adequate Yearly Progress under NCLB in 2005-2006. In the district-generated weighted combination of achievement measures, Desert Springs earned the highest possible z-score of six.

On the AIMS test, fifth graders at Desert Springs mastered state standards at the rate of 89% for Reading, 82% for Writing, and 85% for Math, strongly outperforming the district figures of 61%, 64%, and 64%, respectively. These results placed the school at or near the top of its socioeconomic group in all three tested areas. When fifth grade AIMS scores were disaggregated by race/ethnicity and subjected to the equity achievement driver to evaluate educational equity, the school retained its high rank in all tested areas. These performance measures combined to make Desert Springs a stand-out choice for the study.

In the district's school quality survey, 94% of participating staff members expressed their overall satisfaction with Desert Springs. However, there were several areas in the survey in which over 20% of respondents to the survey registered disagreement. Those areas included student access to computers, adequate rewards and recognition for teachers, quality of school council communications, teacher involvement in school decisions, principal leadership, instructional leadership, and the district textbook adoption.

A visitor to Desert Springs encountered a large complex that was thirteen years old and still appeared new and modern, with sky lights and impressive pitched roofs. Student artwork was displayed in the office area. The principal of Desert Springs was White and worked at Desert Springs for six years before taking another position after the

2005-2006 school year. On the district-administered school quality survey, she was rated as an effective leader by 57% of the staff who participated; this was the lowest endorsement among the four principals in the study. Desert Springs' principal was a highly energetic individual who held high expectations for herself and others and seemed driven to achieve. She related that on Saturday mornings she would take long walks with a principal from another high-achieving school and share ideas for school improvement. She continually asked herself and others, "Have we done all we can do?" when referring to children's learning. She was child-centered and spoke of the need to build relationships with individual students centered on their work. She spoke quickly, as if impatient for results, described her approach as "data-driven," and displayed a greater confidence and understanding regarding national and state accountability systems than other participants in the study. More will be reported about Desert Springs' principal in Chapter 4.

The instructional coach position at Desert Springs was shared by two women who were interviewed together. The first coach was White and had worked as instructional coach at Desert Springs for two years in a four-fifths time position. The second coach was White, worked as an instructional coach one-fifth of the time, and taught fourth graders the rest of the time. She earned a Ph.D. in Education with an emphasis on student motivation. The two coaches had a close relationship since the first coach began her teaching career as the second coach's student teacher. When the researcher met them for the initial interview, the two coaches were concluding a training session with several teachers after contract hours on a district-approved methodology based on the book

