

**DEFINING SCHOOL SUCCESS:  
EDUCATIONAL LEADERS' PERCEPTIONS OF THE  
COMPLEX WORLD OF SCHOOL ACCOUNTABILITY**

---

A Dissertation  
Submitted to  
The Temple University Graduate Board

---

In Partial Fulfillment  
Of the Requirements for the Degree  
DOCTOR OF EDUCATION

---

by  
Michael R. Gogoj  
May 2024

Examining Committee Members:

Judith C. Stull, Ph.D., Advisory Chair

Professor, Department of Policy, Organizational, and Leadership Studies

Christopher W. McGinley, Ed.D.

Professor, Department of Policy, Organizational, and Leadership Studies

Richard M. Englert, Ed.D.

University President and Professor, Department of Policy, Organizational, and Leadership Studies

Joseph P. DuCette, Ph.D.

Professor (ret.), Department of Psychological Studies in Education

©  
Copyright  
2024

By

Michael R. Gogoj  
All Rights Reserved

## ABSTRACT

School accountability is a powerful force in education. Today, schools are within the early years of the *Every Student Succeeds Act (ESSA)* which has introduced a new accountability model to states and districts, in a post-pandemic era in which schools' roles and responsibilities within the larger society have continued to expand, and at the epicenter of complex, and sometimes tumultuous, social and cultural forces. It is within this context that the present study investigates the perceptions of the leaders of 30 Pennsylvania school districts across the continuum of economic disadvantage. It expands on previous research and literature that explored the benefits and challenges of mandated accountability, and that, most recently, has sought to explain how stakeholders understand and make sense of measuring performance. Specifically, this study examines district leaders' perceptions of mandated accountability, the extent to which they value various mandated and unmandated performance measures, and the strategies they use to communicate accountability to their stakeholders. Findings suggest that (1) district leaders' perceptions of accountability are complex and nuanced, (2) that they both believe in the importance of public-facing accountability and, simultaneously, raise concerns about its limitations and potential misrepresentation of schools (especially disadvantaged schools), (3) that they value many of the existing performance measures while also desiring additional, more robust measures, and (4) that individual leaders' approaches to interpreting and communicating accountability measures are not uniform, but are driven by each leader's personal beliefs as well as community interests.

## **DEDICATION**

This work is dedicated to my wife, Ashley, and my children, Evan and Emma. Ashley, on every step of this journey you have given me the gifts of time, support, and encouragement. You helped me to think through my ideas, you read my work, and you provided invaluable insights and wisdom. More than anything, you have been my biggest cheerleader over these three years, and I could not have accomplished this without you by my side. Evan and Emma, you understood that Dad was working toward a goal, and it required time and commitment. You demonstrated maturity, patience, and support in so many ways, and for that I am grateful. I hope that this accomplishment makes you as proud of me as I am of all of you. I love you.

## ACKNOWLEDGEMENTS

It has been a privilege to pursue my doctoral degree from Temple University. Our program of study was enriching, our instructors dedicated, and our cohort an engaged group of professionals with whom I look forward to a lifelong connection.

My deepest appreciation to Dr. Judith Stull, my dissertation chairperson. Dr. Stull is one of the most committed and passionate instructors with whom I've had the pleasure to work. She gave generously of her time, knowledge, and pushed me – and so many of us – to continuously improve our craft. Thanks also to Dr. Christopher McGinley, Dr. Richard Englert, and Dr. Joseph DuCette, who graciously agreed to serve on my committee. Your thoughtful questions, insights, and recommendations made my work better. Thank you, all.

I am truly grateful for the district leaders who volunteered to participate in this study. Though I cannot mention you by name, this endeavor would have been impossible without your time as well as your thought-provoking and insightful perspectives. Thank you for your dedication and commitment to our profession, and I hope that we continue to cross paths in the coming years.

Finally, I could not have accomplished this goal without the support and encouragement of my family and friends. To my parents, who instilled the importance of education and hard work at an early age, my entire family, my friends, and my colleagues who supported me every step of the way. You gave me encouragement and took the time to ask about my work and progress. Even as I saw your eyes glaze over a little bit as I talked about my research, I know that you cared and were in my corner! Thank you!

## TABLE OF CONTENTS

	Page
ABSTRACT .....	iii
DEDICATION.....	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS .....	vi
LIST OF TABLES.....	xi
LIST OF FIGURES .....	xiii
CHAPTER ONE: INTRODUCTION.....	1
A Brief History of School Accountability.....	2
The Elementary and Secondary Education Act (1965) .....	2
The Coleman Report (1966).....	3
A Nation at Risk (1983) .....	5
The 1980s and 1990s: An “Age of Accountability” in Education .....	6
The No Child Left Behind Act of 2001 .....	8
The Every Student Succeeds Act of 2015 (ESSA).....	11
The Central Issues of the Present Study .....	13
CHAPTER TWO: LITERATURE REVIEW.....	24
Underlying Theoretical Perspectives.....	24
Consequential Accountability .....	24
The Principal-Agent Relationship .....	25
Loose Coupling .....	26
Empirical Research and Relevant Literature.....	28

Conceptualizing Accountability in Education.....	28
The NCLB Era of Standardized Testing and High-Stakes Accountability .....	31
Stakeholder Responses to Federal and State Accountability .....	44
Contemporary Literature on the Design of Accountability Systems.....	53
Research Questions .....	57
CHAPTER THREE: METHODOLOGY .....	58
Introduction to the Methodology.....	58
Sample.....	60
Purposeful Identification of Districts .....	60
Recruitment .....	60
Characteristics of the Sample .....	61
The Sample in Relation to the State of Pennsylvania.....	61
The Sample Broken into Tertiles of Economic Disadvantage .....	64
School District Contextual Data.....	65
Participant Contextual Data.....	69
Benefits and Limitations of the Sample .....	73
Data Issues Related to School Measurement and Accountability.....	74
Operationalization and Explanation of Variables.....	74
How Pennsylvania Measures Schools: The Future Ready PA Index.....	77
Research Questions, Survey Design, and Interview Design .....	81
Methodological Procedures.....	85
Data Analysis .....	86
Selected-Response Questions.....	86

Open-Ended Questions .....	88
CHAPTER FOUR: FINDINGS .....	89
Research Question 1: What are Pennsylvania district leaders’ perceptions of mandated accountability? What is the relationship between district contextual attributes and their perceptions? .....	91
Participants’ Positions on Mandated Accountability .....	91
Quantitative Analysis of Participants’ Perceptions of Accountability .....	93
Emergent Themes from Qualitative Data on Leaders’ Perceptions .....	103
Research Question 2: To what extent do Pennsylvania district leaders value various mandated and unmandated (alternative or additional) public-facing accountability measures? What is the relationship between district contextual attributes and the measures they value? .....	110
Quantitative Review of the Extent to Which Leaders Value Measures .....	110
Other Measures Valued by Participants .....	118
Research Question 3: How do Pennsylvania district leaders communicate mandated and/or unmandated (alternative or additional) accountability measures to their stakeholders? What is the relationship between district contextual attributes and their communication behaviors? .....	118
Frequency of Communication of Accountability Measures to Various Stakeholders .....	119
Frequency of Questions on Accountability from Stakeholders .....	121
Methods of Communication on Accountability .....	122
Benefits and Challenges of Publicly Communicating School Accountability .....	122
Practices Employed to Effectively Communicate Accountability .....	125
Conclusion .....	127
CHAPTER FIVE: DISCUSSION .....	130
Findings .....	130

Limitations .....	132
Implications and Recommendations .....	134
Policy.....	134
Leadership .....	138
Recommendations for Future Research .....	141
Conclusion.....	142
REFERENCES .....	144
APPENDICES	
APPENDIX A: INTERVIEW PROTOCOL .....	154
APPENDIX B: SAMPLE TERTILES VS. PENNSYLVANIA TERTILES .....	164
APPENDIX C: PARTICIPANTS' CONTEXTUAL INFORMATION.....	167
APPENDIX D: PARTICIPANTS PERCEPTIONS OF VARIOUS ASPECTS OF MANDATED ACCOUNTABILITY AND THE FUTURE READY PA INDEX (BY TERTILE).....	168
APPENDIX E: MEANS DENOTING DISTRIBUTION OF PARTICIPANTS' PERCEPTIONS OF MANDATED ACCOUNTABILITY (BY TERTILE) .....	169
APPENDIX F: PARTICIPANTS' GENERAL PERCEPTIONS OF MANDATED ACCOUNTABILITY (BY TERTILE).....	170
APPENDIX G: PARTICIPANTS' PERCEPTIONS OF WAYS MANDATED ACCOUNTABILITY ADVANCES OR HINDERS THEIR EDUCATIONAL OBJECTIVES (BY TERTILE).....	171
APPENDIX H: NUMBER OF PARTICIPANTS WHO VALUE VARIOUS MANDATED AND UNMANDATED (ALTERNATE OR ADDITIONAL) ACCOUNTABILITY MEASURES (BY TERTILE) .....	172
APPENDIX I: MEANS DENOTING DISTRIBUTION OF NUMBER OF PARTICIPANTS WHO VALUE VARIOUS MANDATED AND UNMANDATED (ALTERNATE OR ADDITIONAL) ACCOUNTABILITY MEASURES (BY TERTILE).....	173

APPENDIX J: DISTRICT LEADERS' IDENTIFICATION OF OTHER VALUABLE ACCOUNTABILITY MEASURES (BY TERTILE)..... 174

APPENDIX K: MEANS DENOTING DISTRIBUTION OF PARTICIPANTS' RESPONSES ON QUESTIONS RELATED TO COMMUNICATION OF MANDATED ACCOUNTABILITY (BY TERTILE)..... 175

## LIST OF TABLES

		Page
3.1	Sampled Districts' vs. All PA Districts' Contextual Data .....	63
3.2	Approximate Range of Economic Disadvantage within Each Tertile .....	64
3.3	Crosstabulation of ED Tertiles and Various District Characteristics .....	65
3.4	Crosstabulation of ED Tertiles and Various Student Characteristics .....	66
3.5	Crosstabulation of ED Tertiles and Financial Data .....	68
3.6	Crosstabulation of ED Tertiles and District Locale .....	68
3.7	Crosstabulation of ED Tertiles and District Leaders' Roles.....	70
3.8	Crosstabulation of ED Tertiles and District Leaders' Gender .....	70
3.9	Crosstabulation of ED Tertiles and District Leaders' Age .....	71
3.10	Crosstabulation of ED Tertiles and District Leaders' Race/Ethnicity .....	71
3.11	Crosstabulation of ED Tertiles and District Leaders' Highest Degree Earned .....	71
3.12	Crosstabulation of ED Tertiles and District Leaders' Years in Current Position .....	72
3.13	Crosstabulation of ED Tertiles and District Leaders' Years in Any District Level Position .....	72
3.14	Crosstabulation of ED Tertiles and District Leaders' Years in Public Education .....	72
3.15	Crosstabulation of ED Tertiles and District Leaders' Years as a Pennsylvania Educator.....	73
3.16	Pennsylvania Future Ready PA Index Accountability Indicators.....	80
3.17	Research Questions, Connections to Literature, Operationalization, and Associated Interview Questions.....	82
4.1	Data and Findings Concepts .....	90
4.2	Participants' Positions on Mandated Accountability Measures (By Tertile).....	93

4.3	Participants' Frequency of Communication with Various Stakeholders (By Tertile).....	120
4.4	Frequency of Questions or Concerns Received from Public Stakeholders (By Tertile) .....	121
4.5	District Leaders' Strategies for Communicating with Stakeholders (By Tertile).....	122
4.6	Participants' Perceptions of the Benefits and Challenges of the Public Communication of Accountability (By Tertile).....	125
4.7	Participants' Perceptions of Important Practices When Communicating Accountability (By Tertile).....	126

## LIST OF FIGURES

		Page
2.1	Conceptual Model: The Implementation of Accountability Through Multiple Layers .....	28
4.1	Extent to Which Leaders Expressed Understanding of the Measures in the Future Ready PA Index .....	96
4.2	Extent to Which Leaders Expressed That They Value and Utilize the Measures in the Future Ready PA Index .....	98
4.3	Extent to Which Leaders Expressed Confidence in the Data Reported in the Future Ready PA Index .....	100
4.4	Extent to Which Leaders Believed the Future Ready PA Index Has Engendered Positive Outcomes .....	102
4.5	Extent to Which Leaders Believed Schools Have the Needed Resources/ Capacity to Meet Expectations .....	103
4.6	Number of Participants Who Highly Valued Various Mandated and Unmandated Accountability Measures .....	112

## CHAPTER ONE: INTRODUCTION

The public school represents a critical cornerstone in American democracy. The mission of the U.S. Department of Education, formerly a part of the Department of Health Education and Wellness and established by Congress as its own entity in 1980, is “to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access” (U.S. Department of Education, 2024a, para. 1). As public, taxpayer-funded institutions, schools are expected to equitably and effectively serve all students to ensure that they achieve academic success and develop social behaviors and attitudes that both enrich their own lives and serve the greater public good. It naturally follows, then, that school districts, schools, and individuals are held accountable for their performance by evaluating the achievement and other outcomes of the students whom they serve. However, educational accountability has proven to be an especially complex endeavor, evolving considerably since its inception in the 1960s, and shaped by a multitude of internal and external pressures and forces.

Today, schools are situated within the early years of the *Every Student Succeeds Act (ESSA)* which has introduced a new accountability model to states and districts, in a post-pandemic era in which schools’ roles and responsibilities within the larger society have continued to expand, and at the epicenter of complex, and sometimes tumultuous, social and cultural forces. As stipulated in *ESSA*, school districts are held accountable for specific, mandated federal and state performance measures. However, district leaders are also accountable to their local communities and taxpayers, parents and families, students, their employees, school boards, and even to their own core beliefs about their education

mission. This has led to rich and fruitful scholarly discussions on accountability. For what should schools be held accountable? To whom are they accountable? How should accountability be measured and reported? In short, we live in an exciting time to study school accountability, what we value in our schools, and how we should measure and communicate those values to stakeholders.

### **A Brief History of School Accountability**

Before describing the central issues at the heart of this study, it is helpful to understand the evolution of our modern accountability system, hereafter described as *federal and state accountability* or *mandated accountability*. The pathway to modern school accountability is complex, impacted by political, economic, and societal forces.

#### ***The Elementary and Secondary Education Act (1965)***

The 1965 *Elementary and Secondary Education Act (ESEA)*, the federal government's most significant piece of education legislation to date, was situated within civil rights era policy (U.S. Department of Education, 2024b). A product of President Lyndon Johnson's "War on Poverty," the law directed federal funding, via Title I and its associated programs, to educational initiatives that supported the nation's disadvantaged youth. Its objective was to close achievement gaps in reading and mathematics and to ensure equity across America's network of public schools (ESEA Network, 2024). Like a giant cog beginning its motion, the *ESEA* was slow to gain momentum. However, it boasted several accomplishments in its initial decades of implementation. Specifically, it (1) funneled billions of dollars of federal aid to public schools, (2) shone a bright light on racial and economic inequality as a root cause of poor student achievement, (3) promoted

stronger parent engagement in schools, and (4) established a federal foothold in education that has endured for the last 60 years (Halperin, 1975).

Most importantly for the purpose of this study, the *ESEA* included strong language, known as the *Kennedy Amendment*, requiring states to “adopt effective procedures, including provisions for appropriate objective measurements of educational achievement... for evaluating at least annually the effectiveness of the programs in meeting the special educational needs of culturally deprived children” (Halperin, 1975, p. 6). It was natural that increased federal involvement in education ignited new demand for strong accountability policies. Halperin (1975) wrote, “thus began, crudely, the educational accountability movement with its emphasis on measuring educational outcomes” (p. 6). On April 11, 1965, President Johnson visited an elementary school in Johnson City, Texas, and stated, “I believe deeply no law I have signed or will ever sign means more to the future of America” (The American Presidency Project, Remarks in Johnson City, para. 20). Almost six decades later, our American education system continues to grapple with the complexity of bringing this mandate to life by measuring student achievement.

### ***The Coleman Report (1966)***

The *Civil Rights Act of 1964* mandated what became titled the *Equality of Educational Opportunity* report, or, colloquially, *The Coleman Report*. The U.S. Congress sought to understand the educational effects of segregation and poverty in order to make informed policy decisions. Named after its primary researcher, James Coleman, the study painted a damning picture of the unequal educational opportunities and experiences for the nation’s Black students and other students of color relative to the

nation's White students. Perhaps most surprising was the study's assertion that out-of-school factors affecting students' backgrounds had a significantly stronger effect on student achievement than in-school factors. Buried in the middle of the 749-page report is the statement,

One implication stands out above all: That schools bring little influence to bear on a child's achievement that is independent of his background and general social context; and that this very lack of an independent effect means that the inequalities imposed on children by their home, neighborhood, and peer environment are carried along to become the inequities with which they confront adult life at the end of school. (Coleman et al., 1966, p. 325)

This assertion has made a decades-long imprint on the American education system, affecting public perception, legislation, and general attitudes about the capacity of educational reform to effect real change. The central question became how to best hold schools accountable for outcomes that may be shaped largely by outside factors.

However, it is important to acknowledge the academic criticism of *The Coleman Report*. Cain and Watts (1970) contested the researchers' insufficient "theoretical justification" to draw sound conclusions, asserting that they selected and ordered explanatory variables without adequately defending their methodology. Even Coleman himself acknowledged that studying schools' learning *rates* rather than their learning *levels* would have been an "alternative and in some ways preferable method" (Coleman et al., 1966, p. 292).

Despite the ongoing scholarly debate, *The Coleman Report* deeply influenced the evolution of public school accountability. Its assertions about the limited impact of school factors have shaped the narrative around public education for decades. Ravitch (1981) wrote,

It is impossible to assess the damage done to the self-esteem of the education profession and the consequent demoralization of the very teachers dedicated enough to inform themselves about educational research. Whether students did well or poorly in schools seemed determined for the most part by ‘human resources,’ and little, if at all, by anything that teachers and schools did. (p. 719)

The report laid the foundation for federal and state accountability models that measure, sort, and rank schools based mostly on standardized testing data. Paradoxically, it simultaneously strengthened the political and public desire for school accountability while also asserting the limited impact of schools on student achievement.

### ***A Nation at Risk (1983)***

In the decades following *The Coleman Report*, the United States entered a period of increased globalization and competition amongst rival countries. In 1983, the Reagan Administration’s Commission on Excellence in Education released *A Nation at Risk: The Imperative for Educational Reform*. From the report’s opening lines:

Our Nation is at risk... We report to the American people that while we can take justifiable pride in what our schools and colleges have historically accomplished and contributed to the United States and the well-being of its people, the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people. (National Commission on Excellence in Education, 1983, p. 1)

In stark contrast to *The Coleman Report*’s message that student achievement was impacted by factors outside of the school’s control, *A Nation at Risk* placed the responsibility for student outcomes squarely at the schoolhouse steps. Specifically, some researchers argued that there were observable and tangible differences between high and low performing schools – strong administrative leadership, high expectations, an orderly learning environment, emphasis on skill-building, and tracking student progress – that proved that schools could and should improve their practice (Mehta, 2015).

*A Nation at Risk* has been described as the beginning of a “moment of angst” about the nation’s public schools (Kamenetz, 2018, para. 5). Like *The Coleman Report*, *A Nation at Risk* engendered academic criticism. Berliner and Biddle (2001) referred to the report as a “manufactured crisis” that misled the American people, by employing narrow and incomplete data, into believing America’s schools were performing worse than they were. They maintained that this crisis was predicated on “myths, half-truths, and sometimes outright lies” that paved the way for misguided and even damaging educational reforms (p. 423). Yet the report was influential, cementing for generations both the federal government’s involvement in public education as well as the American public’s scrutiny of educational outcomes. Scholars continue to discuss the report’s impact and importance, with one expert stating, “My view of it in retrospect is seldom, maybe never, has a public report been so wrong and done so much good” (Kamenetz, 2018, para. 50).

### ***The 1980s and 1990s: An “Age of Accountability” in Education***

These historic milestones paved the way for increased school accountability. Fueled by the *A Nation at Risk* report, governors across the nation came together in 1986 in a bi-partisan effort to establish a five-year plan, titled *Time for Results*, to improve American education. Amongst a wide-ranging series of recommendations – to include stronger school leadership, increased parent choice, government take-over of poorly-performing schools, better pre-kindergarten preparation, and more – was a recommendation that “the nation, the state, and school districts all need better report cards about results – about what students know and can do” (Alexander, 1986, p. 202). Undergirding the report was the message that state governors would prefer to collaborate

with educators than to battle them, and that they would be prepared to relinquish some regulatory control “if schools and school districts will be accountable for the results” (Alexander, 1986, p. 203).

By 1990, President George H.W. Bush introduced *America 2000: An Education Strategy*. The plan was influenced by these state governors urging him to lead a “populist crusade” for education reform. Specifically, their memo to the President read, “This strategy grows directly out of the Administration’s well-known emphasis on recognizing and rewarding excellence, outcomes rather than inputs, accountability for performance, parent choice, and strong partnerships with the governors” (Manno, 2018, *America 2000* and the GI Bill for Kids section, para. 3). Presented to the American people as an opportunity for educational transparency, Bush stated that stronger implementation of standardized testing would “tell parents and educators, politicians, and employers just how well our schools are doing” (Bush, 1991, para 18).

While President Bush’s proposed legislation failed to receive congressional support, it set the precedent for new iterations of federal involvement in education accountability. President Clinton succeeded in passing both *Goals 2000* and the *Improving America’s Schools Act (IASA)*. Together, these pieces of legislation required schools to create and adopt stronger academic standards, mandated that students take standardized assessments in specific grade levels, and established a reporting system that disaggregated student data by subgroup, to include gender, race, English learners, students with disabilities, and low socioeconomic status. A report by the Hoover Institute (2020), described the impact of the new legislation by stating that it “established the congressional habit of employing terms such as ‘proficiency,’ ‘high standards,’ and

‘higher order skills’ instead of attaching the targets to bona fide readiness for life after high school” (Finn, 2020, p. 10). With these pieces of legislation, our modern accountability system had taken root.

### ***The No Child Left Behind Act of 2001***

President George W. Bush’s *No Child Left Behind (NCLB) Act* of 2001 further expanded the federal government’s reach into the nation’s public schools. In a foreword to the U.S. Department of Education’s (USDE) *NCLB* summary document, Bush (2001) wrote,

This blueprint represents part of my agenda for education reform... Taken together, these reforms express my deep belief in our public schools and their mission to build the mind and character of every child, from every background, in every part of America. (p. 5)

Fundamentally, the purpose of the law was to address lagging student achievement, especially in the nation’s poorest and most disadvantaged areas. According to the USDE’s summary document, it would do this by increasing accountability and transparency related to student achievement results, funneling federal dollars to effective and research-based programs and practices, increasing flexibility for states and local school districts to make educational decisions, and empowering parents by providing performance data and offering school choice (Bush, 2001). For fifteen years, from 2001 to 2015, *NCLB* was the law of the land.

An instructive analogy, Carey (2012) wrote, “While the law marked a high-water mark of federal control over K-12 education, it was still, relatively speaking, not far from the ocean floor” (para. 2). What does this mean? *NCLB*’s most significant and widely recognized provision was its strong federal mandate that *all* students would demonstrate proficiency on state examinations by the year 2014. This mandate represented the zenith

of the federal government's involvement in public education. However, education is primarily a state and local responsibility. This resulted in the "establishment of 52 different accountability systems – one for each state, the District of Columbia, and Puerto Rico – each with different academic standards, levels of student proficiency, and requirements for teachers" (Stecher & Vernez, 2010, p. 1). The natural result was a mixed bag of successes as well as unintended, sometimes harmful, consequences.

*NCLB* was successful in establishing a national accountability model that held schools and individual teachers responsible for student performance, engendering a "flurry of improvement activities" mandated by law, and increasing educational spending within Title I schools (Stecher & Vernez, 2010, pp. 1-2). Additionally, *NCLB* realized statistically significant gains in boosting achievement metrics such as standardized test scores and graduation rates (Spurrier, 2020). In one notable study, Dee and Jacob (2011) concluded that NAEP scores increased in fourth and eighth grade math, though remained relatively stagnant in reading, following the implementation of *NCLB*. By shining a light on academic achievement as well as historically underrepresented subgroups, some students began performing two to three grade levels ahead of their earlier counterparts, yielding "historic, lifechanging progress" in the trajectory of American public education (Petrilli, 2019, Accountability 1.0 section, para. 8). Moreover, *NCLB* accountability was more transparent than ever before, representing the "main vehicle for communicating expectations and spurring action" (Education Trust, 2016, para. 4). *NCLB* had moved the needle of education reform and had shifted the national conversation on educational accountability in some positive ways.

However, its legacy is complex and punctuated with unintended consequences. These include decreases in educator morale resulting from the strict expectations of the law, a shifting away from non-tested subject areas such as social studies, science, and the arts in favor of tested subjects, and administrative challenges in meeting the requirements and mandates of the law (Husband & Hunt, 2015). Moreover, the decoupling of federal oversight and state control paved the way for watered-down standards, inconsistent metrics, and gaming the system in efforts to appear successful in meeting federal requirements (Ravitch, 2005). The dual oversight from both the federal and state levels further left the U.S. Department of Education in the “untenable political position” of exerting enough control to effectively implement the legislation while also avoiding the appearance of undermining state authority (Ramanathan, 2008, p. 307). Fundamentally, the premise behind *NCLB* was to impose clear targets undergirded by rewards, consequences, and sanctions. Largely ignored or underrepresented were the deeply ingrained structural barriers that prevent schools from serving all students equitably and effectively. Darling-Hammond (2007) wrote, “At base, the law has misdefined the problem. It assumes that what schools need is more carrots and sticks rather than fundamental changes” (*NCLB*’s Promise and Problems section, para. 4)

For the purpose of this study, the longstanding effects of *NCLB* on American public education cannot be understated. It fundamentally changed the way in which educators and stakeholders perceive and interpret student, teacher, and school success. Though *NCLB* was replaced almost one decade ago, its imprint remains strong.

### ***The Every Student Succeeds Act of 2015 (ESSA)***

The *Every Student Succeeds Act (ESSA)*, the federal reauthorization of the 1965 *Elementary and Secondary Education Act (ESEA)*, was signed into law on December 15, 2015 under the Obama Administration. The U.S. Department of Education states, “Over time, *NCLB*’s prescriptive requirements became increasingly unworkable for schools and educators,” and, as a result, “the Obama administration joined a call from educators and families to create a better law that focused on the clear goal of preparing all students for success in college and careers” (U.S. Department of Education, 2024b, para. 2). The shift from *NCLB* to *ESSA* brought forth several important changes. Most significantly, it eliminated the Annual Yearly Progress (AYP) system that evaluated schools based almost exclusively on standardized tests scores. *ESSA* continues to require states and districts to submit accountability plans, but the measures used within those plans are more diverse, and states are afforded more control over the measures that are used. These other measures of effectiveness include but are not limited to: achievement and growth of English language learners, graduation rates, participation in college and career readiness activities, postsecondary transition data, participation in advanced courses, and school climate and culture measures. Despite the broader array of measures, standardized test scores are still required to be weighted “much more” than the other indicators. States are required to intervene in poorly performing districts through targeted or comprehensive improvement plans. Other *ESSA* provisions included newly designed funding formulas intended to promote equity, the elimination of a requirement linking student achievement to teacher performance ratings, and increased emphasis on English language learners and special education populations (Klein, 2016).

Presently, we are eight years into the *ESSA* era. The body of scholarly discourse on the law's possibilities, implementation, and effects continues to expand. At the core of this discourse is whether the intentions of the law will become a reality in practice. For example, the Learning Policy Institute published a 2016 report titled *Pathways to New Accountability Through the Every Student Succeeds Act* that served as a call-to-action for states to take advantage of the flexibility and autonomy granted through *ESSA*: "*ESSA* marks an important move toward a more holistic approach to accountability by encouraging multiple measures of school and student success," thus creating "new opportunities for local innovation by giving states the opportunity to create new approaches to accountability and improvement" (Darling-Hammond et al., 2016, p. 1). The authors assert that *ESSA* represents a paradigm shift in measuring school quality, that accountability should "be designed to help leverage improvement, not just to label or sanction schools," and that outcomes can only be interpreted "in relation to inputs and processes" (p. 2). In sum, they advocate for a more holistic school accountability model that captures the complexity of the school ecosystem and emphasizes supportive measures rather than punitive action.

Critics, however, assert that, despite *ESSA*'s new and flowery language, the core of the educational accountability system remains the same: test-driven metrics that drive state response. In a recent Kappan policy report titled *Why ESSA Has Been Reform Without Repair*, Saultz et al. (2019) argued that,

If the feds renounce their tools of control – standards, tests, and accountability measures – they will appear derelict in their duties. But if they preserve these flawed systems, they will appear inept. Their solution has been to enact reform without repair. The new boss is the same as the old boss, just with a different name. (p. 19)

They further argue that federal accountability relies on over-simplified metrics that may be, at best, incomplete and, at worst, inaccurate. However, education leaders remain trapped in the “iron cage of quantification,” forced to respond with new metrics while unable to identify and address the root of educational needs (p. 20). They write, “Consequently, the new version will resemble the old and will eventually require another ‘fix’” (p. 20).

This scholarly debate, continuing to gain momentum as *ESSA* approaches a decade as the law of the land, is at the core of the present study. How do federal laws trickle down to the school district level? How do local leaders receive, interpret, and modify messages from above? How do they make sense of them, and what is their impact on their actual practice?

### **The Central Issues of the Present Study**

Having established the historical context of our modern federal and state accountability model, we enter into the present-day context in which educational accountability is hardly a settled matter. It is deeply woven into the fabric of public education, consequential to almost every aspect of schooling, and remains at the center of discourse and debate among policymakers and educators alike. The following paragraphs highlight the present-day issues and themes that shape accountability, and assert the reasons why accountability is such a relevant and compelling focus of study.

Education is a public institution, is funded by taxpayer dollars, and should be held accountable to the public which it serves. One perspective of school accountability, largely supported by families and other stakeholders, is that our existing systems – *NCLB* and *ESSA* – have been effective in increasing student achievement (Polikoff, 2017).

Moreover, standardized tests and the public presentation of data remain powerful levers to ensure that schools focus on the continuous improvement of student achievement (Huffman, 2017). Staunch defenders of strong federal education and economic policy assert that, while our system may be imperfect, outcomes-based accountability via standardized assessments remains our most reliable method of ensuring student achievement and warn against watering down outcomes-based metrics with other input and process measures that demonstrate weak associations with student performance (Hanushek & Raymond, 2001; 2002). Petrilli (2019) argued that students began performing two to three grade levels ahead of their earlier counterparts and referred to the “historic, lifechanging progress” created by mandated accountability (Accountability 1.0 section, para. 8).

Federal and state accountability has yielded other benefits. Proponents assert that the system is the primary method of communicating progress to stakeholders and engenders community pressure to improve performance (Education Trust, 2016), and that it represents a tool for parents and advocacy groups to ensure that children are being equitably served (Polikoff, 2017). After all, a critical ingredient of modern accountability is to shine light on the underperformance of subgroups of students. Devoid of such a system, schools either fly blind or may sidestep the poor achievement of historically underrepresented subgroups of children. It stands to reason, then, that public education in the United States is stronger because of consistent, transparent metrics to measure student outcomes.

However, these strengths are counterbalanced by a range of issues that call into question the system’s usefulness, integrity, and fairness. These issues emerged forcefully

upon the implementation of *NCLB* (2001), influenced the changes and improvements embedded within *ESSA* (2016), and have shaped and colored the perceptions, behaviors, and responses of stakeholders across the education system. Therefore, these issues exist at the heart of this study. Fundamental questions persist about the underlying assumptions and premises of federal and state accountability models. Even under the more forward-thinking *ESSA* legislation, standardized tests remain deeply entrenched as the engine that drives accountability. Critics argue that tests are not inherently bad, but that the overreliance on test scores as the primary lens through which to view and understand a school's effectiveness, and consequently to implement change efforts, is overly simplistic, misguided, and potentially harmful to schools and students (Koretz, 2018).

This is partially because standardized tests, while perhaps the best option to measure student performance efficiently and consistently, remain deeply flawed as measures of academic knowledge and ability. Braun and Mislevy (2005) presented an informative synopsis of how the practice of using tests to measure knowledge is built upon a house-of-cards of logic. At each stage of a test's life cycle, an array of decisions and presumptions chip away at the test's ability to fully capture a student's learning. For example, there may be poor alignment between what was taught and what was tested, a student's response may be indicative of factors other than content knowledge (i.e. familiarity with the testing format, learning patterns from similar questions, etc.), there may be discrepancies in scoring practices, and the tests likely account for only a small portion of the total body of knowledge in a content area. Unfortunately, the authors asserted that technical decisions around testing and other quantitative measures are being based on this type of "intuitive testing theory," with "untoward consequences as a likely

result” (p. 497). While standardized testing and other outcome measures are important and valuable pieces of data, policymakers and educators need to recognize their limitations and be cautious about leveraging – or weaponizing – them to draw unfounded or inappropriate conclusions.

Moreover, most stakeholders have a poor understanding of schools’ ability to impact student achievement on a scale that would equate to massive swings on standardized test results. In 2017, David Berliner wrote a letter to Philadelphians about the state of their city’s schools. His letter captures well the inherent conflict between the assumptions of federal and state accountability and the realities within actual schools. He wrote:

The education research community clearly knows what politicians and the media don’t fully grasp: teachers simply don’t account for much of the variance in standardized test scores. A reasonable estimate is that teachers account for about 10 percent of the variance in standardized achievement test scores. Research also suggests that outside-of-school factors account for six times more of that variance!... It will be difficult for teachers to show that they can turn Philadelphia’s schools into higher-achieving institutions. Teachers may help their students become stronger and more engaged learners, but they probably won’t be able to demonstrate student learning in the way that most people understand it, namely, through higher standardized achievement test scores. (Ravitch, 2018, para. 2-3).

Mandated accountability, then, places district and school leaders in a precarious position. Lest they appear defensive about standardized test results, they must embrace and integrate federal and state accountability into their operations, communicating to stakeholders the meaning and value of test scores and other metrics while, simultaneously, experiencing the reality that David Berliner depicted: that outside factors contribute more to a student’s achievement than their teachers, principals, or district leaders. This is not to say that measuring school effectiveness is not important. But as

Gill (2021) puts it, asking “How are the students doing?” is a fundamentally different question from “What does the school contribute?” (para. 5). It is within this complex and tumultuous intersection that the present study is situated.

Using standardized test scores in a misguided manner also leads to both expected and unexpected changes in behavior – at the teacher, school, and district levels. A number of studies, discussed later, suggest that school leaders and teachers respond to testing in a variety of ways, to include beneficial changes (teaching more content, teaching harder, working more effectively, and establishing a better alignment of time and resources during the school day, etc.), concerning changes (tailoring instruction to only highly tested content, coaching on test-taking strategies, etc.), and downright unethical changes (cheating) (Koretz et al., 2001). An additional effect of the era of high-stakes accountability has been the increased focus on English Language Arts, mathematics, and science, leading districts to funnel resources and funding to those areas to the detriment of programs and opportunities in the arts, music, civics, literature, and other areas. While schools may experience short-term gains in academic performance on tests, they simultaneously narrow the curriculum and rob students of other, non-tested opportunities (Greene, 2017). These opportunity costs – over many decades and in thousands of schools – represent a significant policy concern.

All of these factors contribute to the common refrain that federal and state accountability is simply out of touch with the real work of schools. Gottlieb and Schneider (2020) wrote, “Schools pursue a broad range of aims: not just to teach academic content but also to cultivate social skills and critical thinking, prepare young people for work and citizenship, foster creativity, and promote emotional and physical

health” (p. 24). Murnane (2021), viewing accountability through an economic lens, argued that the existing measures of a student’s success deemphasize social skill development that serves them in college, at work, and in their personal relationships. The school’s role was further expanded throughout and following the Covid-19 pandemic as schools scrambled to meet student and family needs by expanding food service, tending to students’ social and emotional wellness and connectivity with peers, and monitoring, from afar, signs of physical, mental, and emotional abuse or neglect. The researchers poignantly asked, “For many Americans, the Covid-19 pandemic has brought to light the large number of essential, nonacademic functions of schools, but will our accountability systems take note?” (Gottlieb & Schneider, 2020, p. 24). If not, we risk creating a troubling disconnect between what schools are asked to do and how they are held accountable.

This is both an important and urgent leadership issue because there are extensive ramifications for the equitable education of all students. Scholars argue that current accountability models, presented in the form of public dashboards, measure only small, isolated aspects of an entire school. Further, these metrics tend to relate to the demographics (especially socioeconomic status and race) within schools. These scores, in turn, are accessed by families (disproportionately White, middle-class families) or are embedded into non-governmental dashboards as well as real estate websites. These include sites such as Trulia, Zillow, RedFin, Niche, and GreatSchools, creating new causal relationships among accountability results, the housing market, and the schools in which families enroll their children (Schneider, 2017a; Scheider 2021). The very solutions that the government enacts to improve underperforming schools may, in fact,

have the opposite effect: rather than fostering healthy civic engagement centered on supporting all students to achieve academic success, families simply move to schools with better metrics. Consequently, the problem of segregated neighborhoods and schools is exacerbated.

On a larger scale, Americans tend to hold a pessimistic view of the nation's public schools. Schneider (2017b) asserted that propagating the narratives that schools are in crisis has been a consistent theme since the 1950s, used to incrementally increase the federal government's involvement in education. Interestingly, he argued that Americans' views of their own public schools – the ones which their own children attend – tend to be more positive than their views of the country's public school system as a whole. This may be evidence, he suggested, that the limited available data, in conjunction with the way it is presented, paints an inaccurate picture of a failing system. Conversely, when individuals know about their own schools, or are provided richer, more holistic data, their satisfaction with schools tends to increase.

Few would argue that school accountability is important; however, what we measure, how we measure it, and how it is presented to educational stakeholders continue to be matters of utmost importance. School and district leaders understand that accountability systems yield immense power to shape public opinion, generate parental support or resistance, and enable them to control their educational mission. Unfortunately, they may find themselves navigating rocky ground as they strive to be responsive to mandated accountability measures while also addressing a host of other needs that, while vital to their educational mission, may not move the needle on formal

accountability measures. Moreover, they must contend with difficult matters centered on fairness and equity. As Gagnon and Schneider (2019) wrote,

We must recall that accountability systems in education are... intended to promote equity for our most vulnerable students who deserve a fair and adequate education. For this task, current measurement and accountability systems appear even less up to the task. By stigmatizing and sanctioning low-achieving schools without understanding how well such schools perform across their full mission, we exacerbate inequality of opportunity. Those harmed, as a result, are those most in need of our care. (p. 755)

District and school leaders, then, must confront decisions that are not only technical, but also ethical. Borrowing concepts from Shapiro and Stefkovich (2016), they must balance multiple, sometimes conflicting, ethical paradigms. An ethic of the profession requires them to acknowledge and assume responsibility for achievement on tests and other accountability metrics while an ethic of critique may lead them to question the effect of rigid accountability systems on students, staff, and the larger community.

The convergence of the new *ESSA* legislation with post-pandemic norms and expectations has made this an interesting time to explore the construct of school accountability (Lake & Worthen, 2021). Specifically, the previous eight years have seen a wealth of scholarly discussion, articles, policy reports, and news coverage imagining the future of school accountability. What could this future accountability look like? The Learning Policy Institute describes “next generation accountability” as a shift from “no child is left behind” to a system in which “children are healthy and thriving” (Adams et al., 2017, p. 1). Specifically, they outline several key tenets of reimagined accountability to include: deeper learning to augment rote test-taking; capacity building; shared accountability across government, schools, and families that dethrones the existing top-down accountability framework; formative rather than summative assessments of

schools; an emphasis on retaining and supporting teachers; and the inclusion of multiple indicators of whole system performance (Adams et al., 2017).

It is generally agreed that the future of accountability should continue to include standardized test scores and that, without them, districts and schools essentially fly blind on the path to improvement. However, the weight of standardized assessments should decrease, measures of a school's impact on student growth, rather than their proficiency levels, should be prioritized, and leaders should explore strategies to measure student achievement on performance-based assessments that require deeper, higher-order thinking (Osborne, 2016).

An additional theme is that the future of accountability includes more holistic, robust measures – both direct and indirect – that tell a better, more complete story of a school. The Center for American Progress describes a sophisticated accountability system consisting of state-, district-, and school-level inputs, processes, outputs, and outcomes that provide more detailed information (Jiminez & Sargrad, 2017). Common elements appear across much of the literature. For example, schools could measure school culture and climate, school safety, and student engagement, all viewed as important precursors to learning, via student, parent, and staff surveys. Research has shown that non-cognitive skills have a positive impact on academics, career, and overall well-being, and are worth tracking and measuring. Civic knowledge and engagement, one of the primary and often overlooked objectives of public schooling, might be measured by tracking student participation in voting, volunteerism, and engagement with local community partners (Lake & Worthen, 2021; Osborne, 2016; Steiner & Bjorklund-Young, 2020). Scholars also discuss the importance of providing informational items without linking them to

high-stakes consequences. For example, it may be valuable to provide data on student discipline rates or a school's ability to retain students. However, they should stop short of attaching value judgments to those measures. In other words, those are highly contextualized data points, and there is no simple benchmark to indicate that a school is succeeding or failing. Still, they can provide valuable information with which to initiate public engagement and engender improvement efforts (Osborne, 2016). Moreover, some data are best delivered in a qualitative fashion, to include a district's core beliefs, descriptions of the curriculum, mapping of learning by semester, amount of time devoted to the arts, teacher planning and collaboration time, etc. (Steiner & Bjorklund-Young, 2020).

Scholars also recommend better accountability processes. Fairly evaluating schools may require at least some element of qualitative review, such as audits by outside experts (Osborne, 2016). Additionally, there is emphasis on flexibility: allowing districts and schools to generate and measure items of value to their specific educational missions. These might include data points on STEM courses, dual language immersion, or other key features of their educational program (Osborne, 2016).

At its core, this study is about how district leaders navigate the rocky ground of federal and state accountability and competing accountability demands in the era of *ESSA*. More broadly, it is about understanding how leaders leverage their own strategic and ethical decision-making power to make accountability better and more meaningful within their own communities. By understanding perspectives from leaders across the state about how they perceive, navigate, communicate, and augment mandated

accountability, recommendations for policy change may emerge and future leaders may be in a better position to lead their own districts.

## CHAPTER TWO: LITERATURE REVIEW

Federal and state accountability – or mandated accountability – is a powerful force in public education. Consequently, the literature abounds with studies and analyses that seek to answer critical questions about accountability. Has it improved student outcomes? How has it changed the behaviors of the people within and outside of schools? What are educational stakeholders' attitudes toward accountability, and how have they shifted over time? The present study centers on the ways in which superintendents and other district leaders understand and navigate mandated accountability within their own systems.

### **Underlying Theoretical Perspectives**

This study is built upon three underlying theoretical perspectives. They are (1) consequential accountability, (2) the principal-agent relationship, and (3) loose coupling. Taken together, these theoretical underpinnings help to explain district leaders' role in mediating and implementing accountability policy.

### ***Consequential Accountability***

Mandated accountability's impact has waxed and waned over many decades, but its presence in schools has remained constant. Its theory of action, sometimes called *consequential accountability* and understood as a three-legged stool, requires the implementation of: (1) clear and consistent standards for all schools, (2) a mandated assessment system to measure student achievement of the standards, and (3) a set of incentives and consequences for academic performance (Education Trust, 2016; Kress et al., 2011). With these pieces in place, the publicizing of school performance data would

then incentivize leaders and teachers to change behaviors, improve practices, and ultimately increase student achievement.

This construct is relevant to the present study in two important ways. First, and most obviously, consequential accountability describes the existing federal and state accountability model. While the high-stakes rewards and sanctions of the *NCLB* era have been replaced with softer support and improvement language under *ESSA*, the result remains the same: schools that do not meet standards established by the state are publicly identified and must develop clear plans for improvement. Second, and perhaps less obvious, is that district and school leaders must simultaneously respond to student, family, and community standards, their imposed measures of success, and local rewards and sanctions in the form of school board elections, public feedback, a satisfied or dissatisfied constituency, etc. Therefore, these leaders must navigate two “planes” of accountability, simultaneously.

### ***The Principal-Agent Relationship***

This accountability model is constructed largely in response to the principal-agent problem. The principal-agent problem, rooted in economic theory, asserts that the agent (principals, teachers, schools) may not act in the best interests of the principal (students, families) because their work is not easily observable or monitored. Therefore, an external or independent evaluation system (school accountability) is implemented to provide transparent information to stakeholders about the effectiveness of the school (Figlio & Loeb, 2011).

The advantages and disadvantages of this system have been detailed throughout this document. There are intended and unintended consequences of a mandated

accountability system. In any case, this theoretical underpinning is important to the present study. District and school leaders instinctively live and understand the principal-agent problem even if they do not refer to it by name. They realize that many stakeholders want to know and understand what is happening within the school walls. Therefore, a series of either intentional or unintentional communications ensue.

### ***Loose Coupling***

The third theoretical model underpinning the present study is that of loose coupling. According to Weick (1976), coupled events are “responsive [to each other], *but...* each event also preserves its own identity and some evidence of its physical or logical separateness” (p. 3). This means that various aspects of the school ecosystem may be linked, or connected, strongly or weakly depending upon the actions and processes that occur. Weick identified several examples of coupling, to include, but not limited to: teachers-materials, voters-schoolboards, administrators-classrooms, process-outcome, teacher-teacher, parent-teacher, and teacher-pupil. Within the context of the present study, the concept of coupling is helpful to conceptualize and understand various relationships. For example, accountability requires the following coupled relationships: federal-state (*ESSA* is federal legislation that the states enact, perhaps in 50 unique ways), state-district (states then oversee district performance, the data dashboard, and school improvement plans), district-schools (district leaders manage state expectations and rely on principals to execute improvement actions with fidelity), and schools-teachers (principals who implement improvement plans rely on teachers to implement changes within classrooms, the primary building block of the school ecosystem). At each level, there are opportunities for tightly coupled or loosely coupled relationships.

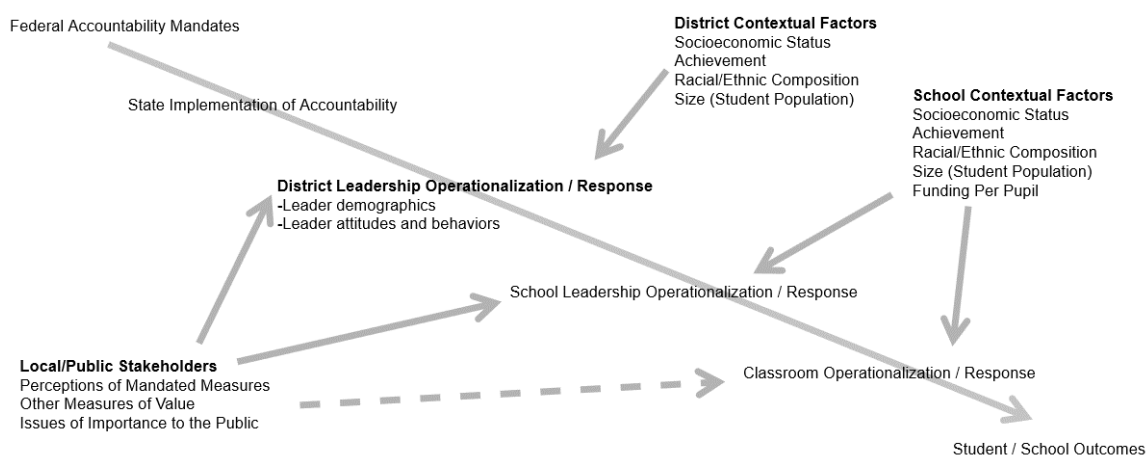
Weick described several lenses through which *loose coupling* may be observed in the school setting, two of which are especially helpful to explain the concept of mandated accountability. Through one lens one may observe “richly connected networks in which influence is slow to spread and/or is weak while spreading.” Through another, “a relative lack of coordination, slow coordination or coordination that is dampened as it moves through a system” (p. 6). These statements help to illuminate the reality that accountability moves slowly through several layers of authority – federal, state, district, and building. As it moves through each layer, information and messaging may be changed, misunderstood, diminished, watered down, resisted, ignored, or any combination thereof (see Figure 2.1). Accountability essentially enters into the messy world of organizational dynamics. Districts and schools engage in myth and ceremony to legitimize their work to stakeholders (Meyer & Rowan, 1977); leaders interact with problems, solutions, and decisions in sometimes organized and sometimes haphazard ways (Cohen et al., 1972); schools engage in isomorphic practices in order to respond to competing stakeholder demands (DiMaggio & Powell, 1983); leaders at each level undergo their own sensemaking processes that shape their perceptions, interpretations, and values related to accountability (Weick, 2001); and teachers themselves engage in complex social behaviors that lead them to adopt, adapt to, or reject change to varying degrees (Coburn, 2004).

An entire body of work could be devoted to the ways that various organizational dynamics affect school accountability. What is critical to understand is that mandated accountability does not exist in a vacuum or a perfectly closed system. Districts and schools do not and cannot focus exclusively on the dashboard of indicators deemed

important by the state. However, they are naturally held accountable for those indicators. Simultaneously, they respond to a range of competing accountability expectations from school boards, communities, parents, students, and others. As a result, the district and school organizations may become loosely coupled with federal and state accountability.

**Figure 2.1**

*Conceptual Model: The Implementation of Accountability Through Multiple Layers*



## Empirical Research and Relevant Literature

### *Conceptualizing Accountability in Education*

Beginning with the *Elementary and Secondary Education Act (ESEA)*, policymakers wrestled with a fundamental question: what does it mean to hold a school accountable? This question has been explored through various lenses. Glass (1972) discussed the complexity of school accountability, asserting that, despite internal accountability mechanisms (teacher to principal, principal to superintendent, etc.), that “fundamental accountability is owed to the public” (p. 637). Similarly, Wescott (1972) asserted that many stakeholders play a role in school accountability, to include national,

state, and local boards of directors; parents and community members; administrators; teachers; and the students themselves. At the center of accountability is the school, which should be held accountable for student outcomes, not processes. This early scholarly conversation, emerging in the years following the *ESEA*, suggests a basic understanding that the school provides an educational service and that it should be evaluated on its effectiveness.

Some scholars, however, conceptualized federal and state accountability more broadly. For example, as mandated accountability reached its apex under *NCLB* in 2001, Behn (2003) questioned *who* should be held accountable in public education. He urged a paradigm shift in how we think about accountability and encouraged a shared approach in which educators, administrators, policymakers, state and federal leaders, community and business leaders, and even students and their parents divide up the burden of accountability. After all, *The Coleman Report* of 1966, he argued, established the importance of outside factors in a student's education, a concept largely ignored when accountability pressures are so acutely targeted at schools. He concluded that "a compact of mutual and collective responsibility is a much more complex institution compared with traditional, unidirectional, hierarchical accountability" (Behn, 2003, p. 63). These perspectives underscore the ongoing dilemma of how to assign responsibility to a child's education. While the literature is consistent that accountability is complex and that many stakeholders impact the educational process, they arrive at divergent conclusions about how to implement shared accountability in practice.

Other researchers have explored not *who*, but *what*, should be accountable in education. Levin (1974) posited that accountability is manifest in four distinct ways: (1)

as performance reporting, (2) as a technical process, (3) as a political process, and (4) as an institutional process. Performance reporting is the act of linking a school's test scores or other outcomes to its effectiveness. As a technical process, accountability evaluates a school's decision-making to allocate money and resources to deliver a quality product. As a political process, schools are held responsible for ensuring adequate success and results for particular groups, especially those holding power conferred through wealth, race, or status. As an institutional process, schools are held accountable for maintaining the "status quo" in terms of the existing order of society.

Decades later, Finn (2002) echoed Levin's (1974) argument that there are various, sometimes conflicting, accountability forces. He identified four types of accountability: (1) compliance (emphasizing rules, programs, procedures, and other inputs), (2) professional norms and expertise (relying on national organizations, educational leaders, and best practices to drive decisions), (3) standards-based reform (using outcomes-based measures, predominantly standardized tests, to evaluate schools), and (4) marketplace accountability (employing school-choice mechanisms such as vouchers to allow families, i.e. the customer, to engender change in schools). With the benefit of three decades of accountability on which to base his conclusions, Finn went a step further by analyzing each type, evaluating the various combinations that exist in schools, and concluded that a purely standards-based conceptualization of accountability had failed, and that a more nuanced blend of standards-based and market-driven accountability would better serve schools, students, and their families. Specifically, he argued that "each offers a promising solution to a big problem besetting the other" (p. 43). The reality is that superintendents and district leaders must respond to all forms of accountability, simultaneously. District

leaders may engage with accountability for various, sometimes conflicting, aims and constituencies.

The critical takeaway from this strand of literature is that school accountability has been conceptualized in various ways. Over decades, federal- and state-level policymakers prioritized standards-based accountability, thus laying the foundation for today's formal, government-sanctioned accountability model. However, as comprehensively reviewed in Finn's (2002) work, other forms of accountability – political, economic, and social – exerted their influence and, sometimes, created tension and conflict. School and district leaders live at the epicenter of this tension, making decisions and taking actions that are responsive to one form of accountability or another.

### ***The NCLB Era of Standardized Testing and High-Stakes Accountability***

Upon the implementation of *NCLB*, standardized testing became the primary, high-stakes mechanism to evaluate school effectiveness. Linn (2000) explained that standardized tests are popular as a measurement tool because they are relatively inexpensive, easily mandated, rapidly implemented, and produce visible results. Consequently, the literature, especially between 2001 and 2010 when *NCLB* was implemented and deeply studied, abounds with research examining standardized tests' impact on schools, educators, and students, and analyzing their validity as measurement tools. As high-stakes testing remains the centerpiece of *ESSA*-era accountability, it is instructive to review the scholarly discussion on testing. Moreover, it is likely that superintendents' and other district leaders' views on federal and state accountability have been shaped by the discussion surrounding standardized testing. The literature appears to be centered on the following intersecting themes: (1) the effect of high-stakes

accountability on student achievement, (2) the intended and unintended consequences of high-stakes accountability on behavior, and (3) the validity of using high-stakes testing as a measurement tool.

**The Effect of High-Stakes Accountability on Student Achievement.** The purpose of *NCLB* was to establish the clear, though lofty, expectation that all students would be proficient in reading and mathematics by 2014. Naturally, a central question was whether the accountability systems introduced throughout the 1990s yielded increased student achievement. Early studies produced mixed results. For example, Carnoy and Loeb (2002) studied the relationship between the strength of a state's accountability system, as measured by a 0-5 indexing system, and student gains on the NAEP mathematics tests in 1996 through 2000. They found that students in states with stronger accountability systems averaged statistically significant gains in grade eight, but not in grade four, and that the positive relationship was evident for Black, White, and Hispanic students. Hanushek and Raymond (2002, 2005) found that NAEP scores increased approximately 1% upon the introduction of a state's accountability system and after controlling for inputs such as parent education, school spending, and racial exposure. They also suggested that accountability systems did not have an impact on special education placements once accounting for normal increases in special education placements over time. Interestingly, they concluded that the racial achievement gap may *widen* because the White subgroup gains were more significant than those of the Black subgroup, an unintended consequence warranting further policy consideration.

Subsequent studies examined the specific impact of *NCLB*. Like Hanushuk and Raymond, Jacob's (2005) study found that increased achievement on reading and

mathematics standardized tests was correlated with the implementation of strong accountability models in Chicago Public Schools. However, they also found that teachers and leaders responded “strategically” to testing by increasing special education placements, preemptively retaining students, and reallocating time away from other subjects such as science and social studies. They concluded that, while stronger accountability models supported general incentive theories, it was “less clear how to evaluate high-stakes testing as a school reform strategy” (p. 791). The literature suggests that this was a common theme: observing some academic gain while simultaneously unable to differentiate legitimate intellectual improvement from test preparation behaviors or, worse, gaming the system. For example, Neal and Schanzenbach (2010) also found that *NCLB*-era accountability increased test scores, but not for the lowest performing students. As a result, this perversely incentivized schools to focus attention on bubble students who could increase proficiency rates. Conversely, some studies (Springer, 2008) found evidence that the lowest-achieving students in failing schools experienced gains in test scores, but still suggested that continued research was needed to further isolate positive effects from “perverse incentives” to game the system to increase scores (p. 8).

A decade into the implementation of *NCLB*, the literature continued to report mixed results. Dee and Jacob (2011) conducted a study of *NCLB*'s impact by examining NAEP scores at the state level. NAEP is considered a “low-stakes” test and therefore more insulated from attempts to game the system or teach to the test. They found that *NCLB* engendered statistically significant increases in fourth grade math achievement (effect size of 0.23 by 2007), some increase in grade eight mathematics, and no evidence

that *NCLB* impacted fourth grade reading achievement levels. Ten years into its implementation, the researchers noted that *NCLB* fell short of its ambitious goals and that policymakers needed to decide whether to “end or mend” the policy (p. 419).

The studies discussed here represent only a sampling of the scholarly literature. This body of research helps to explain how district leaders may perceive accountability’s impact on student performance. More specifically, it is reasonable to expect that the high-stakes, test-driven experiment of *NCLB*-era accountability confirmed what many superintendents, district leaders, school leaders, and teachers knew instinctively: that strong accountability is important, can yield gains in student achievement, but is insufficient as a lone driver of change because student performance can be affected by many factors.

**The Effect of High-Stakes Accountability on Behaviors Within and Outside of the School.** Tied closely with the issue of student achievement is the question of *why* test scores may change in response to high-stakes accountability. This strand of the literature suggests that there have been both intended and unintended consequences of high-stakes accountability on teacher and leader behavior within schools. Koretz et al. (2001) explored how student achievement is measured and reported. For example, a newspaper may write about its local school district’s performance on state testing while overlooking the “array of proficiencies implied by scores” and the “uncertainty and disagreement about intended or appropriate inferences” (p. 11). Put simply, test scores may change based on any number of factors. In an effort to create a rudimentary framework to articulate their point, they identified seven types of test preparation that may lead to increased scores, to include: (1) teaching more, (2) teaching harder, (3)

teaching more effectively, (4) reallocating resources and instructional time, (5) aligning instruction to specific standards or testing elements, (6) coaching on test-taking, and (7) cheating. Some of these behaviors – teaching more, teaching more effectively, etc. – are positive, intended consequences of accountability while other behaviors – teaching to specific standards or testing elements, emphasizing test taking strategies rather than content knowledge, cheating, etc. – are negative behaviors that may yield inaccurate measures of students' actual intellectual knowledge.

The *NCLB* era generated a substantial body of literature that examined this phenomenon of changed behavior – both intended and unintended – in response to high-stakes accountability. Ladd (2001) found that accountability, especially when tied with financial incentives, empowered some principals to drive improvement efforts, but could, conversely, lead teachers and principals to avoid serving in the highest need schools. Rouse et al. (2013) conducted a similar study that suggested that educators in failing schools – in this specific study, Florida schools receiving a grade of “F” in the summer of 2002 – responded favorably by engendering immediate and sustained improvement resulting from curricular and pedagogical changes. Another study (Hamilton et al., 2005) found that teachers in California, Georgia, and Pennsylvania attempted to align instruction to state standards, and that professional development opportunities focused more intensely on tested subjects. Some studies examined principal and teacher perceptions of test score gains, finding that these stakeholders attributed students' familiarity with the tests themselves (rather than increased knowledge) (Koretz, 1996) and explicit test preparation (Stecher et al., 2000) to the increase in test scores. Dee et al.

(2013), a decade into *NCLB*, also asserted that instructional time spent on tested subjects increased at the expense of other subjects.

Other changes in behavior related to high-stakes accountability occurred outside of the classroom. For example, Figlio (2006) found evidence that schools increased disciplinary consequences for low-performing students during testing windows and in grade levels that are tested. In a separate study, Figlio and Getzler (2006) found that the rates of special education designation and classification increased following the introduction of high-stakes accountability models in the late 1990s and early 2000s. Some literature even suggested that a school's food service was modified in response to testing. Figlio and Winicki (2005) found that schools increased servings of food with low nutritional value, but that may provide a short-term boost of cognitive focus. More research would be required to establish a strong linkage between high-stakes accountability and any one of these factors, but there is compelling evidence that adult behavior in schools is altered when attempting to change the conditions around standardized testing.

Some of the literature addresses cheating behaviors. For example, Amrein-Beardsley et al. (2010) studied the extent to which teachers engaged in varying degrees of behavior that would be considered cheating. Using survey and focus group data from educators in Arizona, they found that first-degree cheating (overt and intentional) was uncommon, but that approximately half of the teachers engaged in or knew someone who engaged in some form of second- or third-degree cheating (ignoring some testing rules, providing students with more time, prompting students to rethink an answer, or providing students with copycat questions to prepare for a test). Later, Hibbel and Penn (2020)

studied the widespread cheating episode in Georgia public schools between 2010 and 2012. They concluded that cheating should not be viewed as deviant behavior by individual teachers, but rather as a consequence of organizational dynamics and pressures. In the Atlanta incident, mandates to cheat originated at all levels of the organizational structure. As a result, the researchers urged the careful consideration of test scores as the primary determinant of a school's effectiveness and recommended a more "balanced system of measurement and accountability" (p. 350). Such studies, while largely inconclusive about the complexity, scope, and depth of cheating on high-stakes assessments, provide important context about how leaders and teachers may respond to federal and state accountability.

Researchers also studied the impact of accountability policies on teacher recruitment and retention. Boyd et al. (2008) examined the hiring patterns of fourth grade teachers in New York state to ascertain whether high stakes accountability impacted teacher recruitment and retention. They found no evidence that the turnover rate increased in relation to high-stakes assessment and, in fact, found decreases in the turnover rate in grade four. They did find some changes in the characteristics of new fourth grade teachers, finding that more experienced and highly educated teachers were selected for those positions. They concluded that high-stakes accountability impacted recruitment and hiring, but in some unexpected ways. In another study, Clotfelter et al. (2004) found that schools in North Carolina experienced higher turnover rates in low-performing schools but found little evidence that the quality of the teachers who filled those vacancies declined as a result of high-stakes accountability. More recently, Ingersoll et al. (2016) studied the effect of accountability policies on teacher retention by

examining data from the Schools and Staffing Survey (SASS) and the Teacher Followup Survey (TFS) over two years. They found that low-performing schools experience higher turnover rates. Interestingly, though, schools that prioritized better working conditions – operationalized as strong leadership support, adequate classroom resources, influence in school decisions, and classroom autonomy – were able to mitigate some turnover.

Arguably, federal and state accountability changes behaviors in local communities, as well. For example, a body of literature has documented the effect of high-stakes accountability on local property values, a potentially important factor in the present study of district leaders' attitudes toward accountability. Kane et al. (2003) found evidence that housing values in North Carolina could be related to school accountability measures, and Figlio and Lucas (2004) found a relationship between test scores and property values in a medium-sized Florida district, concluding, "the innocuous-seeming school classifications may have large distributional implications, and that policymakers should exercise caution when classifying schools" (p. 25). Both papers urge caution, however, in drawing widespread conclusions. Approximately a decade later, Imberman and Lovenheim (2016) built upon Figlio and Lucas' (2004) work, examining property values in Los Angeles following the release of standardized and value-added testing results. They found no statistical evidence that test scores impacted housing value. However, they asserted that further research would be necessary to determine residents' perceptions of the importance of the data.

The body of literature on how mandated accountability affects the behavior of people within and outside of the school ecosystem is large. The sheer volume of studies dedicated to understanding shifts in organizational dynamics when faced with high-stakes

accountability demonstrates its importance. Superintendents and district leaders find themselves at the center of these organizational shifts. They must respond to community, parent, and staff perceptions of testing while simultaneously leading change efforts that improve performance. As evidenced by the literature, school systems have exhibited a continuum of responses – intended and unintended, well-intentioned and manipulative, and everything in between – when confronted with mandated accountability. Of importance to the present study is the extent to which changes in behaviors continue to manifest in the *ESSA* era, and how district leaders are responding to accountability pressures.

**The Validity of High-Stakes Testing.** The third interrelated strand within the literature review on high-stakes accountability is the validity of standardized testing as an accurate measurement of student learning and school effectiveness. Just before the implementation of *NCLB*, Linn’s (2000) study captured the complexities and nuances of using standardized testing data in a high-stakes manner. He identified several ways to view and dissect the same student data – current status models, comparisons of different cohorts or grades within the same year, comparisons of the same grade year-over-year, longitudinal comparisons of aggregate scores, and the employment of regression-based adjustments – and concluded that “the choice of constructs matters” (p. 12). Linn wrote, “in most cases the instruments and technology have not been up to the demands that have been placed on them by high-stakes accountability” and, as a result, “lose much of their dependability and credibility” causing a scenario in which “the unintended negative effects often outweigh the intended positive effects” (p. 14). Linn’s conclusion predicts the literature that emerged in the early years of *NCLB*.

Some of the literature on validity centered on the unintended impact on adult behaviors, to include test preparation and gaming the system. For example, Hanushek and Raymond (2002) examined cross-sectional, longitudinal, and status-change models for tracking academic achievement and progress, and concluded that, while schools did respond to the presence of an accountability system, it was difficult to discern, statistically, long-lasting impacts to learning from short-term adjustments that boosted scores.

Other literature explored high-stakes assessments through the lens of statistical validity. Kain and Staiger (2002) raised concerns over the statistical noise that leads to imprecise measurement. In their study of North Carolina public schools, the smallest schools were more likely to appear successful and were 23 times more likely to win a “Top 25” award than the largest schools. This is mostly because the large schools must make greater gains for more students to effect an empirical change to their results. Similarly, racially heterogenous schools were less likely to win performance awards due to the “statistical bias” against such schools, especially smaller ones. They further warned against making inferences based on one-year changes, writing, “even when a school is on the right track, the path to improved student performance is rarely a straight path” (p. 102). Conversely, Greene et al. (2004) compared student results on “low-stakes tests” – which schools had little reason to manipulate – to those on “high-stakes tests,” and found a high correlation between the two. He concluded that stakeholders should have confidence that high-stakes tests appropriately reflect student achievement. Koretz (2008) studied value-added measures (VAMs), a subcomponent of accountability measurement. Value-added measures provide a statistical analysis of the change in student achievement

between two assessments, and further attributes the change to individual teachers or schools. He concluded that value-added measures have improved educational accountability but remain problematic because tests are often imprecise measures of learning due to a combination of error and bias, the incongruity of content between grade levels, and factors outside of the school's control that can interfere in results to an undetermined degree. Similarly, Price's (2010) study of Milwaukee public schools suggested that *NCLB* accountability ineffectively accounted for socioeconomic, transiency, and other factors that could negatively affect student performance. She concluded that "this failure to control for outside influences makes the labeling procedures unfair" (p. 804).

Researchers also analyzed whether high-stakes assessments were useful tools for measuring what is considered important. For example, Meyer et al.'s (2014) analysis of the rise of accountability suggested that *NCLB* was overly reliant upon quantitative measures (test scores) to identify student success and teacher effectiveness. As such, it attempted to "turn education into an all-measurable process, even if these schemes are as opaque to the public as they are easily manipulated by politicians" (p. 9). Referencing Isaiah Berlin's warning that ignoring "all that proves incapable of quantification" is a "form of false consciousness of bureaucrats and administrators," they suggested that school accountability likely requires a more holistic and humanistic approach that more artfully blends quantitative and qualitative metrics to evaluate schools (p.9). Howe and Murray (2015) evaluated states' A through F rating systems, concluding that they were invalid as measures of success, as policy instruments, and as democratic frameworks. Echoing the previous study, they concluded that standardized testing measurement fails

to tell a more holistic and accurate story of the schools. Goldhaber and Özek (2019) examined standardized test scores in relation to later life outcomes. They found evidence that tests accurately predict later life outcomes and are, therefore, an important component of measuring school effectiveness. However, they warned that test scores may not capture the totality of a school's impact on students and their long-term success.

Finally, within the last 15 years, scholars have noted a relationship between socioeconomic conditions and school performance, raising important questions about the tools used to evaluate schools. Tienken et al. (2017) examined whether student performance in middle schools in New Jersey could be explained by various social and economic factors beyond the schools' control. Their results suggested that, given access to a few, targeted pieces of U.S. Census data (e.g., percentage of families in a community with income over \$200,000 per year, or the percentage of people in a community in poverty, or the percentage of people in a community with bachelor's degrees), there is a high probability that they could predict standardized test proficiency results for students in grades six through eight. Monarrez and Chingos (2020) engaged in a similar study of North Carolina schools, finding that publicly available data, when controlled for demographics, produced school performance data that better align to the schools' growth metrics. This suggested that, when this process is used, it produces results that are less biased against schools with high numbers of low socioeconomic students. Even so, they continued to find "considerable bias and noise" (p. 6). They concluded by suggesting that "publishing school quality ratings or making policy decisions based on these measures is fraught" (p. 6).

**Conclusions.** The previous sections on *NCLB*-era, high-stakes accountability addressed three, intersecting themes: (1) the effect of high-stakes accountability on student achievement, (2) the intended and unintended consequences of high-stakes accountability on behavior, and (3) the validity of using high-stakes testing as a measurement tool. This body of literature is consequential to the present study, centered on superintendents' and other district leaders' perceptions of and attitudes toward federal and state accountability, in several important ways. First, there is some evidence that high-stakes accountability improved student achievement for some students; however, there is uncertainty about whether these improvements are indicative of actual learning or of other, targeted attempts to improve test scores. Put simply, federal and state accountability may or may not generate intended academic outcomes. Second, the literature suggests the prevalence of both intended and unintended effects of high-stakes assessment on behavior within the school. While the new *ESSA* legislation expanded the definition of school success from test scores alone to a more diverse array of indicators, the concept of high-stakes accountability shaping human behavior in schools remains as important as ever. Third, issues of validity continue to be relevant. Standardized testing remains in place as the bedrock of the *ESSA* accountability system. Though other metrics and indicators have been introduced, student proficiency and growth on standardized metrics continue to drive the identification of schools targeted for support and improvement plans. As such, scholars' questions and analyses on these metrics remain as vital as ever.

### ***Stakeholder Responses to Federal and State Accountability***

There is a large body of literature exploring stakeholders' – specifically leaders', educators', and families' – perceptions of federal and state accountability. One of the driving purposes of both *NCLB* and *ESSA* has been to increase stakeholder understanding of school performance to engender external pressure and internal change efforts. *ESSA* was designed to expand *NCLB*'s achievement mandate to utilize wider, more diversified metrics that drive instructional decision-making at the district, building, and classroom levels. Moreover, it strongly emphasizes stakeholder engagement by transparently publicizing the indicators to foster family and community understanding of a school's performance and to encourage stronger involvement in education. (Darling-Hammond et al., 2016). Put simply, mandated accountability does not exist in a vacuum. It is received, understood, navigated, and communicated in different ways by different stakeholders. This body of literature informs the present study by examining how key stakeholders perceive mandated accountability and how those perceptions, in turn, impact superintendents' and other district leaders' responses to accountability.

**School Leaders' and Practitioners' Responses to Federal and State Accountability.** In the mid-2000s, a series of studies sought to understand the relationships among district leaders', building leaders', and teachers' perceptions of mandated accountability. Russell and McCombs (2006) conducted a study of stakeholder perceptions in California, Georgia, and Pennsylvania. They found mixed perceptions – both positive and negative – about assessments and the effects of the accountability system. In general, there was evidence that all three groups believed that the *NCLB* accountability model improved instructional practice, curricular alignment, and general

rigor, but that it simultaneously decreased staff morale. In Pennsylvania, the focus of this present study, 30% of superintendents believed that AYP (the accountability metric under *NCLB*) accurately reflected students' performance.

Dessimone (2006) conducted a similar study of the same three stakeholder groups but examined their perceptions through a different lens. He studied specificity (the degree to which the accountability policy is understood and implemented), authority (the degree to which the policy receives the backing and support of those in charge of its implementation), power (the degree to which the policy yields effective rewards and sanctions related to implementation), policy consequences (the degree to which the policy has changed instruction and practice), and barriers (the degree to which various obstacles have impacted the policy's implementation). Key findings included: (1) respondents were in most agreement that students represented an obstacle to reform, (2) respondents demonstrated the most disagreement over the value and usefulness of authority from state and federal agencies, and (3) the three sets of respondents responded in different ways about the successful implementation of the policy within their schools and the authority that school districts wield to successfully implement policy.

Finally, Englert et al. (2007) conducted and synthesized a series of three studies investigating superintendent, principal, and teacher perceptions of accountability via surveys. The study involved four states (Colorado, Kansas, Missouri, and South Dakota), 49 superintendents, 121 principals, and 153 teachers. Each group was nested within the group above it (the teachers worked in the principals' buildings and the principals worked in the superintendents' districts) in order to ascertain patterns and trends across groups in similar environments. Undergirding the study were seven principles of highly effective

assessment systems, from which the survey protocols and interview questions were generated. The researchers emerged with a range of comparative data across the three groups. Notable findings included: (1) in general, all three groups were generally favorable to the accountability system in place, (2) the teacher group generated lower ratings in several components, to include their belief that assessments were of high quality and that data-informed decision-making processes, (3) superintendents indicated the highest concern for having adequate resources to support improved instruction, and (4) principals and superintendents indicated a high level of fidelity in data use and analysis, whereas teachers indicated a more superficial usage because they were required to use the data in prescribed ways.

It is instructive that these types of studies were prevalent during the *NCLB* era. Though similar, they approach high-stakes accountability from various directions, examining the perceptions and attitudes of those in charge of leading and implementing change efforts within schools. Synthesized and taken together as a whole, they reveal themes that are important for the present study. First, there is consensus that accountability is good for education. Second, district leaders and principals maintain, at least outwardly, that the existing accountability measures have represented a fair way to measure student success and school effectiveness. Teachers, however, expressed more pessimistic views on this. Third, accountability was viewed as having a combination of positive effects (increased focus on standards, improved pedagogy, attention to low-performing subgroups, etc.) and negative effects (decreased morale, overemphasis on testing and limited educational objectives, barriers to reaching designated benchmarks, etc.).

While the three previous studies used a very specific research methodology – studying district leaders’, principals’, and teachers’ perceptions in a nested manner – other studies have explored perceptions and attitudes in other ways, many of which led to thought-provoking but ungeneralizable results. For example, Abelman et al. (1999) utilized a case study model to determine how accountability affected the daily work of people in real schools. One of their findings was that school personnel largely bypassed mandated accountability, instead creating their own “internal normative structures that are relatively immune to external influences” (p. 39). These internal normative structures are developed less so from a strong, coherent school or district vision and more from individuals’ beliefs, attitudes, and understandings about education and their students. The result is that these internal accountability mechanisms are only partially aligned with mandated accountability. Similarly, Mintrop (2012) compared five high performing and five low performing schools in California. He explored schools’ responses to mandated accountability through the construct of integrity. He found that some schools make sense of it by increasing their collective integrity, defined as maintaining strong leadership centered on a core set of beliefs and strategies that interweave external obligations with internal mission. Conversely, other schools lack this construct of integrity. Individuals within these schools may assume defensive postures, become fragmented under pressure, or balk against any form of conformity to a common mission, presuming that they must know better than external agencies what is best for students. This study illuminates the different ways that schools might make sense of and respond to mandated accountability.

Other studies examined district and school leaders’ sense-making of federal and state accountability, and how their understanding of the construct affected their

leadership actions. For example, Rorrer and Skrla (2005) studied district and school leader responses to mandated accountability through multiyear, multiphase research projects in North Carolina and Texas. They found that these leaders acted as “policy mediators” who, when effective, implemented a combination of the following strategies: (1) cultivated relationships and interactions that reframe and reconceptualize accountability policies not just for themselves, but for those within the organization, (2) recultured the district to model high expectations that all students can and should succeed, and (3) integrated accountability aims and components so as to ensure that district priorities, goals, resources, and other systems and structures are fully aligned to promote student success. In their conclusion, the researchers stated that the leaders in their studies were “ordinary leaders who accepted the challenge posed by accountability policy and who learned how to use it for the benefit of their districts and schools” (p. 61).

Knapp and Feldman (2012) conducted a multi-case study of principal leadership in 15 urban schools across the United States. They found that principals internalized mandated accountability systems, thus merging them seamlessly into their school culture and expectations. Notably, they found that these principals offered “ways to visualize what it might mean to reconcile what the larger educational system is asking for and what school-based professionals want and need, to make their work fulfilling and productive” (p. 690).

Seashore Louis and Robinson (2012) found a similar leadership response. They examined how principals’ understanding of federal and state accountability affects their leadership responses and actions. They suggested that principals tend to view federal and state accountability as important and valuable when (1) they perceive strong district-level support for accountability-driven improvement efforts and (2) when mandated

accountability aligns with their own beliefs about what is important in schools. Without these components in place, perceptions toward accountability were more negative. When these factors were in place, leaders also tended to be stronger instructional leaders, as evidenced through principal and staff surveys. Like in Knapp and Feldman's (2012) study, they tended to undergo a process in which they "internalized" mandated accountability policies to weave them seamlessly into the unique priorities and needs within their own schools.

School board members' role in implementing federal and state accountability is often overlooked, yet board members represent a key linkage between the mandated model and internal implementation, to include communication to the larger public. Sutherland (2022), via a qualitative case study of school board members and district leaders across multiple districts, studied school board sensemaking of federal and state accountability policies under *NCLB* and *ESSA*. She found that school board members tended to land in one of three predominant categories: resistant to testing as an incomplete measure of a student; ambivalent about testing as an important, but limited, indicator; or strongly supportive of testing as a means to measure school and teacher effectiveness. In turn, these perceptions affected how they communicated and messaged accountability to their constituents. Most notably, school board members' perceptions often differed with the intended goals of federal and state accountability, seemingly more influenced by local politics, perceptions of themselves as a governing body, and their own educational values. This study suggests that federal and state policy may be filtered and morphed in various, unanticipated ways as it permeates individual states and school

districts. While not generalizable, the findings in this study are relevant to other groups, such as district leaders.

Collectively, these studies inform our understanding of what happens when accountability enters real schools. Mandated accountability is not a simple or straightforward process because district and school leaders are responding, simultaneously, to often-divergent institutional pressures from the school board, district administrators, staff, families, the larger community, and students. Any single decision may or may not produce meaningful effects on the mandated accountability model. Moreover, the school is not a blank slate, but rather a complex collection of entrenched beliefs, attitudes, norms, and expectations which the leader must navigate. As related to the present study, this body of literature informs our understanding of how individuals within the school and district setting make sense of accountability, and how both internal and external forces shape attitudes and beliefs. In turn, these attitudes and beliefs impact superintendent and district leader behavior and action in response to accountability.

**Families' Responses to Federal and State Accountability.** One of the major functions of federal and state accountability is to inform families and the larger community about school quality and effectiveness, thus engendering civic participation in public education. In turn, the public response informs superintendents' and other district leaders' actions. Therefore, it is important to examine the smaller, but equally important, body of literature describing families' and the communities' perceptions and attitudes toward accountability.

Some studies sought to understand the information that stakeholders valued. For example, Owens and Peltier (2002) reported on a Nevada statewide survey of 4,900

parents. They found that most (75%) families reported reading most or all of their schools' accountability reports, valuing the diverse array of measurement indicators, and a significant portion (85%) of families stated that the reported information was relevant to them. However, parents indicated that they would like future reports to include information on school or district programs and policy, more detailed classroom information, school comparisons across the district, state, and nation, more financial information, and more contextual information on items such as uniforms, zoning, testing, year-round education, teacher raises, and parent involvement. In a later study, Brewer et al. (2015) studied the extent to which 5,900 survey respondents, both educators and non-educators, in South Carolina perceived standardized testing as a useful tool to drive improved student academic achievement. Unlike Owens and Peltier (2002), they found that most stakeholders do not view state testing as a valid instrument for discerning school quality or as a useful tool to help students learn. Interestingly, disaggregated data revealed that minority populations expressed more optimism than their White counterparts that standardized testing offered the best possibility of school improvement and student achievement. This suggests, perhaps, that tests are seen as a mechanism to shine a light on perceived inequality in formal institutions, such as the school. Studies such as these address fundamental questions about stakeholders' satisfaction of, and belief in, what we measure to ascertain school effectiveness. This is critical to the present study because superintendents and other district leaders must make decisions about what information and data to present to their local constituencies. These decisions, in turn, affect community perception of their schools and their ability to generate support for their administration, decision-making, and change efforts.

Other researchers examined how the physical presentation of public data affects stakeholder perception of their local schools. For example, Jacobsen et al. (2013) explored the relationship between parent satisfaction and the school report card letter grade in New York City. They found that parent satisfaction decreased when student scores dropped immediately following the implementation of higher standards. While the purpose of raising expectations was to increase student achievement, the authors suggested a paradoxical effect in which the public support that is needed to support action and reform is, simultaneously, eroded when test scores decline. In a subsequent study, Jacobsen et al. (2014) investigated how the public presentation of state accountability - as letter grades, performance index rating, percent meeting a goal, or achievement level designation - affects respondent attitudes toward public education. Respondents were asked to rate the public report card along the American Customer Satisfaction Index. They indicated stronger responses to schools with letter grades (greater satisfaction for higher letter grades and lesser satisfaction for lower letter grades) as compared to other types of data. This may be because respondents are familiar with the general concept of a letter grade from their own school and popular culture.

Recently, states have increasingly used value-added measures to assess student learning and measure teacher effectiveness. These data help stakeholders understand a school's impact on student performance over time. Some studies have explored public response to such value-added measures. In one example, Houston et al. (2022) studied the effects of student growth metrics on stakeholder perception of schools. When the researchers provided participants with achievement data, with which they were already familiar, their views on school quality did not change in a meaningful way. Conversely,

when the participants were provided data on academic growth, their views changed to be more in line with the growth measures. This suggested that respondents were less knowledgeable about growth levels than achievement levels, and that growth scores do, in fact, affect public opinion.

These studies highlight not only the importance of chosen accountability metrics, but also their format and presentation, on family and community perception. As demonstrated above by Jacobsen et al. (2013), the seemingly logical and innocuous policy decision of raising standards yielded unintended negative consequences on the perceptions of local schools' effectiveness. Even the physical representation of effectiveness, using letter grades or some other metric, may change stakeholders' perceptions of a school. Whether or not superintendents and other district leaders actively recognize these dichotomies, they exist and are clearly consequential to the work that they do. This strand of the literature informs the present study by focusing attention on the complexity of stakeholder understanding of and response to mandated accountability.

### ***Contemporary Literature on the Design of Accountability Systems***

Schools and districts are complex organizations that are accountable for a multitude of outcomes and to many stakeholders. Formalized accountability systems, however, are used to measure only a selected sample of input, output, and process indicators. Moreover, even once the indicators are chosen, there are many ways to approach the actual measurement. For example, something as seemingly simple as measuring *student academic growth* can be approached in a host of ways: by following individual students from year to year, or cohorts of students from year to year, or monitoring a single grade level over time, and the list goes on. Moreover, different

stakeholders have different beliefs and attitudes about what is worth measuring. Some want measures on academic performance, or career opportunities, or civic engagement, or the qualities of the teaching staff, or access to technology, or other output or outcome measures. Others desire information on input measures: funding per pupil, class sizes, access to curricular opportunities, etc. Consequently, an emerging strand of academic literature has centered on the actual design of our accountability systems. What is included? What do they measure? What is important to stakeholders?

Various studies have explored the functionality and effects of broader, more holistic accountability models. In one example, Gagnon and Schneider (2019) examined stakeholder perceptions – via 10 focus groups consisting of teachers, administrators, parents, and community members - within a mid-sized, urban district in Massachusetts of the perceived value of the state’s existing accountability dashboard relative to an expanded dashboard of indicators that includes opportunity-to-learn measures and social and emotional learning measures. In terms of establishing the overall effectiveness of a school, they found that the state accountability system’s results roughly align with their expanded framework’s results. However, there were other consequential differences. First, expanded metrics were highly valued by stakeholders because they provided a wider and more holistic understanding of a school. Second, the expanded dashboard clearly identified inputs, processes, and outputs that are important, thus providing a pathway to tangible, actionable reform in those areas. Third, they asserted that the state dashboard remains tightly coupled with the school’s socioeconomic status. They suggested that “by stigmatizing and sanctioning low-achieving schools without understanding how well such schools perform across their full mission, we exacerbate

inequality of opportunity. Those harmed, as a result, are those most in need of our care” (p. 755). In a corollary study by some of the same researchers, Schneider et al. (2021) sought to understand the effects of adding student feedback into federal and state accountability models. The researchers found weak relationships between achievement on tests and students’ responses to the feedback survey, as well as a weak correlation between students’ demographics and their survey responses. This is a consequential finding because student feedback is one of the few accountability metrics that does not align with student socioeconomic status, suggesting that the data point does, in fact, capture relevant information about the school itself. Interestingly, schools that serve high populations of disadvantaged students experienced increases in accountability scores when taking into consideration student voice.

Mandinach et al. (2020) expanded on this body of literature by studying the type and format of data that parents valued. They convened 21 focus groups, to include 118 parents, across diverse regions of Missouri. They found that parents want not only quantitative data (e.g., test scores and graduation rates) but also qualitative data that provide contextual information about a school (e.g., a school’s culture and climate, curriculum, homework policies, overall happiness in schools, etc.). Moreover, parents expressed the need for data to be presented in a timely fashion and in a manner that is easy to read and understand. Critically, the researchers stated, “Educators want data that can lead to a process of continuous improvement” whereas parents “want data and information to make informed decisions about their child. That data may well be more qualitative and not easily reside in a data system” (Mandinach et al., 2020, p. 30). This study is limited in its sample (a small group of highly involved parents in one state), and

it may suffer from focus group design issues such as contamination bias. However, the results are thought-provoking and help to illuminate the complexity, nuance, and multi-dimensionality regarding what parents value to evaluate their schools' effectiveness.

**Conclusions.** Several themes emerge from these, and other, studies on new accountability models. First, they underscore the complexity of measuring a school's effectiveness and the diversity of indicators that could be employed to do so. Unlike a private corporation – a car company, for example – whose clear mission is to sell a product and maximize profit – a school's mission is not only diverse, but ever-changing in the eyes of many stakeholders. Any attempt to quantify or describe results, by nature, is exclusionary. Not everything can be included, captured, and presented. But academic literature will inevitably help educators and policymakers to create *better* systems. Second, current accountability models fall short of providing the full range of information that parents and families want. Studies such as Mandinach et al.'s (2020) exploration of family perspectives reveal that families are concerned with more than a limited set of outcome measures; rather, they seek to understand the full school experience, to include inputs, processes, and outputs. Finally, there is increased attention on the importance of separating the school's impact on student outcomes from contextual factors that may be beyond the control of the school. For example, while raw academic proficiency may remain closely coupled with socioeconomic status, students' academic growth, family and student perceptual data, and other input and process indicators are proving to be viable options for better, fairer, and more reliable measures to communicate a school's effectiveness.

## Research Questions

Superintendents and other district leaders find themselves at the epicenter of accountability: receiving their mandate from above while having to implement it within the school district and to communicate it to the larger community. They are policy mediators. Their own experiences shape their perceptions of accountability, how they communicate it to the public and, most interestingly, how they decide to tell their own story of their districts' and schools' quality and effectiveness. The following research questions emerge from this body of literature and aim to expand our evolving understanding of accountability and to provide practical, usable information to contemporary policymakers and leaders wrestling with accountability in their own contexts. The questions are:

- Research Question 1: What are Pennsylvania district leaders' perceptions of mandated accountability? What is the relationship between district contextual attributes and their perceptions?
- Research Question 2: To what extent do Pennsylvania district leaders value various mandated and unmandated (alternative or additional) public-facing accountability measures? What is the relationship between district contextual attributes and the measures they value?
- Research Question 3: How do Pennsylvania district leaders communicate mandated and/or unmandated (alternative or additional) accountability measures to their stakeholders? What is the relationship between district contextual attributes and their communication behaviors?

## CHAPTER THREE: METHODOLOGY

### Introduction to the Methodology

The purpose of this study was to explore the perceptions, attitudes, and behaviors of Pennsylvania superintendents or other district-level leaders as they navigate the complex landscape of educational accountability. Further, it moved beyond simple descriptions and sought to understand how districts' levels of economic disadvantage (hereafter referred to as "ED") may affect those perceptions, attitudes, and behaviors. This deeper analysis lays the groundwork for explanation which, in turn, is needed to support policymaking. Babbie (2020) described explanatory studies as those attempting to explain why things occur or why people have certain beliefs.

Data for this project were derived from two sources: government public education websites and participant interviews. Specifically, school district contextual data were collected from the Pennsylvania Department of Education's Future Ready PA Index, the state's official public-facing accountability dashboard for education. Available data from the Future Ready PA Index include school and district contextual information (size, demographic composition, socioeconomic status, etc.) as well as accountability metrics (standardized test results, graduation rates, college and career readiness indicators, etc.). Additional contextual data were drawn from the U.S. Department of Education's National Center for Education Statistics (NCES). Data from NCES employed in this study include financial data (expenditures and revenues per student), geographic data, and student-to-teacher ratios. Both sets of data are collected annually from every school district in Pennsylvania and the United States, respectively. In both cases, the most recent available data were used (Future Ready PA Index data reflect the 2021-2022 school year,

and the NCES data reflect the 2020-2021 school year). Together, they provide the necessary contextual information for the study.

Subsequently, interview data were collected from 30 Pennsylvania district superintendents or other district-level leaders. The researcher employed a combination of open-ended questions and limited-selection questions in which participants chose an answer from a set number of choices. Some were in the form of Likert-scale questions and others multiple-choice questions. Questions were original and stemmed from information and themes in the existing literature.

This approach is best described as *within-strategy mixed methods data collection*, which Teddlie and Tashakkori (2009) described as the “gathering of both qualitative and quantitative data using the same data collection strategy” (p. 218). Frels and Onwuegbuzie (2013), building upon this work, “contend that the interview, as a natural mode of inquiry, can be enhanced when researchers/interviewers collect data alongside qualitative responses” (p. 188). This allowed the collection of consistent, comparable data across participants while simultaneously exploring topics deeply via open-ended questions. More broadly, Mertens (2019) asserted that a mixed methods approach allows information to be clarified and enhanced to achieve deeper understanding. It is important to note that, due to the limitations of the present study as well as limited access to district leaders (the researcher was fortunate to have about one hour with each participant), the number of questions were narrowed to a moderate sampling of items. Consequently, robust statistical analysis of Likert-scale and other selected-response items was not possible. Interviews ranged from approximately 45 minutes to approximately one hour and fifteen minutes, with an average interview time of about one hour.

## **Sample**

### ***Purposeful Identification of Districts***

There are 499 school districts in Pennsylvania. A sample of 30 school districts was identified using the following process. First, the master file of 2021-2022 Future Ready PA Index data (publicly available) was downloaded. This file contained all contextual and outcome data for every school district in Pennsylvania and is the data file that feeds the public-facing web-based dashboard. Second, the 499 school districts were sorted in order of district-level ED. Economic disadvantage is a district-level data point derived from various possible sources, to include “Temporary Assistance for Needy Families (TANF) cases, census poor, Medicaid, children living in institutions that are neglected or delinquent, those supported in foster homes, or free/reduced price lunch eligibility” (Pennsylvania Department of Education, 2024b). Finally, every 17<sup>th</sup> school district was systematically pulled out of the total list, yielding a representative sample of 30 school districts across the state at recurring intervals of ED.

### ***Recruitment***

Once the specific districts were identified, the researcher sent an email invitation to voluntarily participate in the study to leaders within each district. The invitation was sent in the form of a letter sent via email. The letter introduced the researcher, presented background information on the study, and invited the leaders to participate in one 45-60 minute interview. When appropriate, the email invitation was followed-up with a reminder email in which the information found within the recruitment letter was restated.

If a particular school district was unresponsive or declined the invitation by the requested response date, then the next school district on the list, sorted by percentage of

ED, was invited. For example, the initial invitation may have been sent to school districts 1, 17, 34, 51, 68, etc. If school district 34 declined, then the invitation was transferred to school district 35. An interview was secured with a district leader at each interval or within the interval band, allowing a strong representative sample of the state. The process of systematically ranking, sorting, and inviting school districts in a pre-defined sequence eliminated the risk of selection bias on the part of the researcher.

### **Characteristics of the Sample**

#### ***The Sample in Relation to the State of Pennsylvania***

Table 3.1 shows the contextual information of the 30 sampled school districts in relation to the state of Pennsylvania as a whole. The study's most important variable is each district's percentage of ED. The sample ( $\bar{X}=43.46$ ) closely reflects the whole state ( $\bar{X}=43.51$ ). Moreover, the sample contains districts across the spectrum of ED, from the lowest ED (under 10%) to the highest ED (above 90%).

Table 3.1 also shows other associated district characteristics drawn from the Future Ready PA Index and NCES data sources. Notably, the sample's mean district enrollment ( $\bar{X}=2686.77$ ) is reflective of the whole state ( $\bar{X}=3037.93$ ). This makes sense as Pennsylvania is comprised mostly of small- to medium-sized school districts. The sample is also reflective of the state in terms of students' race and ethnicity (most notably percentages of White, Black/African American, and Hispanic students), gender, and other student identifications such as special education and English learners. Finally, the sample is representative of the state in terms of district revenue and expenditures on a per pupil basis. When funding is further disaggregated, the sample is consistent with the state in that the most revenue comes from local taxes, followed by state sources, and with the

least amount of revenue coming from federal sources. On the expenditure side, the sample is consistent with the state in that instructional spending is most prominent, followed by operational spending, then administrative spending, and with the least amount of per pupil spending on student and staff supports.

**Table 3.1**  
*Sampled Districts' vs. All PA Districts' Contextual Data*

School District Characteristics	Sampled Districts Data (30)		All PA Districts (499)	
	Mean	SD	Mean	SD
Economically Disadvantaged				
Percent Economically Disadvantaged	43.46	18.84	43.51	17.30
District Characteristics				
District Enrollment	2686.77	1867.01	3037.93	5815.45
Geographic Size in Miles	123.24	178.50	90.79	102.26
Number of Schools	4.83	2.76	5.36	10.32
Number of Teachers	190.88	133.62	220.03	397.81
Student:Teacher Ratio	13.86	2.30	13.27	1.84
Charter School Enrollment	155.77	153.20	327.88	3560.56
Enrollment in Partnering Career and Technical Center	105.55	105.35	101.83	100.61
Student Race / Ethnicity				
2 or More Races	4.73	4.12	4.40	3.53
American Indian/Alaskan Native	0.18	0.20	0.14	0.20
Asian	1.53	2.38	2.29	4.04
Black/African American	6.76	15.48	6.70	13.13
Hispanic	5.86	4.84	7.42	10.90
Native Hawaiian/Pacific Islander	0.06	0.07	0.07	0.10
White	80.88	18.79	78.97	21.09
Student Gender*				
Female	48.23	1.10	48.46	1.34
Male	51.77	1.10	51.54	1.34
Student Identifications				
Special Education	17.57	4.30	18.08	3.53
English Learner	1.30	1.28	2.07	3.50
Percent of Gifted Students	2.73	2.08	2.78	1.98
Homeless	1.48	1.27	1.46	1.30
Foster Care	1.00	1.25	0.76	0.70
Military Connected	1.22	2.17	0.59	1.11
District Revenue				
Total Revenue By Student	\$18,962.25	\$3,123.59	\$20,036.24	\$4,247.67
Total Federal Revenue By Student	\$785.21	\$474.84	\$813.77	\$636.03
Total Local Revenue By Student	\$9,745.27	\$3,857.51	\$10,518.38	\$4,437.77
Total State Revenue By Student	\$8,431.77	\$3,567.59	\$8,704.08	\$3,931.43
District Expenditures				
Total Expenditures By Student	\$15,171.05	\$2,319.43	\$15,910.49	\$2,663.71
Instructional Expenditures By Student	\$9,461.16	\$1,359.69	\$9,888.02	\$1,670.22
Student and Staff Support Expenditures By Student	\$1,283.20	\$396.05	\$1,377.57	\$481.84
Administration Expenditures By Student	\$1,645.78	\$401.55	\$1,761.59	\$476.52
Operations Expenditures By Student	\$2,780.91	\$747.66	\$2,883.31	\$721.19

### *The Sample Broken into Tertiles of Economic Disadvantage*

Throughout chapter four, findings will be presented in relationship to districts' level of ED. Specifically, sampled districts have been broken into three tertiles: *lowest ED tertile* (10 districts with the lowest ED percentages), *middle ED tertile* (10 districts in the mid-range of ED percentage), and *highest ED tertile* (10 districts with the highest ED percentage). Table 3.2 shows the range of ED within each tertile using approximate values (specific high and low values are not used in order to maintain the confidentiality of all districts). Moreover, while Table 3.1 presented *all* of the sampled districts in relationship to the state as a whole, a comparison of the *sample's ED tertiles* to the *entire state's ED tertiles* may be found in Appendix B. One will observe similar trends and patterns within each tertile of the sample as across the entire sample. This process provides assurance that the sampled districts, and each respective tertile, is reflective of the state of Pennsylvania as a whole.

**Table 3.2**

*Approximate Range of Economic Disadvantage within Each Tertile*

	Economic Disadvantage	
	Low	High
Lowest ED Tertile	0%	35%
Middle ED Tertile	35%	50%
Highest ED Tertile	50%	95%

Note: Values are approximations.

### *School District Contextual Data*

As discussed, this study's findings will be reported in relationship to the three tertiles of ED. However, there are many important district characteristics beyond ED that help to illuminate a fuller picture of the sample. This section explores the relationships between those other characteristics and the ED tertiles so that, as findings are presented in relation to the ED tertiles in chapter four, richer and more meaningful connections can be made to additional characteristics.

**Districts' Contextual Characteristics In Relation to ED Tertiles.** Table 3.3 examines various district characteristics across the tertiles of ED. Notably, there is a relationship between ED and both district enrollment and geographical size; higher ED districts tend to be smaller in population and larger in terms of geographic size.

**Table 3.3**  
*Crosstabulation of ED Tertiles and Various District Characteristics*

Variable	Economic Disadvantage	Mean	SD	F Stat	$\rho$ value
District Enrollment	Lowest ED Tertile	4132.60	2080.44	8.629	0.001
	Middle ED Tertile	2598.70	1273.76		
	Highest ED Tertile	1329.00	949.43		
Geographic Size (in Square Miles)	Lowest ED Tertile	76.00	66.64	0.622	0.544
	Middle ED Tertile	127.81	85.69		
	Highest ED Tertile	165.90	293.88		
Number of Schools	Lowest ED Tertile	6.70	3.13	5.300	0.011
	Middle ED Tertile	4.60	1.71		
	Highest ED Tertile	3.20	2.20		
Number of Teachers	Lowest ED Tertile	283.22	161.62	6.327	0.006
	Middle ED Tertile	187.91	94.28		
	Highest ED Tertile	101.51	64.49		
Student:Teacher Ratio	Lowest ED Tertile	14.97	1.68	2.639	0.090
	Middle ED Tertile	13.87	1.89		
	Highest ED Tertile	12.73	2.80		
Charter School Enrollment	Lowest ED Tertile	146.80	113.83	0.026	0.974
	Middle ED Tertile	157.90	126.48		
	Highest ED Tertile	162.60	215.69		
Enrollment in Partnering Career and Technical Center	Lowest ED Tertile	125.40	81.44	0.263	0.771
	Middle ED Tertile	92.11	52.32		
	Highest ED Tertile	97.80	157.41		

**Districts' Student Characteristics in Relation to ED Tertiles.** Table 3.4

examines various student characteristics across the tertiles of ED. The highest ED tertile is related to lower White population and higher Black/African American population. The highest ED tertile is also related to the highest percentage of students in special education, who are homeless, and who are in foster care. It is inversely related to the percentage of students in gifted education.

**Table 3.4**  
*Crosstabulation of ED Tertiles and Various Student Characteristics*

Variable	Economic Disadvantage	Mean	SD	F Stat	p value
White (%)	Lowest ED Tertile	84.11	7.23	1.932	0.164
	Middle ED Tertile	86.77	9.19		
	Highest ED Tertile	71.77	29.30		
Black / African American (%)	Lowest ED Tertile	2.20	1.55	2.621	0.091
	Middle ED Tertile	2.65	2.39		
	Highest ED Tertile	15.44	25.27		
Hispanic (%)	Lowest ED Tertile	6.25	3.10	0.090	0.914
	Middle ED Tertile	5.99	6.08		
	Highest ED Tertile	5.33	5.32		
Female (%)	Lowest ED Tertile	48.63	0.70	1.489	0.244
	Middle ED Tertile	47.80	1.44		
	Highest ED Tertile	48.28	0.97		
Male (%)	Lowest ED Tertile	51.38	0.70	1.489	0.244
	Middle ED Tertile	52.21	1.44		
	Highest ED Tertile	51.72	0.97		
Special Education (%)	Lowest ED Tertile	14.80	2.75	22.179	0.000
	Middle ED Tertile	15.65	2.73		
	Highest ED Tertile	22.25	2.74		
English Language Learners (%)	Lowest ED Tertile	1.54	1.06	0.628	0.541
	Middle ED Tertile	0.93	0.97		
	Highest ED Tertile	1.43	1.72		
Gifted (%)	Lowest ED Tertile	4.68	2.13	11.779	0.000
	Middle ED Tertile	1.99	1.28		
	Highest ED Tertile	1.51	1.12		
Homeless (%)	Lowest ED Tertile	0.79	0.53	5.752	0.008
	Middle ED Tertile	1.25	0.74		
	Highest ED Tertile	2.42	1.69		
Foster Care (%)	Lowest ED Tertile	0.35	0.26	3.395	0.048
	Middle ED Tertile	0.97	1.49		
	Highest ED Tertile	1.69	1.31		

**Districts' Financial Data in Relation to ED Tertiles.** Table 3.5 examines financial data across the tertiles of ED. Notably, the highest ED tertile received the most revenue per pupil, but the source of funding varies across tertile. Whereas the highest ED districts receive most of their funding through federal and state sources, the lowest ED districts receive most of their funding through local revenues. This is unsurprising as lower ED districts tend to be more affluent with wealthier property tax bases, and higher ED districts tend to have lower local tax bases with augmented funding from state and federal agencies via supplementary funding such as Title I. Interestingly, the highest ED tertile spends the most per student. This is likely due to increased state and federal supplementary funds directed toward additional teachers, specialists, and support staff to meet students' needs. This observation may challenge the misconception that more affluent districts spend more per pupil and is also relevant within the context of Pennsylvania's ongoing litigation around equitable funding formulas.

**Table 3.5**  
*Crosstabulation of ED Tertiles and Financial Data*

Variable	Economic Disadvantage	Mean	SD	F Stat	p value
Total Revenue Per Student	Lowest ED Tertile	18349.61	2531.43	3.771	0.036
	Middle ED Tertile	17593.04	2636.34		
	Highest ED Tertile	20944.09	3349.04		
Total Federal Revenue Per Student	Lowest ED Tertile	417.75	286.33	20.428	0.000
	Middle ED Tertile	659.79	155.65		
	Highest ED Tertile	1278.09	427.64		
Total Local Revenue Per Student	Lowest ED Tertile	12699.64	2289.12	6.442	0.005
	Middle ED Tertile	8919.18	2701.94		
	Highest ED Tertile	7616.98	4462.98		
Total State Revenue Per Student	Lowest ED Tertile	5232.22	1289.93	23.646	0.000
	Middle ED Tertile	8014.06	1719.86		
	Highest ED Tertile	12049.02	3206.69		
Total Expenditures Per Student	Lowest ED Tertile	14898.00	1651.28	2.440	0.106
	Middle ED Tertile	14240.65	2590.15		
	Highest ED Tertile	16374.50	2290.26		
Total Instructional Expenditures Per Student	Lowest ED Tertile	9540.03	887.90	0.730	0.491
	Middle ED Tertile	9057.20	1827.79		
	Highest ED Tertile	9786.24	1233.80		
Total Staff/Support Expenditures Per Student	Lowest ED Tertile	1279.93	352.41	0.238	0.790
	Middle ED Tertile	1222.14	290.18		
	Highest ED Tertile	1347.54	536.91		
Total Administrative Expenditures Per Student	Lowest ED Tertile	1544.21	180.92	5.484	0.010
	Middle ED Tertile	1452.01	276.53		
	Highest ED Tertile	1941.11	510.17		
Total Operations Expenditures Per Student	Lowest ED Tertile	2533.83	775.96	4.479	0.021
	Middle ED Tertile	2509.30	609.81		
	Highest ED Tertile	3299.61	615.21		

**Districts' Locale in Relation to ED Tertiles.** Table 3.6 examines district locale across the tertiles of ED. The lowest ED tertile districts tend to be more suburban while the highest ED districts tend to be more rural.

**Table 3.6**  
*Crosstabulation of ED Tertiles and District Locale*

	Locale			Total
	Suburb	Town	Rural	
Lowest ED Tertile	8 (80%)	0 (0%)	2 (20%)	10 (100%)
Middle ED Tertile	3 (30%)	3 (30%)	4 (40%)	10 (100%)
Highest ED Tertile	4 (40%)	0 (0%)	6 (60%)	10 (100%)
	15 (50%)	3 (10%)	12 (40%)	30 (100%)

$\chi^2=10.8, \rho=0.029$

### *Participant Contextual Data*

The study's participants were Pennsylvania superintendents or other district-level leaders. This section describes the leaders' characteristics in relation to the ED tertiles. Like the district contextual characteristics, the leaders' characteristics provide valuable context around their perceptions, attitudes, and behaviors.

Two of the 30 interviews involved two participants rather than one participant. Consequently, there were 32 unique participants within the sample of 30 school districts. In these two unique situations, only the demographic information of the leader with more district-level experience in the school district was included in the demographic calculations that follow. This is purely related to clarity and consistency of presentation by maintaining a focus on 30 school districts. Fortunately, for all quantitative interview questions, the two participants in each interview agreed upon Likert-scale and other ratings. Still, the unique contributions of the second participant in these two interviews are highly valuable and will be interwoven into the qualitative findings presented in chapter four.

**Participants' Roles in Relation to ED Tertiles.** Table 3.7 shows that most of the participants (24) in this sample are superintendents. The remaining six participants are other district leaders, to include assistant superintendents or assistants to the superintendent (4), a director of curriculum and instruction (1), and a director of operations (1). (Note: The two participants not included in these calculations were both district-level leaders within their respective school districts.)

**Table 3.7**  
*Crosstabulation of ED Tertiles and District Leaders' Roles*

	Current Position		Total
	Superintendent	Other District Leader	
Lowest ED Tertile	8 (80%)	2 (20%)	10 (100%)
Middle ED Tertile	7 (70%)	3 (30%)	10 (100%)
Highest ED Tertile	9 (90%)	1 (10%)	10 (100%)
	24 (80%)	6 (20%)	30 (100%)

$\chi^2=1.250$ ,  $p=ns$

**Participants' Demographics in Relation to ED Tertiles.** Tables 3.8, 3.9, 3.10, and 3.11 show participants' gender, age, race/ethnicity, and highest degree earned within each ED tertile. Most of the participants (22) were male, most (20) were between 50 and 59 years old, almost all (29) were White (there was one African American superintendent, who was also female), and most (19) earned a doctorate degree. In terms of educational attainment, fewer district leaders within this sample hold a doctorate degree in the highest ED tertile.

**Table 3.8**  
*Crosstabulation of ED Tertiles and District Leaders' Gender*

	Gender		Total
	Female	Male	
Lowest ED Tertile	3 (30%)	7 (70%)	10 (100%)
Middle ED Tertile	3 (30%)	7 (70%)	10 (100%)
Highest ED Tertile	2 (20%)	8 (80%)	10 (100%)
	8 (27%)	22 (73%)	30 (100%)

$\chi^2=.341$ ,  $p=ns$

**Table 3.9**  
*Crosstabulation of ED Tertiles and District Leaders' Age*

	Age		Total
	40-49	50-59	
Lowest ED Terstile	2 (20%)	8 (80%)	10 (100%)
Middle ED Terstile	5 (50%)	5 (50%)	10 (100%)
Highest ED Terstile	3 (30%)	7 (70%)	10 (100%)
	10 (33%)	20 (67%)	30 (100%)

$\chi^2=2.1$ ,  $p=ns$

**Table 3.10**  
*Crosstabulation of ED Tertiles and District Leaders' Race/Ethnicity*

	Race/Ethnicity		Total
	White	Black/African American	
Lowest ED Terstile	10 (100%)	0 (0%)	10 (100%)
Middle ED Terstile	10 (100%)	0 (0%)	10 (100%)
Highest ED Terstile	9 (90%)	1 (10%)	10 (100%)
	29 (97%)	1 (3%)	30 (100%)

**Table 3.11**  
*Crosstabulation of ED Tertiles and District Leaders' Highest Degree Earned*

	Highest Degree Earned		Total
	Doctorate Degree	Master's Degree	
Lowest ED Terstile	8 (80%)	2 (20%)	10 (100%)
Middle ED Terstile	8 (80%)	2 (20%)	10 (100%)
Highest ED Terstile	3 (30%)	7 (70%)	10 (100%)
	19 (63%)	11 (37%)	30 (100%)

$\chi^2 = 7.177$ ,  $p=.028$

**Participants' Experience Levels in Relation to ED Tertiles.** Finally, Tables 3.12, 3.13, 3.14, and 3.15 delineate participants' educational and leadership experience. Participants have been in their current positions from one month to 15 years. They held *any* district-level position from one to 24 years. Overall, this is an experienced group of educators with between 16 and 33 years in public education, and between seven and 33

years as an educator within Pennsylvania. There were no statistically significant differences among the tertiles in any of these areas.

**Table 3.12**

*Crosstabulation of ED Tertiles and District Leaders' Years in Current Position*

	Participant Years in Current Position			Total
	Lowest Third (0-1 year)	Middle Third (1.1-5 years)	Highest Third (5.1 to 15 years)	
Lowest ED Tertile	2 (20%)	4 (40%)	4 (40%)	10 (100%)
Middle ED Tertile	4 (40%)	3 (30%)	3 (30%)	10 (100%)
Highest ED Tertile	4 (40%)	3 (30%)	3 (30%)	10 (100%)
	10 (33.3%)	10 (33.3%)	10 (33.3%)	30 (100%)

$\chi^2=1.2$ ,  $p=ns$

**Table 3.13**

*Crosstabulation of ED Tertiles and District Leaders' Years in Any District Level Position*

	Participant Years in Any District-Level Position			Total
	Lowest Third (0-6 years)	Middle Third (7-13 years)	Highest Third (14-24 years)	
Lowest ED Tertile	2 (20%)	4 (40%)	4 (40%)	10 (100%)
Middle ED Tertile	6 (60%)	2 (20%)	2 (20%)	10 (100%)
Highest ED Tertile	5 (50%)	2 (20%)	3 (30%)	10 (100%)
	13 (43%)	8 (27%)	9 (30%)	30 (100%)

$\chi^2=3.667$ ,  $p=ns$

Note: Column totals are not perfect thirds because some data were duplicated across thirds.

**Table 3.14**

*Crosstabulation of ED Tertiles and District Leaders' Years in Public Education*

	Participant Years in Public Education			Total
	Lowest Third (0-24 years)	Middle Third (25-29 years)	Highest Third (30-33 years)	
Lowest ED Tertile	3 (30%)	4 (40%)	3 (30%)	10 (100%)
Middle ED Tertile	4 (40%)	1 (10%)	5 (50%)	10 (100%)
Highest ED Tertile	3 (30%)	6 (60%)	1 (10%)	10 (100%)
	10 (33%)	11 (37%)	9 (30%)	30 (100%)

$\chi^2=6.321$ ,  $p=ns$

Note: Column totals are not perfect thirds because some data were duplicated across thirds.

**Table 3.15**  
*Crosstabulation of ED Tertiles and District Leaders' Years as a Pennsylvania Educator*

	Participant Years in Pennsylvania			Total
	Lowest Third (0-22 years)	Middle Third (23-29 years)	Highest Third (30-33 years)	
Lowest ED Tertile	2 (20%)	5 (50%)	3 (30%)	10 (100%)
Middle ED Tertile	4 (40%)	1 (10%)	5 (50%)	10 (100%)
Highest ED Tertile	4 (40%)	5 (50%)	1 (10%)	10 (100%)
	10 (33%)	11 (37%)	9 (30%)	30 (100%)

$\chi^2=6.376$ ,  $p=ns$

Note: Column totals are not perfect thirds because some data were duplicated across thirds.

### Benefits and Limitations of the Sample

The benefit of using an economic disadvantage indicator to identify school districts is that socioeconomic status has been shown to be related to a range of other academic outputs. In a meta-analysis of 20 years of education data, Sirin (2005) asserted that socioeconomic status is both directly linked to academic achievement and indirectly linked to a range of other student and family characteristics, to include racial and ethnic background, school and neighborhood location, access to home resources and social capital, and students' likelihood of attending schools that are adequately staffed and funded. A more contemporary factsheet published by the American Psychological Association affirmed that little has changed in the past two decades. The organization noted that the literacy skills of children from low socioeconomic backgrounds is five years behind their high socioeconomic peers (Reardon et al., 2013); that the National Center for Education Statistics shows that the low socioeconomic dropout rate was 11.6% compared to only 2.8% for high socioeconomic students; and that success rates in science, technology, engineering, and math are lower for students with low socioeconomic status (American Psychological Association, 2017; Doerschuk et al., 2016). Moreover, researchers' understanding of the relationship among socioeconomic

status, brain plasticity, and the learning process is evolving rapidly and will impact both educational practice and policy development (Farah, 2018). Therefore, the socioeconomic status variable represents a logical way to identify districts for this study because it relates directly to metrics within federal- and state-mandated accountability systems.

Of course, there are limitations to this approach, as well. Identifying 30 school districts based on socioeconomic status does not necessarily account for other important differences among districts. For example, large, urban districts tend to be clustered at the lower end of the socioeconomic ranking, yielding a lower representation of these districts in the study. Moreover, this type of a systematic sample does not guarantee that other important district characteristics will be equally represented. While the data above provide information about how other district characteristics relate to economic disadvantage, they are not perfect correlations. Each of the contextual factors, and a host of others not addressed in this study, could yield a different sample of school districts.

## **Data Issues Related to School Measurement and Accountability**

### ***Operationalization and Explanation of Variables***

This study is situated within the contemporary literature exploring how we publicly measure schools in the *ESSA*-era. Bae (2018) described key features of “next phase” accountability, arguing that, to be competitive and relevant in a 21st century world with increasing complexity and challenges, “schools and districts must be accountable for more than just testing and reporting on a narrow set of outcomes aimed at minimum levels of competency” (p. 3). Specifically, a better form of accountability would encompass five key features: (1) a broader set of outcome measures, (2) a mix of

state and local indicators, (3) measures of opportunities to learn, (4) data dashboards, and (5) school quality reviews by diverse, external stakeholders that allow the deeper and more comprehensive study of a school's strengths and weaknesses. She provided specific examples – California's three-tiered state and local accountability system, New York City's inclusion of school and family climate data, and Alberta, Canada's development of an interactive data dashboard that measures an array of indicators – as promising exemplars of next phase accountability.

Other researchers have sought to understand the implementation of accountability at the state level. For example, Portz (2021) studied the various ways in which states executed *ESSA*'s mandate for a "school quality or student success" indicator (example: school climate data). He investigated each state's *ESSA* plan to ascertain the weighting it assigned to this new indicator. He found that most states allocate less than 10% of their accountability model to this indicator, while only two states (California and Maryland) allocated 15% or more. He determined that states have taken a mostly cautious approach to new measures, reverting to straightforward indicators such as chronic absenteeism and college and career readiness metrics instead of alternate measures. He further concluded that the states at the forefront of "next generation" accountability tend to be more liberal states.

It is important to acknowledge that the academic discussion on how to measure a school's performance is not a new one. For decades following the implementation of *ESEA* and the publication of *The Coleman Report*, scholars discussed and debated how accountability systems could be changed and potentially improved. For example, Benviste's (1985) policy analysis asserted that accountability's three primary functions

are to inform, reorient action, and justify what is done. Further, it is a complex undertaking to choose a combination of input, process, and output measures that include both objective (mostly quantitative) and subjective (mostly qualitative) data. Most importantly, he concluded that there is “good accountability” and “bad accountability.” Good accountability is “the careful selection of specific measures that are or can be available,” that “measure what is significant,” and that “can readily translate into new patterns of action.” It does not measure “what is beyond teachers’ and schooling’s ability to change.” By contrast, bad accountability “measures what is difficult to measure” and “provides considerable information about what is wrong, and little about what is right or what can be done to improve the endeavor” (p. 265).

Other researchers have evaluated specific accountability indicators. Färe et al. (1989) studied alternate models of accountability by statistically measuring both student outcomes (in this study, on Missouri’s standardized assessment, the BEST test) as well as input measures, to include factors such as teacher experience, training opportunities, credentials of administrative staff, library facilities, curriculum, etc. They suggested that the results of such modeling provide a more sensitive, statistically accurate, and holistic picture of a school’s effectiveness. Similarly, Oakes (1989) asserted the importance of what she called *context indicators* – specifically, access to knowledge, the press for achievement, and professional teaching conditions – in formal accountability models. Including such measures acknowledges that stakeholders value more than outcomes alone, reduces the likelihood of gaming or altering the system to increase a few isolated outcomes, and provides important information to policymakers and educators about what may be driving results.

Taken together, this body of research undergirds the present study. The literature suggests how we measure schools is an unsettled and contested matter, and that states and individual districts interpret and respond to federal and state accountability in divergent ways. The flexibility afforded by *ESSA* has created a ripe opportunity to think differently about how students and schools are measured, and how we communicate quality and effectiveness to stakeholders. This study aims to contribute to our collective knowledge about *ESSA*-era accountability by examining Pennsylvania educational leaders' attitudes and actions.

### ***How Pennsylvania Measures Schools: The Future Ready PA Index***

Following the enactment of *ESSA*, each state developed and submitted to the U.S. Department of Education a state-level plan. As part of this plan, states were tasked with the development of a new accountability model that included both required federal performance indicators as well as additional indicators at the discretion of each state. Required measures include academic achievement, graduation rates for high schools, academic progress for elementary and middle schools, progress in attaining English language proficiency, and at least one state-selected indicator of “school quality or student success,” each identified as a “substantial indicator.” (U.S. Department of Education, 2024c, p. 2).

Taken together, these “substantial indicators” are employed by the state to identify schools in need of improvement. While *ESSA* represents a departure from *NCLB* in that student performance on standardized test scores is not the sole measure of success, the accountability system continues to rely on tests as its flagship indicators. To ensure that states remain committed to properly measuring and weighting academic proficiency,

the U.S. Department of Education stipulates that schools that avoid comprehensive improvement status because of the state-selected measures must at least demonstrate “significant progress” on the substantial indicators (U.S. Department of Education, 2024c, p. 2). This is designed to guard against states gaming the system with their own, possibly less rigorous, measures to avoid the consequences of poor performance on standardized metrics. The result is an accountability system that is broader and more holistic than ever before while, simultaneously, staying true to *NCLB*-era concepts on standardized testing. In this manner, *NCLB*’s influence is apparent in the transition to *ESSA*.

It is within this context that Pennsylvania developed its public-facing dashboard of indicators called the Future Ready PA Index. According to the Pennsylvania Department of Education’s (PDE) website (2024a), the Index “is the result of discussions with thousands of educators, parents, advocates, policymakers, and business leaders across Pennsylvania about how communities should evaluate schools” (para. 2).

Table 3.16 shows the indicators that comprise the Future Ready PA Index. The table includes *Indicator*, showing the title of each performance metric, *Description*, providing a description of each metric, *Federal*, indicating whether the metric is a required component of *ESSA*, and *Approach*, showing how Pennsylvania categorizes the metric.

Four approaches to the public presentation of data are utilized. The first, called *2033 Goal and Targets*, measures both a 2033 goal and interim, yearly targets for the indicators of Achievement, English Learner Progress, and Graduation Rate. The second, *Academic Growth*, presents the school’s impact on year-to-year changes in academic

performance on the PSSA and Keystone English/Literature examinations, PSSA and Keystone Mathematics/Algebra I examinations, and PSSA and Keystone Science/Biology examinations. The third, *Performance Standards*, depicts norm-referenced presentations of school performance employing cut scores at the 75<sup>th</sup> and 25<sup>th</sup> percentiles. This method is utilized for the indicators of Regular Attendance, Career Standards Benchmark, Industry-Based Learning, and Rigorous Courses of Study. The final approach, *Informational*, presents additional data as a percentage of students who attain specific indicators. This approach is utilized for the indicators of Advanced on State Assessments, Industry-Based Learning, Advanced on Industry Competency Assessment, Rigorous Courses of Study, Postsecondary Transitions to School, Military or Work, and Early Indicators of Success. The indicators are further broken down into three sections on the public-facing dashboard, to include: (1) Statewide Assessments, (2) On Track Measures, and (3) College and Career Readiness Measures (Pennsylvania Department of Education, 2024a).

**Table 3.16**  
*Pennsylvania Future Ready PA Index Accountability Indicators*

Indicator	Description	Federal	Approach
<b>Statewide Assessments</b>			
Percent Proficient or Advanced	Percentage of students who scored Proficient or Advanced on the PSSA, Keystone, or PASA examinations	Yes	2033 State Goal
Meeting Annual Growth Expectations (PVAAS)	Growth index demonstrating extent to which students achieved, exceeded, or fell short of one year's worth of academic growth	Yes	Academic Growth
Percent Advanced	Percentage of students who scored Advanced on the PSSA, Keystone, or PASA examinations	No	Informational
<b>On-Track Measures</b>			
English Language Proficiency	Percentage of English Learners (EL) who met their interim growth target or attainment level ahead of time	Yes	2033 State Goal
Regular Attendance	Percentage of students enrolled for 90 or more school days and present 90 percent or more school days	No	Performance Standard
Early Indicators of Success	Percentage of students who attain proficiency on the Grade 3 Reading PSSA and the Grade 7 Math PSSA	No	Informational
<b>College and Career Readiness Measures</b>			
Graduation Rate	Percentage of students in a cohort graduating with a high school diploma on or before the reporting year	Yes	2033 State Goal
Career Standards Benchmark	Percentage of students completing the required number of career-based activities at specific grade bands	No	Performance Standard
Industry-Based Learning	Percentage of students who, between grades seven and twelve, scored competent on an industry-based assessment, earned at least one industry-recognized credential, or completed a work-based learning experience	No	Informational
Advanced on Industry-Based Competency Assessment	Percentage of students who scored advanced on an industry-based competency assessment	No	Informational
Rigorous Courses of Study	Percentage of grade twelve students who participated in AP, IB, or dual credit courses in a concentrated CTE program of study	No	Informational
Postsecondary Transition to School, Military, and Work	Percentages of graduates who enrolled in an institution of higher education, enlisted in the military, or entered the workforce	No	Informational

### **Research Questions, Survey Design, and Interview Design**

Data were collected via one interview with each participant that lasted approximately one hour. The interview contained a combination of open-ended questions and selected-response questions in which participants chose an answer from a set number of choices. Some of the questions were in the form of Likert-scale questions and others multiple-choice questions. According to Babbie (2008),

The particular value of this format is the unambiguous ordinality of response categories. If respondents were permitted to volunteer or select such answers as ‘sort of agree,’ ‘pretty much agree,’ ‘really agree,’ and so forth, you would find it impossible to judge the relative strength of agreement intended by the various respondents. The Likert format solves this problem. (p. 188)

This interweaving of both qualitative and quantitative questions allowed the researcher to collect consistent data across participants while also exploring topics deeply via open-ended questions. The questions were original and stemmed from information and themes in the existing literature. The full interview protocol may be found in Appendix A. Table 3.17 depicts the connections among this study’s research questions, the literature, and the associated interview questions.

**Table 3.17***Research Questions, Connections to Literature, Operationalization, and Associated Interview Questions*

Element	Description
	<p data-bbox="315 453 1398 510">Research Question 1: What are Pennsylvania district leaders' perceptions of mandated accountability? What is the relationship between district contextual attributes and their perceptions?</p>
<p data-bbox="298 905 521 961">Critical Connections to Literature Review</p>	<ul data-bbox="574 579 1409 1350" style="list-style-type: none"> <li>• Accountability is conceptualized in various ways by various stakeholders (Behn, 2003; Finn, 2002; Glass, 1972; Hanushek &amp; Raymond, 2001; Levin, 1974; Wescott, 1972).</li> <li>• Studies show mixed results on academic gains following standardized test implementation (Carnoy &amp; Loeb, 2002; Dee &amp; Jacob, 2011; Hanushek &amp; Raymond, 2002, 2005; Jacob, 2005; Neal &amp; Schanzenbach, 2010; Springer, 2008).</li> <li>• In-class and instructional behaviors change because of standardized testing and mandated accountability (Amrein-Beardsley, 2010; Dee et al., 2013; Hibel &amp; Penn, 2020; Koretz, 1996, 2001; Ladd, 2001; Rouse, 2013; Stecher, 2000).</li> <li>• Outside-of-school and other school behaviors change because of standardized testing and mandated accountability (Figlio, 2006; Figlio &amp; Getzler, 2006; Figlio &amp; Winicki, 2005).</li> <li>• Standardized testing and mandated accountability can impact larger educational systems, such as teacher recruitment and retention (Boyd, 2008; Clotfelter, 2004; Ingersoll, 2016).</li> <li>• Standardized testing and mandated accountability can impact communities and housing markets (Figlio &amp; Lucas, 2004; Imberman &amp; Lovenheim, 2016; Kane et al., 2003).</li> <li>• Research reveals mixed results of the statistical validity of standardized tests as an effective way to measure student and school performance (Goldhaber &amp; Ozek, 2019; Green, 2004; Hanushek &amp; Raymond, 2002; Howe &amp; Murray, 2015; Kain &amp; Staiger, 2002; Koretz, 2008; Meyer et al., 2014; Monarrez &amp; Chingos, 2020; Price, 2010; Tienken, 2017).</li> </ul>
<p data-bbox="298 1444 500 1472">Operationalization</p>	<p data-bbox="574 1419 1409 1528">The interview protocol contained Likert-scale and multiple-choice questions that asked participants to rate their agreement with or assessment of a series of statements related to mandated accountability. These were coupled with open-ended interview questions.</p>

**Table 3.17 continued***Research Questions, Connections to Literature, Operationalization, and Associated Interview Questions*

Element	Description
Example Questions	<ul style="list-style-type: none"> <li>• (OE) Please describe your general impression of the federally- and state-mandated accountability measures, such as those found in the Future Ready PA Index.</li> <li>• (OE) In what ways do federally- and state-mandated accountability measures, such as those in the Future Ready PA Index, positively affect/impact your district, or advance your educational goals?</li> <li>• (OE) In what ways do federally- and state-mandated accountability measures, such as those in the Future Ready PA Index, negatively affect/impact your district, or hinder your educational goals?</li> <li>• (SR) Rate these statements using a Likert-scale. Your options are: you “strongly agree”, “agree,” are “neutral” on the statement, “disagree,” or “strongly disagree.” You may also answer that the statement is “not applicable” or that you are “undecided.” I will also ask you to briefly explain your rationale for each selection. [Followed by 11 statements related to the federal and state accountability system.]</li> <li>• (OE) How do you use federally- and state-mandated accountability measures, such as those found in the Future Ready PA Index, in your leadership capacity?</li> </ul>
<p>Research Question 2: To what extent do Pennsylvania district leaders value various mandated and unmandated (alternate or additional) public-facing accountability measures? What is the relationship between district contextual attributes and the measures they value?</p>	
Critical Connections to Literature Review	<ul style="list-style-type: none"> <li>• Accountability is conceptualized in various ways by various stakeholders (Behn, 2003; Finn, 2002; Glass, 1972; Hanushek &amp; Raymond, 2001; Levin, 1974; Wescott, 1972).</li> <li>• Standardized testing and mandated accountability can impact communities and housing markets (Figlio &amp; Lucas, 2004; Kane et al., 2003; Imberman &amp; Lovenheim, 2016).</li> <li>• School and district leaders have various perceptions of mandated accountability and the relative importance of various performance measures (Abelman et al., 1999; Dessimone, 2006; Englert et al., 2007; Knapp &amp; Feldman, 2012; Mintrop, 2012; Robinson, 2012; Rorrer &amp; Skrla, 2005; Russell &amp; McCombs, 2006; Sutherland, 2022).</li> <li>• Families and communities have various perceptions of mandated accountability and the relative importance of various performance measures (Brewer et al., 2015; Mandinach et al., 2020; Owens &amp; Peltier, 2002).</li> <li>• Researchers are actively exploring various effects of alternate and/or additional performance measures, suggesting opportunities for better, fairer, or more holistic school performance profiles (Gagnon and Schneider, 2019; Sare et al., 1989; Schneider et al., 2021).</li> </ul>

**Table 3.17 continued***Research Questions, Connections to Literature, Operationalization, and Associated Interview Questions*

Element	Description
Operationalization	<p>The interview protocol contained open-ended questions that allowed the participant to describe performance measures of importance to them. Additionally, participants were asked to rate the extent to which they believe particular performance measures are valuable as public-facing accountability metrics.</p>
Example Questions	<ul style="list-style-type: none"> <li>• (OE) Think of your ideal public-facing accountability dashboard. What measures, data, or information should be publicly presented to “tell the story” of a school’s performance?</li> <li>• (SR) In this question, I will identify several areas from the existing Future Ready PA Index. Please rate the extent to which you believe the following accountability measures are valuable public-facing accountability measures. Your choices are: <i>highly valuable, somewhat valuable, not valuable</i>. I will ask you to explain your answer for some of these. You are welcome to explain your answer for any of them. [Followed by a listing of the indicators from the current Future Ready PA Index.]</li> <li>• (SR) The following group of accountability measures are not currently mandated measures. Please rate the extent to which you believe the following would be a valuable public-facing accountability measure. For this exercise, you may presume they would be presented similarly to any other measure on the Future Ready PA Index and disaggregated by subgroup. Your choices are: <i>highly valuable, somewhat valuable, or not valuable</i>. [Followed by a listing of indicators <i>not</i> currently included in the Future Ready PA Index.]</li> <li>• (OE) Are there other measures, not listed here, that you believe would be valuable or highly valuable, regardless of whether or not you employ them in your own district? Why do you find them valuable?</li> <li>• (OE) Outside of the mandated federal and state accountability measures, what other measures, if any, do you publicly share with your stakeholders?</li> </ul>
<p>Research Question 3: How do Pennsylvania district leaders communicate mandated and/or unmandated (alternate or additional) accountability measures to their stakeholders? What is the relationship between district contextual attributes and their communication behaviors?</p>	
Critical Connections to Literature Review	<ul style="list-style-type: none"> <li>• Families and communities have various perceptions of mandated accountability and the relative importance of various performance measures (Brewer et al., 2015; Mandinach et al., 2020; Owens &amp; Peltier, 2002).</li> <li>• Families’ and communities’ perceptions of accountability are affected by how it is presented (Houston et al., 2022; Jacobsen et al., 2013, 2014; Mandinach et al., 2020).</li> </ul>

**Table 3.17 continued**

*Research Questions, Connections to Literature, Operationalization, and Associated Interview Questions*

Element	Description
Operationalization	<p>The interview protocol contained Likert-scale or multiple-choice questions that asked participants to rate their agreement with or assessment of a series of statements related to their communication of school and district performance measures to their communities. These were coupled with open-ended interview questions.</p>
Example Questions	<ul style="list-style-type: none"> <li>• (OE) In your district, who is the primary communicator of federal and state accountability results to the school board?</li> <li>• (OE) In your district, who is the primary communicator of federal and state accountability results to the parents/families?</li> <li>• (SR) On average, which best describes how often you (or someone else) present federal and state accountability results (the Future Ready PA Index) to the following stakeholders? Your options are <i>never, less than annually, annually, bi-annually, quarterly, more than quarterly</i>. [Followed by a listing of five stakeholder groups.]</li> <li>• (SR) This is a multiple-choice question: Which best describes the frequency of questions, inquiries, or comments from families / parents regarding federal and state accountability? [Followed by 3 choices.]</li> <li>• (SR) This is a multiple-choice question: Which best describes the frequency of questions, inquiries, or comments from community members (not families / parents) regarding federal and state accountability? [Followed by three choices.]</li> <li>• (OE) How do you publicly share information on accountability measures to families? (What communication practices do you employ?)</li> <li>• (OE) What benefits have you experienced related to the public presentation of mandated accountability measures?</li> <li>• (OE) What challenges have you experienced related to the public presentation of, or public understanding of, mandated accountability measures?</li> <li>• (OE) What practices have you employed to overcome these challenges?</li> </ul>

### **Methodological Procedures**

To conduct this research, data were gathered and analyzed using an interview protocol designed to obtain the perspectives of educational leaders. Procedures included:

*Step 1:* The listing of all Pennsylvania school districts was downloaded from the Future Ready PA Index website. (The Index is described above.) A probability systematic sample was chosen from this list.

*Step 2:* Once the specific districts were identified, the researcher sent an email invitation to participate in the study to leaders within each district. The invitation was in the form of a letter sent via email. The letter introduced the researcher, presented background information on the study, and invited leaders to participate in one 45-60 minute interview. When appropriate, the email invitation was followed-up with a reminder email in which the information found within the recruitment letter communication was restated.

*Step 3:* District leaders who agreed to participate were sent several dates and times to schedule the interview via a Zoom meeting.

*Step 4:* Data were collected via one interview. The interview contained a combination of open-ended questions and selected-response questions in which participants chose an answer from a set number of choices. When appropriate, questions were displayed on the computer screen.

*Step 5:* Data were analyzed according to the procedures in the subsequent section.

## **Data Analysis**

### ***Selected-Response Questions***

Participants' selected-response answers were aggregated on spreadsheets and analyzed to identify emergent trends and patterns. The inclusion of these questions successfully generated measurable data that could be compared and analyzed across numerous participants. Grim et al. (2006) referred to this technique as *quanti-qualitative*

*methodology*, which involves “the strategic incorporation of quantitative techniques into a qualitative method in such a way as to make the results more empirically transparent” (p. 517). These researchers, who described the strategy in relation to focus groups, sought, “both the subjective opinions of those we interviewed and some empirical measurements related to the opinions expressed” (p. 517). This aptly describes the objective in the present study. The quantitative measures, while not statistically generalizable, do, in fact, ground participants’ open-ended responses in empirical data.

Data were analyzed by employing univariate and bivariate statistics. Univariate measures captured the frequency of responses to questions across the entire sample of superintendents or district leaders (example: do most superintendents identify state mandated measures as being an important driver of their strategic planning?). Bivariate measures illuminated patterns and trends within the data (example: do more superintendents within the highest ED tertile believe that state mandated measures are indicative of their schools’ effectiveness?).

In some instances, a *mean value* will be provided in relation to Likert-scale or other selected-response questions. In this study, the mean value *does not* indicate the strength of agreement or disagreement (it is understood that Likert data cannot be *averaged* to determine how strongly the group agreed with something). The mean value simply expresses, in mathematical terms, a single value to capture the intensity and distribution of participants’ responses. When each selected-response item is converted to a number (highly agree=5, agree=4, etc.), then a higher mean indicates that, in the aggregate, participant responses trended higher on the Likert- or other scale *in relation to* another group (such as one tertile in relation to another tertile). Mean values will most

often be found in the appendices and may be used simply to reinforce the frequency data that are observable through the embedded tables and bar charts.

### *Open-Ended Questions*

Open-ended responses were recorded via note-taking during the interviews. The researcher captured major themes and ideas of the participants. Written notes were then transferred and organized within master documents. The researcher employed a process of organizing the data within each question into major ideas, open coding to identify important clusters or groups of ideas, and subsequently axial coding to capture higher-level, emergent themes that illuminate the quantitative data and help to answer the research questions as a whole (Babbie, 2008).

The combination of the selected-response data and open-ended response data provided a more robust and holistic picture of district leaders' perceptions and attitudes toward school accountability.

## CHAPTER FOUR: FINDINGS

The purpose of this study was to explore the perceptions, attitudes, and behaviors of Pennsylvania district leaders as they navigate the complex landscape of educational accountability. More specifically, it examined the degree to which districts' levels of economic disadvantage (hereafter referred to as "ED") affect those perceptions, attitudes, and behaviors. As detailed in chapter three, the sample is comprised of leaders from Pennsylvania districts with varying levels of student ED, purposefully sampled to represent the entire state. The research questions being addressed are as follows:

- Research Question 1: What are Pennsylvania district leaders' perceptions of mandated accountability? What is the relationship between district contextual attributes and their perceptions?
- Research Question 2: To what extent do Pennsylvania district leaders value various mandated and unmandated (alternate or additional) public-facing accountability measures? What is the relationship between district contextual attributes and the measures they value?
- Research Question 3: How do Pennsylvania district leaders communicate mandated and/or unmandated (alternate or additional) accountability measures to their stakeholders? What is the relationship between district contextual attributes and their communication behaviors?

Table 4.1 presents a series of notes as a guide for readers.

**Table 4.1**  
*Data and Findings Concepts*

Term	Description
Sample and Tertiles	Data are presented in terms of the whole sample (n=30) or as comparisons between tertiles of ED (each tertile, n=10).
Economic Disadvantage Abbreviation	For ease of reading, “economic disadvantage” will be presented as “ED.” It is important to emphasize that the “lowest ED tertile” essentially means “lowest student poverty.” Conversely, the “highest ED tertile” essentially means “highest student poverty.”
Frequencies and Means	Data are presented predominantly as frequencies. However, sometimes a mean will be noted to capture a relationship between tertiles or among variables. When a mean is presented, it is only for comparative purposes to demonstrate the distribution and intensity of participant responses <i>in relation to</i> another group. The mean carries no other statistical meaning. This was described at length in chapter three.
Participant Qualitative Data	When participant qualitative data are presented, the individual will be identified with the letter “P” (to designate “participant”) and a number. P1 through P10 are in the lowest ED tertile. P11 through P20 are in the middle ED tertile. P21 through P30 are in the highest ED tertile. Therefore, it can be assumed that P29, for example, is the leader of a district with especially high student ED.
Participant Contextual Data	For ease of reading, a chart describing basic participant data is found in Appendix C. The reader may quickly cross reference any participant comment with their years of experience, gender, age, race/ethnicity, the highest degree earned, and the district enrollment. Hopefully, this helps to provide context that enriches the participants’ comments.

**Research Question 1: What are Pennsylvania district leaders' perceptions of mandated accountability? What is the relationship between district contextual attributes and their perceptions?**

In 2018, Pennsylvania implemented a new and more holistic public-facing dashboard of educational outcomes called the Future Ready PA Index. District leaders are the policy mediators responsible for communicating and explaining the accountability dashboard to their stakeholders at the local level. This research question explored leaders' perceptions of mandated accountability within this context. Perceptions were examined through a variety of lenses that, taken together, paint a multi-faceted picture. First, data depicting participants' current positions on the measures will be presented. Second, quantitative data will show leaders' perceptions within the following areas: understanding of the system, the value and utility of the system, confidence in the accuracy of the system, the effects and outcomes of the system, and their beliefs on whether current resources and capacity can meet state expectations. Finally, qualitative interview data will be presented to augment and further illuminate the findings.

***Participants' Positions on Mandated Accountability***

Participants were asked to identify the statement, from among a set of choices, that best reflected their present position on mandated accountability (see Table 4.2). Half of the participants (15 out of 30) indicated that they are *satisfied with the existing measures, but would like to add additional measures* to more holistically tell a school's story. Specifically, these participants indicated that they value the more comprehensive dashboard and its consistency across school districts, but that broader measures are

needed to capture what schools do, especially in terms of responding to socioeconomic and student wellness needs.

No other choice exceeded five participants. Participants who indicated that they *like the measures and would not change them* (2) generally cited the rigor and consistency of the process as positive attributes. Participants who indicated that they *did not like the measures and would change them* (4) tended to cite overly narrow, simplified, or confusing metrics, and the need to better reflect individual communities' values. Participants who indicated that they *don't believe publicly reporting measures is the right approach to accountability* (5) argued that the measures do not appropriately reflect what schools do, that comparisons are inappropriate, and that it siphons time and resources away from what is most important. The remaining participants (4) were either neutral or expressed an alternate option.

Participant selections were similar across tertiles with no notable trends. This finding is interesting given that one may hypothesize that higher ED district leaders would be more strongly against the existing mandated accountability system. On the contrary, no leaders from the highest ED tertile indicated that they do not like the measures. Notably, the two participants with the highest ED indicated "other" as a selection. One stated that the measures were fine, but that they took too long to retrieve. The other argued that the current model engenders unequitable comparisons, stating that we are not comparing apples to apples, but rather "apples to pears to strawberries to blueberries." See Table 4.2 for further details.

**Table 4.2**  
*Participants' Positions on Mandated Accountability Measures (By Tertile)*

	Position Statement	Number of Participants
Lowest ED Tertile	Likes the Measures / Would Not Change	0
	Satisfied / But Would Add Measures	5
	Do Not Like Measures / Would Change	3
	Do Not Agree with Public Reporting	1
	Neutral or No Preference	1
	Other	0
Middle ED Tertile	Likes the Measures / Would Not Change	1
	Satisfied / But Would Add Measures	6
	Do Not Like Measures / Would Change	1
	Do Not Agree with Public Reporting	2
	Neutral or No Preference	0
	Other	0
Highest ED Tertile	Likes the Measures / Would Not Change	1
	Satisfied / But Would Add Measures	4
	Do Not Like Measures / Would Change	0
	Do Not Agree with Public Reporting	2
	Neutral or No Preference	1
	Other	2

### *Quantitative Analysis of Participants' Perceptions of Accountability*

**Overview.** Participants were asked to rate the extent of their agreement to various statements on mandated accountability and Pennsylvania's public-facing accountability dashboard, the Future Ready PA Index (see Appendix D for a full table of results). The statements are as follows:

- You have a strong understanding of the accountability measures in the Future Ready PA Index.
- You value the mandated accountability measures in the Future Ready PA Index.
- Your district leadership prioritizes strong performance on standardized tests as a major district priority.
- Your district relies upon the accountability measures in the Future Ready PA Index as a major driver of strategic planning.

- You believe that the accountability measures in the Future Ready PA Index effectively capture student performance in schools.
- You believe that the accountability measures in the Future Ready PA Index provide reliable information/data on a school's overall effectiveness or performance.
- You believe the accountability measures in the Future Ready PA Index reflect the quality of instruction in schools (i.e. higher results mean that better instruction is taking place in particular schools).
- You believe the accountability measures in the Future Ready PA Index reflect the quality of leadership in schools (i.e. higher results mean that better leadership is taking place in particular schools).
- In general, you believe the accountability measures in the Future Ready PA Index have had a positive effect on school climate.
- In general, you believe that your schools perform better in the areas that are publicly reported in the Future Ready PA Index.
- You believe that schools have the resources/capacity to meet state benchmarks/expectations.

In the sample (n=30), across all of the questions, there appeared to be more agreement/strong agreement (156 responses) than disagreement/strong disagreement (117 responses) to these statements. The middle ED tertile expressed stronger agreement in the aggregate ( $\bar{X}$ =3.35) than did the lowest ED tertile ( $\bar{X}$ =2.91) or the highest ED tertile ( $\bar{X}$ =3.05).

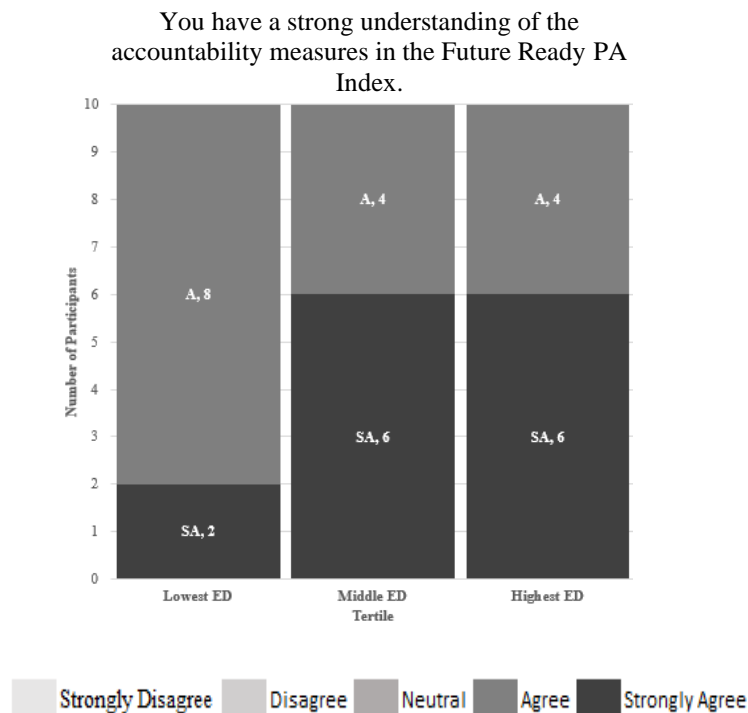
Leaders expressed the strongest agreement that they *understand the measures*, that they *use the measures for strategic planning*, and that they *value the measures* (see Appendix E for a full summary of means denoting the distribution and intensity of participant responses across the tertiles). The subsequent sections quantitatively analyze leaders' responses to the above statements and will further highlight emergent patterns amongst the tertiles.

#### **Perception of Their Own Understanding of Mandated Accountability.**

Participants were asked to rate, using a Likert scale, the extent to which they agreed that they *have a strong understanding of the measures in the Future Ready PA Index*. Of all the perception questions, participants expressed the strongest agreement with this statement. In fact, all participants reported that they strongly agreed or agreed. This is unsurprising as school accountability is ever present in the lives of Pennsylvania district leaders; however, the data establish that leaders consider themselves knowledgeable about the topics that follow. See Figure 4.1.

**Figure 4.1**

*Extent To Which Leaders Expressed Understanding of the Measures in the Future Ready PA Index*

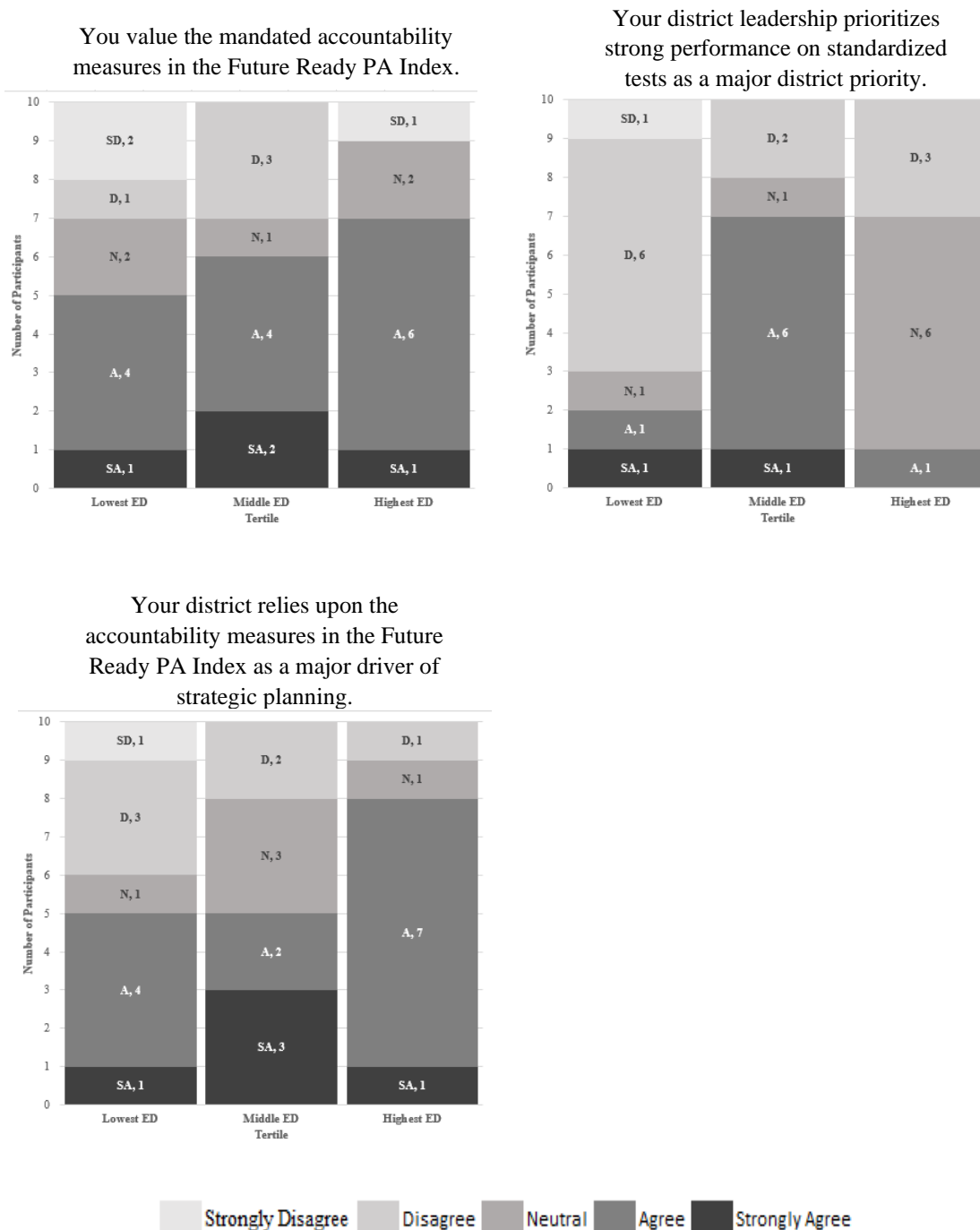


**Perception of the Value and Utility of School Accountability.** Figure 4.2 shows leaders' perceptions on various statements related to the value and utility of school accountability. As a sample, participants agreed/strongly agreed more frequently than they disagreed/strongly disagreed that they *value the measures in the Future Ready PA Index* and that they *use the measures for strategic planning*. The highest ED tertile showed the strongest agreement that it *uses the measures for strategic planning*. This may suggest that state accountability becomes more consequential as a district's ED increases because there is a stronger likelihood that a high ED district is placed on a state improvement plan to improve accountability data. Put simply, they are required to attend to the data more acutely than their lower ED counterparts.

Interestingly, the full sample expressed weaker agreement that standardized tests are prioritized as important (though there was stronger agreement in the middle ED tertile). This could suggest that standardized testing is no longer district leaders' singular focus in relation to mandated accountability. Alternately, this observation may be driven by individual districts' culture and priorities as they relate to standardized testing.

**Figure 4.2**

*Extent To Which Leaders Expressed That They Value and Utilize the Measures in the Future Ready PA Index*



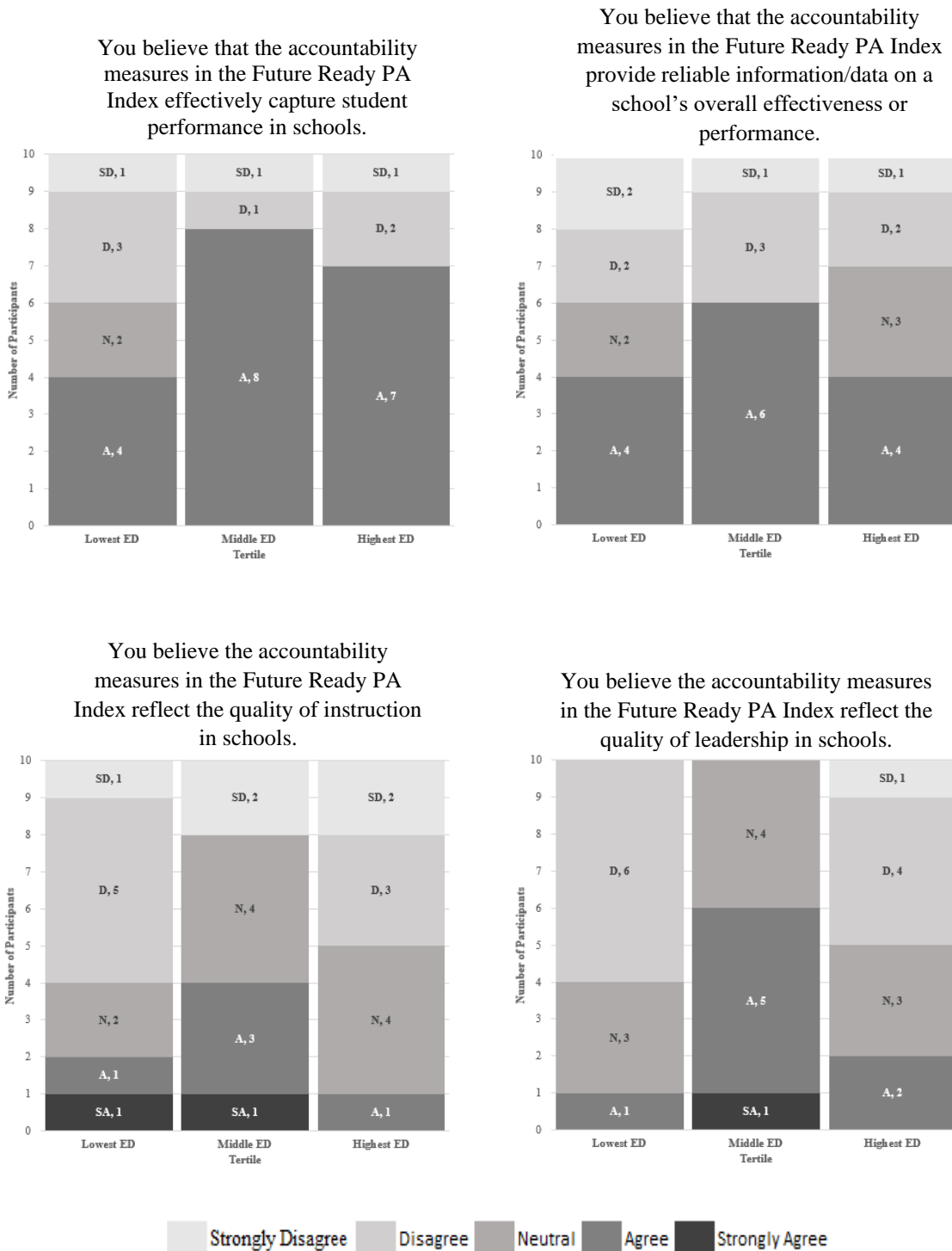
**Leaders' Confidence in Mandated Accountability Measures.** Figure 4.3 shows the extent to which leaders expressed confidence that the mandated accountability measures successfully captured various aspects of school and student performance. The data suggest the sample was conflicted about what, exactly, mandated accountability tells us about *student performance*, *school performance*, the *quality of instruction in a school*, and the *quality of leadership in a school*.

There was stronger agreement that the accountability measures successfully capture *student* achievement on standardized assessments than that it captures a *school's* performance, the quality of instruction, or the quality of leadership within the schools. This aligns with leaders' anecdotal comments throughout the interviews. They were mostly supportive of mandated accountability as a fair and consistent way to gauge *student* performance across the state; however, they believe the model goes a step too far in unfairly highlighting and penalizing *schools* with high poverty rates or other related challenges. This concept will be discussed at length later in this chapter.

Additionally, the lowest ED tertile and highest ED tertile more strongly disagreed that mandated accountability captures the quality of instruction and leadership in their schools. However, leaders' comments suggested different reasons for this. While the lowest ED schools value strong instruction and leadership, they also tend to experience strong student performance *despite* the instruction and leadership taking place. The highest ED schools, by contrast, struggle to overcome academic barriers *even with* strong instruction and leadership taking place. Participants expressed that this issue is complicated, and that these data cannot be blindly correlated to school quality. The middle ED tertile maintained consistent strength of agreement across these questions.

**Figure 4.3**

*Extent To Which Leaders Expressed Confidence in the Data Reported in the Future Ready PA Index*



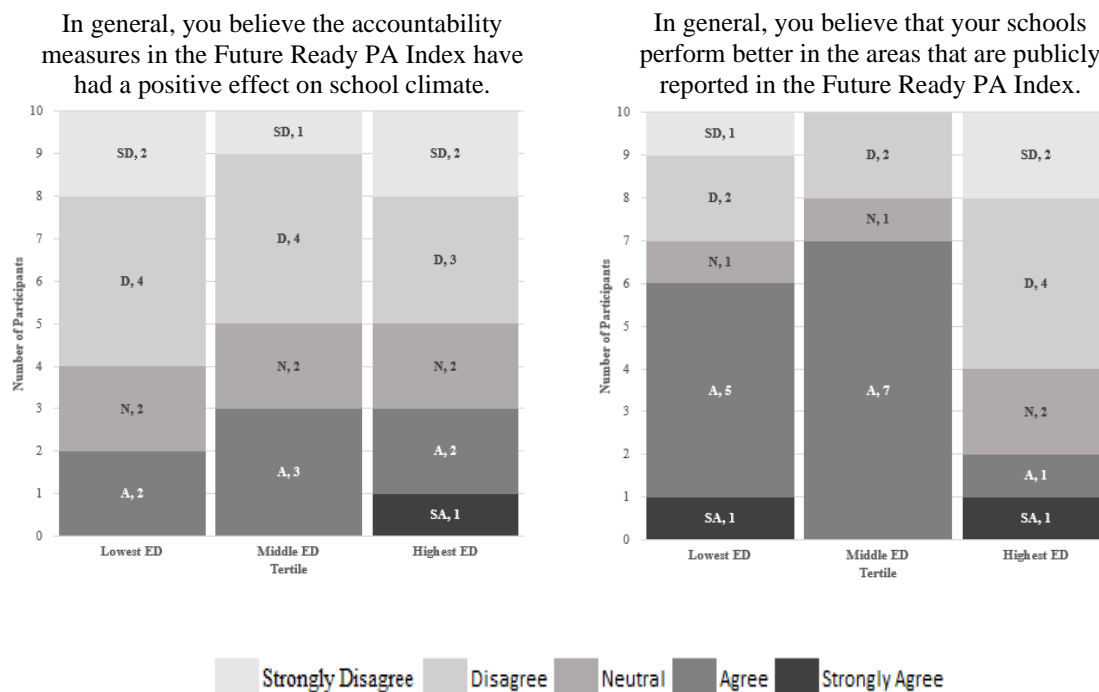
**Extent To Which Leaders Perceive that State Accountability Generates Positive Outcomes.** Figure 4.4 shows leaders' perceptions of whether mandated accountability generates positive outcomes.

The full sample was conflicted about whether the *public reporting of measures* engenders better results. However, there was nuance within the tertiles. Perhaps unsurprisingly, the highest ED tertile expressed stronger disagreement that *publicly reporting scores in particular areas increases achievement in those areas*. Anecdotal comments suggest that leaders and teachers in these districts are attending to a variety of challenges and needs that extend beyond what is reported on the accountability dashboard. Moreover, leaders across the tertiles stated that their principals and teachers are committed to high achievement regardless of whether it is publicly presented, and offered no indication that effort is lower in subject areas that are *not* publicly reported.

Participants expressed almost twice as much disagreement/strong disagreement as agreement/strong agreement that mandated accountability *has had a positive effect on school climate*. There were no detectable differences among tertiles. While some participants expressed that mandated accountability negatively impacted climate and morale, others described a more neutered effect, conveying that the accountability information is effectively out of sight and out of mind, thereby having no real effect.

**Figure 4.4**

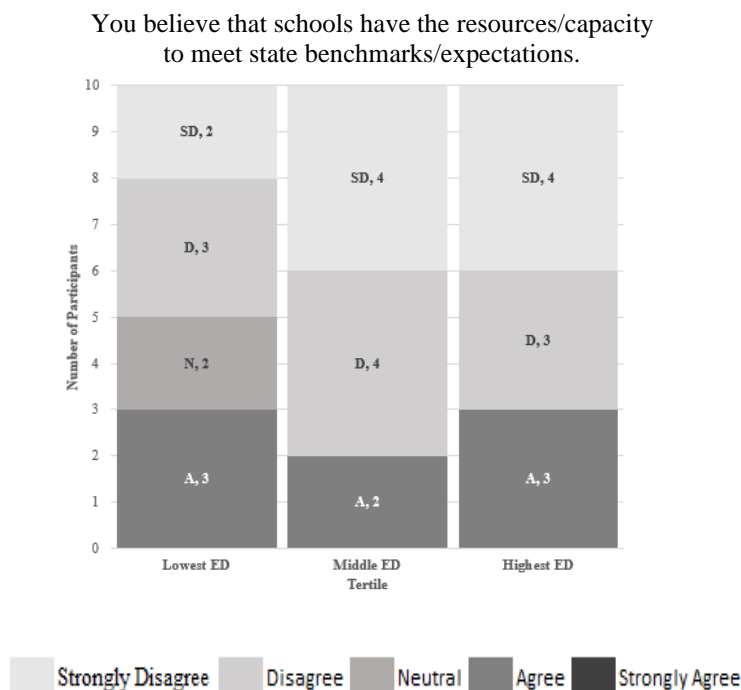
*Extent To Which Leaders Believed the Future Ready PA Index Has Engendered Positive Outcomes*



**Extent to Which Leaders Believe Schools Have the Necessary Resources and Capacity.** The full sample more strongly disagreed than agreed that *schools have the necessary resources and capacity to meet state expectations*. When participants did agree, they frequently clarified that they were speaking in terms of their *own* district having sufficient resources, and that districts with more significant economic challenges did not have appropriate resources. This finding relates to statewide discussions on equitable and appropriate funding. See Figure 4.5.

**Figure 4.5**

*Extent To Which Leaders Believed Schools Have the Needed Resources/Capacity to Meet Expectations*



### *Emergent Themes from Qualitative Data on Leaders' Perceptions*

Participants' qualitative comments further illuminate the quantitative responses described above. At the start of the interview, participants were asked to share their general perceptions of state accountability and the Future Ready PA Index. Their responses reflect their immediate, and most raw, views on mandated accountability, and therefore make the themes that emerged especially insightful. All comments were sorted, analyzed, and coded into overarching, axial themes. Appendix F shows all the themes that emerged. The following themes emerged most prominently.

First, the most frequent theme (13 out of 30 participants) was that the *measures are too narrow and that the current accountability system provides an incomplete*

*representation of the school.* They indicated that the current accountability system, despite improvement from previous versions, falls short of capturing the totality and complexity of what schools do. For example, Participant 5B said that it is “better than it’s ever been since we’ve had accountability,” but it is “not necessarily a good and accurate picture of what schools are doing.” P8 said, “The role of public education has expanded so far beyond basic content knowledge. School is the center of most communities – counseling, social work, health, social-emotional learning, coping strategies, financial planning.” P13, a district leader who had the opportunity to participate in the development of the measures, added, “Some of the benchmarks are not reasonable for [some schools]. The goal was just to get kids to school, to feed kids... Not one-size-fits-all.” P28, the superintendent of a higher ED district, echoed these concerns:

It does not look at mental health, social-emotional learning. None of that is a part of it. We know that in school districts like mine there are a lot of barriers. Kids have a lot of barriers. To remove them is difficult. When we can remove them it’s easier for kids to learn.

Finally, P14A said, “It’s never been a complete picture of what happens. Some things are not able to be reported statistically – not quantitatively.” The underlying tone of their responses was frustration that the public-facing measures against which they are held accountable fall short of capturing so much of the actual work that they, their administrators, teachers, and staff do each day. There were no notable differences amongst the tertiles.

The next most frequent theme (12 out of 30 participants) was that *mandated accountability is important for consistency and equity across school districts.* Participants expressed general support for a uniform mechanism through which to measure schools and hold them accountable. P7 said, “I think that federal and state accountability is

absolutely critical.” P15 “appreciates the transparency that it creates for all districts,” and P17 believes it “gives us a standard to which they can aspire.” Others expressed that, while the system may be imperfect, it meets the need of providing a consistent measuring stick for all districts. P13 said, “I don’t agree with everything,” but we “need to understand what our goals are.” P24 said, “I get a lot of complaints... Whole career is this accountability thing,” but “you have to have something... You have to look at something.” Interestingly, more leaders in the lowest ED and middle ED tertiles expressed this theme than in the highest ED tertile. This may suggest that higher ED districts view mandated accountability as misaligned with their daily challenges, and that there is a strong association with a subsequent theme related to socioeconomic variables (see below).

The third most frequently cited theme (9 out of 30 participants) was that *current accountability, via the Future Ready PA Index, is an improved and more comprehensive system than previous versions of accountability*. Previously, school districts were accountable almost solely for achievement test scores under the *No Child Left Behind (NCLB)* legislation. However, under the *Every Student Succeeds Act (ESSA)*, accountability has broadened in scope, as described in detail in chapter three. Participants generally saw this as a positive shift. For example, P3 said, “Compared to SPP [School Performance Profile], definitely a step in the right direction,” and P18 said that this is “probably their best attempt to take and reform that system and ask, how can we look at school systems not just from state tests for accountability?” P19 refers to the current model as a “more holistic approach to evaluating learners,” and P24 believed that, “when

put into proper context, they unfold in a way that's healthy for a district." P25

specifically addressed the new system as it relates to poverty:

I like them because we are about 60% poverty. It evens the playing field a little bit. Talks about industry-based a little bit more. A little less on the standardized testing. More things we have a little more control over with kiddos. I think it's going in the right direction, anyway.

Participants indicated that the new model more closely captures the complexity of what their schools do.

Finally, the fourth most cited theme (7 out of 30 participants) was that *the data are most closely related to socioeconomic factors*. Participants in the highest ED tertile were the most passionate about this theme. This is, perhaps, unsurprising as these leaders are most affected by the implications and effects of high poverty rates. Notably, though, participants across the socioeconomic continuum expressed concern about the relationship between a district's level of economic disadvantage and how they are publicly presented through mandated accountability. P1, leading the most affluent district in this sample, said that the data are,

misleading and unhelpful. When we look at accountability measures and testing in general, most of the testing we do is more correlated with family wealth than any other factor. It seems to be a hammer for schools in high poverty. Do my kids start on third base here? Yes. Did they start on third base in *x district*? No.

P14B echoed the concept that, in some districts, students may achieve *despite* the actions of the school: "If a school has high achievement, it doesn't necessarily mean it's because the school was responsible for that." Participants in the highest ED tertile spoke most directly to the relationship among demographics, opportunity, and outcomes. P27 said, "They're a poor measure of student success. They can vary greatly depending on student demographics, socioeconomics and, quite frankly, geography and zip code." P28 said,

“Locally, there are two districts that are high achieving. Most are professional families, parents highly educated. Demographics predominantly Caucasian. Scores are higher. But no one talks about that.” And P29 said, “If you look at demographics in economically disadvantaged areas, you see the correlation. When there’s no or lack of opportunity, children tend to score lower. They are disheartened.”

Participants were further prompted to identify the ways in which the current model either advances and/or hinders their leadership efforts. Appendix G summarizes all of the themes that emerged. Two particular themes are worthy of discussion here.

First, 19 out of 30 leaders said that the accountability system *engenders analysis, goal-setting, and improved practices*. Generally, participants indicated that the data provide direction for growth and achievement. P3 stated that it “opens the eyes of administrators” and that they “have these ‘aha’ moments.” P6 stated that it “makes sure we put goals in place that we’re actually working toward.” P12 echoed that the system “shines a light on things that need to be looked at, things that demand attention,” and P16 went further by explaining, “from data we generate ‘whys.’ Why are we experiencing lackluster graduation percentages? We can start working to build that up. There’s commonality.” P18 put it plainly: “To summarize, it forces school systems to create systems to produce results.” Participants in higher ED districts seemed to emphasize the importance of growth rather than achievement. For example, P25 said, “We want to make sure we’re accountable to the growth of kids. We *try* to get up the achievement level, but with extra emphasis on growth, all of that is good.” There were no significant differences amongst the tertiles.

In contrast to the previous theme, 13 out of 30 leaders indicated that the *accountability system is misunderstood, that it oversimplifies the work of schools, and consequently engenders poor public understanding*. P3 said, “We’re in a world now where life is political. Districts have to defend how they do things. Everyone has access without understanding what went into the data.” P7 went further by saying, “We need scores, but must be careful. When they report publicly that schools are failing, it’s political. And we have to spend time fixing that because the state colored a red box.” Other leaders expressed concern about the complex calculations that generate a score, and the limited ability to educate the community on the process. P14A said, “We would do PD with staff, teachers, admin to understand what that means. The public gets no professional development on what that means.” P11 said: “The general public is not looking at a continuous improvement lens, but a lens of who’s better and who’s worse.” Again, higher ED district leaders conveyed dissatisfaction that the accountability measures do not capture what is being done in schools. P26 said that there are,

a lot of really hard workers as far as students, teachers, paras. Assessments hurt education when you’re not making the achievement level that matches the energy and effort to educate the whole child. A lot of people on the outside don’t understand what teachers do to support students.

P28, a leader of a high ED district said succinctly: “When you publish just the scores, it gives us a black eye.” In summary, participants addressed this concept of misunderstanding or misuse from a variety of angles: making measures political, misunderstanding mathematical calculations, lack of understanding of what schools actually do, and neglecting to take into account the added challenges and barriers of schools in poverty. There were no notable differences across tertiles.

Across these questions, other themes emerged. On the positive side, participants believed that the new model *affirms and motivates good work; provides necessary consistency in measuring and evaluating schools; and provides a simple and clear form of communication*. On the negative side, they believed the system *is an exercise in compliance; is problematic in terms of specific measures or calculations; is unsuccessful in engaging the public; that it negatively impacts teachers; and encourages teaching to the test*. Each provides valuable insight into issues related to school accountability.

The themes described above capture the essence of the participants' beliefs and are interwoven throughout this dissertation report. They convey a complexity and nuance that illuminates their collective perceptions on accountability. In general, the findings suggest that district leaders are supportive of the act of being held accountable. They affirm the importance of ensuring transparency and equity across the state. Simultaneously, though, they expressed concern that the system fails to capture the totality of what schools do, is overly simplistic, and that this creates misunderstanding and misuse within the public realm. Further, while the system is an improvement over the previous, achievement-heavy model, it remains troublingly related to socioeconomic status, and specifically poverty and race. Most participants expressed that several of these themes coexist with one another, and they were often addressed even within a single sentence. The findings also suggest that the current system is not inherently "wrong" or "bad," but that continuous improvements may lead to even stronger support, buy-in, and understanding. The themes, and the relationships among them, are relevant from educational and policy perspectives. Again, a full summary of themes may be found in Appendices F and G.

**Research Question 2: To what extent do Pennsylvania district leaders value various mandated and unmandated (alternative or additional) public-facing accountability measures? What is the relationship between district contextual attributes and the measures they value?**

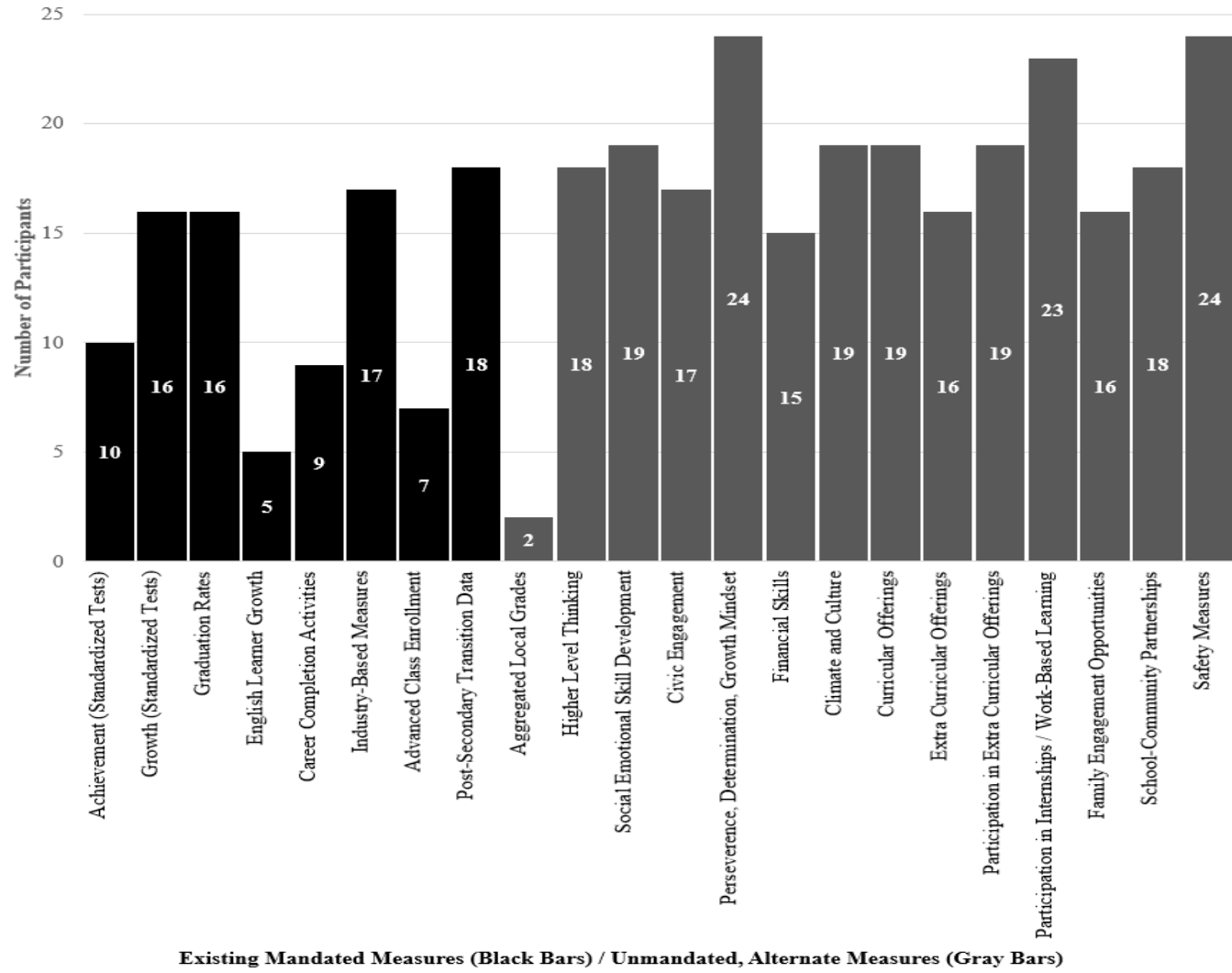
The second research question explored the extent to which Pennsylvania district leaders value various mandated and unmandated (alternate or additional) public-facing accountability measures. These data provide insight into what leaders perceive as most important within the *ESSA*-era, post-Covid 19 pandemic educational context. Moreover, the research may support strategies for leaders to develop their own accountability models to supplement the state mandated measures.

***Quantitative Review of the Extent to Which Leaders Value Measures***

First, leaders were asked to rate various measures that comprise the current, public-facing accountability dashboard (examples: student achievement on tests, student growth on tests, graduation rates, acquisition of industry-based credentials, etc.). They were then asked to rate another 14 measures that do not currently exist on the current dashboard (examples: culture and climate ratings, civic engagement levels, extra-curricular participation rates, safety measures, etc.). Participants were asked to rate each individual measure as *highly valuable*, *somewhat valuable*, or *not valuable* as a public-facing accountability indicator. Figure 4.6 shows the number of participants across the sample that found each measure *highly valuable*. For more detail, Appendix H presents a table of participant responses across ED tertiles, and Appendix I presents a table that shows the means denoting the distribution of participant responses (again, mean values

are intended as a comparative tool, not to confer deeper meaning as to the strength of responses). Notable findings follow.

**Figure 4.6**  
*Number of Participants Who Highly Valued Various Mandated and Unmandated Accountability Measures*



**Findings Related to Mandated Measures.** As a sample (n=30), the greatest number of participants expressed that they highly valued the *post-secondary transition* measure and the *industry-based learning* measure. These items were not publicly reported until 2018. This finding suggests that district leaders highly prioritize students' readiness for post-secondary life, and that these are valuable additions to the public-facing accountability dashboard. The *graduation rate* measure is also notable as zero participants indicated that they did not value it as an important measure. Conversely, the fewest participants indicated that they highly valued the *participation in advanced classes* measure, the *career completion activities* measure, and the *achievement on standardized tests* measure. (The *English Learner growth* measure is inconsequential as many participants had only a small or no English learner population.)

There were some differences among the ED tertiles. First, the data illuminate the ongoing discussion about students' academic achievement (percent proficient or advanced on a standardized assessment) versus students' academic growth (extent to which students grew from one standardized assessment to the next). While the data show stronger support for *growth* measures than *achievement* measures, there is more nuance within the tertiles. Both the lowest ED tertile and the middle ED tertile expressed that they value *achievement* and *growth* to the same degree of strength. The highest ED tertile, however, valued *growth* more than *achievement*. This is a logical finding; as ED increases, there is a correlational, inverse relationship to achievement. It makes sense, then, that higher poverty districts who struggle with *achievement* believe that *growth* is a more valuable, and perhaps fairer, measure. P26 stated,

This is a real critical one. We can't necessarily control what skills students arrive to us with, barriers with mental health, poverty, disabilities, etc. If you can show every student is growing every year, then we are doing the best we can for that student. The big factor is time. You can only do so much in a day and school year.

It is important to note that there was not universal support for the *growth* measure; some participants disagreed with the fundamental calculations undergirding the measure.

Second, the lowest ED tertile valued the existing *graduation rate* measure less than the other two tertiles. This may suggest that low poverty districts experience greater success with graduation rates and they are, therefore, less of a concern. Conversely, middle and higher ED districts strive to increase graduation rates because they are frequently cited as a leading indicator of a school's success both within and outside of educational circles. As such, this finding makes sense. P16, a middle tertile district leader, said this is "the most important thing. End goal. If I can't hit that, then I need to think about the 'why.' Why aren't they graduating? Prioritize those areas." Other leaders echoed this statement, expressing that the graduation measure reflects the culmination of a student's academic experience, and is a measure that is understandable to both practitioners and the general public.

Third, there is an interesting relationship between the measure of students *earning an industry-based credential* and the measure of student *participation in advanced-level classes*. Overall, more participants expressed that they highly value the *industry-based* measure. However, whereas the lowest ED tertile expressed less value for the *industry-based* measure, the highest ED tertile expressed less value for a measure of students *participating in advanced classes*. This finding may suggest that lower ED districts attend more closely to traditional academic pathways and college-bound students.

Conversely, higher ED districts with potentially fewer college-bound students are leveraging internship and industry-based experiences to prepare graduates for post-secondary success. In reference to advanced classes, P29, superintendent of a high ED district, also raised an important concern related to districts' *opportunity* to compete with other districts. She said, "That's access. We only have two advanced classes. No! Now I'm venting. We would love to offer more but I can't." This finding further intersects with the *post-secondary transition* indicator. Curiously, the lowest ED tertile valued this measure less strongly than the highest ED tertile. This may suggest that the lowest ED districts who already enjoy high post-secondary transition rates, predominantly to college, are less interested in tracking this measure closely. However, further study would be required to better understand this observation.

**Findings Related to Unmandated Measures.** As a sample (n=30), the greatest number of participants expressed that they highly value measures related to *safety, participation in internships, externships, and work-based experiences*, and students' *perseverance, determination, and growth mindset*.

Additional findings emerged. First, participants did not rate the *reporting of aggregated local grades* (conveying the percentages of students who earned As, Bs, Cs, etc.) as a valuable measure. Almost universally, participants cited the subjective nature of local grades that makes them almost meaningless as a consistent, transparent accountability measure. For example, P18 said, "[Only] if I had a statewide standardized curriculum full of assessments and all grading procedures, then yes." P11 was more direct: "Grades are stupid. Not valuable." While participants were generally supportive of

fair and equitable accountability on a statewide level, reporting local grades was not perceived as a pathway to achieve that objective.

Second, measures denoting *social-emotional skill development* as well as students' *perseverance, determination, and growth mindset* were strongly valued across the tertiles. This suggests that leaders perceive that students across all socioeconomic levels require similar supports and dispositions to achieve success in school and in the future. P8 said, "Incredibly valuable. Research indicates that these things can lead a student to make something positive out of their life."

Third, a measure denoting *curricular offerings* was highly and consistently valued across the three tertiles. This is in contrast to the measure of *participation in advanced classes* in the state's current accountability dashboard. Participants seemed to appreciate that this measure would capture the totality of curricular offerings (art, music, etc.) while the other measure more closely reflected opportunity and access for a selected subgroup of students (AP, IB classes, etc.). In both instances, participants expressed concern about inequitable opportunity. P21 said, "Could be good, but it's going to hurt smaller schools," and P5 elaborated, "The more you can explain the better. Looks good in a report, but you take x district with a tax base that is higher. They can afford more teachers and courses." P4 summed it up by simply stating, "Haves and have nots."

Fourth, to elaborate on this theme of access and opportunity, participants valued measuring students' *participation in extra-curricular activities* slightly more strongly than they valued measuring things like *extra-curricular offerings* and *parent engagement opportunities*. There was some divergence of thought on this matter. Some leaders preferred to measure what the school *offers* while others preferred to measure *actual*

*participation*. P9 said, “Not so much opportunities, but [if we can measure] actual engagement and participation.” Conversely, P13 said, “If [we provide] just a list of opportunities, then yes,” and P19 stated, “We can offer, but if they don’t come...” These reflect important considerations when designing measures for public consumption.

Fifth, measures related to *school safety* were ascribed the most value. Every participant either rated this measure as *highly valuable* or *somewhat valuable*. However, detailing what and how to report on measures related to *safety* introduced a new and more nuanced discussion that is beyond the scope of this report.

**Aggregate Data.** Taken together, Table 4.6 suggests that there are dimensions of student and school success that would be valued as accountability measures, but that are not part of the existing dashboard. In fact, the majority of the unmandated, alternate, or additional measures received as much or more support as the *most supported* mandated measure. While this is an interesting observation, it is also limited. It is important to consider that leaders were evaluating existing, known accountability measures alongside *the promise* of new measures. For example, while educational leaders tend to support the construct of measuring academic achievement, they may have a negative impression of how it is operationalized through standardized testing. Similarly, they are supportive of the construct of cultivating students’ perseverance and growth mindset, but we have not yet operationalized that measure. Several participants responded that they would *highly value* a measure “if done well” or “if done correctly.” This indicates that the *way* it is measured matters. It is also important to note that several leaders credited federal and state government with attempting to operationalize complex measures of success. Still,

these data suggest that there is interest in broader measures that capture a more holistic picture of student and school success.

### ***Other Measures Valued by Participants***

Participants were also asked to identify additional measures that they believed are valuable and should be considered as an accountability measure. Measures on *climate and culture* were most frequently noted (frequency=7) with a majority of references originating from the lowest ED tertile (frequency=6). Participants also mentioned *post-secondary* measures, *socio-economic factors/context*, *financial* measures, indicators of *well-rounded student experiences*, *local assessments*, and *student discipline* measures. All responses are summarized in Appendix J. Taken together, this provides additional insight into the types of measures of value to district leaders.

### **Research Question 3: How do Pennsylvania district leaders communicate mandated and/or unmandated (alternative or additional) accountability measures to their stakeholders? What is the relationship between district contextual attributes and their communication behaviors?**

The third and final research question explored how district leaders communicate mandated and/or unmandated (alternative or additional) accountability measures to their stakeholders. This question will be viewed through a variety of lenses that, taken together, will illuminate district leaders' behaviors around communication. Further, it investigated the extent to which a district's degree of ED affects leaders' behaviors by disaggregating the 30-district sample into tertiles in a consistent fashion as the previous two research questions.

### *Frequency of Communication of Accountability Measures to Various Stakeholders*

Table 4.3 displays the frequency of communication of accountability measures to various stakeholders. Frequency is operationalized as a selection of choices ranging from *never* to *more than quarterly*. In general, the data are unremarkable in terms of differences by tertile. However, the aggregated data provide some insight into district leaders' communication patterns. Almost every leader reported communicating accountability to the school board either *annually* or *bi-annually*. Two participants reported that they communicate to the school board more frequently than that, and one reported *never* communicating accountability to the board. The frequency of communication to parents closely resembles the frequency of communication to the school board. This makes sense as family and community communication often follows a public presentation at the board level.

In terms of frequency of communication to district leaders and principals, most participants indicated communication occurring *quarterly* or *more than quarterly*, with some exceptions. It is important to note that, within this sample, the distinction between district administrators and principals was often blurred because small school districts tend to utilize principals as their district-level planning team. Frequency of communication to teachers is more dispersed across frequencies, but most participants reported that they present accountability information to teachers *bi-annually*, *quarterly*, or *more than quarterly*. For district administrators, principals, and teachers, participants referenced a combination of formal presentations and other types of communication, such as work sessions.

These data indicate that the most formalized and consistent communication on school accountability is to the school board, but that the most frequent and ongoing communication occurs internally with district administrators, principals, and teachers. See Appendix K for a table of means denoting the distribution of frequency of leaders' communication with various stakeholders.

**Table 4.3**  
*Participants' Frequency of Communication with Various Stakeholders (By Tertile)*

		Communication To...				
		District				
		Board	Administration	Principals	Teachers	Parents
Lowest ED Tertile	Never	1	0	0	1	1
	Less Than Annually	0	1	1	0	0
	Annually	5	1	1	1	7
	Bi-Annually	4	2	1	4	1
	Quarterly	0	3	3	1	1
	More Than Quarterly	0	3	4	3	0
Middle ED Tertile	Never	0	0	0	0	0
	Less Than Annually	0	0	0	0	1
	Annually	6	1	0	3	7
	Bi-Annually	2	0	2	3	2
	Quarterly	1	5	3	1	0
	More Than Quarterly	1	4	5	3	0
Highest ED Tertile	Never	0	0	0	0	0
	Less Than Annually	0	0	0	0	0
	Annually	7	3	3	3	7
	Bi-Annually	3	0	0	1	3
	Quarterly	0	5	5	2	0
	More Than Quarterly	0	2	2	4	0
All Sample	Never	1	0	0	1	1
	Less Than Annually	0	1	1	0	1
	Annually	18	5	4	7	21
	Bi-Annually	9	2	3	8	6
	Quarterly	1	13	11	4	1
	More Than Quarterly	1	9	11	10	0

### *Frequency of Questions on Accountability from Stakeholders*

Table 4.4 presents the frequency of questions on accountability that participants receive from parents and families, as well as from the larger community (non-parents and families). The data suggest that questions from stakeholders are rare as most participants indicated that they *never* or *infrequently* receive questions from those stakeholders. Only one participant indicated that their district *frequently* receives questions from parents and families.

When participants indicated that they did receive questions from the larger community, they typically came from other leaders, business groups, older citizens, or other sources. See Appendix K for a table showing means denoting the distribution of frequency of questions received from public stakeholders.

**Table 4.4**  
*Frequency of Questions or Concerns Received from Public Stakeholders (By Tertile)*

		Questions From...	
		Parents/Families	Larger Community
Lowest ED Tertile	Never	6	6
	Infrequently	3	4
	Frequently	1	0
Middle ED Tertile	Never	4	7
	Infrequently	6	3
	Frequently	0	0
Highest ED Tertile	Never	8	2
	Infrequently	2	8
	Frequently	0	0
All Sample	Never	18	15
	Infrequently	11	15
	Frequently	1	0

### ***Methods of Communication on Accountability***

Table 4.5 shows the most utilized methods of communication with stakeholders. The most prevalent methods are board or committee presentations (15 responses), website postings (12 responses), and communication via parent communication portals or emails (12 responses). No significant trends emerged among tertiles.

**Table 4.5**

*District Leaders' Strategies for Communicating with Stakeholders (By Tertile)*

	Economic Disadvantage			Total
	Lowest ED Tertile	Middle ED Tertile	Highest ED Tertile	
Board or committee presentation	4	7	4	15
Website	6	3	3	12
Parent communication portals or emails	3	3	6	12
Building-level communication	2	3	4	9
Social media	3	2	3	8
Newsletters	2	2	4	8
Videos and screencasts	1	1	1	3
Annual or other report	2	0	0	2
Parent groups	1	0	0	1
Comprehensive planning teams	0	1	0	1
Newspaper	0	1	0	1
In-person events	0	0	1	1

### ***Benefits and Challenges of Publicly Communicating School Accountability***

An important component of understanding district leaders' behaviors in relation to publicly communicating accountability is their perception of the benefits and challenges associated with such communication. Table 4.6 provides a review of all of the benefits and challenges that the participants identified, organized into emergent themes, and sorted in order of the frequency of the total number of times that theme was identified.

Two benefits emerged as particularly relevant. First, several participants (12) indicated that the *public presentation of accountability measures engenders communication and connection amongst stakeholders*. More participants in the highest

ED tertile expressed this theme than did participants in the other tertiles. For example, P28 said, “The more people you get involved, people have ideas... ideas are good as they can help to solve problems.” P29 said, “We can strategize together. Get the input of community and stakeholders. Where *they* see barriers to success. What are the barriers? We can start as a school to mitigate those barriers.”

Second, several participants (10) expressed that the public presentation of accountability measures *engenders analysis, goal-setting, and improved practices*. P6 said that this “helps us to know we’re on the right track.” P16 said, “Transparency. I have to own the data. Got to own your data. Tell me what you’re doing to fix the data.” P18 said, “Because I share data, what I now have is, ‘this isn’t my thoughts, not admin team’s thoughts... it helps me to get things that our system needs to improve. I always start under that premise.” The participants expressed that it is healthy to remain open and transparent when sharing strengths and weaknesses. This theme was slightly more prevalent among the lowest and middle ED tertiles than the highest ED tertile. In the context of the other data, this could suggest that the highest ED tertile spends more time addressing challenges and barriers unrelated to formal accountability, and therefore finds less value in analysis and goal-setting using these measures. However, further study would be required to ascertain a strong relationship.

Conversely, one challenge emerged as most prominent. Thirteen participants, almost equally distributed across the tertiles, reemphasized the theme that the *system oversimplifies the work of schools and engenders poor public understanding*. However, participants’ reasons for this perception varied across tertiles. P1, leading an affluent district, said, “We are high achieving. We get beat up. [There is] misunderstanding.” This

leader spoke of community perception that the district *overemphasizes* academic achievement which seemed contrary to the district's almost total lack of communication about achievement. Other participants discussed misunderstanding related to the metrics themselves. For example, P9 said,

People make assumptions that aren't necessarily true. We have a small elementary with 100 students. Those students can have wide variances from year to year. Basing a judgment on an entire school on 20-30 students is flawed. Not statistically valid.

P25, representing a higher ED district, said that it is challenging when people "don't understand growth and achievement. Kids may have had a great year, but still not at that achievement line. There is confusion. It is hard to explain over email or newsletter. It depends on how you define achievement." In summary, participants expressed support for the concept of accountability, but also found it challenging to adequately communicate the nuance required to interpret the data, especially in this post-pandemic period in which the data have been skewed in new and unforeseen ways.

Beyond the themes identified here, participants identified several others. Benefits included *helping to tell the story of the school's performance, promoting transparency and equity, allowing valuable comparisons across school districts, and providing "cover" from the state for change efforts*. Additional challenges included *data that is "cherry-picked" and used against the district, incomplete or inaccurate media coverage, lacking resources to implement change, and lagging and outdated data*.

**Table 4.6**

*Participants' Perceptions of the Benefits and Challenges of the Public Communication of Accountability (By Tertile)*

	Economic Disadvantage			Total
	Lowest ED Tertile	Middle ED Tertile	Highest ED Tertile	
<b>Benefits</b>				
Engenders communication and connection amongst stakeholders	3	3	6	12
Engenders analysis, goal-setting, and improved practices	3	5	2	10
Helps to tell the story of district or school performance	3	2	2	7
Promotes transparency and equity	2	2	2	6
Affirms, motivates, and encourages the work of schools	0	2	0	2
Allows valuable comparisons across schools and districts	0	0	2	2
Enhances trust of leadership	1	0	0	1
Provides a simple and clear form of communication on school performance	1	0	0	1
Provides "cover" for change efforts (PDE)	0	0	1	1
Improved and more comprehensive than previous versions	1	0	0	1
<b>Challenges</b>				
System is misunderstood / Oversimplifies the work of schools / Poor public understanding	5	4	4	13
Data can be "cherry-picked" and used against district	2	3	1	6
Incomplete or inaccurate media coverage	1	1	1	3
Community is disengaged or apathetic	0	2	1	3
Demonstrating response to data	2	0	0	2
Measures are too narrow / Incomplete representation of a school district	1	0	1	2
Challenging to improve the data	0	2	0	2
Overuse of data with harmful consequences	1	0	0	1
Reflects poorly conceived / implemented law	1	0	0	1
Can negatively spotlight some student groups	1	0	0	1
Preparing to explain, contextualize data	0	1	0	1
Staying focused on the positive	0	0	1	1
Lack of resources or money to enact change	0	0	1	1
Engenders state oversight and pressure	0	0	1	1
Data is lagging and outdated	0	0	1	1
Encourages unhealthy competition amongst districts	0	0	1	1
Unimpactful to district work	0	1	0	1

### ***Practices Employed to Effectively Communicate Accountability***

As described in chapter three, this sample of participants contains leaders with many years of experience as superintendents or in other district-level leadership. Table 4.7 identifies emergent themes that capture strategies these leaders employ to mitigate communication challenges and communicate effectively on accountability. The most prevalent theme (9 responses) was to be *honest and transparent about areas for growth*. P5A said, "Come out first with honesty. Don't try to hide it, doesn't do any good. Here's what's going to be put in place." P17 emphasized the importance of "underselling the positive," instead telling the public, "here's what we have, here's what went well, here's

a challenge, and here’s what we’re trying to do to address it.” P24 emphasized the importance of “being true to your word,” and P27 said, “Don’t try to bury it. Bring it to light.” It is not surprising that more participants in the highest ED tertile expressed this theme given the reality that their accountability metrics may be lower or harder to explain than their counterparts.

Participants identified other strategies, to include *engaging collaboratively with stakeholders, focusing on a holistic view of a strong educational system, dedicating time to preparation to share results, highlighting positive outcomes, and ensuring that teachers have a strong understanding of the data.* The full list of themes with their associated frequencies may be found in Table 4.7.

**Table 4.7**  
*Participants’ Perceptions of Important Practices When Communicating Accountability (By Tertile)*

	Economic Disadvantage			Total
	Lowest ED Tertile	Middle ED Tertile	Highest ED Tertile	
Be honest and transparent about areas for growth	3	1	5	9
Engage collaboratively with individual stakeholders	3	1	1	5
Focus on a holistic, strong education system	1	3	1	5
Tell a holistic story of the district using varied metrics	4	0	1	5
Dedicate time to preparation	1	2	0	3
Maintain a results-oriented narrative	2	1	0	3
Highlight positive outcomes	2	1	0	3
Dedicate time to educating the community	1	0	1	2
Involve parents and the community in change efforts	1	0	1	2
Listen to stakeholders	0	0	2	2
Do not try to please everyone	1	0	0	1
Ensure teachers have a strong understanding of the data	0	1	0	1
Tailor data to the needs of various stakeholders	0	1	0	1
Approach school as a business	0	1	0	1

## Conclusion

This study's findings provide valuable insight into Pennsylvania superintendents' and other district leaders' perceptions of and behaviors around accountability and performance measures.

The findings within Research Question 1 suggest that participants' perceptions are complex and nuanced. In general, the data show that the highest concentration of participants support public-facing accountability, but that they would recommend that the model is augmented or changed in various ways. They tend to believe that public accountability increases transparency, provides consistency through common measurement, and promotes equity by shining a light on the outcomes of all subgroups of students. Further, they value the measures for analysis, strategic planning, goal-setting, and to develop improved practices (strongest agreement by the highest ED tertile, perhaps because accountability metrics are tied to state improvement plans).

In general, there was stronger agreement that public-facing accountability measures describe *student* performance than a *school's* performance, a distinction that is consequential when considering policy and leadership implications (more on that in chapter five). This distinction is partly explained by two concerns that emerged most prevalently: first, that the measures, while more holistic than ever before, are still too narrow to completely and holistically tell the story of school performance and, second, that outcomes are problematically correlated with both student socioeconomic status as well as the wealth of a school or district (a problem of "haves" and "have nots").

Consequently, mandated accountability can lower staff morale, misinform the public, or

provide an incomplete or inaccurate representation of a school. As stated, participant perceptions were nuanced across the sample and within the individual tertiles.

Research Question 2 provided insight into the extent to which participants valued a range of mandated as well as unmandated, alternate, or additional performance measures. In the aggregate, the majority of the unmandated measures received as much or more support as the *most supported* mandated measure. While this suggests interest in new and expanded measures, this finding is limited in that the mandated measures have already been operationalized whereas the alternate measures are abstractions. Further research would be needed following the implementation of alternate measures to determine whether participant perception would change. While the conclusions are complex and nuanced, in general participants most supported measures that were future-facing (examples: *post-secondary transition, graduation rate, industry-based learning, internships and externships*), that addressed student dispositions (*perseverance, determination, growth mindset and social-emotional skills*), and measures on *school safety*. Conversely, they least supported the measures of *advanced classes offered* (perceived as correlated to opportunity and access), *career completion activities* (viewed as a checkbox task unrelated to real outcomes), and *aggregated local grades* (a subjective and unusable measure). The data suggest some differences across tertiles, most notably that the highest ED tertile expressed stronger support for measures of *academic growth*, where they could demonstrate progress, than *academic achievement*, which is highly correlated to socioeconomic status and difficult to change.

Finally, Research Question 3 provided insight into participants' behaviors around the communication of accountability. While this question was valuable for contextual

purposes, it was less fruitful in terms of conclusions. The data suggest that district leaders communicate least frequently, but most formally, with their school boards and families on mandated accountability. Most participants stated that their families and the larger community either infrequently or never ask questions or make comments about the performance measures. Participants identified two outcomes of the public communication of accountability as most beneficial: that it allows *communication and connection amongst stakeholders* and that it *engenders analysis, goal-setting, and improved practices*. Conversely, one challenge, that the *system oversimplifies the work of schools and therefore engenders poor public understanding*, emerged as most prominent. Participants most frequently cited the following practice as important for the effective communication of accountability: *be honest and transparent about areas for growth*.

## CHAPTER FIVE: DISCUSSION

The purpose of this study was to explore the perceptions, attitudes, and behaviors of Pennsylvania district leaders as they navigate the complex landscape of educational accountability. Specifically, it examined the degree to which districts' levels of economic disadvantage affect those perceptions, attitudes, and behaviors. This was a mixed methods study that integrated quantitative components into a qualitative interview of 30 school district leaders. While the study's findings are not generalizable to the larger population, they may be viewed through the lenses of policymaking as well as school and district leadership. This chapter addresses this study's limitations, implications, and suggests areas for future research.

### **Findings**

This study yielded several notable findings and insights. First, participants revealed complex and nuanced perceptions related to accountability. They acknowledged that accountability advances transparency, equity, and consistency across the state, and that it engenders analysis, goal-setting, and strategic planning. Conversely, though, they expressed concern that the performance measures are too narrow to capture a holistic picture of a school's performance and responsibilities, and that the available metrics are too closely associated with socioeconomic status at the individual student and district levels. These data contribute to the extant research (Dessimone, 2006; Englert, et al., 2007; Russel & McCombs, 2006; Seashore Louis & Robinson, 2012; Sutherland, 2022) on superintendent, principal, teacher, and school board perceptions of various aspects of school accountability.

Second, district leaders value the existing accountability measures to varying degrees but placed higher value on measures that addressed students' future success, such as *post-secondary transition*, *industry-based* measures, and the *graduation rate*. There was stronger support, especially at the highest ED tertile, for a public measure of academic *growth* than *achievement*.

Third, leaders highly valued a range of unmandated, alternate, or additional performance measures, especially those that captured the array of responsibilities and student outcomes for which modern schools are responsible. These include measures of *perseverance*, *determination*, and *growth mindset*; *participation in internships and externships*; and measures of *safety*. Given participants' general perception about the need for expanded metrics to tell a more holistic story, these findings make sense. Further information on trends and patterns within the ED tertiles may be found in detail in chapter four.

These data reinforce the existing literature, predominantly centered on parents and families as the primary consumer of accountability data (Brewer et al., 2015; Jacobsen et al., 2013; Jacobsen et al., 2014; Owens & Peltier, 2002), that suggest that stakeholders value a diverse set of performance indicators and that their perceptions of a school's quality are influenced by the way indicators are presented. More contemporary research (Gagnon & Schneider, 2019; Houston et al., 2022; Mandinach et al., 2020; Schneider et al., 2021) explores educators' and other stakeholders' perceptions of state-mandated dashboards versus expanded dashboards with alternate measures. They further affirm that stakeholders value expanded measures, that such measures provide pathways for tangible, actionable reform in those areas, and that it may be important to include data, such as

student perception surveys, that have been found to *not be* correlated with a school's socioeconomic status.

Finally, findings suggest that district leaders' communication of mandated and/or unmandated accountability measures seems to be related to both their personal beliefs about the importance of the metrics as well as the demands placed upon them by their individual boards, families, and communities. While some leaders limit their communication to infrequent and highly formal board or committee presentations, others more systematically embed communications about accountability into their leadership and communication plan. In general, most participants stated that their families and the larger community either infrequently or never ask questions or make comments about the performance measures. Participants who valued the public communication of performance measures tended to believe that it introduced opportunities for connection amongst stakeholders and that it guided analysis, goal-setting, and improving practices. Conversely, the perception that the model oversimplifies the work of schools and creates public misunderstanding emerged as the most prominent challenge.

### **Limitations**

This study was limited in several ways. First, while the sample of 30 school districts was statistically representative of the 499 school districts across Pennsylvania in terms of economic disadvantage, the sample is too small to draw generalizable conclusions. Participant responses may be affected by levels of economic disadvantage but leaders' unique characteristics and experiences, as well as districts' social and cultural contexts, also inform their responses.

Second, time with each participant was limited to approximately one hour. While each interview reflected a robust and engaging discussion, topics could not be explored in great depth.

Third, while the interview protocol included quantitative elements (example: Likert-type items), they were not able to be analyzed using statistical methods. There were not enough questions in each category of questioning to draw statistically sound conclusions. While this would have been preferable, more questions and more participant time outside of the interview would have been required. Given superintendents' and other district leaders' limited availability, the researcher would have been unlikely to be granted this level of time and access. Therefore, a decision was made to focus on a variety of themes and concepts in one targeted interview that would capture a broad overview of participant perceptions.

Fourth, this is a predominantly qualitative interview rooted in perception data. The nature of the questions may naturally elicit answers that are derived from the most recent or vivid experiences that the participant has had. For example, if the leader has just experienced a decrease in standardized test scores from the previous year, they may be less likely to say that such an accountability metric is reflective of the quality of their schools. Along these lines, there may be some *self-selection bias* as leaders who have vivid experiences or deeply entrenched beliefs may have been more likely to accept the interview invitation in the first place. Similarly, *reactivity* is an issue in that participants may have reacted to the question being asked not having had an opinion prior to the interview.

Finally, data were collected via face-to-face interviews, making anonymity impossible. While the researcher assured the participants that their names, school districts, and other identifying information would not be included in the final dissertation, district leaders, occupying vulnerable and public leadership positions, may naturally exercise caution and restraint when answering questions.

## **Implications and Recommendations**

### ***Policy***

School accountability is a powerful force in public education. From the onset of strong federal accountability in the 1960s to the modern era of *NCLB* and *ESSA*, accountability policy has been shaped and reshaped across many decades.

Superintendents and other school district leaders are the policy mediators who must receive, interpret, communicate, and enact federal and state policy at the local level. Therefore, their perceptions of mandated accountability represent one important lens through which to analyze whether the policy is achieving its desired outcomes. This study elicits several recommendations that might further advance accountability policy.

First, policymakers should consider a requirement that accountability dashboards contain more detailed and nuanced contextual information about schools and districts. Participants expressed that, given their limited access to community stakeholders, explaining and contextualizing the state-mandated metrics can be challenging, in turn allowing misinformation to harm the school-community relationship. By embedding contextual information into the dashboard itself, those stakeholders investing time in exploring public accountability measures (interested parents, future community residents, a local business, etc.) have immediate access to more robust information about their

schools, which further helps to illuminate the performance data. Both quantitative information (rates of student transiency, student-to-teacher ratios, numbers of specialized programs and classrooms, revenue and expenditure figures, etc.) as well as qualitative information (descriptions of curricular and extra-curricular programming, explanations of student supports, narratives about the district’s vision, mission, and beliefs, etc.) would be valuable additions.

A corollary issue is that the Pennsylvania dashboard utilizes a color-coding system – blue, green, and red – to designate the extent to which a school is meeting performance requirements. While this is a clear and simple way to present data to the public, it is critical to consider the impact on the most disadvantaged schools. One superintendent noted the “increased emotion” that accompanies color-coding achievement, and another, a leader of a highly economically disadvantaged district, described the publishing of scores in this manner as getting “a black eye.” Put simply, extant studies (Houston et al., 2022; Jacobsen et al., 2014) suggest that *how* information is presented affects stakeholders’ attitudes toward the school, a conclusion supported in this study and one worthy of deeper consideration from a policy perspective.

Second, it is recommended that accountability policy require a more diverse set of measures that tell a holistic story of a school. Participants expressed that the measures used to evaluate schools, despite being better and wider than ever before, fall short of capturing the totality of what schools do. This is most problematic for schools that have high rates of economic disadvantage, poverty, or other barriers that restrict access and opportunity. Those schools expend more money, time, and resources on student supports – physical and mental health, social-emotional skills, removing barriers to attendance,

responding to substance abuse or other crises within households, etc. – to meet basic needs. These efforts, however, are not reflected in the public accountability dashboard, where outcomes and outputs are naturally prioritized over inputs and processes. While the accountability measures are not *inaccurate*, they may be *incomplete*, in turn setting the stage for public misunderstanding, political scrutiny, and a potentially misleading representation of the very thing that the system is attempting to measure: the effectiveness of the school. Unfortunately, as supported by the literature (Gagnon & Schneider, 2019), signaling that high-need schools are failing schools may, in turn, diminish public support and encourage families to exit those schools, having an unintentionally adverse effect on them.

Third, it is recommended to conduct a thorough and scientific review of the scope and limitations of quantitative performance measures, especially as they relate to advantaged and disadvantaged districts. For over two decades, the extant literature (Kain & Staiger, 2002; Monarrez & Chingos, 2020; Price, 2010; Tienken, 2017) suggests that performance outcomes may result from a complex and even blurry cluster of factors that are outside of the school's direct control, to include socioeconomic factors, community demographics, and family background. As discussed earlier in this paper, David Berliner noted in his letter to Philadelphia citizens that, when it comes to student performance, outside-of-school factors are *six times* more impactful than a teacher (Ravitch, 2018). Moreover, as of the writing of this dissertation, Pennsylvania finds itself amidst a court ruling that has deemed its funding of public schools unconstitutional, and the state is wrestling with the difficult problem of how to fix the system to adequately fund schools. Participants noted that some schools and districts are better positioned to offer advanced

classes, work-based learning experiences, and supports to assist students as they transition to post-secondary life, all of which cost money in terms of staffing and programming, and which require access in terms of geographic location and community partnerships. Adequate school funding can address some, though not all, of these issues.

Within this context, it may be unsurprising that participants expressed stronger agreement that the accountability dashboard effectively captures *student performance* in school than it does a *school's effectiveness, instructional quality, or leadership quality*.

At first glance, a reader may interpret this to mean that district leaders are eschewing their responsibility for student performance, simply ascribing performance to the uncontrollable factors around them. However, the evidence suggests otherwise.

Participants took a more nuanced approach to this concept: while leaders valued high-quality instruction and leadership, they asserted that it is possible for a school with an effective leader and strong teachers to appear poorly performing on the accountability dashboard. Conversely, a more advantaged school at which students experience fewer barriers to learning and, as the superintendent of the lowest ED school in the study stated, “arrive to us on third base,” may appear to be strongly performing on the accountability dashboard *regardless* of leadership and instructional quality.

In other words, they do not perceive a singular or direct correlation that can be established between what appears on the dashboard and the actual quality of a school. As this nuance and complexity is difficult to explain, it is mostly left unaddressed and uncontextualized within public-facing accountability. As a result, what was meant to be simplified and straightforward communication with the public can become a minefield of assumptions – some founded and some not – that are highly consequential for schools,

teachers, principals, and district leaders. This point centers on establishing clarity as to the purpose of public-facing accountability. After all, asking “How are the students doing?” is a fundamentally different question from “What does the school contribute?” (Gill, 2021, para. 5). As accountability policy is intended to generate productive community discourse, it is crucial to ensure that the data are appropriately defensible, reliable, and nuanced.

The study provided invaluable insights from Pennsylvania district leaders. Their perceptions support the call for new conceptualizations of public accountability. This suggests practitioner support for Bae’s (2018) “next phase” accountability, inclusive of a broader set of outcome measures, a mix of state and local indicators, measures of opportunities to learn, data dashboards, and school quality reviews by diverse, external stakeholders. Similarly, given the complexity of modern schooling’s broad range of aims, Jiminez and Sargrad (2017) recommend a more robust system consisting of state-, district-, and school-level inputs, processes, outputs, and outcomes that, taken together, provide a more holistic picture of a school. The leaders in this study recognize education as a public good and defend strong accountability as one policy lever that can make school better for children. Their insights, supported by the literature, create a roadmap to improve accountability policy.

### ***Leadership***

Soon after the implementation of *ESSA* and the new accountability model, public education entered a period of turbulence stemming from both the Covid-19 pandemic as well as increased political strife at the school board level. Both teaching and administrative positions are increasingly difficult to fill, and superintendent turnover has

risen. Participants indicated that their local communities, which have always shaped and steered district priorities and focus areas, have played an even more prominent role in recent years. Moreover, communities differ in their values, priorities, and interests. Some value a high ranking on the U.S. News and World Report list of best schools. Others are more interested in student activities, experiences, and opportunities for work-based learning and future success. Still others, predominantly in locations with higher poverty and crime rates, weigh safety and security measures as most important.

This means that leaders are already responding to these local imperatives, either intentionally or unintentionally, by creating their own accountability systems that respond to their unique contexts. In doing so, they describe a complex landscape in which the mandated accountability measures established at the federal and state levels align or collide with local needs and values to varying degrees. Consequently, they are left to navigate and respond to coexisting, sometimes conflicting, planes of accountability.

The question that emerges from this study and others, then, is: how might leaders develop systems that bring these various planes of accountability into better, more tightly coupled alignment, in strategic and intentional ways that move their mission forward, tell a holistic story of their schools, and improve education for students?

As policy mediators, district leaders wield immense influence to affect the implementation and communication of accountability in their own spheres. The findings in this study and others suggest that effective leaders interweave tenets of strong leadership with modern conceptions of public accountability. Marzano and Waters (2009) describe the importance of district leadership that establishes clear goals, collaborates with stakeholders, and allocates resources accordingly. It is, after all, the leader's role and

responsibility to transition from a *reactive* to a *proactive* approach in synthesizing the values, priorities, and viewpoints of diverse stakeholders into a well-defined and mission-driven accountability model.

To do this, leaders are encouraged to employ progressive practices in accountability. Bae and Jiminez and Sargrad, referenced above, advocate blending mandated accountability with other important, local contextual data that inform the public of current challenges, diagnose systems in need of improvement, and highlight areas of interest to the local community. Many leaders in the present study are already “telling their stories” in their own ways: by sharing state accountability metrics; other local metrics; surveys such as the Pennsylvania Youth Survey (PAYS) that address substance abuse, alcohol use, and violence within communities; climate and culture data; discipline statistics; extra-curricular and athletic news; career readiness opportunities; vignettes of student success; and progress toward local goals. The superintendent of the most economically disadvantaged school district noted that she communicates through pictures, her most effective way of sharing and describing the work of the schools.

It is recommended that leaders examine these strategies and adapt them to their own district’s needs and contexts. The most successful leaders will develop their own communication plan, perhaps to include a broader and wider local accountability dashboard, that weaves together quantitative and qualitative data to elicit deeper understanding, richer discussion, and tangible reform. This is possible only when accountability and performance metrics are clearly defined and articulated. In a post-pandemic, highly politicized educational environment, this is challenging to accomplish but also critical for success.

## **Recommendations for Future Research**

This field of study would benefit from quantitative research on larger samples of district-level leaders to determine whether the findings in the present study are generalizable across the population. For example, an in-depth survey of all Pennsylvania superintendents would contribute to a valuable program evaluation of *ESSA*-era mandated accountability and performance measures. Similarly, a nationwide survey of district-level leaders would provide equally valuable insights about the state of school performance measurement across the country (a national instrument would, of course, need to be sufficiently broad to allow for differences across 50 unique statewide plans). This type of quantitative analysis would be beneficial to ascertain the perspectives of other stakeholders (principals, teachers, parents, students) as well, and to determine significant relationships among them. Such research would inform the development of useful measurement instruments that could effectively meet policy objectives.

Conversely, qualitative case study research has the potential to provide insight into the internal organizational dynamics that drive accountability. For example, an in-depth qualitative analysis of a single school district could examine how mandated and unmandated accountability and performance measures are interpreted, communicated, and acted upon as they traverse multiple layers of authority and various stakeholder groups. While this type of a study may not yield the policy implications of a broad quantitative study, it could provide invaluable information for district and building leaders and practitioners navigating accountability in their own contexts. Understanding these formal and informal pathways of communication would allow future leaders to

make better decisions, faster, and to enhance and maximize their own public communication on accountability.

Yet another untapped area for future research is to analyze various secondary data to determine relationships between districts' or schools' accountability outcomes with other data. For example, district strategic plans – the documents that encapsulate districts' values, mission, and improvement actions in the eyes of the state – could be collected, coded, and analyzed to determine the extent to which they align with areas of need identified in mandated accountability. These two data sources represent the “input” and “output” of districts' and schools' performance efforts. Understanding the extent to which they interweave would be valuable from both policy and leadership perspectives.

### **Conclusion**

Schools and school districts should be held accountable for their quality and effectiveness through the measurement of specific performance metrics that are publicly communicated. However, accountability policy is a complex endeavor, expanding and contracting over many decades in response to new scientific understandings about learning, political pressures and forces, and evolving conceptions about the role of school in society and what it is expected to deliver. Today, our educational system is in the early years of the implementation of the *Every Student Succeeds Act*'s (ESSA) newly expanded accountability dashboards, is evaluating the progress of student learning as we emerge from the Covid-19 pandemic, and is experiencing the impact of strong political, cultural, and social forces. Public education is being pulled in often divergent directions as it wrestles with complex questions about not only *how* students and schools are performing, but *what* should be measured in the first place.

It is within this context that the present study investigates the perceptions of the leaders of 30 Pennsylvania school districts across the continuum of economic disadvantage. It expands on previous research and literature that explored the benefits and challenges of mandated accountability and that, most recently, has sought to understand how stakeholders receive and make sense of performance measurement. Findings suggest that 1) district leaders' perceptions of accountability are complex and nuanced, 2) that they both believe in the importance of public-facing accountability and, simultaneously, raise concerns about its limitations and potential misrepresentation of schools (especially disadvantaged schools), 3) that they value many of the existing performance measures while also desiring additional, more robust measures, and 4) that individual leaders' approaches to interpreting and communicating accountability measures are not uniform, but are driven by the leader's personal beliefs as well as community interests.

It is to the credit of our federal and state governments that they have striven to enhance the school accountability model. As Gagnon and Schneider (2019) wrote, if we are not thoughtful in our approach to accountability, "those harmed, as a result, are those most in need of our care" (p. 755). Therefore, it is the hope that this study contributes, even in a small way, to the continued pursuit of school accountability that thoughtfully measures what is most important, recognizes both local needs as well as our shared educational mission, and improves the educational experience for students.

## REFERENCES

- Abelmann, C., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). When accountability knocks, will anyone answer? *Consortium for Policy Research in Education*.
- Adams, C. M., Ford, T. G., Forsyth, P. B., Ware, J. K., Olsen, J. J., Lepine Sr, J. A., & Mwavita, M. (2017). Next generation accountability: A vision for school improvement under ESSA. *Learning Policy Institute*.
- Alexander, L. (1986). "Time for Results:" An overview. *The Phi Delta Kappan*, 68(3), 202-4.
- American Psychological Association. (n.d.). *Education and Socioeconomic Status*. <https://www.apa.org/pi/ses/resources/publications/education>.
- Amrein-Beardsley, A., Berliner, D. C., & Rideau, S. (2010). Cheating in the first, second, and third degree: Educators' responses to high-stakes testing. *Education Policy Analysis Archives*, 18(14).
- Babbie, E. R. (2008). *The practice of social research*. The Thomas Corporation.
- Babbie, E. R. (2020). *The practice of social research*. Cengage Learning.
- Bae, S. (2018). Redesigning systems of school accountability: A multiple measures approach to accountability and support. *Education Policy Analysis Archives*, 26(8).
- Behn, R. D. (2003). Rethinking accountability in education: How should who hold whom accountable for what? *International Public Management Journal*, 6(1), 43-74.
- Benveniste, G. (1985). The design of school accountability systems. *Educational Evaluation and Policy Analysis*, 7(3), 261-279.
- Berliner, D., & Biddle, B. (2001). Thinking about education in a different way. *The Jossey-Bass Education Series*, 407.
- Boyd, D., Lankford, H., Loeb, S., & Wyckoff, J. (2008). The impact of assessment and accountability on teacher recruitment and retention: Are there unintended consequences? *Public Finance Review*, 36(1), 88-111.
- Braun, H. I., & Mislevy, R. (2005). Intuitive test theory. *Phi Delta Kappan*, 86(7), 488-497.

- Brewer, C., Knoepfel, R. C., & Lindle, J. C. (2015). Consequential validity of accountability policy: Public understanding of assessments. *Educational Policy*, 29(5), 711-745.
- Bush, G. H. W. (1991). Address to the nation on the national education strategy. *Gov Info*. <https://www.govinfo.gov/content/pkg/PPP-1991-book1/html/PPP-1991-book1-doc-pg395-2.html>.
- Bush, G. W. (2001). No Child Left Behind. *Department of Education*.
- Cain, G. G., & Watts, H. W. (1970). Problems in making policy inferences from the Coleman Report. *American Sociological Review*, 35(2), 228-242.
- Carey, K. (2012). Requiem for a failed education policy: The long slow death of No Child Left Behind. *The New Republic*. <https://newrepublic.com/article/104960/requiem-failed-education-policy-the-long-slow-death-no-child-left-behind>.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Clotfelter, C. T., Ladd, H. F., Vigdor, J. L., & Diaz, R. A. (2004). Do school accountability systems make it more difficult for low-performing schools to attract and retain high-quality teachers? *Journal of Policy Analysis and Management*, 23(2), 251-271.
- Coburn, C. E. (2004). Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom. *Sociology of Education*, 77(3), 211-244.
- Cohen, M. D., March, J. G., & Olsen, J. P. (1972). A garbage can model of organizational choice. *Administrative Science Quarterly*, 72(1), 1-25.
- Coleman, J. S. (1966). Equality of educational opportunity. *U.S. Department of Health*.
- Darling-Hammond, L. (2007). Evaluating No Child Left Behind. *The Nation*. [http://www.leland.stanford.edu/~hakuta/Courses/Ed205X%20Website/Resources/LDH\\_%20Evaluating%20'No%20Child...pdf](http://www.leland.stanford.edu/~hakuta/Courses/Ed205X%20Website/Resources/LDH_%20Evaluating%20'No%20Child...pdf).
- Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., & Stosich, E. L. (2016). Pathways to new accountability through the Every Student Succeeds Act. *Learning Policy Institute*. [https://learningpolicyinstitute.org/media/176/download?inline&file=Pathways\\_New-Accountability\\_Through\\_Every\\_Student\\_Succeeds\\_Act\\_04202016.pdf](https://learningpolicyinstitute.org/media/176/download?inline&file=Pathways_New-Accountability_Through_Every_Student_Succeeds_Act_04202016.pdf).

- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.
- Dee, T. S., Jacob, B. A., Hoxby, C. M., & Ladd, H. F. (2010). The impact of No Child Left Behind on students, teachers, and schools. *Brookings papers on economic activity*, 149-207.
- Dee, T. S., Jacob, B., & Schwartz, N. L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 35(2), 252-279.
- Desimone, L. M. (2006). Consider the source: Response differences among teachers, principals, and districts on survey questions about their educational policy environment. *Educational Policy*, 20(4), 640-76.
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 48(2), 147-160.
- Doerschuk, P., Bahrim, C., Daniel, J., Kruger, J., Mann, J., & Martin, C. (2016). Closing the gaps and filling the STEM pipeline: A multidisciplinary approach. *Journal of Science Education and Technology*, 25, 682-695.
- Education Trust. (2016). New school accountability systems in the states: Both opportunities and peril. <https://edtrust.org/new-school-accountability-systems-in-the-statesboth-opportunities-and-peril/>.
- Englert, K. S., Fries, D. A., Martin-Glenn, M. L., & Douglas, B. B. (2007). Accountability systems: A comparative analysis of superintendent, principal, and teacher perceptions. *International Journal of Education Policy and Leadership*, 2(4).
- ESEA Network. (n.d.) *About ESEA*. ESEA Network. <https://www.eSeanetwork.org/about/esea>.
- Farah, M. J. (2018). Socioeconomic status and the brain: prospects for neuroscience-informed policy. *Nature Reviews Neuroscience*, 19(7), 428-438.
- Färe, R., Grosskopf, S., & Weber, W. L. (1989). Measuring school district performance. *Public Finance Quarterly*, 17(4), 409-428.
- Figlio, D. N. (2006). Testing, crime and punishment. *Journal of Public Economics*, 90(4-5), 837-851.
- Figlio, D. N., & Getzler, L. S. (2006). Accountability, ability and disability: Gaming the system? *Improving School Accountability*, 14, 35-49.

- Figlio, D., & Loeb, S. (2011). School accountability. *Handbook of the Economics of Education*, 3, 383-421.
- Figlio, D. N., & Lucas, M. E. (2004). What's in a grade? School report cards and the housing market. *American Economic Review*, 94(3), 591-604.
- Figlio, D. N., & Winicki, J. (2005). Food for thought: the effects of school accountability plans on school nutrition. *Journal of Public Economics*, 89(2-3), 381-394.
- Finn, C. E. (2002). Real accountability in K-12 education. School accountability. *The Hoover Institute*.
- Finn, C. (2020). School accountability: Past, present, and future: Findings and recommendations for state and district policymakers. *The Hoover Institute*.
- Frels, R. K., & Onwuegbuzie, A. J. (2013). Administering quantitative instruments with qualitative interviews: A mixed research approach. *Journal of Counseling & Development*, 91(2), 184-194.
- Gagnon, D. J., & Schneider, J. (2019). Holistic school quality measurement and the future of accountability: Pilot-test results. *Educational Policy*, 33(5), 734-760.
- Gill, B. (2021). Making sure school performance measures provide the right diagnosis to improve student outcomes. *Education Next*.  
<https://www.educationnext.org/making-sure-school-performance-measures-provide-the-right-diagnosis-to-improve-student-outcomes/>.
- Glass, G. V. (1972). The many faces of 'educational accountability'. *The Phi Delta Kappan*, 53(10), 636-639.
- Goldhaber, D., & Özek, U. (2019). How much should we rely on student test achievement as a measure of success? *Educational Researcher*, 48(7), 479-483.
- Gottlieb, D., & Schneider, J. (2020). Educational accountability is out of step – now more than ever. *Phi Delta Kappan*, 102(1), 24-25.
- Greene, J. P. (2017). Futile accountability systems should be abandoned. *Education Next*, 17(3), 51-54.
- Greene, J. P., Winters, M. A., & Forster, G. (2004). Testing high-stakes tests: Can we believe the results of accountability tests? *Teachers College Record*, 106(6), 1124-1144.
- Grim, B., Gromis, J., & Harmon, A. H. (2006). Focused group interviews as an innovative quanti-qualitative methodology (QQM): Integrating quantitative elements into a qualitative methodology. *The Qualitative Report*, 11(3), 516-537.

- Halperin, S. (1975). ESEA ten years later. *Educational Researcher*, 4(8), 5-9.
- Hamilton, L. S., Berends, M., & Stecher, B. M. (2005). Teachers' responses to standards-based accountability. [Working paper]. *RAND: Changes to Instruction*.
- Hanushek, E. A., & Raymond, M. E. (2001). The confusing world of educational accountability. *National Tax Journal*, 54(2), 365-384.
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24(2), 297-327.
- Hanushek, E. A., & Raymond, M. E. (2002). Lessons about the design of state accountability systems. [Paper presentation]. *Taking Account of Accountability: Assessing Policy and Politics Conference*.
- Hibel, J., & Penn, D. M. (2020). Bad apples or bad orchards? An organizational analysis of educator cheating on standardized accountability tests. *Sociology of Education*, 93(4), 331-352.
- Houston, D. M., Henderson, M., Peterson, P. E., & West, M. R. (2022). Status, growth, and perceptions of school quality. *Educational Evaluation and Policy Analysis*, 44(1), 105-126.
- Howe, K. R., & Murray, K. (2015). Why school report cards merit a failing grade. *National Education Policy Center*.
- Huffman, K. (2017). If parents push for it, accountability can work. *Education Next*, 17(3), 51-55.
- Husband, T., & Hunt, C. (2015). A review of the empirical literature on No Child Left Behind from 2001 to 2010. *Planning and Changing*, 46(1/2), 212.
- Imberman, S. A., & Lovenheim, M. F. (2016). Does the market value value-added? Evidence from housing prices after a public release of school and teacher value-added. *Journal of Urban Economics*, 91, 104-121.
- Ingersoll, R., Merrill, L., & May, H. (2016). Do accountability policies push teachers out? *Educational Leadership*, 73(8), 44.
- Jacob, B. (2005). Accountability, incentives and behavior. *Journal of Public Economics*, 89(5-6), 761-796.
- Jacobsen, R., Saultz, A., & Snyder, J. W. (2013). When accountability strategies collide: Do policy changes that raise accountability standards also erode public satisfaction? *Educational Policy*, 27(2), 360-389.

- Jacobsen, R., Snyder, J. W., & Saultz, A. (2014). Informing or shaping public opinion? The influence of school accountability data format on public perceptions of school quality. *American Journal of Education*, *121*(1), 1-27.
- Jimenez, L., & Sargrad, S. (2017). A new vision for school accountability. *Center for American Progress*, *3*.
- Johnson, L. B. (1965). Lyndon B. Johnson: Remarks in Johnson City, Texas, Upon signing the Elementary and Secondary Education Bill. *The American Presidency Project*. <https://www.presidency.ucsb.edu/documents/remarks-johnson-city-tex-upon-signing-the-elementary-and-secondary-education-bill>.
- Kamenetz, A. (2018). What 'A Nation At Risk' got wrong, and right, about U.S. schools. *National Public Radio*.
- Kane, T. J., & Staiger, D. O. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, *16*(4), 91-114.
- Kane, T. J., Staiger, D. O., Samms, G., Hill, E. W., & Weimer, D. L. (2003). School accountability ratings and housing values. *Brookings-Wharton Papers on Urban Affairs*, 83-137.
- Klein, A. (2016). The Every Student Succeeds Act: An ESSA overview. *Education Week*. <https://www.edweek.org/policy-politics/the-every-student-succeeds-act-an-essa-overview/2016/03>.
- Knapp, M. S., & Feldman, S. B. (2012). Managing the intersection of internal and external accountability: Challenge for urban school leadership in the United States. *Journal of Educational Administration*, *50*(5), 666-694.
- Koretz, D. (1996). Using student assessments for educational accountability. *The National Academies*.
- Koretz, D. (2008). A measured approach. *American Educator*, *32*(2), 18-39.
- Koretz, D. (2018). Moving beyond the failure of test-based accountability. *American Educator*, *41*(4), 22-26.
- Koretz, D. M., McCaffrey, D. F., & Hamilton, L. S. (2001). Toward a framework for validating gains under high-stakes conditions. *Center for the Study of Evaluation, National Center for Research on Evaluation*. University of California.
- Kress, S., Zechmann, S., & Schmitten, J. M. (2011). When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Journal on Legislation*, *48*, 185.

- Ladd, H. F. (2001). School-based educational accountability systems: The promise and the pitfalls. *National Tax Journal*, 54(2), 385-400.
- Lake, R., & Worthen, M. (2021). State accountability systems in the COVID era and beyond. *Center on Reinventing Public Education*.
- Levin, H. M. (1974). A conceptual framework for accountability in education. *The School Review*, 82(3), 363-391.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
- Mandinach, E. B., Miskell, R. C., & Gummer, E. S. (2020). Parental educational decision making: The information they seek and what they want from data systems. *Teachers College Record*, 122(1), 1-42.
- Manno, B. V. (2018). George H.W. Bush: The education president. *The Thomas B. Fordham Institute*. <https://fordhaminstitute.org/national/commentary/george-hw-bush-education-president>.
- Marzano R. & Waters, T. (2009). District leadership that works: Striking the right balance. *Solution Tree*. [Book]
- Mehta, J. (2015). Escaping the shadow: “A Nation at Risk” and its far-reaching influence. *American Educator*, 39(2), 20.
- Mertens, D. M. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage Publications.
- Meyer, J. W., & Rowan, B. (1977). Institutionalized organizations: Formal structure as myth and ceremony. *American Journal of Sociology*, 83(2), 340.
- Meyer, H. D., Tröhler, D., Labaree, D. F., & Hutt, E. L. (2014). Accountability: Antecedents, power, and processes. *Teachers College Record*, 116(9), 1-12.
- Mintrop, H. (2012). Bridging accountability obligations, professional values and (perceived) student needs with integrity. *Journal of Educational Administration*, 50(5), 695-726.
- Monarrez, T., & Chingos, M. (2020). Can we measure school quality using publicly available data? *Center on Education Data Policy*.
- Murnane, R. J. (2021). Can policy interventions reduce inequality? Looking beyond test scores for evidence. *William T. Grant Foundation*.

- National Commission on Excellence in Education. (1983). A nation at risk: The imperative for educational reform. <https://edreform.com/wp-content/uploads/2013/02/ANationAtRisk1983.pdf>.
- Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2), 263-283.
- Oakes, J. (1989). What educational indicators? The case for assessing the school context. *Educational Evaluation and Policy Analysis*, 11(2), 181-199.
- Osborne, D. (2016). Creating measurement and accountability systems for 21st century schools: Guide for state policymakers. *Progressive Policy Institute*.
- Owens, T., & Peltier, D.G. (2002). Communicating with parents: Parental perceptions of state school accountability reporting. *Education*, 122(3), 595-599.
- Pennsylvania Department of Education. (2024a). Future Ready PA Index. (n.d.). Glossary. <https://www.education.pa.gov/K-12/ESSA/FutureReady/Pages/default.aspx>.
- Pennsylvania Department of Education. (2024b). Future Ready PA Index. (n.d.). Glossary. <https://futurereadypa.org/Home/Glossary>.
- Petrilli, M. J. (2019). A new era of accountability has barely just begun. *The Thomas B. Fordham Institute*.
- Polikoff, M. S. (2017). Why accountability matters, and why it must evolve. *Education Next*, 17(3), 51-54.
- Portz, J. (2021). Beyond test scores: "School Quality or Student Success" in state ESSA plans. *Teachers College Record*, 123(6), 1-36.
- Price, H. E. (2010). Does No Child Left Behind really capture school quality? Evidence from an urban school district. *Educational Policy*, 24(5), 779-814.
- Ramanathan, A. (2008). Paved with good intentions: The federal role in the oversight and enforcement of the Individuals with Disabilities Education Act (IDEA) and the No Child Left Behind Act (NCLB). *Teachers College Record*, 110(2), 278-321.
- Ravitch, D. (1981). The meaning of the new Coleman report. *The Phi Delta Kappan*, 62(10), 718-720.
- Ravitch, D. (2005, November 7). Every state left behind. *The New York Times*. <https://www.nytimes.com/2005/11/07/opinion/every-state-left-behind.html>.

- Ravitch, D. (2018). David Berliner offers advice to the citizens of Philadelphia about their public schools. <https://dianeravitch.net/2017/12/21/david-berliner-offers-advice-to-the-citizens-of-philadelphia-about-their-public-schools/>.
- Reardon, S. F., Valentino, R. A., Kalogrides, D., Shores, K. A., & Greenberg, E. H. (2013). Patterns and trends in racial academic achievement gaps among states, 1999-2011. Unpublished Working Paper. *Center for Education Policy Analysis, Stanford University*.
- Rouse, C. E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy*, 5(2), 251-281.
- Rorrer, A. K., & Skrla, L. (2005). Leaders as policy mediators: The reconceptualization of accountability. *Theory into Practice*, 44(1), 53-62.
- Russell, J., & McCombs, J. S. (2006). Superintendent, principal, and teacher understanding of and attitudes about accountability. *RAND*.
- Saultz, A., Schneider, J., & McGovern, K. (2019). Why ESSA has been reform without repair. *Phi Delta Kappan*, 101(2), 18-21.
- Schneider, J. (2017). The urban school stigma. *The Atlantic*. <https://www.theatlantic.com/education/archive/2017/08/the-urban-school-stigma/537966/>.
- Schneider, J. (2017). Why Americans think so poorly of the country's schools. *The Atlantic*.
- Schneider, J. (2021). Ratings, rankings, and segregation: The failure of measurement and accountability in education. *Poverty and Race Research Action Council*, 30(3), 9-14.
- Schneider, J., Noonan, J., White, R. S., Gagnon, D., & Carey, A. (2021). Adding "student voice" to the mix: Perception surveys and state accountability systems. *AERA Open*, 7.
- Seashore Louis, K., & Robinson, V. M. (2012). External mandates and instructional leadership: School leaders as mediating agents. *Journal of Educational Administration*, 50(5), 629-665.
- Shapiro, J. P., & Stefkovich, J. A. (2016). *Ethical leadership and decision making in education: Applying theoretical perspectives to complex dilemmas*. Routledge.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.

- Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, 27(5), 556-563.
- Spurrier, A., Aldeman, C., Schiess, J. O. N., & Rotherham, A. J. (2020). The impact of standards-based accountability: Making next year count. *Summer 2020 Accountability Series*. Bellwether Education Partners.
- Stecher, B. M., Barron, S. L., Chun, T., & Ross, K. (2000). The effects of the Washington State Education Reform on schools and classrooms. *CSE Technical Report*.
- Stecher, B. M., & Vernez, G. (2010). What can we learn from the implementation of “No Child Left Behind”? *RAND Corporation*.
- Steiner, D. & Bjorklund-Young, A. (2020). The future of school accountability. *Hoover Institute*.
- Sutherland, D. H. (2022). School board sensemaking of federal and state accountability policies. *Educational Policy*, 36(5), 981-1010.
- Teddlie, C., & Tashakkori, A. (2009). Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences. Sage. [Book]
- Tienken, C. H., Colella, A., Angelillo, C., Fox, M., McCahill, K. R., & Wolfe, A. (2017). Predicting middle level state standardized test results using family and community demographic data. *RMLE Online*, 40(1), 1-13.
- U.S. Department of Education. (2024a). *Mission*. U.S. Department of Education. <https://www2.ed.gov/about/overview/mission/mission.html>.
- U.S. Department of Education. (2024b). *Every Student Succeeds Act (ESSA)*. U.S. Department of Education. <https://www.ed.gov/essa?src=rn>.
- U.S. Department of Education. (2024c). *ESSA Fact Sheet*. U.S. Department of Education. <https://www2.ed.gov/policy/elsec/leg/essa/essafactsheet170103.pdf>.
- Weick, K. E. (1976). Educational organizations as loosely coupled systems. *Administrative Science Quarterly*, 1-19.
- Weick, K.E. (2001). *Making sense of the organization*. Blackwell Publishing Ltd.
- Wescott, J. P. (1972). Accountability: For whom, to whom, for what? [Speech]. Annual Convention of American Association of School Administrators.

## APPENDIX A: INTERVIEW PROTOCOL

### Introductory Remarks

*Thank you very much for taking the time to participate in this interview. My name is Michael Gogoj, and I am completing dissertation research to fulfill part of the requirements for the doctoral degree program in the College of Education and Human Development at Temple University. This interview seeks to explore your perceptions of mandated accountability measures (via the Future Ready PA Index), how you communicate accountability outwardly to your public stakeholders, and those mandated as well as unmandated accountability measures that you believe are important to your public stakeholders or that are important to you as a leader. **I've reached out to superintendents across PA with diverse demographic and socioeconomic profiles.***

*Your thoughts are important, and I know that your responses will help me to better understand the experiences of educational leaders regarding school accountability. Thank you for being willing to participate in this interview.*

### Interview Procedures

*Before we get started, I would like to go over a few details of the procedures for this session. I have a set of questions related to school and district accountability. Some of the questions are open-ended and some ask you to respond given a set of choices.*

*I am collecting this information for research purposes. This interview will not be recorded. I will take written notes that capture your responses. I emphasize that no personally identifying information nor any identifying district information will be used in my dissertation. Your responses are confidential. Only aggregate data and patterns will be analyzed in relation to publicly available data. In the event that a direct quotation is used, no identifying information will be associated with the direct quotation. Pseudonyms will be used.*

*I will now go over the formal consent procedure and ask you to confirm your consent verbally.*

### Informed Consent

*Again, thank you for your time and your interest in this study. As part of the scheduling process, I sent an electronic copy of the Informed Consent Form. I will now display it on the screen and give you a moment to familiarize yourself with it if you need it.*

*(Wait a few minutes)*

*For protocol purposes, I need to make sure we cover specific information with respect to your rights so let me review key information from the Informed Consent Form before we begin.*

*The Informed Consent Form provides information regarding your rights as a participant. (1) Participation is voluntary; (2) Participants may choose not to participate at any time and there is no penalty if you do not participate; (3) Participants may choose not to provide responses to any or all questions/topics of discussion; (4) Data collected will be treated confidentially; (5) Data will be compiled, analyzed and reported only at the group level and aggregated. No individually identifying information about you or your organization will be included in the dissertation; (6) All paper and electronic files will be stored in areas with locked storage; where applicable stored using encrypted and password-protected files, computers, and software accounts; and (8) All hard copy notes will be destroyed upon completion of my degree.*

*Let me ask you to verbally verify your response to the following:*

1. *You have read and understood the procedures described in the Informed Consent Form.*
2. *You voluntarily agree to participate in this research.*
3. *You give your permission for me to take notes during this interview session.*

*Do you have any questions for me at this time either in terms of the content or process for our interview or this research effort?*

### **Background Questions**

*In order to establish context about you as a leader, I would like to ask some general questions about your personal characteristics. For any of these questions, you may answer that you “prefer not to say.”*

What is your current position?

Is your position considered a district-level position?

How many years have you been in your current position?

How many years total have you been in a district-level position?

How many years have you been in public education?

How many years of your public education experience has been in Pennsylvania?

How do you define your gender?

Which best describes your age?

Under 30

30-39 years old

40-49 years old

50-59 years old

60 or older

Prefer not to say

What is your race/ethnicity?

What is the highest level of education you have received?

In terms of state accountability, are you the person or one of the people in charge of overseeing and/or presenting state accountability measurements in your district?

Briefly describe your role in relation to mandated accountability measures / the Future Ready Index.

***Part I: The following questions solicit your general impressions of federal and state mandated accountability as it currently exists.***

Please describe your general impression of the federally- and state-mandated accountability measures, such as those found in the Future Ready Index.

In what ways do federally- and state-mandated accountability measures, such as those in the Future Ready Index, positively affect/impact your district, or advance your educational goals?

In what ways do federally- and state-mandated accountability measures, such as those in the Future Ready Index, negatively affect/impact your district, or hinder your educational goals?

Rate these statements using a Likert-scale. Your options are: you “strongly agree”, “agree,” are “neutral” on the statement, “disagree,” or “strongly disagree.” You may also answer that the statement is “not applicable” or that you are “undecided.” I will also ask you to briefly explain your rationale for each selection.

Statement	Selection	Explanation
You believe that the accountability measures in the Future Ready Index effectively capture <u>student performance</u> in schools.		
You believe that the accountability measures in the Future Ready Index provide reliable information/data on a <u>school's</u> overall effectiveness or performance.		
You believe the accountability measures in the Future Ready Index reflect the quality of instruction in schools (i.e. higher results mean that better instruction is taking place in particular schools).		
You believe the accountability measures in the Future Ready Index reflect the quality of leadership in schools (i.e. higher results mean that better leadership is taking place in particular schools).		
You value the mandated accountability measures in the Future Ready Index.		
Your district leadership prioritizes strong performance on standardized tests as a major district priority.		
In general, you believe the accountability measures in the Future Ready Index have had a positive effect on school climate.		
In general, you believe that your schools perform better in the areas that are publicly reported in the Future Ready Index.		
You have a strong understanding of the accountability measures in the Future Ready Index.		

Your district relies upon the accountability measures in the Future Ready Index as a major driver of strategic planning.		
You believe that schools have the resources/capacity to meet state benchmarks/expectations.		

Would you like to make additional comments about any of these statements?

How do you use federally- and state-mandated accountability measures, such as those found in the Future Ready Index, in your leadership capacity?

***Part II: The next set of questions explore your personal experiences and beliefs related to accountability measures, both mandated (via the Future Ready Index) and unmandated via local interests and your own personal beliefs.***

This is a multiple-choice question: The Future Ready Index is designed to be a public-facing representation of every school in PA. Which statement best describes your position on accountability measures included in the Future Ready Index?

- A – You like the existing measures and would not change them.
- B – You are satisfied with the existing measures but would add additional measures.
- C – You do not like the existing measures and would change them.
- D – You do not think publicly reporting measures is the right approach to accountability.
- E – Neutral or no preference

Please explain your selection.

In this question, I will identify several areas from the existing Future Ready Index. Please rate the extent to which you believe the following accountability measures are valuable public-facing accountability measures. Your choices are: *highly valuable, somewhat valuable, not valuable*. I will ask you to explain your answer to some of these. You are welcome to explain your answer to any of them.

<b>Statement</b>	<b>Selection</b>	<b>Explanation</b>
Measure(s) of student achievement on state standardized tests		Required:
Measure(s) of student growth on state standardized tests		Required:
Measure(s) of English learners who meet state benchmarks for language growth / attainment		Optional:
Measure(s) of graduation rates		Required:
Measure(s) of completion of career readiness activities		Optional:
Measure(s) of participation/completion of industry-based programs		Optional:
Measure(s) of participation rates in advanced classes		Optional:
Measure(s) of student transition to post-secondary education, work, or the military		Optional:

Would you like to make additional comments about any of these metrics?

Think of your ideal public-facing accountability dashboard. What measures, data, or information should be publicly presented to “tell the story” of a school’s performance?

The following group of accountability measures are not currently mandated measures. Please rate the extent to which you believe the following would be a valuable public-facing accountability measure. For this exercise, you may presume they would be presented similarly to any other measure on the Future Ready Index and disaggregated by subgroup. Your choices are: *highly valuable*, *somewhat valuable*, or *not valuable*.

<b>Indicator</b>	<b>Selection</b>	<b>Explanation</b>
Measure(s) of aggregated grades (A, B, C, D, F, etc.) in local classes (for example: sharing percentages of students who earn As, Bs, Cs, etc.)		Optional:
Measure(s) of students' higher-level thinking skill development		Optional:
Measure(s) of students' social-emotional skill development		Optional:
Measure(s) of students' civic engagement and participation		Optional:
Measure(s) of students' perseverance, determination, and growth mindset		Optional:
Measure(s) of students' financial skill development		Optional:
Measure(s) of stakeholder(s) perception of climate and culture		Optional:
Measure(s) of curricular offerings		Optional:
Measure(s) of extra-curricular activities/experiences offered		Optional:
Measure(s) of student participation rates in extra-curricular activities/experiences		Optional:
Measure(s) of student participation in internships, externships, work-based learning		Optional:
Measure(s) of family / parent engagement opportunities		Optional:

Measure(s) of school-community partnerships		Optional:
Measure(s) related to school safety		Optional:

Are there other measures, not listed here, that you believe would be valuable or highly valuable, regardless of whether or not you employ them in your own district? Why do you find them valuable?

Would you like to make additional comments about any of these metrics?

To whom or to what do you feel most accountable? Explain.

***Part III: The following questions explore your interactions with your public stakeholders around accountability measures. Public stakeholders are defined as parents, families, community members, and school board members (as they are representative of the larger community).***

In your district, who is the primary communicator of federal and state accountability results to the school board?

In your district, who is the primary communicator of federal and state accountability results to the parents/families?

On average, which best describes how often you (or someone else) present federal and state accountability results (the Future Ready Index) to the following stakeholders? Your options are *never, less than annually, annually, bi-annually, quarterly, more than quarterly*.

To the Board of Directors?

To district leadership?

To principals?

To teachers and staff?

To parents/families?

This is a multiple-choice question: Which best describes the frequency of questions, inquiries, or comments from families / parents regarding federal and state accountability?

- A – Never
- B – Infrequently (A few times per year)
- C – Frequently (Monthly or more frequently)

This is a multiple-choice question: Which best describes the frequency of questions, inquiries, or comments from community members (not families / parents) regarding federal and state accountability?

- A – Never
- B – Infrequently (A few times per year)
- C – Frequently (Monthly or more frequently)

What are the top issues (ranked in order of importance) for which you are held most accountable by your school board?

What are the top issues (ranked in order of importance) for which you are held most accountable by your parents/families?

Outside of the mandated federal and state accountability measures, what other measures, if any, do you publicly share with your stakeholders?

How would you describe the relationship between the state-mandated accountability measures and these other measures? (Do they complement each other or hinder each other?) How do you navigate this relationship as a district leader?

How do you publicly share information on accountability measures to families? (What communication practices do you employ?)

What benefits have you experienced related to the public presentation of mandated accountability measures?

What challenges have you experienced related to the public presentation of, or public understanding of, mandated accountability measures?

What practices have you employed to overcome these challenges?

Is there anything else you would like to share about any of today's topics?

### **Conclusion**

*Thank you again for participating in this interview.*

*Again, no identifying information will be presented in my dissertation. This includes individuals' names, school district names, or any other identifying information. Data will be presented only in aggregated form. Again, in the event that a direct quotation is used, no identifying information will be associated with the direct quotation.*

*Thank you!*

## APPENDIX B: SAMPLE TERTILES VS. PENNSYLVANIA TERTILES

*Lowest Economic Disadvantage Tertile of Sampled School Districts vs.  
Lowest Economic Disadvantage Tertile of All Pennsylvania School Districts*

School District Characteristics	Sample Tertile (10)		State Tertile (166)	
	Mean	SD	Mean	SD
<b>Economically Disadvantaged</b>				
Percent Economically Disadvantaged	24.13	9.92	25.40	8.56
<b>District Characteristics</b>				
District Enrollment	4132.60	2080.44	3725.36	2676.77
Geographic Size in Miles	76.00	66.64	61.03	51.17
Number of Schools	6.70	3.13	5.78	3.19
Number of Teachers	283.22	161.62	268.39	191.40
Student:Teacher Ratio	14.97	1.68	13.75	1.54
Charter School Enrollment	146.80	113.83	106.99	100.38
Enrollment in Partnering Career and Technical Center	125.40	81.44	108.75	83.43
<b>Student Race / Ethnicity</b>				
2 or More Races	3.62	1.65	3.93	2.04
American Indian/Alaskan Native	0.14	0.11	0.13	0.11
Asian	3.61	3.25	4.54	5.45
Black/African American	2.20	1.55	3.23	2.81
Hispanic	6.25	3.10	6.04	5.18
Native Hawaiian/Pacific Islander	0.07	0.06	0.07	0.08
White	84.11	7.23	82.07	10.48
<b>Student Gender</b>				
Female	48.63	0.70	48.53	1.09
Male	51.38	0.70	51.47	1.09
<b>Student Identifications</b>				
Special Education	14.80	2.75	16.14	3.11
English Learner	1.54	1.06	1.83	1.81
Percent of Gifted Students	4.68	2.13	4.25	2.36
Homeless	0.79	0.53	0.82	0.61
Foster Care	0.35	0.26	0.41	0.35
Military Connected	1.00	0.99	0.70	1.06
<b>District Revenue</b>				
Total Revenue By Student	\$18,349.61	\$2,531.43	\$19,885.84	\$3,176.38
Total Federal Revenue By Student	\$417.75	\$286.33	\$406.97	\$194.51
Total Local Revenue By Student	\$12,699.64	\$2,289.12	\$13,562.84	\$3,988.16
Total State Revenue By Student	\$5,232.22	\$1,289.93	\$5,916.03	\$2,113.26
<b>District Expenditures</b>				
Total Expenditures By Student	\$14,898.00	\$1,651.28	\$16,110.71	\$2,550.09
Instructional Expenditures By Student	\$9,540.03	\$887.90	\$10,195.99	\$1,700.64
Student and Staff Support Expenditures By Student	\$1,279.93	\$352.41	\$1,489.29	\$430.98
Administration Expenditures By Student	\$1,544.21	\$180.92	\$1,725.98	\$474.86
Operations Expenditures By Student	\$2,533.83	\$775.96	\$2,699.44	\$644.81

*Middle Economic Disadvantage Tertile of Sampled School Districts vs.  
Middle Economic Disadvantage Tertile of All Pennsylvania School Districts*

School District Characteristics	Sample Tertile (10)		State Tertile (166)	
	Mean	SD	Mean	SD
<b>Economically Disadvantaged</b>				
Percent Economically Disadvantaged	43.00	3.61	42.95	3.48
<b>District Characteristics</b>				
District Enrollment	2598.70	1273.76	1994.25	1467.19
Geographic Size in Miles	127.81	85.69	118.26	97.53
Number of Schools	4.60	1.71	3.96	1.95
Number of Teachers	187.91	94.28	149.79	106.00
Student:Teacher Ratio	13.87	1.89	12.99	1.72
Charter School Enrollment	157.90	126.48	94.74	105.93
Enrollment in Partnering Career and Technical Center	92.11	52.32	84.75	65.39
<b>Student Race / Ethnicity</b>				
2 or More Races	3.66	1.70	3.32	2.13
American Indian/Alaskan Native	0.18	0.18	0.12	0.14
Asian	0.69	0.47	1.15	2.36
Black/African American	2.65	2.39	2.96	5.70
Hispanic	5.99	6.08	5.55	7.42
Native Hawaiian/Pacific Islander	0.07	0.08	0.07	0.09
White	86.77	9.19	86.84	13.14
<b>Student Gender</b>				
Female	47.80	1.44	48.43	1.40
Male	52.21	1.44	51.57	1.40
<b>Student Identifications</b>				
Special Education	15.65	2.73	17.82	3.02
English Learner	0.93	0.97	1.29	2.25
Percent of Gifted Students	1.99	1.28	2.41	1.21
Homeless	1.25	0.74	1.40	1.00
Foster Care	0.97	1.49	0.77	0.58
Military Connected	1.85	3.50	0.71	1.34
<b>District Revenue</b>				
Total Revenue By Student	\$17,593.04	\$2,636.34	\$19,289.88	\$2,837.45
Total Federal Revenue By Student	\$659.79	\$155.65	\$732.14	\$252.58
Total Local Revenue By Student	\$8,919.18	\$2,701.94	\$9,628.64	\$3,549.52
Total State Revenue By Student	\$8,014.06	\$1,719.86	\$8,929.09	\$2,495.67
<b>District Expenditures</b>				
Total Expenditures By Student	\$14,240.65	\$2,590.15	\$15,582.89	\$2,284.00
Instructional Expenditures By Student	\$9,057.20	\$1,827.79	\$9,655.24	\$1,438.68
Student and Staff Support Expenditures By Student	\$1,222.14	\$290.18	\$1,305.58	\$345.70
Administration Expenditures By Student	\$1,452.01	\$276.53	\$1,736.96	\$378.16
Operations Expenditures By Student	\$2,509.30	\$609.81	\$2,885.11	\$588.92

*Highest Economic Disadvantage Tertile of Sampled School Districts vs.  
Highest Economic Disadvantage Tertile of All Pennsylvania School Districts*

School District Characteristics	Sampled Districts Data (30)		All PA Districts (499)	
	Mean	SD	Mean	SD
<b>Economically Disadvantaged</b>				
Percent Economically Disadvantaged	63.24	13.51	62.06	11.78
<b>District Characteristics</b>				
District Enrollment	1329.00	949.43	3392.04	9513.61
Geographic Size in Miles	165.90	293.88	93.07	132.93
Number of Schools	3.20	2.20	6.34	17.40
Number of Teachers	101.51	64.49	242.34	649.50
Student:Teacher Ratio	12.73	2.80	13.08	2.13
Charter School Enrollment	162.60	215.69	779.19	6140.32
Enrollment in Partnering Career and Technical Center	97.80	157.41	111.61	137.49
<b>Student Race / Ethnicity</b>				
2 or More Races	6.93	6.41	5.95	4.99
American Indian/Alaskan Native	0.22	0.28	0.18	0.30
Asian	0.29	0.31	1.19	2.53
Black/African American	15.44	25.27	13.88	19.98
Hispanic	5.33	5.32	10.65	16.09
Native Hawaiian/Pacific Islander	0.03	0.08	0.07	0.12
White	71.77	29.30	68.08	29.38
<b>Student Gender</b>				
Female	48.28	0.97	48.43	1.50
Male	51.72	0.97	51.57	1.50
<b>Student Identifications</b>				
Special Education	22.25	2.74	20.26	3.18
English Learner	1.43	1.72	3.07	5.17
Percent of Gifted Students	1.51	1.12	1.68	1.13
Homeless	2.42	1.69	2.16	1.68
Foster Care	1.69	1.31	1.11	0.87
Military Connected	0.80	1.08	0.37	0.86
<b>District Revenue</b>				
Total Revenue By Student	\$20,944.09	\$3,349.04	\$20,926.72	\$5,888.27
Total Federal Revenue By Student	\$1,278.09	\$427.64	\$1,296.85	\$839.31
Total Local Revenue By Student	\$7,616.98	\$4,462.98	\$8,394.80	\$4,023.01
Total State Revenue By Student	\$12,049.02	\$3,206.69	\$11,235.07	\$4,631.51
<b>District Expenditures</b>				
Total Expenditures By Student	\$16,374.50	\$2,290.26	\$16,038.32	\$3,077.30
Instructional Expenditures By Student	\$9,786.24	\$1,233.80	\$9,815.12	\$1,812.08
Student and Staff Support Expenditures By Student	\$1,347.54	\$536.91	\$1,338.76	\$611.33
Administration Expenditures By Student	\$1,941.11	\$510.17	\$1,821.25	\$556.30
Operations Expenditures By Student	\$3,299.61	\$615.21	\$3,063.20	\$857.64

## APPENDIX C: PARTICIPANTS' CONTEXTUAL INFORMATION

Participant Number	Tertile	Years in Current Position	Years in District-Level Position	Gender	Age	Race/Ethnicity	Highest Education Level Achieved	District Enrollment
1	Lowest ED Tertile	11-15	16-20	Female	50-59	White	Doctorate	3001-4000
2	Lowest ED Tertile	0-5	21-25	Male	50-59	White	Doctorate	3001-4000
3	Lowest ED Tertile	0-5	5.1-10	Male	50-59	White	Master's	7001-8000
4	Lowest ED Tertile	0-5	5.1-10	Male	50-59	White	Doctorate	4001-5000
5A	Lowest ED Tertile	6-10	5.1-10	Male	50-59	White	Doctorate	3001-4000
5B	Lowest ED Tertile	6-10	5.1-10	Male	50-59	White	Master's	3001-4000
6	Lowest ED Tertile	0-5	11-15	Male	40-49	White	Master's	2001-3000
7	Lowest ED Tertile	0-5	5.1-10	Female	40-49	White	Doctorate	3001-4000
8	Lowest ED Tertile	0-5	5.1-10	Female	50-59	White	Doctorate	3001-4000
9	Lowest ED Tertile	11-15	16-20	Male	50-59	White	Doctorate	1001-2000
10	Lowest ED Tertile	6-10	11-15	Male	50-59	White	Doctorate	7001-8000
11	Middle ED Tertile	6-10	5.1-10	Female	40-49	White	Doctorate	3001-4000
12	Middle ED Tertile	6-10	5.1-10	Male	50-59	White	Doctorate	1001-2000
13	Middle ED Tertile	0-5	5.1-10	Male	40-49	White	Doctorate	3001-4000
14A	Middle ED Tertile	0-5	0-5	Male	50-59	White	Master's	2001-3000
14B	Middle ED Tertile	0-5	0-5	Female	50-59	White	Doctorate	2001-3000
15	Middle ED Tertile	0-5	0-5	Male	50-59	White	Doctorate	1001-2000
16	Middle ED Tertile	0-5	5.1-10	Male	40-49	White	Doctorate	4001-5000
17	Middle ED Tertile	6-10	5.1-10	Male	40-49	White	Doctorate	1001-2000
18	Middle ED Tertile	0-5	11-15	Male	40-49	White	Doctorate	1001-2000
19	Middle ED Tertile	0-5	5.1-10	Female	50-59	White	Doctorate	4001-5000
20	Middle ED Tertile	0-5	11-15	Female	50-59	White	Master's	1001-2000
21	Highest ED Tertile	0-5	0-5	Male	50-59	White	Master's	0-1000
22	Highest ED Tertile	0-5	0-5	Male	40-49	White	Master's	0-1000
23	Highest ED Tertile	0-5	5.1-10	Male	50-59	White	Master's	1001-2000
24	Highest ED Tertile	0-5	0-5	Male	40-49	White	Master's	0-1000
25	Highest ED Tertile	11-15	11-15	Male	50-59	White	Doctorate	1001-2000
26	Highest ED Tertile	0-5	0-5	Male	40-49	White	Master's	3001-4000
27	Highest ED Tertile	6-10	16-20	Male	50-59	White	Master's	0-1000
28	Highest ED Tertile	6-10	11-15	Male	50-59	White	Doctorate	2001-3000
29	Highest ED Tertile	0-5	0-5	Female	50-59	African American	Doctorate	0-1000
30	Highest ED Tertile	0-5	21-25	Female	50-59	White	Master's	1001-2000

**APPENDIX D: PARTICIPANTS PERCEPTIONS OF VARIOUS ASPECTS OF  
MANDATED ACCOUNTABILITY AND THE FUTURE READY PA INDEX (BY  
TERTILE)**

		Lowest ED Tertile					Middle ED Tertile					Highest ED Tertile					All Sample				
		SA	A	N	D	SD	SA	A	N	D	SD	SA	A	N	D	SD	SA	A	N	D	SD
Understanding	You have a strong understanding of the accountability measures in the Future Ready Index.	2	8	0	0	0	6	4	0	0	0	6	4	0	0	0	14	16	0	0	0
	You value the mandated accountability measures in the Future Ready Index.	1	4	2	1	2	2	4	1	3	0	1	6	2	0	1	4	14	5	4	3
Value	Your district leadership prioritizes strong performance on standardized tests as a major district priority.	1	1	1	6	1	1	6	1	2	0	0	1	6	3	0	2	8	8	11	1
	Your district relies upon the accountability measures in the Future Ready Index as a major driver of strategic planning.	1	4	1	3	1	3	2	3	2	0	1	7	1	1	0	5	13	5	6	1
	You believe that the accountability measures in the Future Ready Index effectively capture student performance in schools.	0	4	2	3	1	0	8	0	1	1	0	7	0	2	1	0	19	2	6	3
	You believe that the accountability measures in the Future Ready Index provide reliable information/data on a school's overall effectiveness or performance.	0	4	2	2	2	0	6	0	3	1	0	4	3	2	1	0	14	5	7	4
Confidence	You believe the accountability measures in the Future Ready Index reflect the quality of instruction in schools (i.e. higher results mean that better instruction is taking place in particular schools).	1	1	2	5	1	1	3	4	0	2	0	1	4	3	2	2	5	10	8	5
	You believe the accountability measures in the Future Ready Index reflect the quality of leadership in schools (i.e. higher results mean that better leadership is taking place in particular schools).	0	1	3	6	0	1	5	4	0	0	0	2	3	4	1	1	8	10	10	1
	In general, you believe the accountability measures in the Future Ready Index have had a positive effect on school climate.	0	2	2	4	2	0	3	2	4	1	1	2	2	3	2	1	7	6	11	5
Outcomes	In general, you believe that your schools perform better in the areas that are publicly reported in the Future Ready Index.	1	5	1	2	1	0	7	1	2	0	1	1	2	4	2	2	13	4	8	3
Resources	You believe that schools have the resources/capacity to meet state benchmarks/expectations.	0	3	2	3	2	0	2	0	4	4	0	3	0	3	4	0	8	2	10	10

**APPENDIX E: MEANS DENOTING DISTRIBUTION OF PARTICIPANTS’  
PERCEPTIONS OF MANDATED ACCOUNTABILITY (BY TERTILE)**

		Lowest ED Tertile	Middle ED Tertile	Highest ED Tertile	All Sample
Understanding	You have a strong understanding of the accountability measures in the Future Ready Index.	4.20	4.60	4.60	4.47
	You value the mandated accountability measures in the Future Ready Index.	3.10	3.50	3.60	3.40
Value	Your district leadership prioritizes strong performance on standardized tests as a major district priority.	2.50	3.60	2.80	2.97
	Your district relies upon the accountability measures in the Future Ready Index as a major driver of strategic planning.	3.10	3.60	3.80	3.50
Confidence	You believe that the accountability measures in the Future Ready Index effectively capture student performance in schools.	2.90	3.50	3.30	3.23
	You believe that the accountability measures in the Future Ready Index provide reliable information/data on a school’s overall effectiveness or performance.	2.80	3.10	3.00	2.97
	You believe the accountability measures in the Future Ready Index reflect the quality of instruction in schools (i.e. higher results mean that better instruction is taking place in particular schools).	2.60	3.10	2.40	2.70
	You believe the accountability measures in the Future Ready Index reflect the quality of leadership in schools (i.e. higher results mean that better leadership is taking place in particular schools).	2.50	3.70	2.60	2.93
	In general, you believe the accountability measures in the Future Ready Index have had a positive effect on school climate.	2.40	2.70	2.70	2.60
Outcomes	In general, you believe that your schools perform better in the areas that are publicly reported in the Future Ready Index.	3.30	3.50	2.50	3.10
Resources	You believe that schools have the resources/capacity to meet state benchmarks/expectations.	2.60	2.00	2.20	2.27

**APPENDIX F: PARTICIPANTS' GENERAL PERCEPTIONS OF MANDATED  
ACCOUNTABILITY (BY TERTILE)**

Theme	Economic Disadvantage			Total
	Lowest ED Tertile	Middle ED Tertile	Highest ED Tertile	
<b>Positive Perceptions</b>				
Important for consistency and equity across districts	4	6	2	12
Improved and more comprehensive than previous versions	2	3	4	9
Provides direction and motivation for improvement	1	1	3	5
<b>Negative Perceptions</b>				
Measures are too narrow / Incomplete representation of a school district	4	4	5	13
Data most closely relate to socioeconomic factors	2	1	4	7
Public misunderstanding / Public draws incorrect conclusions about school	0	1	3	4
Exercise in compliance	2	2	0	4
Dislikes or questions specific measures or calculations	2	0	0	2
System unsuccessful in engaging the public	1	0	1	2

**APPENDIX G: PARTICIPANTS' PERCEPTIONS OF WAYS MANDATED  
ACCOUNTABILITY ADVANCES OR HINDERS THEIR EDUCATIONAL  
OBJECTIVES (BY TERTILE)**

Theme	Economic Disadvantage			Total
	Lowest ED Tertile	Middle ED Tertile	Highest ED Tertile	
<b>Advances</b>				
Accountability engenders analysis, goal-setting, and improved practices	6	7	6	19
Allows valuable comparisons across schools and districts	1	4	0	5
Affirms, motivates, and encourages the work of schools	0	3	1	4
Provides consistency in measuring / evaluating schools	1	0	2	3
Provides a simple and clear form of communication on school performance	0	1	1	2
Increases public awareness of school performance	1	0	0	1
<b>Hinders</b>				
System is misunderstood / Oversimplifies work of schools / Poor public understanding	4	5	4	13
Misleading or erroneous data and calculations	2	1	1	4
Engenders harmful comparisons across schools and districts	2	0	2	4
Negatively impacts teachers / Creates tensions with teachers	0	2	2	4
Data most closely relate to socioeconomic factors	0	2	0	2
Reflects poorly conceived / implemented law	1	1	0	2
Encourages teaching to the test	2	0	0	2
Lack of professional understanding (teachers, administration) of the data	0	2	0	2
Challenging to improve the data	0	2	0	2
Measures are too narrow / Incomplete representation of a school	1	0	0	1
Too much data	0	1	0	1
Engenders state oversight and pressure	0	0	1	1
Community is disengaged or apathetic	1	0	0	1
Insufficient time to analyze data	1	0	0	1
Does not provide contextual or causal information about outcomes	0	1	0	1

**APPENDIX H: NUMBER OF PARTICIPANTS WHO VALUE VARIOUS MANDATED AND UNMANDATED (ALTERNATE OR ADDITIONAL) ACCOUNTABILITY MEASURES (BY TERTILE)**

		Lowest ED Tertile			Middle ED Tertile			Highest ED Tertile			All Sample		
		HV	SV	NV	HV	SV	NV	HV	SV	NV	HV	SV	NV
Measures in the Current System	Achievement (Standardized Tests)	3	5	2	6	3	1	1	6	3	10	14	6
	Growth (Standardized Tests)	3	6	1	6	3	1	7	2	1	16	11	3
	Graduation Rates	2	8	0	7	3	0	7	3	0	16	14	0
	English Learner Growth	1	7	2	2	4	3	2	7	1	5	18	6
	Career Completion Activities	2	7	1	3	7	0	4	5	1	9	19	2
	Industry-Based Measures	3	5	2	7	3	0	7	2	1	17	10	3
	Advanced Class Enrollment	3	6	1	4	4	2	0	7	3	7	17	6
	Post-Secondary Transition Data	4	4	2	7	0	3	7	3	0	18	7	5
Measures Not in the Current System	Aggregated Local Grades	1	6	3	1	1	8	0	5	5	2	12	16
	Higher Level Thinking	5	5	0	5	2	3	8	1	1	18	8	4
	Social Emotional Skill Development	6	4	0	6	3	1	7	2	1	19	9	2
	Civic Engagement	5	4	1	5	5	0	7	2	1	17	11	2
	Perseverance, Determination, Growth Mindset	7	2	1	8	1	1	9	0	1	24	3	3
	Financial Skills	5	5	0	6	4	0	4	4	2	15	13	2
	Climate and Culture	5	4	1	6	4	0	8	1	1	19	9	2
	Curricular Offerings	7	2	1	6	4	0	6	4	0	19	10	1
	Extra Curricular Offerings	5	3	2	5	5	0	6	3	1	16	11	3
	Participation in Extra Curricular Offerings	5	4	1	7	3	0	7	2	1	19	9	2
	Participation in Internships / Work-Based Learning	7	3	0	8	2	0	8	1	1	23	6	1
	Family Engagement Opportunities	3	6	1	7	1	2	6	2	2	16	9	5
	School-Community Partnerships	4	5	1	7	2	1	7	2	1	18	9	3
	Safety Measures	5	5	0	10	0	0	9	1	0	24	6	0

Note 1: Total for English Learner Growth, Middle ED Tertile does not equal 10 as one participant indicated "NA."

**APPENDIX I: MEANS DENOTING DISTRIBUTION OF NUMBER OF PARTICIPANTS WHO VALUE VARIOUS MANDATED AND UNMANDATED (ALTERNATE OR ADDITIONAL) ACCOUNTABILITY MEASURES (BY TERTILE)**

		Lowest ED	Middle ED	Highest ED	All Sample
Measures in the Current System	Achievement (Standardized Tests)	2.10	2.50	1.80	2.13
	Growth (Standardized Tests)	2.20	2.50	2.60	2.43
	Graduation Rates	2.20	2.70	2.70	2.53
	English Learner Growth	1.90	1.89	2.10	1.96
	Career Completion Activities	2.10	2.30	2.30	2.23
	Industry-Based Measures	2.10	2.70	2.60	2.47
	Advanced Class Enrollment	2.20	2.20	1.70	2.03
	Post-Secondary Transition Data	2.20	2.40	2.70	2.43
Measures Not in the Current System	Aggregated Local Grades	1.80	1.30	1.50	1.53
	Higher Level Thinking	2.50	2.20	2.70	2.47
	Social Emotional Skill Development	2.60	2.50	2.60	2.57
	Civic Engagement	2.40	2.50	2.60	2.50
	Perseverance, Determination, Growth Mindset	2.60	2.70	2.80	2.70
	Financial Skills	2.50	2.60	2.20	2.43
	Climate and Culture	2.40	2.60	2.70	2.57
	Curricular Offerings	2.60	2.60	2.60	2.60
	Extra Curricular Offerings	2.30	2.50	2.50	2.43
	Participation in Extra Curricular Offerings	2.40	2.70	2.60	2.57
	Participation in Internships / Work-Based Learning	2.70	2.80	2.70	2.73
	Family Engagement Opportunities	2.20	2.50	2.40	2.37
	School-Community Partnerships	2.30	2.60	2.60	2.50
Safety Measures	2.50	3.00	2.90	2.80	

**APPENDIX J: DISTRICT LEADERS' IDENTIFICATION OF OTHER  
VALUABLE ACCOUNTABILITY MEASURES (BY TERTILE)**

Indicators	Economic Disadvantage			Total
	Lowest	Middle	Highest	
	ED Tertile	ED Tertile	ED Tertile	
Climate and Culture Indicators	6	1	0	7
Post-Secondary Success Indicators	3	2	1	6
Socio-Economic Factors / District Context	2	2	1	5
Financial Indicators	2	2	0	4
Student Opportunities / Well-Rounded Experiences Indicators	1	2	1	4
Local or Other Assessments	1	2	1	4
Student Discipline Indicators	2	1	0	3
Student Dispositions Indicators	0	2	1	3
Safety and Crime Indicators	1	0	2	3
Drug and Alcohol Use Indicators	1	1	0	2
Broader Narrative of District	1	0	1	2
Access to Health Services Indicators	0	2	0	2
Teacher Satisfaction Indicators	0	1	1	2
Social-Emotional Indicators	1	0	0	1
Academic Achievement by Gender	1	0	0	1
Perception of the Quality of School Indicators	1	0	0	1
Adherence to Professional Standards Indicators	0	1	0	1
Disciplinary Equity Indicators	0	0	1	1
Teacher Rating Indicators	1	0	0	1
Comparisons to National or International Measures	0	1	0	1

**APPENDIX K: MEANS DENOTING DISTRIBUTION OF PARTICIPANTS'  
RESPONSES ON QUESTIONS RELATED TO COMMUNICATION OF  
MANDATED ACCOUNTABILITY (BY TERTILE)**

*Means Denoting Distribution of Frequency of Leaders' Communication with Various Stakeholders*

	Communication To...				
	Board	District Administration	Principals	Teachers	Parents
Lowest ED Tertile	3.20	4.60	4.80	4.30	3.10
Middle ED Tertile	3.70	5.20	5.30	4.40	3.10
Highest ED Tertile	3.30	4.60	4.60	4.70	3.30
All Sample	3.40	4.80	4.90	4.47	3.17

*Means Denoting Distribution of Frequency of Questions from Stakeholders*

	Questions From...	
	Parents/Families	Larger Community
Lowest ED Tertile	1.50	1.40
Middle ED Tertile	1.60	1.30
Highest ED Tertile	1.20	1.80
All Sample	1.43	1.50