

2016

Toward an Advanced System of Accountability for Improved Student Learning Outcomes: A Mixed-Methods Analysis of Test-Based Accountability in Louisiana

Susan Nicole Kahn

Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations



Part of the [Education Commons](#)

Recommended Citation

Kahn, Susan Nicole, "Toward an Advanced System of Accountability for Improved Student Learning Outcomes: A Mixed-Methods Analysis of Test-Based Accountability in Louisiana" (2016). *LSU Doctoral Dissertations*. 132.

https://digitalcommons.lsu.edu/gradschool_dissertations/132

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.

TOWARD AN ADVANCED SYSTEM OF ACCOUNTABILITY FOR IMPROVED
STUDENT LEARNING OUTCOMES: A MIXED-METHODS ANALYSIS OF TEST-
BASED ACCOUNTABILITY IN LOUISIANA

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The School of Education

by

Susan N. Kahn

B.A., University of Louisiana, 2006

M.Ed., University of Louisiana, 2011

Ed.S., Louisiana State University, 2014

August 2016

@ 2016 copyright
Susan Fontenot Kahn
All rights reserved

To Eli and Otto,
for coloring my world and this work, at times literally, with love, joy, and laughter.

It is my hope that the future you will inherit will be one in which excellence in student learning outcomes will be sought after by a generation using the most advanced means to ensure a healthy and happy populace for all.

ACKNOWLEDGEMENTS

This project would never have begun had it not been for the students at Ossun Elementary, L. Leo Judice Elementary, and J.W. Faulk Elementary who taught me about teaching—who I wanted to be as a teacher and who I wanted to become as an educational leader to impact society. You inspired me then with your magic and you inspire me today as I see you working and learning throughout our community. The impact you have made on my life is poured into the lives of future teachers as I refer to many of our moments when I teach pre-service teacher candidates. I am amazed at your tenacity and dedication as I run into you as you working the register at a local grocery store (A.G.) or local restaurant (A.G.), when you invite me to your soccer (R.N.) and basketball games (D.M.), when you list me as your favorite teacher in your high school honors ceremony guide (B.G.), or when you send me a keepsake postcard that I mailed to you (J.M.S.). I am eternally grateful for the time we had together, as teacher and student, and hope all the best for all my former K-12 students and your futures.

I have been fortunate to have been engaged in tremendously important and I daresay, fun, work within the K-12 setting. My classroom experiences with Amy Bartlett, Danielle Babineaux, and Charmaine Jones laid the groundwork for this study. Serving the public in the administration of large-scale assessment systems during a time of change has been the professional highlight of my career, and was made robust and richly satisfying because of my mentors Thomas “Tom” Spencer and Dr. William “Bill” Morrison. Your candor and support have taught me about leadership in practice, and more so about the value of friendship within professional relationships.

I am grateful for the opportunity to have engaged with the incredible scholars in the College of Education under the leadership of Dr. Nathan Roberts and Dr. Peter Sheppard. The opportunity to instruct pre-service teacher candidates expanded my teaching experiences and personal perspective on teacher preparation in the United States, for which I am ever grateful.

Dr. Dana Bickmore and Dr. Dianne Taylor—this journey began with you. I recall sitting in each of your offices interviewing for the program and many times thereafter discussing the complexities and challenges facing K-12 education. Your experiences and insight have influenced my thinking and growth as a scholar of educational leadership and research. I am thankful to Dr. Kenneth Fasching-Varner for his example of scholarship and collegiality. I could not have asked for a better compilation of minds on my committee to sharpen this work and hone my understanding of its implications Dr. Fasching-Varner, Dr. Lori Martin and Dr. Roland Mitchell—as a committee, your perspective and support have made this work possible.

I am grateful for the amazing staff at Acadian Baptist Center, for providing a quiet place of respite and the space to ponder, reflect, and grapple with the details of data analysis. James Newsom, Peggy Watson, Rod King, and Guy Atkinson—your lives ministered to my own during the final stages of this project. Thank you for your example and life of faith.

I am grateful to the family of faith at Christ Church and to our pastors who lead not to change policy, but to change culture through the preaching and teaching of the Holy Bible and the renewing of lives through counsel in wisdom and truth. Kirk

Alexander, Stuart Amidon, Joe Garner, and Brandon Nealy—your lives have touched this work, and I thank you for your leadership.

Finally, to Elliott. You attracted me first by your wisdom and insight, and you attract me still by the power of love. I am better because of you, and ever so grateful for the opportunity to know you, to love you, and to be loved by you.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	ix
LIST OF FIGURES.....	xiii
ABSTRACT.....	xiv
CHAPTER 1. INTRODUCTION.....	1
Test-Based Accountability in Louisiana.....	1
Louisiana Context.....	3
Statement of the Problem.....	13
Purpose Statement.....	13
Research Questions.....	14
Theoretical Framework.....	14
Taxonomy for Learning.....	14
Educational Assessment.....	18
Knowledge-Based Economy.....	19
Goal(s) of Formal Education	22
Significance of the Study.....	23
Every Student Succeeds Act.....	23
School Governance Models.....	24
Policy Window.....	24
Summary.....	25
CHAPTER 2. REVIEW OF THE LITERATURE.....	28
Historical Perspective on the Emergence of Test-Based Accountability in America: 1965-2015.....	28
Elementary and Secondary Education Act: Closing the Achievement Gap.....	30
Improving America’s Schools Act: Standards.....	42
No Child Left Behind Act: Assessments, High Stakes, Sanctions, and Subgroups.....	44
Implications of NCLB for School Improvement.....	68
Test-Based Accountability: Cognitive Domain for Learning.....	69
History of Educational Testing in the United States.....	69
Design.....	72
Validity.....	75
Unintended Consequences.....	75
Non-Cognitive Domains for Learning	76
Affective.....	76
Psychomotor.....	81
Innovation in State Assessment Systems.....	82
Authentic Assessments for Deeper Learning.....	83

Performance-Based Assessments.....	84
Need for Advanced System of Test-Based Accountability.....	84
Summary.....	86
CHAPTER 3. METHODOLOGY.....	89
Philosophical Underpinnings of Educational Research.....	89
Quantitative.....	90
Qualitative.....	91
Mixed Methods.....	99
Study Design: Mixed Methods.....	101
Phase I: Quantitative.....	103
Phase II: Qualitative.....	107
Ethical Considerations.....	108
Anticipated Implications for Future Research.....	110
Axiology.....	111
CHAPTER 4. DATA COLLECTION AND ANALYSIS.....	114
Phase I: Quantitative.....	114
Phase II: Qualitative.....	160
Limitations of this Study.....	168
Future Research.....	170
CHAPTER 5. RESULTS AND CONCLUSIONS	172
REFERENCES.....	174
APPENDIX A- SURVEY INSTRUMENT.....	203
APPENDIX B- INTERVIEW PROTOCOL.....	208
APPENDIX C- REQUEST FOR SURVEY.....	209
APPENDIX D- REQUEST FOR INTERVIEW.....	210
APPENDIX E- CONSENT FORMS FOR INTERVIEWS.....	211
APPENDIX F- INSITUTIONAL REVIEW BOARD APPROVAL.....	212
VITA.....	213

LIST OF TABLES

1.1	Mandated Annual Standardized Tests, 2015.....	6
1.2	Summary of NAEP Results for Louisiana.....	8
1.3	Cognitive Domain for Learning.....	15
1.4	Cognitive Process.....	16
2.1	21 st Century Skills.....	79
4.1	Survey Question 1.....	115
4.2	Survey Question 2.....	116
4.3	Survey Question 3: Descriptive Statistics.....	119
4.4	Survey Question 3: Descriptive Statistics.....	119
4.5	Survey Question 3: Test of Homogeneity of Variance.....	120
4.6	Survey Question 3: Oneway ANOVA.....	121
4.7	Survey Question 3: Robust Tests of Equality of Means.....	121
4.8	Survey Question 3: Cross Tabulations.....	121
4.9	Survey Question 3: Chi Square Tests.....	122
4.10	Survey Question 4: Descriptive Statistics.....	123
4.11	Survey Question 4: Descriptive Statistics.....	123
4.12	Survey Question 4: Test of Homogeneity of Variance.....	124
4.13	Survey Question 4: Oneway ANOVA.....	125
4.14	Survey Question 4: Cross Tabulations.....	125
4.15	Survey Question 4: Chi Square Tests.....	125
4.16	Survey Question 5: Descriptive Statistics.....	126
4.17	Survey Question 5: Descriptive Statistics.....	127

4.18	Survey Question 5: Test of Homogeneity of Variance.....	128
4.19	Survey Question 5: Oneway ANOVA.....	128
4.20	Survey Question 5: Cross Tabulations.....	128
4.21	Survey Question 5: Chi Square Tests.....	129
4.22	Survey Question 6: Descriptive Statistics.....	130
4.23	Survey Question 6: Descriptive Statistics.....	130
4.24	Survey Question 6: Test of Homogeneity of Variance.....	131
4.25	Survey Question 6: Oneway ANOVA.....	132
4.26	Survey Question 6: Robust Tests of Equality of Means.....	132
4.27	Survey Question 6: Cross Tabulations.....	132
4.28	Survey Question 6: Chi Square Tests.....	133
4.29	Survey Question 7: Descriptive Statistics.....	134
4.30	Survey Question 7: Test of Homogeneity of Variance.....	134
4.31	Survey Question 7: Oneway ANOVA.....	135
4.32	Survey Question 7: Cross Tabulations.....	135
4.33	Survey Question 7: Chi Square Tests.....	135
4.34	Survey Question 8: Descriptive Statistics.....	136
4.35	Survey Question 8: Test of Homogeneity of Variance.....	137
4.36	Survey Question 8: Oneway ANOVA.....	137
4.37	Survey Question 8: Cross Tabulations.....	138
4.38	Survey Question 8: Chi Square Tests.....	138
4.39	Survey Question 9: Descriptive Statistics.....	139
4.40	Survey Question 9: Test of Homogeneity of Variance.....	140

4.41	Survey Question 9: Oneway ANOVA.....	140
4.42	Survey Question 9: Cross Tabulations.....	141
4.43	Survey Question 9: Chi Square Tests.....	141
4.44	Survey Question 10: Descriptive Statistics.....	142
4.45	Survey Question 10: Test of Homogeneity of Variance.....	143
4.46	Survey Question 10: Oneway ANOVA.....	144
4.47	Survey Question 10: Cross Tabulations.....	144
4.48	Survey Question 10: Chi Square Tests.....	144
4.49	Survey Question 11: Descriptive Statistics.....	145
4.50	Survey Question 11: Test of Homogeneity of Variance.....	146
4.51	Survey Question 11: Oneway ANOVA.....	146
4.52	Survey Question 11: Cross Tabulations.....	147
4.53	Survey Question 11: Chi Square Tests.....	147
4.54	Survey Question 12: Descriptive Statistics.....	148
4.55	Survey Question 14: Descriptive Statistics.....	150
4.56	Survey Question 16: Word Frequencies.....	153
4.57	Survey Question 16: Themes.....	154
4.58	Survey Question 17: Word Frequencies.....	156
4.59	Survey Question 17: Themes.....	157
4.60	Survey Question 18: Word Frequencies.....	158
4.61	Survey Question 19: Word Frequencies.....	159
4.62	Code Frequencies.....	165
4.63	Co-Occurring Code Frequencies.....	166

4.64	Co-Occurring Code Power.....	168
------	------------------------------	-----

LIST OF FIGURES

1.1	Louisiana's Test-Based Accountability System: SPS Scales 1998-2015.....	4
1.2	NAEP At/Above Basic and Proficient: Reading, grade 4.....	9
1.3	NAEP At/Above Basic and Proficient: Mathematics, grade 4.....	9
1.4	NAEP At/Above Basic and Proficient: Reading, grade 8.....	10
1.5	NAEP At/Above Basic and Proficient: Mathematics, grade 8.....	10
1.6	OECD countries publically posting achievement data for accountability.....	12
1.7	OECD countries comparing schools for accountability.....	12
1.8	Trends in Routine and Nonroutine Task Input, 1960-1998.....	20
1.9	Fortune 500 Most Valued Skills.....	21
2.1	Louisiana Achievement in English language arts and Mathematics.....	51
4.1	Histogram of Survey Question 3.....	120
4.2	Histogram of Survey Question 4.....	124
4.3	Histogram of Survey Question 5.....	127
4.4	Histogram of Survey Question 6.....	131
4.5	Bar Graph of Survey Question 7.....	134
4.6	Bar Graph of Survey Question 8.....	137
4.7	Bar Graph of Survey Question 9.....	140
4.8	Bar Graph of Survey Question 10.....	143
4.9	Bar Graph of Survey Question 11.....	146
4.10	Longitudinal Plot of Survey Question 12.....	148
4.11	Longitudinal Plot of Survey Question 14.....	151

ABSTRACT

The nature of federal and state policies regarding accountability testing narrows the taught curriculum to content tested, thereby changing what is officially valued as student learning. The over-emphasis on standardized test scores has narrowed the curriculum to content that is tested (Amrein & Berliner, 2002; Baker, 2008; Herman, 2008; Koretz & Hamilton, 2003; Linn, 2000). In Louisiana, test-based accountability defines local school quality through a letter grade rating scale that uses an index for school performance scores (SPS), which in turn impacts traditional and nontraditional (e.g. charter) school expansion, closure, and takeover, administration of opportunity scholarships (e.g. voucher) for private school enrollment, and parent petition actions (Louisiana State Board for Elementary and Secondary Education [SBESE], December 2015, §301). Louisiana's current test-based accountability system is not designed to provide information on student achievement beyond the cognitive domain of learning for accountability purposes, though skills in the affective and psychomotor are integral to life success (Rothstein, 2004).

This study examined the test-based accountability system in Louisiana and the extent to which the system provides meaningful and actionable data for stakeholders. The intent of this two-phase, concurrent, mixed-methods study was to discover the educational values most prioritized by Louisiana stakeholders and the extent to which current policymakers were willing to pilot an advanced system for test-based accountability, including tests of the higher levels of the cognitive domain as well as indicators for learning in the affective and psychomotor domains. In the first phase, quantitative research questions addressed the comparison of perceptions about

Louisiana's test-based accountability system and educational values of stakeholders through the administration of a digitally based statewide survey. Over 500 survey responses were collected and interview data from two participating lawmakers were collected. Stakeholders indicate positive interest in holding schools accountable for values other than standardized tests scores and lawmakers indicated interest in piloting an advanced system, while also indicating the political complications for advancing the system in such a way.

CHAPTER 1. INTRODUCTION

What we don't know is: Do these improvements on high-stakes tests represent real learning gains? And do they make students better off in the long run? In fact, we know very little about the impact of test-based accountability on students' later success (Deming, Cohodes, Jennings, & Jencks, 2016, p. 71).

Test-Based Accountability in Louisiana

Researchers have long recognized problems associated with an over-emphasis on standardized tests of student achievement. Over 40 years ago, Ralph Tyler, renowned curriculum specialist and chair of the committee that developed the National Assessment of Educational Progress [NAEP] (Carr, 2004; Shepard, 2008) wrote that “standard achievement tests in common use do not give a dependable measure of what children have learned. They are not constructed to do so” (Tyler, 1972, pp. 5-6). Subsequently, in 1987, the National Association of Educators (NAE) identified a phenomenon later referred to as “goal distortion” (Rothstein, Jacobsen, & Wilder, 2008, p. 45), explaining that when schools focus exclusively on standardized achievement measures “a subtle shift occurs in which fallible and partial indicators of academic achievement are transformed into major goals of schooling” (p. 24). The value placed on educational assessments in K-12 education can be seen by applying variations of the business principle “what gets measured gets managed” (Willcocks & Lester, 1996, p. 466) on the K-12 educational landscape: “what gets measured gets mastered” (Louisiana Department of Education, 2016), “what is measured is treasured” (Pederson, 2007, p. 291), and “what gets tested gets taught; who gets tested gets taught” (Burgess & Kennedy, 1998, p. 1). However, Kornhaber (2004) noted that “the research that has investigated how well students' gains on tests generalize to other tests of the same content does not lend much support to the idea that test-intensive reforms are promoting learning” (p. 57). Moreover, Lindle (2009) reported that, “other than

reports on implementation status or unintended consequences, to date no definitive claims link assessment policy to improved student outcomes” (p. 327). This laser-sharp focus on mandated standardized assessments, to the exclusion of non-tested subjects such as the sciences, histories, and creative arts, and the practice of removing students from non-tested content for remediation in tested content (Herman & Baker, 2009), is at odds with assuring that all students receive a high quality education (Kornhaber, 2004).

Whereas, a budget represents the priorities of an organization for a business (Heifetz & Linksky, 2002), so our accountability systems represent the priorities for K-12 education. The NRC reported on the application of high-stakes, or “accountability systems [that] link rewards and punishments to demonstrated student performance in an effort to transform the quality of schooling” (Hess, 2002, p. 70), to test results as invalid due to the inappropriate interpretation and/or use of test results:

Policy and public expectations of testing generally exceed the technological capacity of the tests themselves. One of the most common reasons for this gap is that policymakers, under constituent pressure to improve schools, often decide to use existing tests for purposes for which they were neither intended nor sufficiently validated. So, for example, tests designed to produce valid measures of performance only at the aggregate level—for schools and classrooms—are used to report on and make decisions about individual students. In such instances, serious consequences (such as retention in grade) may be unfairly imposed on individual students” (Heubert & Hauser, 1999, p. 30)

The nature of federal and state policies regarding accountability testing narrows the taught curriculum to content tested, thereby changing what is officially valued as student learning. Polikoff (2012) noted that state accountability tests serve as the motivating factor for teachers to teach the state standards and “with appropriate supports and accountability measures, the theory [of change] proposes that teachers will align their instruction with the standards and assessments, and student learning will improve” (p. 341). There is vast evidence of teachers

focusing their instruction on content for a high-stakes tests compared with other content (Linn, 2003). Kornhaber (2004) noted that learning is restricted when teachers “...teach to the demands of the test rather than the demands of the academic disciplines” (p. 59). Indeed, where some states have recognized this practice of reducing the curriculum to that which is measured on standardized achievement, some states have increased the testing programs to ensure the teaching of additional content beyond the NCLB requirement of reading, mathematics, and science (Linn, 2003).

Moreover, the over-emphasis on standardized test scores has narrowed the curriculum to content that is tested (Amrein & Berliner, 2002; Baker, 2008; Herman, 2008; Koretz & Hamilton, 2003; Linn, 2000). Herman (2004) found that at-risk students are more likely to receive restricted curriculum as additional time is spent on efforts to remediate academic deficiencies. Further, there is substantial evidence to support the claim of misclassification of some low-performing students as students with disabilities in an effort to exclude their performance scores from accountability calculations (Cullen & Reback, 2006; Figlio & Getzler, 2007)

Louisiana Context

In Louisiana, test-based accountability defines local school quality through a letter grade rating scale that uses an index for school performance scores (SPS), which in turn impacts traditional and nontraditional (e.g. charter) school expansion, closure, and takeover, administration of opportunity scholarships (e.g. voucher) for private school enrollment, and parent petition actions (Louisiana State Board for Elementary and Secondary Education [SBESE], December 2015, §301). In Louisiana, the SPS rating system uses letter grades A, B, C, D, or F, the scales of which have been adjusted over time (see Fig. 1.1).

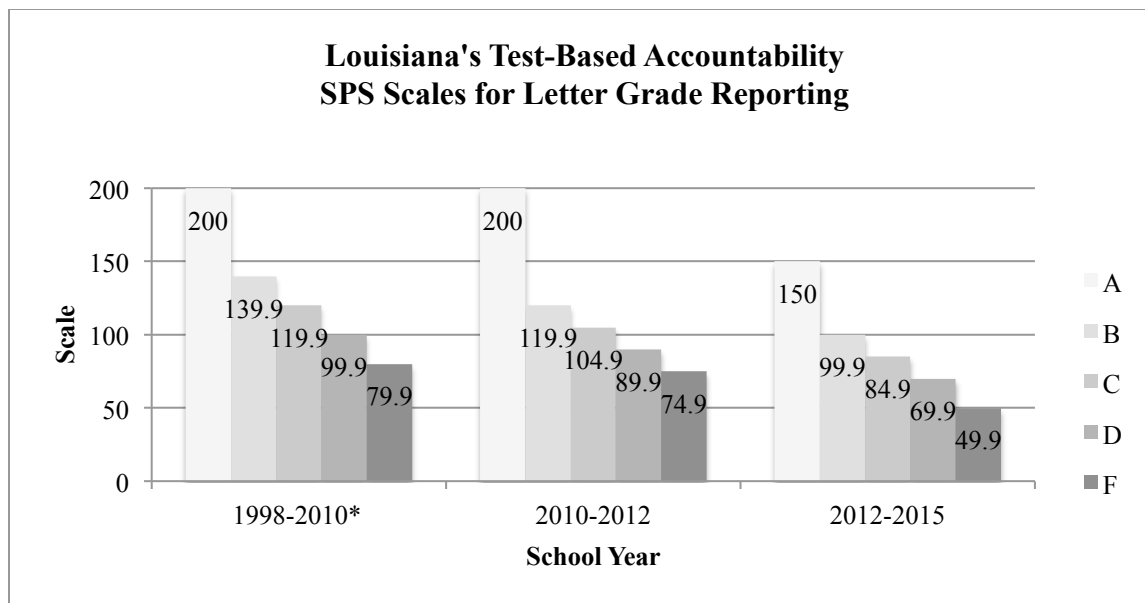


Figure 1.1 Louisiana's Test-Based Accountability System: School Performance Score Scales 1998-2015 (SBESE, 2015).

*Note: From 1998-2010, Louisiana's school performance labels utilized a 1-5 star rating, with 5-star being the highest attainable rating.

According to Louisiana law, the state education agency (SEA) applies rewards and sanctions to schools based on the rating calculated for the school (SBESE, December 2015, §1101). This system, in turn, determines district and school performance scores for local education agencies (LEA) (i.e. traditional school districts) and charter management organizations (CMO) (SBESE, December 2015, §4301). More importantly, sanctions for low performance ratings include the requirement for districts to provide school choice options for students attending 'failing' schools (SBESE, December 2015, §2501), as well as the threat of takeover by the state for consistently low performing schools (SBESE, December 2015, §2401).

While other top-performing countries on international standardized achievement tests require little to no standardized achievement test in K-12 education (Darling-Hammond, Wilhoit, & Pittinger, 2014; Tucker, 2014), the US outranks them all by mandating seventeen tests across a student's K-12 experience (NCLB, P.L. 107-110, §1116). In 2015, the federal mandate for

standardized achievement tests totaled seventeen tests for students to take across the K-12 experience in public schools. In that same year, Louisiana mandated 62¹ tests in pre-kindergarten through 12th grade (see Table 1.1). In addition to the required state tests, Louisiana incentivizes additional standardized tests by rewarding schools in the state's accountability formula for student achievement on academic measures including Advanced Placement (AP), International Baccalaureate (IB), College Level Equivalency Program (CLEP), and the ACT series WorkKeys (SBESE, December 2015, §515). Furthermore, local education agencies (LEAs) often required district-level assessment practices for benchmarking and monitoring progress, along with school-based required tests and teacher-made classroom tests for daily learning.

¹ Required tests for students in Louisiana PK-12 public schools during the 2014-2015 school year.

Table 1.1

MANDATED ANNUAL STANDARDIZED TESTS, 2015

Grade Level	No Child Left Behind	Louisiana
PK		Readiness (1)
K		Readiness (1)
		ELA (3)
1		ELA (3)
2		ELA (3)
3	ELA (1)	ELA (4)
	Math (1)	Math (1)
	Science (1), once in grades 3-5	Science (1)
		Social Studies (1)
4	ELA (1)	ELA (2)
	Math (1)	Math (2)
		Science (2)
		Social Studies (1)
5	ELA (1)	ELA (1)
	Math (1)	Math (1)
		Science (1)
		Social Studies (1)
6	ELA (1)	ELA (1)
	Math (1)	Math (1)
	Science (1), once in grades 6-8	Science (1)
		Social Studies (1)
7	ELA (1)	ELA (1)
	Math (1)	Math (1)
		Science (1)
		Social Studies (1)
8	ELA (1)	ELA (3)
	Math (1)	Math (3)
		Science (3)
		Social Studies (1)
9		ELA (1)
		Math (2)
		Science (2)
10		ELA (2)
		Math (2)
		Science (1)
11	ELA (1)	ELA (2)
	Math (1)	Math (1)
	Science (1), once in grades 10-12	Science (1)
		Social Studies (1)
12		
TOTAL	17	62

Note. Required tests in Louisiana reflect the 2014-2015 school year. The term “ELA” as used above represents facets of English language proficiency including reading, writing, and language arts. High-stakes for promotion are applied at grades 4 and 8, and high stakes for graduation are applied at grades 10, 11, and 12. Adapted from Louisiana State Board of Elementary and Secondary Education, 2015b, Bulletin 118: Statewide assessment standards and practices.

Note. Not including alternate assessments (LAA1, LAA2, ELDA) or additional assessments optionally used for accountability such as AP, IB, CLEP, and Workkeys.

Despite the numerous mandated tests, Louisiana students' performance on national measures of standardized achievement tests remains stagnant (see Table 1.2). A trademark of the NCLB legislation is the requirement to aggregate data by subgroup², which illuminates achievement gaps over time. This emphasis on the expected achievement for all students (Linn, 2003) clearly articulated the aim and goal of this law—that no child would be left behind. In 2009, Lindle noted that the “analyses of accountability policy often point to the general intent that accountability assures equity, especially for historically underserved and low-performing students” (p. 327) (Delandshere, 2001; Fulton, 2007; McDermott, 2007). Despite the effort to improve student learning through test-based accountability (Jones, Jones, & Hargrove, 2003), test scores on national measures of achievement have been slow to increase. Since the late 1990s, Louisiana has trailed the national average at both the basic and proficient achievement levels in reading and mathematics on NAEP, a nationally representative assessment of student achievement at grades 4, 8, and 12 (see Figures 1.2, 1.3, 1.4, and 1.5).

² Mandatory reporting by subgroup includes gender, education classification, socio-economic status, and race/ethnicity (include NCLB citation).

Table 1.2

SUMMARY OF NAEP RESULTS FOR LOUISIANA

Assessment			Average Scale Score				Achievement Level					
Subject	Grade	Year	State		National public		at or above Basic		at or above Proficient		at Advanced	
			Avg.	SE	Avg.	SE	Pct.	SE	Pct.	SE	Pct.	SE
Mathematics	4	2015	234	(1.1)	240	(0.3)	78	(1.3)	30	(1.9)	4	(0.6)
		2013	231	(1.2)	241	(0.2)	75	(1.4)	26	(1.8)	3	(0.6)
		2011	231	(1.0)	240	(0.2)	73	(1.4)	26	(1.4)	2	(0.3)
		2009	229	(1.0)	239	(0.2)	72	(1.3)	23	(1.4)	2	(0.4)
		2007	230	(1.0)	239	(0.2)	73	(1.4)	24	(1.3)	2	(0.3)
		2005	230	(0.9)	237	(0.2)	74	(1.3)	24	(1.3)	2	(0.4)
		2003	226	(1.0)	234	(0.2)	67	(1.8)	21	(1.2)	2	(0.4)
		2000	218	(1.4)	224	(1.0)	57	(2.0)	14	(1.3)	1	(0.2)
		20001	218	(1.4)	226	(1.0)	57	(2.0)	14	(1.4)	1	(0.2)
		19961	209	(1.1)	222	(1.0)	44	(1.8)	8	(0.9)	#	(†)
	8	2015	268	(1.4)	281	(0.3)	57	(1.8)	18	(1.5)	3	(0.5)
		2013	273	(0.9)	284	(0.2)	64	(1.2)	21	(1.0)	3	(0.4)
		2011	273	(1.2)	283	(0.2)	63	(1.4)	22	(1.2)	3	(0.4)
		2009	272	(1.6)	282	(0.3)	62	(1.6)	20	(1.8)	4*	(1.1)
		2007	272	(1.1)	280	(0.3)	64	(1.8)	19	(1.2)	2	(0.4)
		2005	268	(1.4)	278	(0.2)	59	(2.1)	16	(1.4)	2	(0.4)
		2003	266	(1.5)	276	(0.3)	57	(1.8)	17	(1.3)	2	(0.5)
		2000	259	(1.5)	272	(0.9)	47	(1.9)	11	(1.1)	1	(0.3)
		20001	259	(1.5)	274	(0.8)	48	(1.8)	12	(1.2)	1	(0.4)
		19961	252	(1.6)	271	(1.2)	38	(2.0)	7	(1.1)	#	(†)
Reading	4	2015	216	(1.5)	221	(0.4)	63	(1.9)	29	(1.6)	6	(1.0)
		2013	210	(1.3)	221	(0.3)	56	(1.7)	23	(1.3)	4	(0.6)
		2011	210	(1.4)	220	(0.3)	55	(1.8)	23	(1.4)	4	(0.5)
		2009	207	(1.1)	220	(0.3)	51	(1.5)	18	(1.5)	2	(0.6)
		2007	207	(1.6)	220	(0.3)	52	(2.0)	20	(1.4)	3	(0.6)
		2005	209	(1.3)	217	(0.2)	53	(1.8)	20	(1.4)	3	(0.6)
		2003	205	(1.4)	216	(0.3)	49	(1.7)	20	(1.1)	4	(0.5)
		2002	207	(1.7)	217	(0.5)	50	(1.9)	20	(1.4)	4	(0.5)
		1998	200	(1.6)	213	(1.2)	44	(1.8)	17	(1.2)	3	(0.5)
		19981	204	(1.5)	215	(0.8)	48	(1.6)	19	(1.4)	3	(0.5)
	8	2015	255	(1.2)	264	(0.2)	66	(1.8)	23	(1.4)	2	(0.3)
		2013	257	(1.0)	266	(0.2)	68	(1.4)	24	(1.3)	2	(0.4)
		2011	255	(1.5)	264	(0.2)	66	(2.2)	22	(1.4)	1	(0.3)
		2009	253	(1.6)	262	(0.3)	64	(1.8)	20	(1.8)	1	(0.3)
		2007	253	(1.1)	261	(0.2)	64	(1.6)	19	(1.2)	1*	(0.5)
		2005	253	(1.6)	260	(0.2)	64	(2.2)	20	(1.5)	1	(0.4)
		2003	253	(1.6)	261	(0.2)	64	(1.9)	22	(1.4)	2	(0.3)
		2002	256	(1.5)	263	(0.5)	68	(1.8)	22	(1.5)	1*	(0.4)
		1998	252	(1.4)	261	(0.8)	63	(1.9)	17	(1.5)	1	(0.3)
		19981	252	(1.5)	261	(0.8)	64	(1.9)	18	(1.4)	1	(0.2)

Note: Standard Errors (SE) are shown in parentheses. All scores reported are lower than the National public average, except where indicated by an asterisk (*). In these cases, scores do not differ significantly from the National public sample.

1Accommodations were not permitted for this assessment.

Rounds to zero.

† Not applicable.

This report was generated using the State Profiles. <http://nces.ed.gov/nationsreportcard/states/>

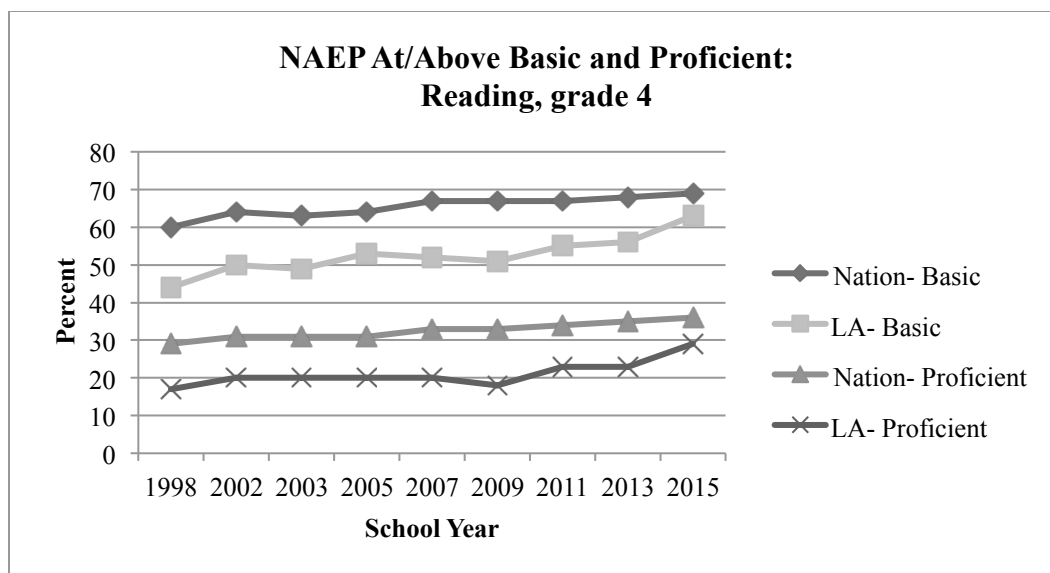


Figure 1.2 NAEP achievement levels attained in Reading for Louisiana and the national average at grade 4. Number reported includes percentage at or above the reported achievement level. Adapted from the National Center for Education Statistics NAEP State Profiles report at <http://nces.ed.gov/nationsreportcard/states/> and the Condition of Education, 2015c.

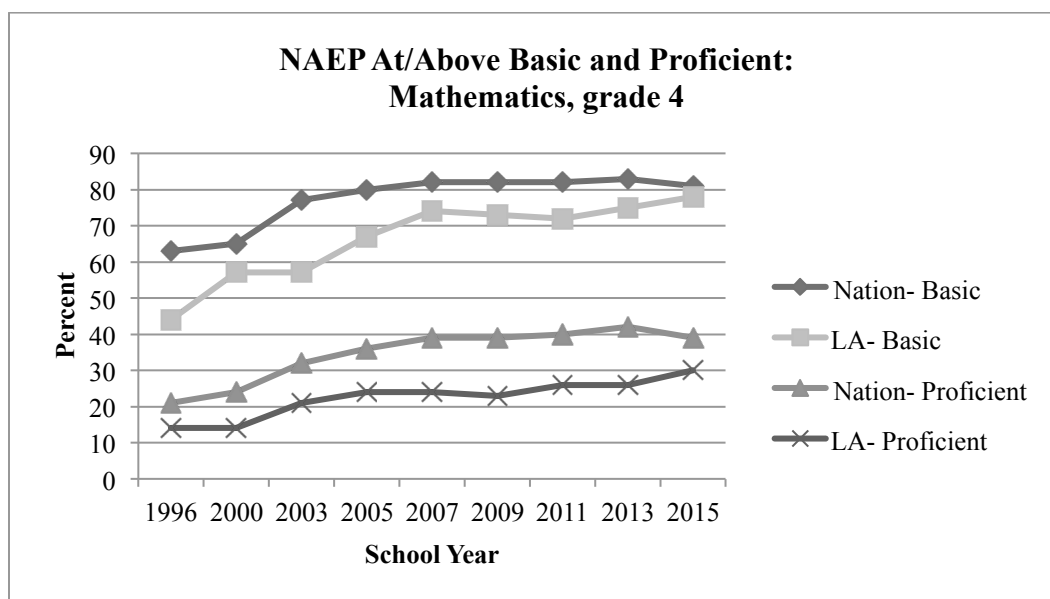


Figure 1.3 NAEP achievement levels attained in Reading for Louisiana and the national average at grade 8. Number reported includes percentage at or above the reported achievement level. Adapted from the National Center for Education Statistics NAEP State Profiles report <http://nces.ed.gov/nationsreportcard/states/> and the Condition of Education, 2015c.

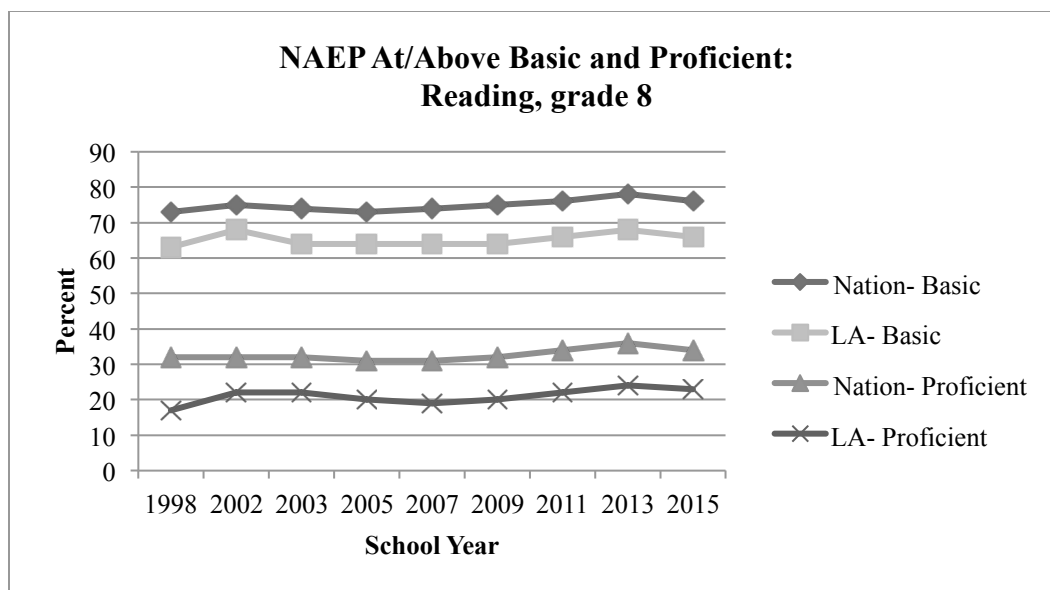


Figure 1.4 NAEP achievement levels attained in Mathematics for Louisiana and the national average at grade 4. Number reported includes percentage at or above the reported achievement level. Adapted from the National Center for Education Statistics NAEP State Profiles report <http://nces.ed.gov/nationsreportcard/states/> and the Condition of Education, 2015c.

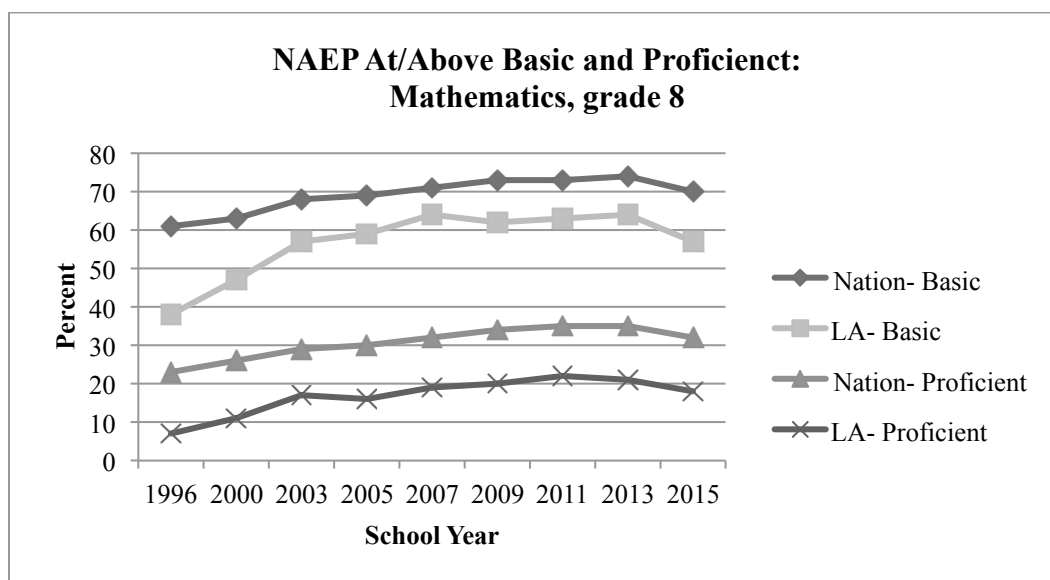


Figure 1.5 NAEP achievement levels attained in Mathematics for Louisiana and the national average at grade 8. Number reported includes percentage at or above the reported achievement level. Adapted from the National Center for Education Statistics NAEP State Profiles report <http://nces.ed.gov/nationsreportcard/states/> and the Condition of Education, 2015c.

Additionally, no significant gain has been made in closing the achievement gap between black and white students in Louisiana on NAEP. In 2013, black students in Louisiana averaged a NAEP score 26 points lower than white students in 4th grade reading achievement (United States Department of Education [USDOE], 2015a). In 2015, that gap narrowed by three points (USDOE, 2015a). For both grade levels and subjects, the achievement gap ranged between 18 and 26 scaled score points between black and white students and showed no statistical significance from previous administrations of the assessment (USDOE, 2015a).

There may be some gain in failure, however, for certain populations and institutions. “All organizations are perfectly designed to get the results they are now getting. If we want different results, we must change the way we do things” (Northup, 2008). Fasching-Varner, Mitchell, Martin, and Bennett-Haron (2014) “reject the discourse of crisis” in education (p. 412).

We suggest, following earlier arguments in our own work, that the systems in place that organize both prisons and schools are far from a broken complex—they are well-oiled machines furthering the economic imperatives of the free market. Further, we believe that by bolstering the economy, they continue to benefit those with significant wealth and access and that school failure and expanding prisons themselves represent remarkably stable and predictable market opportunities, no doubt oppressive to working-class communities of color caught in the collective grindhouse, but the perfect environment to incubate the growth of the market (p. 412).

Moreover, of all Organisation for Economic Cooperation and Development (OECD) countries, the United States (US) ranks highest in publically posting achievement data for accountability purposes (see Figure 1.6) and in using student assessment data to compare school performance (see Figure 1.7). Notable for its high-ranking status on international test scores (National Center on Education and the Economy, 2015), Finland ranks last in publically posting

achievement data for accountability purposes and decreased its use of student assessment data to compare school performance since 2003.

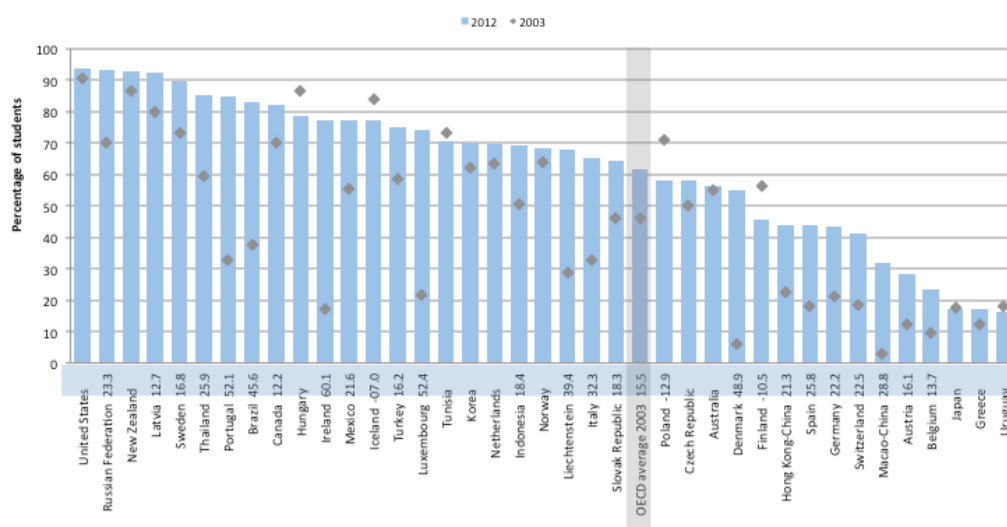


Figure 1.6 Ranking of OECD countries in publically posting achievement data for accountability purposes. Adapted from OECD, PISA 2012 Database, Table IV.4.31.

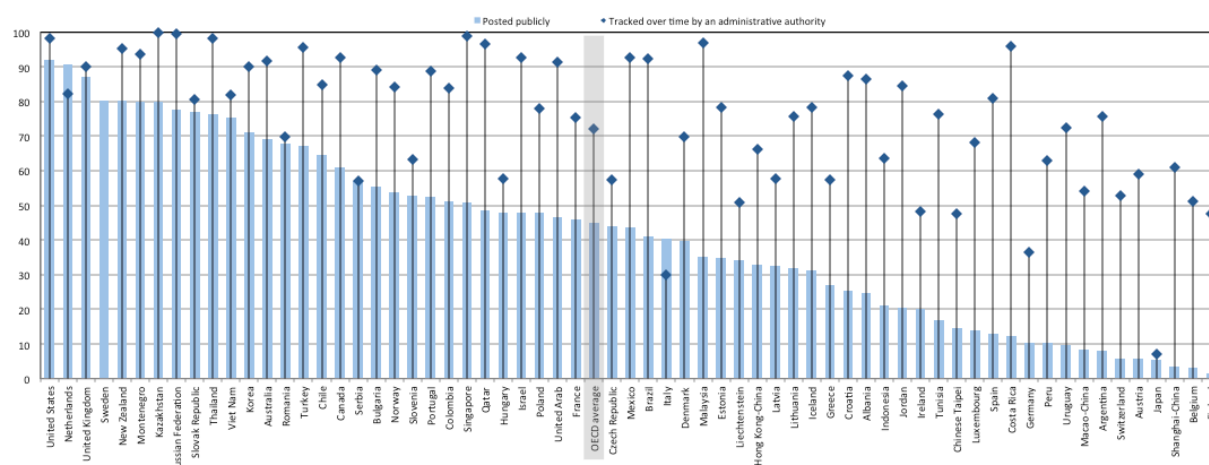


Figure 1.7 Ranking of OECD countries in comparing schools for accountability purposes. Adapted from OECD, PISA 2012 Database, Table IV.4.31

Statement of the Problem

Louisiana's current test-based accountability system is not designed to provide information on student achievement beyond the cognitive domain of learning for accountability purposes, though skills in the affective and psychomotor are integral to life success (Rothstein, 2004). Due to the over-reliance on testing, teachers do not receive quality information on students' learning strengths and weaknesses that can inform the direction of teaching for improved student learning. Moreover, the standardized tests used in the accountability system measure mostly lower levels of knowledge and cognitive processing (Conley & Darling-Hammond, 2013; Glasser, 1990; Miller, Linn, & Gronlund, 2013; Rothstein, 2004). Standardized accountability tests used for test-based accountability are summative in nature and limit the function of informing teacher practice in response to student learning needs (Conley & Darling-Hammond, 2013). Furthermore, high-stakes standardized accountability tests currently being used do not measure student learning and development in the affective, psychomotor, and higher-levels of the cognitive domains. Consequently, these tests inadequately represent students' learning (Rothstein, 2004).

Purpose Statement

This study examined the test-based accountability system in Louisiana and the extent to which the system provides meaningful and actionable data for stakeholders. The intent of this two-phase, concurrent, mixed-methods study was to discover the educational values most prioritized by Louisiana stakeholders and the extent to which current policymakers were willing to pilot an advanced system for test-based accountability, including tests of the higher levels of the cognitive domain as well as indicators for learning in the affective and psychomotor domains. In the first phase, quantitative research questions addressed the comparison of

perceptions about Louisiana's test-based accountability system and educational values of stakeholders through the administration of a digitally based statewide survey. Stakeholders included parents, teachers, teacher/instructional leaders, principals, local community members, LEA staff, LEA superintendents, SEA staff, elected officials, members of institutions of higher education, and the business community. Using hierarchical linear modeling, survey responses were correlated along a line of regression between the dependent and independent variables. Information from this first phase was explored further in the second, qualitative phase. In the second phase, qualitative interviews were used to probe significant comparative relationships of the survey data by exploring aspects of Louisiana's test-based accountability system with lawmakers serving on the House and Senate Education Committees at the Louisiana Legislature. The follow up with qualitative research in the second phase allowed better understanding and explanation of the data captured by the first phase.

Research Questions

Phase I: To what extent does Louisiana's current test-based accountability system deliver results that are valuable to stakeholders?

Phase II: To what extent are policymakers in Louisiana receptive to piloting an advanced system of accountability for Louisiana that includes reported measures in the affective and psychomotor domain to improve student learning outcomes?

Pending results of this study, an immediate subsequent research question may include: What is the impact on student achievement of measuring and reporting indicators in all three domains for learning on the annual state standardized summative test?

Theoretical Framework

Taxonomy for Learning

In 1949, Ralph Tyler, regarded by some as the “father of behavioral objectives” (Fishbein, 1973, p. 55), published the seminal work on curriculum design, *Basic Principles of Curriculum and Instruction*. This work led to later work that influenced curriculum design and theory of learning. Benjamin Bloom, one of Tyler’s students, published a book that outlined the concept of learning across three domains for learning: cognitive, affective, and psychomotor. In this text dedicated to Tyler, Bloom delineated a taxonomy, or classification system, of objectives in the cognitive domain. He included six levels in order of cognitive complexity: knowledge, comprehension, application, analysis, synthesis, and evaluation (see Table 1.3).

Table 1.3

		Cognitive Domain for Learning
Cognitive Complexity		Classification
	High	Evaluation
		Synthesis
	Moderate	Analysis
		Application
		Comprehension
Low	Knowledge	

Almost fifty years later, Anderson and Krathwohl, (2001) expanded on Bloom’s earlier work, updating the taxonomy to include two dimensions of learning in the cognitive domain: knowledge (i.e. content) and cognitive process (i.e. behavior) (see Table 1.4). The knowledge dimension includes four levels of understanding: factual, conceptual, procedural, and meta-cognitive. Factual knowledge is defined as “the basic elements students must know to be acquainted with a discipline or solve problems in it” (Anderson & Krathwohl, 2001, p. 29). Conceptual knowledge is defined as “the interrelationships among the basic elements within a

larger structure that enable them to function together” (Anderson & Krathwohl, 2001, p. 29).

Procedural knowledge is defined as “how to do something, methods of inquiry, and criteria for using skills, algorithms, techniques, and methods” (Anderson & Krathwohl, 2001, p. 29).

Finally, metacognitive knowledge is “knowledge of cognition in general as well as awareness and knowledge of one’s own cognition” (Anderson & Krathwohl, 2001, p. 29). The cognitive process dimension includes six levels of cognitive processing:

- 1.) Remember (i.e. retrieve relevant knowledge from long-term memory) (p. 31),
- 2.) Understand (i.e. construct meaning from instructional messages, including oral, written, and graphic communication) (p. 31),
- 3.) Apply (i.e. carry out or use a procedure in a given situation) (p. 31),
- 4.) Analyze (i.e. break material into constituent parts and determine how parts relate to one another and to an over-all structure or purpose) (p. 31),
- 5.) Evaluate (i.e. make judgments based on criteria and standards) (p. 31), and
- 6.) Create (i.e. put elements together to form a coherent or functional whole; reorganize elements into a new pattern or structure) (p. 31).

Table 1.4

		Cognitive Process					
		Remember	Understand	Apply	Analyze	Evaluate	Create
Knowledge	Factual	List	Summarize	Respond	Select	Check	Generate
	Conceptual	Recognize	Classify	Provide	Differentiate	Determine	Assemble
	Procedural	Recall	Clarify	Carry Out	Integrate	Judge	Design
	Meta-Cognitive	Identify	Predict	Use	Deconstruct	Reflect	Create

Note. From the Center for Excellence in Learning and Teaching at Iowa State University, <http://www.celt.iastate.edu/wp-content/uploads/2015/09/RevisedBloomsHandout-1.pdf>. Adapted from Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., and Wittrock, M.C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom’s Taxonomy of Educational Objectives* (Complete edition). New York: Longman.

It is important to firmly establish learning objectives in our quest to ensure schools provide the education we value and expect from such a system. As teaching and assessing are integrally entwined in the learning process (Black & Wiliam, 2004), adequate and appropriate testing practices are conditioned on the instruction delivered in the classroom.

Binet and Simon (1916), developers of the earliest large-scale educational assessments, “admit of other things than intelligence, to succeed in his studies, one must have qualities which depend especially on attention, will, and character” (p. 254). Bloom’s taxonomy classified these skills in the affective domain for learning. This domain is characterized by “interest, attitudes, and values” (Bloom, 1956, p. 7). Admittedly, Bloom (1956) confirmed the difficulty in delineating learning across the affective domain. These skills are highly personal and manifest within the realm of social-emotional skills. Educational and cognitive psychologists have explored this area of study as non-cognitive skills (Duckworth & Yeager, 2015), social-emotional skills (Durlak, Dymnicki, Taylor, Weissberg, & Schellinger, 2011), 21st century skills (National Research Council, 2012; Conley & Darling-Hammond, 2013), character development (Fadel, Bialik, & Trilling, 2015), soft-skills (Brill, Gilfoil, & Doll, 2014), and executive functioning skills (Zimmerman, Bandura, & Martinez-Pons, 1992). Finally, the psychomotor domain includes the “manipulative or motor-skills” area (Bloom, 1956, p. 7), including “imitating, manipulation, precision, articulation, and naturalization” (Anderson & Krathwohl, 2001, p. 21).

Research has shown that non-cognitive qualities can predict academic achievement (Duckworth & Yeager, 2015) and that incorporating learning activities across the cognitive, affective, and psychomotor domains for learning positively impact academic achievement

(Egalite, Mills, & Green, 2015; Greenberg, Weissberg, O'Brien, Sins, Fredericks, Resnick, & Elias, 2003; Zins, Bloodworth, Weissberg, & Walberg, 2004). Thus, evidence suggests a mutually beneficial relationship (Rothstein, 2004) across the three learning domains.

Educational Assessment

Assessment is “a process of gathering information for the purpose of making judgments about a current state of affairs” (Pellegrino, 2002, p. 49) and is used to evaluate “academic, behavior, and physical problems” (Salvia & Ysseldyke, 1995, pp. 21-23). Tests can be classified according to the intended use of results, including formative or summative (Miller, Linn, & Gronlund, 2013, p. 37).

The critical element that most directly impacts student learning in tests designed to inform classroom instruction is feedback provided to the teacher and the student (Haertel & Herman, 2005; Miller, Linn, & Gronlund; 2013). Accordingly, diagnostic and formative assessments are designed to provide information that can more directly impact teaching and learning in a timely fashion. Research has shown that formative assessments have a positive impact on student learning (Black & Wiliam, 1998), providing valuable information on student achievement to inform decision-making by educators. Tests used for test-based accountability purposes, on the other hand, are objective in nature and summative in design, and therefore are limited in the type of feedback they provide for instructional purposes. The Gordon Commission on Future Assessment in Education (2013) issued the following policy statement on the imperative for ensuring that assessments provide meaningful and valuable information of student learning:

At the most general level, the emphasis in our educational systems needs to be on helping individuals make sense out of the world and how to operate effectively within it. It is important that assessments do more than document what students

are capable of and what they know. To be as useful as possible, assessment should provide clues as to why students think the way they do and how they are learning as well as the reasons for misunderstandings (p. 7).

Knowledge-Based Economy

Over the past forty years, the terrain of skills-demand for the workforce has shifted. In 2003, Artor, Levy, and Murnane, economists at Harvard and MIT, published a study on the shift in workforce demand due to computerization and investigated the reason for high demand of educated workers (i.e. workers with college-level degrees) even within a computerized workforce. Where forty years earlier employers had valued routine cognitive and manual tasks, now value for tasks classified as non-routine interactive, (e.g. complex communication) and non-routine analytic (e.g. expert thinking skills) increased sharply into the new millennium (see Figure 1.8). In a society that increasingly rewards students more for what they can do with what they know, than simply for what they know, (Conley & Darling-Hammond, 2013; OECD, 2014), the goals of our educational accountability systems should be well aligned with those skills most valued by stakeholders.

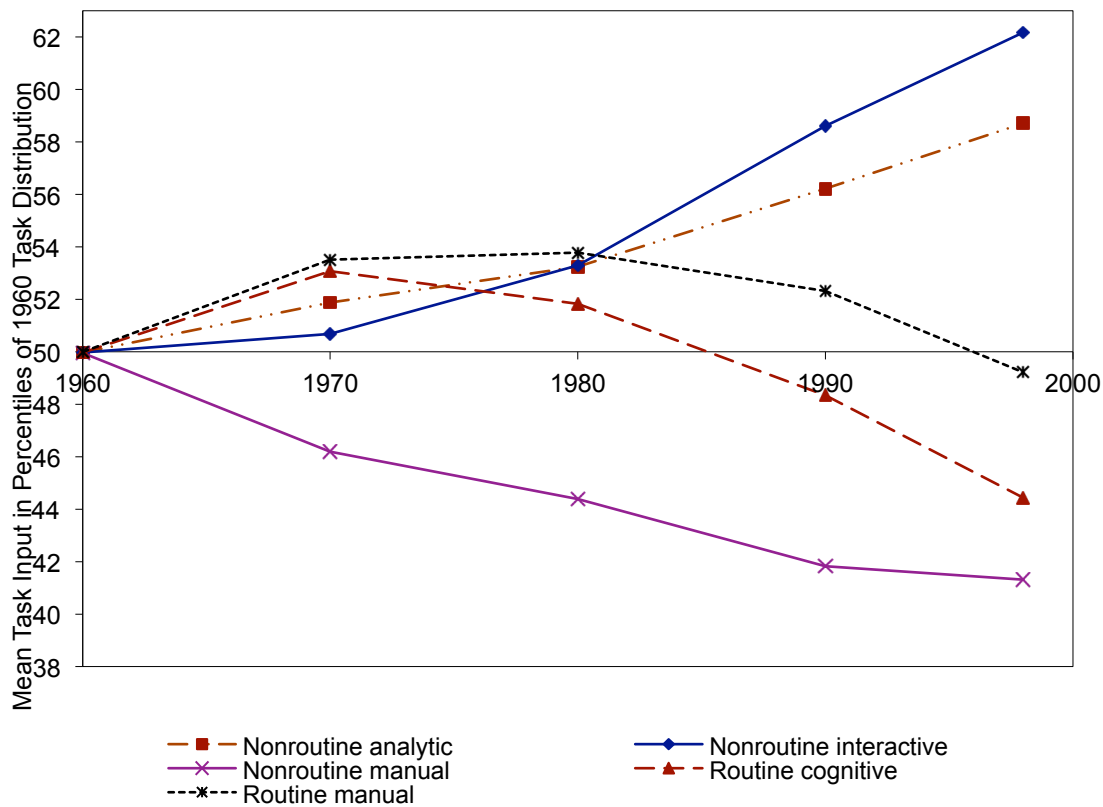


Figure 1.8 Trends in Routine and Nonroutine Task Input, 1960 to 1998 Adapted from Autor, Levy, & Murnane (2003). The skill content of recent technological change: An empirical exploration. *Quarterly Journal of Economics*, 118(4), 1279-1334.

Student achievement as measured by test scores in ELA and math—a “privileged” sect (Figlio & Ladd, 2015, p. 196)—is only a small portion of the educational outcomes valued by stakeholders in this study. In 2008, Rothstein, Jacobsen, and Wilder “conducted a survey of the general public on goals for public education and reported their analysis of findings from the poll in rank order: basic academic skills, critical thinking, social skills and work ethic, physical health, emotional health, preparation for skilled work, citizenship, and the arts and literature” (p. 43). By these measures, test-based accountability systems that use only standardized achievement tests meet only one of the priorities for education valued by stakeholders: basic

academic skills. Measures for the remaining goals are not captured by cognitive achievement tests; therefore, appropriate measures in the corresponding domain(s) are necessary.

Darling-Hammond, Herman, Pellegrino, Abedi, Aber, Baker et al. (2013) reported similar findings. The skills most valued by Chief Executive Officers of Fortune 500 Companies indicated that non-cognitive and higher-order cognitive skills were most valued over basic skills and writing ability (see Figure 1.9).

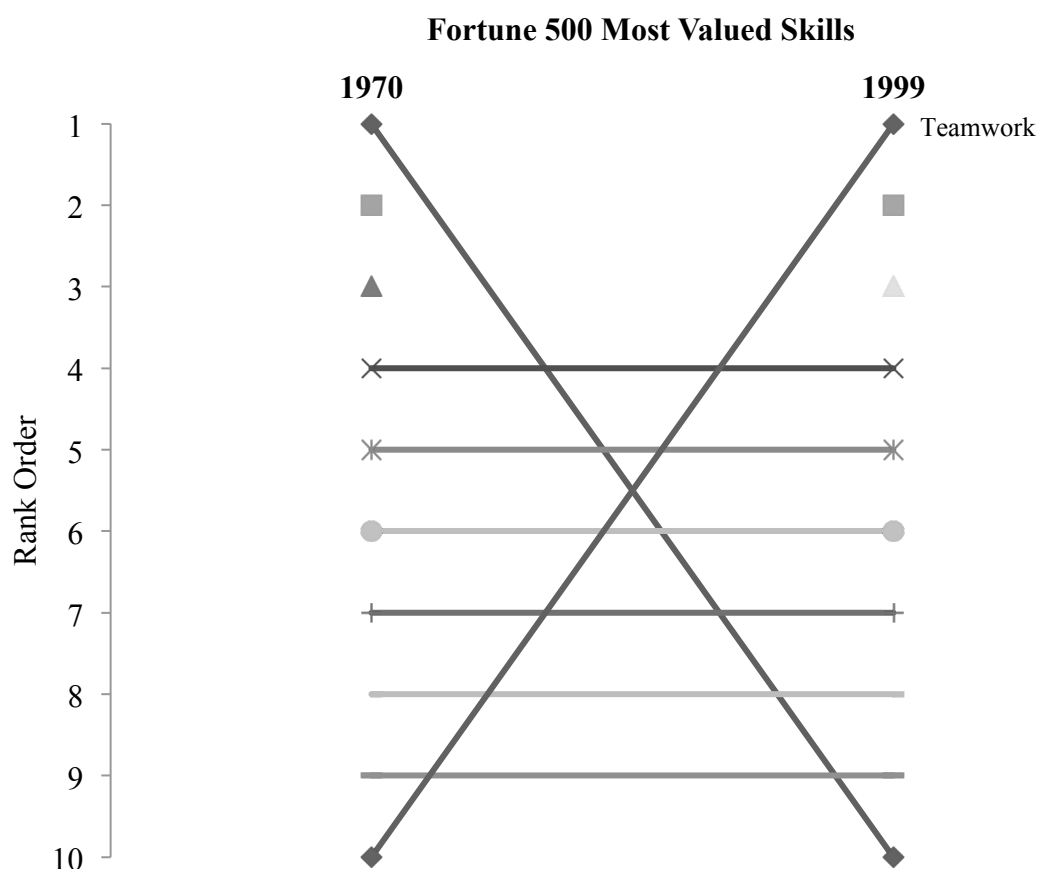


Figure 1.9. Fortune 500 Most Valued Skills. Adapted from Cassel & Kolstad (1999) and Creativity in Action (1990). Cassel, R. N., and Kolstad, R. (1999). The critical job-skills requirements for the 21st century: Living and working with people. *J. Instructional Psychology*, 25(3), 176-180; Creativity in Action (1990). *Skills desired by Fortune 500 companies (in order of importance)*. Buffalo, NY: Creative Education Foundation.

Goal(s) of Formal Education

What then, is the aim of a formal education in America's public schools? A goal of "school learning" is the "transfer of learning, that is, the application or use of what is learned in one domain or context to that of another domain or context" (Amrein & Berliner, 2002, p. 13). Because one cannot directly perceive students' mental processes, educational assessment involves making judgments about what students know based on "a process of reasoning from evidence" (National Research Council (NRC), 2001, p. 53). Thorndike (1918) observed that "education is concerned with changes in human beings" (p. 16) "and its effectiveness could be judged by differences in student behavior" (Haertel & Herman, 2005, p. 4). Amrein and Berliner (2002) described learning as "the process by which education is achieved" (p. 12). Learning has also been described as "changes in the behavior of an organism that are the result of regularities in the environment of that organism" (De Houwer, Barnes-Holmes, & Moors, 2013, p. 633).

If a primary goal of education is learning, then it follows that a goal of educational accountability should be to ensure that learning is effectively facilitated by our agents of this change: teachers in schools. Siegel (2004) argued that, "although some current testing and accountability practices and policies are perfectly legitimate, many of them are largely inimical to the achievement of our most defensible educational ends" (p. 52). Education professionals recognize and acknowledge that the purpose for education extends beyond purely academic gains (Rutledge, Cohen-Vogel, Osborne-Lampkin, & Roberts, 2015). However, the design of current test-based accountability systems that apply high-stakes to the test results encourages states and schools to respond in ways that privilege achievement on standardized tests over deeper learning and does not account for learning in the affective and psychomotor domains.

Significance of the Study

Every Student Succeeds Act

The state of educational assessments in the US garnered national attention from President Barak Obama when he boldly declared:

I am calling on our nation's governors and state education chiefs to develop standards and assessments that don't simply measure whether students can fill in a bubble on a test, but whether they possess 21st-century skills like problem-solving and critical thinking, entrepreneurship and creativity (The White House, 2009).

At the time of this speech, the federal education law *No Child Left Behind* [NCLB] (Public Law 107-110, 2001), a reauthorization of the 1965 Elementary and Secondary Education Act (ESEA) that was due for reauthorization in 2007, remained stagnant in its policies of test-and-punish. To comply with the law, the USDOE began issuing waivers to states to relieve them of the pressures of the law, which required 100% proficiency by 2014. The July 2015 headline of Emily Cadei's story in *Newsweek* read, "No Child Left Behind: the education law everyone wants to fix". In the fall of 2015, the 114th US Congress set to work during the 2015 legislative session and in December, President Obama reauthorized ESEA by signing into law the *Every Student Succeeds Act* [ESSA]. The law, which will go into effect in 2017-2018, maintains the 17 mandated annual state assessments, however, states are allowed flexibility in the utilization of the results. Additionally, the bill allows for the creation of a state's "Innovation Assessment System" using competency-based learning and performance assessment measures. The law also allows states to include a non-cognitive indicator in its accountability system. While some states are comfortable including measures of school climate, usually including information around school discipline (suspension and expulsion) and attendance rates, others are experimenting with including non-cognitive aspects such as social-emotional learning and physical health.

School Governance Models

The opportunities available to parents for school choice in Louisiana are largely driven by the information provided in the state's accountability system and include traditional public schools, non-traditional public schools (charter), private schools, private school opportunity scholarship (i.e. vouchers), and parent trigger laws. Politics promises to have a continued influence over education policy, including interest group mobilization, increasingly connected interest groups and provider networks, as well as realigned interests and weakening the structures of the dominant monopoly on education policy (McDonnell, 2012). Potential changes due to political effect include changes in institutional rules and structures, equity for resources and providing the opportunity to learn, and stakeholder engagement, as well as emerging policies changing the landscape of education delivery systems (e.g. vouchers, tuition-donation rebate or tax rebate programs, online learning, etc.).

The emergence of hybrid structures for systems including the Education Achievement Authority in Michigan, the Recovery School District in Louisiana, and the Achievement District in Tennessee, operationalize flexibility to contract diverse providers, to close and/or consolidate school buildings, and to function outside a collective bargaining framework. Another key factor emerging in importance is the structure for data governance. In a survey conducted by Data Quality Campaign, eight states tracked students from prekindergarten through college or into the workforce arena and fewer than half of all states provided aggregated data reports to key stakeholders (Wong, 2013).

Policy Window

With standards, assessments, and accountability systems so integrally intertwined, education leaders may have access to 'borrowing strength,' Manna's (2006) idea defined as that

which “occurs when policy entrepreneurs at one level of government attempt to push their agendas by leveraging the justification and capabilities that other governments elsewhere in the federal system possess” (p. 5). Marshall and Gerstl-Pepin (2005) described the process by which the conditions are prime for policy approval as the “policy window” (p. 15), in which three factors are recognized: problem recognition, policy proposals, and politics. As accountability systems are allowed greater flexibility under ESSA, perhaps the policy window for providing more coherence, comprehensiveness, and continuity (Herman & Baker, 2009) in test-based accountability system is opening.

Summary

Schmidt and Maier (2009) asserted that the opportunity to learn (OTL) is “perhaps the single most important factor related to student learning” (p. 555) as it represents one’s degree of learning as a function of the actual time spent learning content divided by the time needed for learning the content. Standardized assessment measures, however, emphasize a performance orientation, which may diminish the OTL for some students (Herman & Baker, 2009; Darling-Hammond, 2007), negatively impacting the student’s achievement on standardized assessments. Performance orientation can have a negative effect on the intent of achievement tests—reducing motivation to learn and decreasing students’ overall capacity for education (Herman & Baker, 2009). Koretz (1996) warned against such corruptible measures in highly results-based testing environments as inflation of scores and degradation of instruction. Additionally, analytics allow for school leaders to calculate basic statistics for maintaining or increasing a school performance rating, so as to prioritize those students who are most likely to demonstrate the greater return on investment of time. The preparation for and participation in the standardized testing process infuses high levels of stress and anxiety in students, as well as teachers. Students, therefore, may

experience consequences for results received by a system to which they were disadvantaged. Vulnerable populations, including students with disabilities, English language learners, and students from economically disadvantaged backgrounds, comprise a critical mass for implications of policy considerations (Herman & Baker, 2009; McDonnell, 2009).

Failure on the test is not perceived or undertaken as an opportunity to learn, but rather as a judgment of worth. Furthermore, achievement tests represent only small samples from large domains of achievement. Koretz (1996) asserted that an accountability system is insufficient if it does not account for outcomes beyond those typically measured on tests, such as attitudes and habits. The outcomes of state accountability systems drive public perception of educational progress (Gong, 2002) in and across states, in addition to the various high stakes policies directly impacting students, parents, teachers, and leaders. Intense competing interests remain in accountability and the risk that the “polarization of school accountability is embedded in our interest-based political structure” (Wong, 2013, p. 417).

The definitions of school performance and academic achievement should more closely align with our actual goals in education (Kim & Sunderman, 2005; Haney, 2008). Alternatives for systems of accountability include incorporating multiple measures of student achievement in state accountability ratings of school performance, as well as factoring in student growth rates. Value-added and growth models may provide greater strength in estimating causal inferences (Kim & Sunderman, 2005). Linn (2008) identified student background and academic skills prior to the start of school as the best factors in estimating growth targets for students.

The use of standardized assessments for accountability should result in authentic measures of student learning, as well as provide actionable data for teachers for impacting

instruction. In current practice, student performance may remain uncovered, hidden behind the data, as adults control access to and reporting of student results, with the potential to manipulate the data to portray the narrative of any agenda of priority. For a more accurate measure of student cognitive development, structures should be in place for students to demonstrate what they know and can do through measures that accurately and adequately reflect their learning.

Currently, the design of test-based accountability systems that apply rewards and sanctions to the results of these high-stakes standardized tests encourages states and schools to respond in ways that privilege achievement on high-stakes standardized tests over deeper learning and a well-rounded education. The exclusion of learning that incorporates the affective and cognitive domains warrants learning in those domains to be irrelevant. This shift in pedagogical behavior has the potential to shape the purpose for schooling and the very outcome of education for school children in across the US.

CHAPTER 2. REVIEW OF THE LITERATURE

There is a saying that American students are the most tested, and the least examined, of any in the world. (Darling-Hammond, 2013).

Historical Perspective of the Emergence of Test-Based Accountability in America

1965-2015

“At the heart of NCLB [No Child Left Behind] and its ladder of sanctions is the idea that accountability improves performance” (Hemelt, 2011, p. 702) and has “direct” economic effects (Hanushek & Raymond, 2006, p. 51). NCLB catapulted test-based accountability for student achievement into the mainstream of education systems, requiring, among others, that states identify and intervene in low-performing schools based on student achievement test scores (NCLB, P.L., 2001, pp. 107-110). More critical than the cost of financing current accountability systems is the cost of forsaking authentic learning opportunities in the classroom in pursuit of accountability goals. Interestingly, nearly twenty years prior to the enactment of federal legislation on accountability, Levin (1974) identified parallels between educational systems of accountability and those often observed in business and industry: “cost-accounting systems, employee productivity ratings, contracting for services, cost-effectiveness analysis, and information systems for management decisions making” (p. 368).

A dominating narrative of educational accountability is that improved educational outcomes positively impact the economy (Hanushek & Raymond, 2006). Murnane, Willet, Duhaldeborde, and Tyler (2000) found that an increase in performance in mathematics at the end of high school in the US is the equivalent of higher annual earnings by 12%. Furthermore, the economic impact of productivity gains of human capital have a compounding effect on the national Gross Domestic Product (GDP) (Hanushek & Raymond, 2006). In the US, “if the

economy grew by 1% per year starting in 2000, GDP per capita would increase by 65 percent by 2050” (Hanushek & Raymond, Federal Reserve Bank, 2006, p. 53). Additionally, Figlio and Lucas (2004) noted the impact of information provided by school accountability systems on the choices parents make about schools and affiliated housing market patterns.

Test-based accountability is based on the behaviorist paradigm assumption of causality between external accountability through incentives and sanctions and school improvement (Jacob, 2005), and forms a system of school improvement weighted primarily on motivation of its leaders, teachers, and students (Fuhrman, 2004). Conley and Darling-Hammond (2013) provided a working definition of “system” of accountability as “a set of commitments, policies, and practices that are designed to:

- 1) increase the probability that schools will use good practices on behalf of students;
- 2) reduce the likelihood that schools will engage in harmful practices; and
- 3) encourage ongoing assessment on the part of schools and educators to identify, diagnose, and change courses of action that are harmful or ineffective” (pp. 34-35).

Holmstrom and Milgrom (1991) showed that incentive schemes based on objective criteria led agents to focus on the most easily observable aspects of a multi-dimensional task. Based on similar logic, critics have argued that such policies will cause teachers to shift resources away from low-stakes subjects, neglect infra-marginal students, and ignore critical aspects of learning that are not explicitly tested” (Jacob, 2005, p. 762). It is under this paradigmatic thinking that educational assessment has become more widely used as a policy lever (Miller, Linn, & Gronlund, 2013; Shavelson, Young, Ayala, Brandon, Furtak, Ruiz-Primo, et al., 2008) set within a political context (Jones, Jones, & Hargrove, 2003). Shepard (2008) noted the susceptibility to “politicization” (p. 30) of accountability tests, i.e. “...uses of tests for externally mandated

accountability purposes” (Baker & Linn, 2004, p. 51). According to the NRC (2001), Shepard (2008) noted these tests were

derived from early theories that characterize learning as a step-by-step accumulation of facts, procedures, definitions, and other discrete bits of knowledge and skill. Thus, the assessments tend to include items of factual and procedural knowledge that are relatively circumscribed in content and format and can be responded to in a short amount of time (p. 26).

“Assessment thus is used as an instrument of policy to directly and indirectly advance education and social foals and as a tool to determine the effectiveness of educational policies, practices, programs, and individuals as well as institutions” (Herman & Baker, 2009, p.177). Moreover, Elmore and McLaughlin (1988) identified that “Policies, as we’ve seen, are useful, but blunt, instruments. Under the best of circumstances, they can influence that allocation of resources, the structure of schooling, and the content of practice; but those changes take time and often have unexpected effects” (p. 60). Therefore, these policies are inappropriate for use as a “barometer and lever of reform” (Miller, Linn, & Gronlund, 2013, p. 3).

Elementary and Secondary Education Act

In 1635, the first public school in the US was established. Public education, by federal law [cite], did not become compulsory for another 250 years. From 1890-1930, public high school attendance increased from 10% to 70% enrollment (Kress, Zechmann, & Schmitten, 2011). This period in education history could be characterized by efforts aimed at increasing the quantity of education offered to the American public by increasing student attendance and educational programming.

In the mid-twentieth century, the focus of American public education shifted from efforts to increase the quantity of education to influencing factors affecting the quality of education. The

Supreme Court ruling in *Brown v. Board of Education* (1954) ended racial segregation in the public school system and resonated throughout the country. This landmark ruling, considered by Reber (2007) as “perhaps the most important innovation in US education policy in the 20th century” (p. 1), had “dramatic effects on the characteristics of the schools that black Louisianans attended” (p. 8).

With their own schools shuttered, black students were uprooted from familiar environments and distributed as necessary; they would provide the statistical proof of significant progress. Having lived their whole lives with the same kids in the same neighborhood, black children found themselves divvied up and bused off in opposite directions. They lost their clubs, their teams, their student groups. “At the age of fourteen, it was like someone took a knife and cut off everyone you ever knew,” said one young black student from Texas. And because of white flight and defections to private schools, the number of white students in these systems was plummeting with each passing year. So to meet the needs of racial balance, black students had to be shuffled around every fall, seemingly at random. At the most extreme, a black student might attend four different schools in four different years. (Colby, 2012, p. 44).

This narrative does not sound altogether different from similar experiences of present-day education reform initiatives in Louisiana—school closures and takeovers, inequitable enrollment lottery system, and busing away from neighborhood schools (Buras, 2011; Salazar, Perez, & Cannella, 2011).

Efforts to implement the *Brown* ruling faced years of resistance and did not realize meaningful progress before 1965 (Reber, 2007). Even in 1968, “the average black Louisianan was in a school that was 8 percent white and whites comprised over 60 percent school enrollment” (Reber, 2007, p. 5).

In 1965, then-President Johnson signed into law the Elementary and Secondary Education Act (ESEA) as a part of his “War on Poverty” to provide free and fair public education to American schoolchildren (Kress, Zechmann, & Schmitten, 2011). For the first time in the US, federal funding for public education was tied to federal policy specifically to support

educational opportunity for students from high poverty communities (Forte, 2013). This landmark law was the nation's first effort to close the achievement gap between minority and non-minority school children, as well as between those from disadvantaged and non-disadvantaged backgrounds (Allen, Altwerger, Edelsky, Larson, Rios-Aguilar, Shannon, & Yatvine, 2007; Kim & Sunderman, 2005).

ESEA included a provision for its reauthorization every five years (Superfine, 2005), and provided grants of federal dollars, known as Title I funding, to states' local school systems to provide equitable educational services (Kim & Sunderman, 2005; Jaiani & Whitford, 2010). Levin (1974) asserted that ESEA could be construed as the alteration of resources from the programs of the middle and upper classes to augment the lack of resources in schools attended by students from the lower class. In studies comparing the strength of accountability systems across states, a positive correlation was found indicating that states measured to have a strong accountability system along with large low income populations or large residentially segregated populations may have gained control over resources (Carnoy & Loeb, 2002; Lee & Wong, 2004).

Effective schools research. In 1966, James Coleman published the *Equality of Educational Opportunity* report, also known as the Coleman Report. The report included the adverse claim about the negligible impact of schooling on academic outcomes for students and the detrimental trajectory for long-term life outcomes of these failures.

That schools bring little influence to bear on a child's achievement that is independent of his background and general social context; that this very lack of an independent effect means that the inequalities imposed on children but their home, neighborhood, and peer environment are carried along to become the inequalities with which they confront adult life (Coleman, Campbell, Hobson, McPartland, Mood, Weinfield, et al., 1966, p. 325).

Reminiscent of the classic debate on nature versus nurture, the previously stated claim incited the age-old debate around nature versus nurture. This debate seeks to identify which factor has greater influence on one's overall life outcomes: natural endowment (i.e. home-based factors) or environmental conditioning (i.e. school-based factors)? Proponents of the findings of the Coleman report might argue *nature*, or home-based factors. However, one key finding from the report, that the achievement of students in the minority population was more greatly influenced by the school attended than that of the achievement of students in the majority population, was an early indicator that schools do in fact make a difference in the academic achievement for students.

Diverging from the notion that home-based factors determine academic outcomes for students, the effective schools research provides evidence of school-based factors that contribute to academic achievement of students in spite of social factors. Among others, Weber (1971), Edmonds (1979), Rutter (1983), Sizemore, Brossard, and Harrigan (1983), and Teddlie and Stringfield (1993) found that schools make a difference. After controlling for socioeconomic factors, Klitgaard and Hall (1975) found evidence of schools and districts that consistently produced extraordinary student learning. They focused on statistical outliers of regression analysis rather than the central tendency of four large educational data sets and found that effective schools comprised 2 to 9 percent of the sample population.

Edmonds (1979) defined an effective school as being one wherein there was essentially no relationship between family background and achievement. For the purpose of this literature review, the label of an effective school is based on the premise that the primary function of the school is to provide quality and equitable programs for teaching and learning evidenced by

successful outcomes for all students (Lezotte & Bancroft, 1985; Edmonds, 1979; Teddlie & Stringfield, 2007).

Extensive research on effective schools and their practices has been conducted over the past 50 years. What follows here is a literature review of seminal research studies on effective schools from across the US, including California (Murphy & Hallinger, 1988), Kansas (Weber, 1971), Louisiana (Teddlie & Stringfield, 1993), Michigan (Brookover, Schweitzer, Schneider, Beady, Flood, & Wisenbaker, 1978), New York (Edmonds, 1979; Weber, 1971), Pennsylvania (Sizemore, Broussard, & Harrigan, 1983), and Wisconsin (Witte & Walsh, 1990). Overall, findings indicate that school-based factors such as school leadership, teacher behavior, student engagement, and parent-school relationships contribute to successful academic outcomes for students, despite home-based factors.

Weber's (1971) study of four inner-city schools in New York, Los Angeles, and Kansas City launched an effort to identify school-based factors as contributors to student learning outcomes as measured by academic achievement tests. He studied a significant number of economically disadvantaged students scoring at or above national norms on third grade reading achievement and found that, controlling for home-based factors, school was a determinant to success. Especially telling of his study was his comparison of the sample population to that of students in average-income schools. Students in his study were found to have higher academic achievement than those of average-income populations with similar home-based factors as the sample population. A significant finding of this seminal work was the impact of the school-based factor: leadership.

Edmonds (1979) studied nine elementary schools of economically disadvantaged urban minority students in New York City. In addition to finding that school-based factors contributed

to students' academic achievement, he found that a strong and supportive principal along with teachers' planning coordination were integral factors. He described the strength of these school-based factors as one of the most tangible and indispensable characteristics of effective schools, without which "the disparate elements of good schools could be neither brought together nor kept together" (p. 22).

Sizemore, Broussard, and Harrigan (1983) studied three predominantly high achieving schools attended primarily by black students. Of the three schools, one school seemed to be in decline during the 1979-1980 study, yet was found to be the highest achieving black economically disadvantaged elementary school in the city five years later, although a greater percentage of students were classified as economically disadvantaged; the formerly highest achieving school of the three in 1979-1980 study was found to be in decline five years later, even after desegregation resulted in a smaller percentage of black students. The greatest change experienced by both schools was a change in the principal.

Characteristics of effective schools. Effective schools are led by principals with a strong belief that economically disadvantaged minority students have both the ability and capacity to learn (Sizemore, Broussard, & Harrigan, 1983; Edmonds, 1979; Brookover & Lezotte, 1979; Weber, 1971). Murphy and Hallinger (1988) found effective schools included high responsibility assumed by principals and teachers for academic factors including teaching basic reading and math (Brookover & Lezotte, 1979), as well as non-cognitive factors such as student attendance, discipline, and resolve of parental conflict for all students (Sizemore, Broussard, & Harrigan, 1983). The resolve for academic and non-cognitive success displayed the ability of effective districts to view problems in light of solutions, rather than as an obstacle or barrier to the end goal (Murphy & Hallinger, 1988). Brookover and Lezotte (1979) found that staff members at

more effective schools were more likely to report tension and dissatisfaction with existing conditions; however those tensions did not extend from administrator/teacher relations (Murphy & Hallinger, 1988), but rather from dissatisfaction with the status quo characterized by a drive for continuous improvement. Moreover, staff at high-achieving schools tended to report a greater degree of acceptance for accountability (Brookover & Lezotte, 1979; Mackenzie, 1983).

Brookover and Lezotte (1979) noted a stark contrast in the reports of teachers and principals from high-achieving and low-achieving schools. Both teachers and principals from high-achieving schools reported a belief that all students can master basic objectives, a belief supported by a climate of high expectations for student achievement. Edmonds (1979) described these schools as “instructionally effective” (p. 16), suggesting that high expectations extended to teachers and their work. Effective schools, including college and career aptitude schools (Brookover & Lezotte, 1979), were led by principals who assured that distractions due to behavioral misconduct were minimized (Sizemore, Broussard, & Harrigan, 1983; Edmonds, 1979; Weber, 1971; Joyce, Hersh, & McKibbin, 1983; Mackenzie, 1983; Brookover et al., 1978; Murphy & Hallinger, 1988). Additionally, teachers and principals of effective schools reported higher and increasing levels of student ability. Principals were responsible for establishing a school climate with an orderly atmosphere conducive to learning, using standard operating procedures that were not overly rigid (Weber, 1971; Edmonds, 1979; Joyce, Hersh, & McKibbin, 1983; Mackenzie, 1983), and included the choice of functional routines, scenarios, and processes (Sizemore, Broussard, & Harrigan, 1983), where academic achievement took precedence over other all other school activities (Edmonds, 1979; Sizemore, Broussard, & Harrigan, 1983), and teachers practiced common approaches to discipline (Rutter, 1983). Effective principals helped establish school wide sets of value and norms for behavior by facilitating shared decision-

making with teachers (Mackenzie, 1983) in reaching agreed upon levels of proficiency (Lezotte & Bancroft, 1985).

Principals in effective schools were more likely to be an instructional leader (Brookover & Lezotte, 1979; Murphy & Hallinger, 1988) providing a coherently organized curriculum (Joyce, Hersh, & McKibbin, 1983), making curriculum decisions, prioritizing instructional time (Sizemore, Broussard, & Harrigan, 1983; Rutter, 1983) with an emphasis more time in direct reading, including the use of phonics and individualization in reading (Weber, 1971; Sizemore, Broussard, & Harrigan, 1983) and math instruction (Brookover & Lezotte, 1979; Edmonds, 1979; Joyce, Hersh, & McKibbin, 1983). Teddlie, Kirby, and Stringfield (1989) noted some instructional leaders actually engaged in the delivery of lessons. Principals at effective schools provided the necessary supports for instruction (Murphy & Hallinger, 1988) including additional reading personnel (Weber, 1971) and when necessary, diverted resources and funding to support academic work (Edmonds, 1979). Murphy and Hallinger (1988) noted a preference for instructional approaches and curriculum design within effective school districts where corresponding support and development provided quality assurance. Sizemore, Broussard, and Harrigan (1983) identified the fierce commitment of principals supporting teachers for student success through the willingness to disagree with superior officers around the choices of routines and their implementation, including the use of materials that provide functional for elevating achievement even without approval by the higher authority.

Effective schools were results oriented (Sweeney, 1982) and prioritized the process of goal setting for students in setting clear and attainable academic and social behavior goals (Joyce, Hersh, & McKibbin, 1983) including goal-focused activities bent toward clear, attainable, and relevant objectives with continuous diagnosis, evaluation, and feedback on

student performance (Mackenzie, 1983). Effective schools ensured the frequent and consistent monitoring of students' progress in reading and math, and the supervision of instruction was directed toward the students' mastery of these skills (Sizemore, Broussard, & Harrigan, 1983; Edmonds, 1979; Weber, 1971; Joyce, Hersh, & McKibbin, 1983; Murphy & Hallinger, 1988) including data analysis that provided clarity of impact for instructional decision-making and reflection for improving student learning (Lezotte & Bancroft, 1985; Murphy & Hallinger, 1988). Lezotte and Bancroft (1985) suggested those educational outcomes of greatest priority for districts can be identified by the way in which districts choose to monitor student outcomes.

Close monitoring of progress was not limited to students' academic achievement. Effective schools included principals who provided evaluation and support of teacher performance through rigorous supervision and daily visitations of classrooms and programs (Edmonds, 1979), with the prompt evaluation of teacher and staff performances and provision of assistance, help, and in-service where necessary (Sizemore, Broussard, & Harrigan, 1983; Mackenzie, 1983). Moreover, Sizemore, Broussard, and Harrigan (1983) found principals in high-achieving schools effectively utilizing staff and teacher expertise, skills, information, and knowledge.

Effective schools incorporated the mobilization of consensus among the school and community actors around high achievement as the highest priority goal and the involvement of parents in some participatory and meaningful way in the school's program (Sizemore, Broussard, & Harrigan, 1983; Joyce, Hersh, & McKibbin, 1983; Witte & Walsh, 1990; Irvine, 1988), including public rewards and incentives celebrating academic success (Joyce, Hersh, & McKibbin, 1983; Mackenzie, 1983). Brookover and Lezotte (1979) reported more parent-initiated involvement in high-achieving schools, and Murphy and Hallinger (1988) reported the

ability of effective schools to establish communitywide acceptance of school activities even among diverse subcommunities. Irvine (1988) described mutually beneficial partnerships between the school and community, in which community groups saw the partnership as in their best interest to actively support the local schools.

Finally, effective schools displayed a decreased emphasis on compensatory education programming (Brookover & Lezotte, 1979) where strategies were employed to avoid nonpromotion of students as well as to de-emphasize strict ability grouping (Mackenzie, 1983; Weber, 1971). Student classification and categorization for educational programming was allowed only after all provisions for regular education services had been exhausted (Sizemore et al., 1983). Brookover et al. (1978) found that teachers maintained a full year's growth as the academic goal, regardless of specialized categorization.

In effective schools, teachers displayed commitment to student achievement by spending a larger percent of class time on instruction (Brookover et al., 1978; Sizemore, Broussard, & Harrigan, 1983; Rutter, 1983). Teachers maintained autonomy in the classroom (Brookover & Lezotte, 1979; Joyce, Hersh, & McKibbin, 1983; Mackenzie, 1983), were more task-oriented with more evidence of applied appropriate principles of learning (Rutter, 1983; Brophy, 1988), and employed a variety of teaching strategies with opportunities for student responsibility (Joyce, Hersh, & McKibbin, 1983). Teddlie, Kirby, and Stringfield (1989) found marked differences in teacher behavior at effective schools compared with their lesser counterpart. Mackenzie (1983) noted the behavior of teachers in effective schools regarding the amount of intensity and engagement in school learning and providing well-structured classroom activities in which instruction was guided by content coverage. She also noted the school-wide emphasis on basic and higher order skills, opportunities for individualized work, the number and variety of

opportunities to learn, appropriate levels of difficulty for learning tasks, and collaboration with more accomplished peers. Teachers in effective schools also reported greater satisfaction in their work (Edmonds, 1979; Mackenzie, 1983).

Criticisms of the effective schools research. Purkey and Smith (1983) identified concerns over methodological practices in drawing causal inferences from the research: 1.) narrow and small samples used for intensive study; 2.) error in identification of outlier schools; 3.) aggregating achievement data at the school level; 4.) inappropriate comparisons; and 5.) subjective criteria for determining school success. Furthermore, they questioned whether it was altogether surprising that schools experience higher achievement in areas upon which their faculty and staffs have agreed to focus and emphasize.

Generalizations from the effective schools research are context-bound to snapshots of urban, reading/math, mostly lower grades, as no methodological systematic sampling of different types of schools were employed (Purkey & Smith, 1983). Furthermore, no longitudinal studies indicate long-term outcomes for life success beyond the scope of the K-12 schools' research (Purkey & Smith, 1983). Additionally, the surplus of variables are susceptible to interactive effects and therefore further context-dependent (Sweeney, 1982; Witte & Walsh, 1990). Therefore, limitations on the generalizability of the effective schools research due to varying contextual school-based factors such as socioeconomic status, school type (rural, urban, suburban), as well as grade configuration (elementary, secondary), have been widely noted (Teddle & Stringfield, 1993; Purkey & Smith, 1983; Rowan, Bossert, & Dwyer, 1983). Lezotte and Bancroft (1985) suggested beginning at the classroom level when considering variation in the school building, as school level analyses treat effects too broadly with no clear indicator for

organization and processes that impact student learning (Rowan, Bossert, & Dwyer, 1983; Sizemore, Broussard, & Harrigan, 1983).

Multiple researchers have identified the sparse data on how the operation in the school building made a difference (Sizemore, 1985; Lezotte & Passalacqua, 1978; Sweeney, 1982; Rowan, Bossert, & Dwyer, 1983; Sizemore, Broussard, & Harrigan, 1983) and Brookover et al. (1978) argued for future researchers to “consider variables which are descriptive of the leadership style, climate, and instructional strategies operating in the individual school building” (p. 285). Firestone and Herriott (1980) identified particular differences in the grade configuration context finding that elementary schools were more like rational bureaucracies, formally organized social structure with clearly defined patterns in which every series of actions was functionally related to the goals of the organization. High schools, however, were more like natural systems, functioning as a coherent whole (Rutter, 1983)—actions were not clearly related to goals and individual interests substituted for goals as the primary motivating force. In this context, high schools’ experiences reduced interdependence and dispersed control.

Implications for effective schools. Rowan, Bossert, and Dwyer (1983) boldly asserted that measures of effectiveness are unreliable and invalid because they ignore the variety of school goals and focus only on academic achievement measured by standardized achievement tests. This laser-sharp focus fails to incorporate other non-cognitive desirable school-based factors such as social and emotional development. Achievement data by way of standardized test scores, however, are limited in utility for describing effective schools and quality schooling (Murphy & Hallinger, 1988; Klitgaard & Hall, 1975; Sizemore, 1985). The basis for the effective schools research is that if the measure of success is high academic achievement, then some schools display evidence of greater success than others (Klitgaard & Hall, 1975). Sizemore

(1985) acknowledged a similar limitation in her findings on effective schools, describing education as a three-phase process of training, socialization, and enlightenment; her studies focused solely on the training element and not on the totality of quality education.

The decline in the effective schools research coincided with the publication of *A Nation at Risk* (1983), which glaringly asserted the far-reaching effects of education on society with implications for social welfare and national security. *A Nation at Risk* decried this focus on low-level basic skill outputs as the “rising tide of mediocrity that threatens our very future as a nation and a people” (National Commission on Excellence in Education [NCEE], 1983, p. 5). This report, in turn, spurred the standards-based reform movement, along with school improvement efforts that paved the way for alternative school programming, including a shift in governance structure (charter schools) and funding allocations (vouchers). Fifty years after its onslaught (Teddle & Stringfield, 2007), the effective schools research can be furthered by advancing a more comprehensive system of feedback, analysis, and evaluation on school quality, including indicators for academic, social/emotional, and behavioral outcomes for students.

Improving America’s Schools Act

Under the first reauthorization of ESEA, the *Improving America’s Schools Act* (IASA) of 1994 constituted a paradigm shift towards accountability for student learning by: requiring states to establish common statewide standards for all students in reading and mathematics in grades three through eight, and high school grades; encouraging states to implement statewide assessments aligned to these standards in at least three grades each for reading and mathematics; and implementing a statewide accountability system for evaluating school level performance (Forte, 2013). Spurred, in part, by *A Nation at Risk*, the IASA, followed another national education initiative: *Goals 2000: Educating America Act*. Goals 2000 provided federal grants for

states to develop standards, non- high-stakes assessments and accountability systems by creating structures under which systems develop (Goals 2000, Public Law 1804, 1994; Superfine, 2005). ISIA embedded assessments into the fabric of accountability systems across America prompting a paradigm shift in accountability for resource allocation to accountability for performance (McDonnell, 2012). This key shift was characterized by focusing more on student learning outcomes (i.e. outputs) rather than inputs. The focus of judging school quality had been practiced by accreditation; however, this shift aimed to focus on student learning as the judgment for school quality.

IASA provided fiscal supports through Title I for development of the systemic reforms. Hallmarks of this legislation included common, statewide high standards for student learning, high-quality teacher preparation, fiscal flexibility for local innovation with accountability, and school-family-community partnerships. IASA required standards and standards-based tests in reading and mathematics, at least once in each of the grade spans of 3-5, 6-8, and 9-12 for all students, not limited to those served under Title I (IASA, Public Law 103-382, 1994, U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2007). Although states were required under the law to submit an improvement plan, the program was wrought with a lack of compliance by states and little enforcement by the federal government (Superfine, 2005; Wanker & Christie, 2005). Between 1994 and 1999, the US saw a 58% increase in the number of states that adopted content standards, as required by IASA, for a total of 98% compliance (Superfine, 2005). Wanker and Christie (2005) noted full to partial compliance by all states by 2004, however, only 17 states submitted improvement plans and even fewer established accountability systems based on state assessments.

No Child Left Behind

The second reauthorization of ESEA, the *No Child Left Behind Act* of 2001 (*NCLB Act*) (Public Law 107-110, 2001), ushered in sweeping reforms to standards, assessment, and teacher effectiveness. It mandated that states implement assessment and accountability systems based on “challenging state academic standards” for student learning for all students. Three factors contributed to the “policy window” (Marshall & Gerstl-Pepin, 2005, p. 15) opening, which allowed for the passage of such sweeping reforms. The continuing problem of underperforming schools, i.e. “problem recognition” (Marshall & Gerstl-Pepin, 2005, p. 15), heightened by the publication of *A Nation at Risk* (Shepard, 2008), allowed for the operationalization of the “policy proposals” (Marshall & Gerstl-Pepin, 2005, p. 15), which federal and state legislation had begun years earlier through policies around standards, assessments, accountability, and school choice. Finally, the “politics” (Marshall & Gerstl-Pepin, 2005, p. 15) of the 2000 presidential election supported the convergence of each of the streams. Then- Governor George W. Bush, campaigned on a platform of education reform and was quoted as saying

...this nation of ours must challenge what we like to call the soft bigotry of low expectations. Every child can learn. It starts with raising people’s sights and raising expectations and refusing to yield, refusing to accept a curriculum that won’t work. (Rosenbaum, 2000, p. A14)

Within his first months in office, Bush garnered bipartisan support for the passage of NCLB (Linn, 2008).

NCLB further entrenched test-based accountability as a key policy lever by extending policies established by the IASA to include assessments in reading and mathematics for grades three through eight, adding requirements for standards and assessments in at least three grades for science, and establishing a specific set of rules for states’ accountability systems (Linn,

2008). NCLB mandated state testing of student achievement for accountability purposes and reporting of results by subgroup, along with school improvement goals determined by scientifically based research (Linn, Baker, & Betebenner 2002). NCLB was guided by the logic of clearly defining what students should know and be able to do as well as the level to which students should be able to demonstrate what they know and can do; standardized assessments are utilized to gather data on the extent to which the established content and performance standards have been met, and the results of assessments are then used to inform accountability decisions meant to improve school functioning and improve student achievement (Forte, 2013). NCLB mandated that states reach 100 percent proficiency rates in English language arts and mathematics by the year 2014. The primary metric for calculating school performance is the adequate yearly progress (AYP) target. The law also requires reporting of results by subgroup, such as race/ethnicity, education classification, and socioeconomic status, along with school improvement objectives determined by scientifically based research (Linn, Baker, & Betebenner 2002).

NCLB significantly entrenched test-based accountability systems by expanding the use of state assessments, requiring the setting of targets for attaining AYP, including a provision for high-stakes, identifying schools for improvement, providing parents with public school choice, mandating supplemental educational services, applying corrective actions, restructuring of continuously failing schools, requiring highly-qualified teachers, and using research-based practices (U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2007). NCLB extended IASA by adding supplementary educational services, restructuring, and highly-qualified teachers, as well as the requirement for participation by all public school students in state assessment systems (Goertz,

2005), and the requirement that states set targets for academic achievement that would lead to the target of 100 percent proficiency by 2014 (U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, 2007). Although standardized achievement testing programs were in place in some states prior to NCLB (Haertel & Herman, 2005), NCLB mandated this practice in all states across the US.

Prior to NCLB, several states had begun the work of reforming education under earlier versions of ESEA. Spurred in part by civil rights and efforts by the governor to ensure equal access to high-quality education, as early as 1979 North Carolina led the nation in accountability design in what was initially designed under minimum competency tests (Baker, 2015). Texas began reporting accountability results as early as 1994. Its systems for assessment and accountability served largely as a model for NCLB (Nelson, McGhee, Meno, & Slater, 2007). Louisiana began implementing its accountability system in 1997, and included accountability for student test scores, student attendance rates, and high school dropout rate in its system (Louisiana District and School Accountability Advisory Commission, 1998).

While key systemic advancements have been made for the education of students, such as including all students being included in the test-based accountability system, problems with implementation persisted (Wong, 2013). In 2014, no state had reported meeting expectations for proficiency established by NCLB. Under President Barack Obama, Secretary of Education Arne Duncan provided waivers to the requirement for 100 percent proficiency, based on alternate factors for accountability. Drawing on the NCLB corrective action framework, the Obama administration used financial incentives to mobilize state and local support for building an infrastructure for reform through grant programs like Race to the Top and Investing in Innovation (i3) to ‘transform’ current policy and practices centered around educator

accountability, charter schools, and turnaround school processes. In December 2010, through the awarding of School Improvement Grants (SIG) for 730 schools in 44 states, Secretary Arne Duncan's proposal for ESEA reauthorization, titled the Student Success Act, included strategies for sanctions due to failure under the accountability system including Turnaround Schools, Transformation Schools, Restarted Schools, or School Closure (Wong, 2013).

Non-funded state mandates, such as school choice, require the district level management of resources and are often time-consuming and repeated work across programs (Cohen & Spillane, 1992). With the threat of state take-over, turnaround, or reconstitution as a charter school (Linn, Baker, & Betebenner, 2002), district officials are responsible for compliance with federal and state regulations, along with ensuring secure testing administration amid increasing reports of compromised test administration. At times school performance accountability measures are utilized as a form of program evaluation (Herman & Baker, 2009) and implications for remediation strategies associated with response to intervention. Furthermore, Lee and Wong (2004) found the function of accountability policies to emphasize regulations over support, thereby limiting capacity building for implementation.

Advantages. NCLB emerged from the standards-based reform movement of the 1980s (Polikoff, 2012; Superfine, 2005) during a time when research on school effectiveness was burgeoning. NCLB focused attention and spending on the students most in need of support by requiring states to disaggregate achievement data by subgroup, allowing for targeted intervention, assistance, and education programming (Goertz, 2005; Linn, 2003). It extended fiscal flexibility for states to spend Title I dollars at the school level for those students served under Title I, rather than spending those dollars at the student level (U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and

Regional Assistance, 2007). NCLB streamlined the dual accountability systems that were in place for Title-I students and non-Title-I pre-IASA (Superfine, 2005) that were unable to be streamlined, despite effort, under IASA (Goertz, 2005).

Limitations. The impact of test-based accountability on student achievement varies by subgroup. While accountability policies generally have been shown to produce positive effects for higher achievement and narrow the achievement gap between white and Hispanic students, they have done little to narrow the achievement gap between white and black students (Hanushek & Raymond, 2006). Deming, Cohodes, Jennings, and Jencks (2016) found that, in Texas, low-scoring students, generally of minority and poor backgrounds, attending schools under pressure of receiving low-performing ratings, benefited more from test-based accountability than did low-scoring students attending higher-performing schools under pressure to attain recognition status. Moreover, the authors found “negative long-term impacts” (p. 72) for the latter, citing nuances in the variable of pressure faced by each respective school type.

Under NCLB, schools and students were subjected to high-stakes sanctions. Schools faced the loss of federal funding, could be required to restructure, or could be taken over by the state education agency. High-stakes sanctions for students included such practices as grade retention and program placement. Proponents of high-stakes consequences for performance on tests claim its virtue for motivating teachers to improvement and prioritizing the most important content to teach, while for students, the high-stakes aspect of tests promotes optimum performance and a sense of achievement from success (Amrein & Berliner, 2002). In 2000, Louisiana became the first state in the US to use test scores as high-stakes to determine student promotion (Amrein & Berliner, 2002). Allen et al. (2007) highlighted the risks of high-stakes

failure—those failing a grade one year are 50 percent less likely to graduate from high school and the percentage increases to 90 percent for those failing two years (p. 460).

Teacher opinion on the use of standardized tests in education is varied. Some teachers claim that NCLB does the opposite of what its name purports to do with its ever-entrenched measure of proficiency attainment (Allen et al., 2007, p. 460). Others claim that NCLB standardizes not only the expectation for achievement of all students, but the methods by which they come to learn through standardized teaching (Allen et al., 2007). Another claim is that NCLB encourages teachers to “game the test” (Allen et al., 2007, p. 457) as teachers strategize test preparation for students to perform at their highest ability (Kortez, 2008). In this way and others, “teachers are forced into complicity in harming the very lives they are dedicated to enriching” (Allen et al., 2007, p. 460).

Senechal (2013) highlighted the changing paradigm of education from process-oriented learning to product-oriented achievement. Allen et al. (2007) identified the false label of achievement as intended learning. Senechal (2013) described the responsibility of the creative artist to “delve” into their work as an integral part of the learning process; the author claimed that not only is this responsibility unaccounted for in accountability systems, but also that the very system is an impediment to the student’s cognitive development and progress as the system refuses to acknowledge the role of failure as an inherent component of developing excellence.

Another liability of NCLB is the inadequacy of AYP as a mean proficiency measure. AYP focuses on one narrow calculation and neglects important gains made by students who nevertheless fail to meet the proficiency standard (Kim & Sunderman, 2003; Linn, 2003). Variability in methods for setting cut scores for proficiency to make AYP across states contributes to its liability (Goertz, 2005). Finally, implementation challenges (Superfine, 2005)

and a lack of local capacity to sustain the required reforms (Goertz, 2005; Wanker & Christie, 2005) further exacerbate efforts to translate the law into policy and practice.

In reporting proficiency rates by promotional standards in Louisiana, the Louisiana Department of Education (LDOE) combined the achievement levels for proficiency in English language arts and mathematics. Proficiency rates on the Louisiana state tests indicate minimal growth in 4th and 8th grade achievement since 2010 after a decline in scores in 2009 (see Figure 2.1). Moreover, proficiency standards in Louisiana do not meet NAEP standards for proficiency in 4th grade reading. According to the National Center for Education Statistics (U.S. Department of Education, 2015b), a ‘basic’ achievement level in 4th grade English language arts in Louisiana maps to the ‘below basic’ achievement level on NAEP. This variance in state-determined levels of proficiency has been noted as a weakness of NCLB (Linn, Baker, & Betebenner, 2002) because it inhibits the comparison of educational effectiveness across states and does not reliably correlate with national measures of student achievement such as NAEP (Herman & Baker, 2009; Kress, Zechmann, & Schmitten, 2011). Linn, Baker, and Betebenner (2002) found that state-determined levels of proficiency vary greatly. In 2001, Louisiana reported proficiency rates in 8th grade mathematics at 7 percent, while Mississippi reported 39 percent proficient, and Texas reported 92 percent proficient.

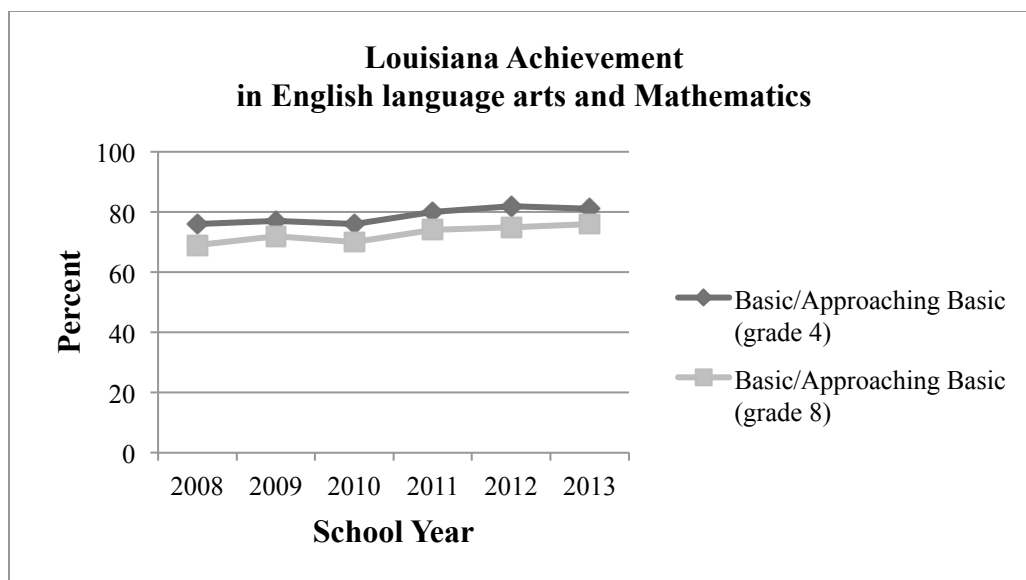


Figure 2.1 Louisiana achievement levels attained in English language arts and mathematics for grade 4 and 8. Number reported includes the percentage at basic or approaching basic achievement level, which is the promotional standard for Louisiana. Data for 2014 and years prior to 2008 are not publically available as of this printing. Adapted from Louisiana Department of Education Test Results <http://www.louisianabelieves.com/resources/library/test-results>

Adequate yearly progress. For publicly funded schools in the US, AYP was used as the central mechanism for improving school performance and academic achievement of all students. AYP was determined by student performance on the state assessment, along with attendance and dropout rates. It followed a conjunctive model for calculating reading and mathematics proficiency rates (Kim & Sunderman, 2005). Depending on the state, the specific number of variables required to meet standards under NCLB may have differed; however, the more the diverse the school was in its demographic makeup, the more hurdles the school likely had to meet to avoid the label of *In Need of Improvement*, the moniker of a school that did not meet one or more of its AYP targets in two consecutive years. If and when a school missed one or more of its AYP targets in each of two consecutive years, the school was been labeled as *In Need of Improvement*. For example, during school year 2006-2007, school A met its AYP target for all

subgroups. However, in school year 2007-2008, the school failed to meet target in reading for education classification (English language learners). In school year 2008-2009, the school met its target for education classification (English language learners), but failed to meet its target for race/ethnicity (Hispanic/Latino) students. Thus School A would be labeled as a school *In Need of Improvement* (Forte, 2013).

Most states use the *current-status* or *school-mean performance*, models of accountability. The successive-cohorts approach to current-status accountability tests the same students each year, successively, and their scores are compared as a group to subsequently tested groups. This type of system loses some credibility when students transfer outside the district and are unable to provide subsequent scores for comparison. Another approach to current-status accountability is longitudinal-tracking in which the same students are tested and their results are individually compared with the test results from the prior year, facilitating the analysis of growth and estimation of gains (Linn, 2008). Each approach operates under the assumption that school improvement efforts lead to increases in student learning (Forte, 2013). With no clear incentive for success, (Kim & Sunderman, 2005), state accountability under NCLB has failed to provide the level of student proficiency intended by the 2002 law.

Researchers have asserted that NCLB belies the accurate measure of school quality. Haney (2008) identified the weakness of a rating system for school quality based simply on reading and math test scores, as fundamental goals of public education extend beyond the teaching of reading and math. He argued that those who value summary judgments on school quality miss the paradox of value in economics in which useful commodities such as air and water have low exchange values, whereas less useful commodities such as diamonds and gold have higher exchange values. Kim and Sunderman (2005) asserted that if mean performance

remains the primary indicator for success, student achievement will not improve and further federally mandated sanctions will be disproportionality applied to minority students and their schools.

Kim and Sunderman (2005) further asserted that, despite the safe harbor provision, high-poverty and racially diverse schools were placed at a disadvantage for meeting the requirements of AYP under NCLB due to the expectations for subgroup performance, including the selection bias of proficiency scores. The AYP calculation did not account for initial differences among students due to background characteristics; instead it portrayed the process of selection bias in practice. Multiple subgroups classifications for individual students may increase the likelihood of failing to make AYP. Value-added modeling, an alternative approach to measuring proficiency based on growth targets, attempts to isolate factors that contribute to student learning.

Forte (2013) sought to investigate the validity of AYP for school improvement by first asking, “Does AYP identify the schools that actually need to improve and would benefit from state intervention to do so?” She found that the AYP algorithm involved an untenable number of conjunctive decisions and probably over- and mis- identifies schools as *In Need of Improvement*. The algorithm, she claimed, is an achievement, not an effectiveness metric, that compares school level results to pre-specified annual targets that increase regularly to equally increase the number of false positives. AYP developers attempted to triangulate data points, however, the resulting algorithm was simply an overall increase in the number of calculations run using the same data set. The results of this process now require a positive outcome in each of 5 to 37 separate comparisons. Forte (2013) further asserted the use of the percent proficient statistic in current-status models is a poor indicator of school quality and that efforts to improve school quality are more likely to succeed if the consequences, sanctions, resources, and support associated with

school improvement status are indeed improvement efforts appropriately assigned and effectively implemented.

Sanctions for failure to meet AYP. Consequences for failure to meet standards under NCLB include the application of sanctions including school choice transfer, supplementary education services (SES), and school reconstitution (Forte, 2013). During a school's first year in improvement status, the school was required to offer to its students the option to transfer out of the failing school to another school with higher achievement status within the local district. This school choice transfer option was included in the NCLB legislation as one of several political compromises that ensured passage of the legislation and was not included on the basis of evidence that it improves the quality of schools or students in achievement (Forte, 2013). Although school choice is the first line of sanction for school improvement, this option has not been widely used by parents of school children for whom this option was available (Forte, 2013; Zimmer, Gill, Razqui, Booker, & Lockwood, 2007; Fusarelli, 2007; DeBray-Pelot, 2007). Across nine large urban districts of those eligible to receive the school choice option in 2004-2005, participation rates at the elementary level were between 0.6 and 1.0%, and 0.2 and 0.4% and the high school level (Zimmer, et al., 2007). Out of the nearly seven million students who were eligible for school choice in the almost 11,000 schools in improvement status during school year 2006-2007, only one percent actually chose and attended a different school (Fusarelli, 2007).

One factor contributing to the dismal participation rate in school choice was the timing of execution (DeBray-Pelot, 2007; Fusarelli, 2007; Jimerson, 2005; Wong, 2013). Schools were notified in the fall of eligibility and parents received notification of eligibility within days of the

start of school. Another factor contributing to low participation rates may have been resistance in implementation by local authorities (Fusarelli, 2007; Jimerson, 2005; Wong, 2013).

The second applied sanction, supplementary education services (SES), was required to be provided by the failing school during the second year of a school labeled as under improvement status. SES was more widely used than the school choice option, with Fusarelli (2007) noting a participation rate of 17%. Zimmer et al. (2007) noted participation rates at more than 20% at the elementary level and less than 5% at high school level. SES may constitute the most legitimate improvement effort, however, it received the least amount of press (Forte, 2013). SES could be offered through the local education agency (LEA), or through a private, for-profit entity. Approval as a SES organization must be granted by the USDOE. One complication with implementation of SES is lack of service providers. No federal funds were provided for the administrative costs of the program. Other factors include logistics such as parental access to transportation and after-school scheduled activities (Fusarelli, 2007). The effectiveness of this programming was further limited by an imbalance in the components necessary for supporting meaningful change, namely the role of the SEA in supporting the local school improvement plan for SES. Alabama, Kentucky, Ohio, Tennessee, and Washington, however, have mandated the supporting role of the SEA for school improvement through sophisticated coordinated systems of support (Forte, 2013).

The third and final stage of applied sanctions was school reconstitution or school takeover. Reconstitution reforms work under the assumption that teachers and administrators are responsible for school performance, and do not account for additional variables that affect student learning (Malen, Croninger, Muncey, & Redmond-Jones, 2002).

Jimerson (2005) claimed that NCLB sets up rural schools for failure and that educational opportunities of students attending rural schools would decrease under NCLB, which would defeat the purpose of the law (i.e. closing the achievement gap). In his 2005 study, Jimerson identified some of the factors contributing to challenges of implementation of applied sanction in rural areas including seven primary obstacles to implementing NCLB: 1) rural districts tend to be small, especially in Vermont and Montana; 2) rural schools in many locations are poor and often have large concentrations of minority students such as in Louisiana and Mississippi; 3) many rural districts are in financial distress including Nebraska, Georgia, and North Carolina; 4) rural schools in many states are situated in remote areas; 5) there is a strong tradition of local control in many rural areas; 6) many rural areas are experiencing depopulation and declining enrollment; and 7) other rural areas are experiencing rapid population increases and rapid ethnic diversification (pp. 212-213).

Small population size was of greatest concern in calculating AYP, as small sample sizes (N) may produce drastic changes over time due to slight variations in one or more students. Furthermore, most states established a minimum number of approximately 40 students for AYP to be calculated for subgroups. This N is significantly lower than the suggested N of around 150 for subgroup cell size. This calculation error could lead to false positive, as well as false negative conclusions for status under NCLB.

Jimerson (2005) used the term “placism” to describe the discrimination against people based on where they live and its associated limitation of the provision for ‘highly qualified’ teachers as mandated by NCLB. Challenges in the implementation of NCLB in rural areas include difficulty in staffing schools because teacher salaries are an average of 13% average lower salary than salaries for teachers in non-rural areas. In addition, the ‘specialist’ mandate is

misaligned with the often-necessary assignment to multiple-subjects for teachers in rural areas. Furthermore, he claimed the emphasis on teacher testing to predict the probability of a teacher's success in teaching does not support longevity in staffing decisions for teachers in rural areas where researchers have identified a deep understanding of local culture to be a contributing factor to teacher retention and commitment.

Although rural flexibility was issued by the USDOE to aid rural areas with implementation of NCLB 26 months after its enactment, only 26 percent of rural and small town districts in the country met the criteria for definition of 'rural' by these flexibility guidelines and 75 percent of rural schools ineligible for this flexibility are located in the South. Additionally, intersections of placism with racism and classism are demonstrated through the implications of this narrow definition of 'rural' for the flexibility waiver.

Using actor-network theory (ANT) as a conceptual framework, Koyama (2012) studied actors via interviews and ethnographic research to "to examine what happens to school failure under NCLB when actors set about eliminating [the label of school failure] in NYC" (p. 876) for one particular school under study: PS 100. Actors included public systems, for-profit educational companies, and authorized policymaking institutions.

By all accounts, even those established by NCLB, PS 100 was a successful school. However, due to miscalculation by the state education agency, PS 100 was labeled a School In Need of Improvement (SINI) and subject to sanctions under NCLB law. Two subgroups (SPED and ELL) were administered the end-of-year assessment separately from other students in the grade level and received additional provisions. "Though the students' scores were well above benchmark set for their subgroups (greater than 52% were at or above proficiency), they were not added to the school's AYP calculations for the ELA test—and thus the school failed to meet

one of the testing requirement [sic] of NCLB” (Koyama, 2012, p. 879). Once the error was identified, the school administration sought correction; however, “the test scores had already coalesced into fact” (p. 879).

Following the miscalculation that placed PS 100 under NCLB sanctions for failing schools, the school was turned into an Empowerment School, in which greater autonomy was coupled with increased accountability (typically inclusive of performance assessments). School administrators struggled to reconcile the failure identification with promotion in standing due to success rate. This conflict imposed confusion on teachers, in which intentions for reporting were often unknown, and even scaled to unimportant. As one teacher said, “...just put what goes on here wherever they think it fits with the story they’re trying to tell that day of whatever” (Koyama, 2012, p. 885). More importantly, as one parent noted “Failure is the Scarlet S [or perhaps ‘F’]. Once you get it, you wear it and can’t get away from it” (p. 885). The school complied with the appropriate mandates for SES, although less advertised than before the miscalculation became known. Teachers adjusted instructional practices and supplemental services were included as extensions to the school day. One actor noted the “real [issues] only in the paper trail, not in the classrooms” (Koyama, 2012, p. 883).

Koyama (2012) concluded “what unfolded at PS 100 points to the arbitrariness of the measures we now rely on to signify success and failure” (p. 886). While this school actually “succeeded” according to NCLB, it was designated as failing. Failure, though fabricated, was made real by legitimization of designation and heeding mandates of the sanctions. This examination contributed to the assessment literature in three ways: 1) demonstrating how standardized testing and the data it generates become somewhat sturdy “as a consequence of the relations in which they are located and performed” (Fenwick & Edwards, 2010, p. 86), 2)

revealing heterogeneous interpretations that are possible in quantitative calculations and numeracy, and 3) illuminating how standards become translated into tests and scores, which in turn are translated in AYP measures—and which are ultimately translated into a success.

Malen et al.'s (2002) study focused on the impact of school reconstitution as an applied sanction for failure to meet standards under NCLB. The authors aimed to “identify and inspect the underlying premises of the reform and offer provisional, ‘analytic generalizations’” (p. 117) of district-mandated reconstitution practices. This study was characterized by the distinction between the “‘espoused theories’ and ‘theories in use’ to compare the ‘official version of how the program or organization operates’ with ‘what really happens’” (Malen et al., p. 114). The authors noted their decision to sacrifice some specificity in numerical profiles to prioritize narrative description, thereby limiting the generalizability of the results about the viability of reconstitution reforms that may be designed and/or implemented differently.

In this case, reconstitution was a final attempt to make drastic changes in student achievement due to increased pressure by state officials and fear of loss of local governance of the schools to the state department (Malen et al., 2002). The superintendent announced plans to reconstitute in May of the school year prior to reconstitution, allowing for three months of preparing and planning which was characterized by teachers as chaotic, confusing, and stressful. The decision to reconstitute the school was based on two primary assumptions: that reconstitution would meet the immediate goal of attracting more capable and committed faculty and staff to advance the immediate aim of redesigning schools, and that the redesigned schools would realize the ultimate aim of improving student achievement.

The study highlighted some of the complexities of implementation. The reconstitution carried with it the loss of reputedly effective and experienced teachers. The school experienced the greatest loss in its teaching force during the year prior to reconstitution, some loss during year one of reconstitution, and more loss during year two of reconstitution. Reconstitution failed to create the cadres of master teachers that were envisioned during the design of the reconstitution, but rather resulted in an influx of new, inexperienced teachers and administrators, causing disruption and working in survival mode (Malen et al., 2002). “On every critical count, the dominant patterns of implementation we discovered in the three schools we studied in depth ran counter to the major premises (and promises) of the policy” (p. 119). The central impact of reconstitution was change in school culture, especially trust and commitment. Secondary to the impact on school culture was the impact on teacher quality and the realized imperative of supporting teacher professional development.

Some interviewees commented that they spent the entire year of reconstitution trying to get back to where they were before the reconstitution, which led to marginal adjustments in school design, rather than the intended reconstitution. The hope of transformation was lost—schools cited using the same curriculum and pedagogies as before reconstitution. Interviewees also described the process of reconstitution as characterized by unfulfilled promises such as smaller class sizes, more master teachers, and additional support and resources. When not accompanied by resources, especially support and time to make the required changes, reconstitution may make it even more difficult for ‘unsuccessful’ or ‘failing’ schools to make gains in student achievement through structural and systemic reform (Malen et al., 2002). In these cases, the authors suggested that personnel changes might not be sufficient to overhaul a school. Malen et al. (2002) identified the intermediate

aim to redesigning the school and also emphasized the difference between redesigning schools and restaffing schools in further understanding the relationship of reconstitution initiatives and institutional supports, to better understand the factors affecting school turnaround.

Another study investigated organizational learning as one district attempted to improve student outcomes under accountability policy sanctions using a theoretical framework of social processing of knowledge and organizational learning. Following a mixed methods approach, Finnigan and Daly (2012) examined the internal conditions of schools under sanctions and the larger district context. Data were collected through surveys at the school and district levels interviews, observation, and document review. These data were analyzed to develop in-depth case studies of the schools and the district. Quantitative data analyses included social network data.

The authors found limited evidence of the technical aspects of learning in schools under sanction. Schools were more likely to revisit previous practices, rather than identify and define underlying assumptions or developing a formal process for evaluating programs. Schools under sanction were more likely to utilize only single-feedback loops. Qualitative data suggested that the pressures of high-stakes accountability policies contributed to negative aspects in these schools climates. They also found variability in perceptions and low levels of trust among district leaders.

Finnigan and Daly (2012) concluded the lack of organizational learning contributed to the applied sanction and term of sanction. “Our findings in combination suggest that organizational learning in these highly turbulent contexts is extremely challenging given both the lack of technical aspects of organizational learning and limited exchanges around

expertise” (p. 65). Single-loop learning was most likely the cause for routine and/or incremental changes, where applicable. Accountability policies target technical aspects of teaching and learning, but ignore the aspects of learning as a social process (Finnigan & Daly, 2012).

DeBray-Pelot (2007) highlighted the conceptual conflict between local education priorities in light of increasingly high stakes federal education mandates by investigating the challenge of offering school choice transfers when the transfers disrupt unitary status for racial integration under federal law. In such a case in Colorado, the USDOE responded that court ordered desegregation was not a sufficient reason for not providing school choice. Federal district judges found that the constitution trumps the statute, with neither modifying the standing court orders for desegregation, and cited the provisions about choice and desegregation that are not only contradictory with principles of constitutional law, but also with other sections of the law that affirm desegregation as a continuing federal goal. Two key implications emerged from this case study: that the intersection among performance-based accountability funding disparities and the end of court-ordered desegregation, and that local school board attorneys have broadly acceptable options as they navigate the complex terrain of implementing federal law. While “this tension between the federal push for innovation and local realities is likely to persist” (Wong, 2013, p. 414), regulations notwithstanding, the USDOE has gradually recognized and respected local school systems’ need to maintain the terms of their court orders, highlighting the practice of trade-offs in the realization of policies (DeBray-Pelot, 2007).

Roda and Wells (2013) studied the extent to which policies written without consideration for effects on racial balance in schools, e.g. colorblind policies, impacted

stratification by race and ethnicity. Since, generally, white, economically advantaged parents are more likely to enroll their children in the highest-status schools regardless of the school choice policies in place, they investigated how these parents interacted with ‘colorblind’ school choice policies and whether they would support changes to the policies that would lead to less segregation across schools. Furthermore, they sought to “examine the contradictions between what advantaged parents say and what they do when confronted with segregated schools and school programs” (p. 266).

Roda and Well (2013) found that while white families with higher-socioeconomic status claim to value diversity and interest in schooling for their children within diverse environments, these parents often selected schools in which their privilege was protected, in which white students remained in the majority, and their decisions further exacerbated the problem of racial segregation under colorblind choice policies. Parents chose schools that characterized their common position in the social hierarchy as they may have feared downward mobility if their children did not have the ‘right’ educational credentials. The authors, however, pinpointed the insufficiencies in the writing, regulation, and implementation of the colorblind policies in effecting racial stratification, indicating that the inadequacies have strong implications for the kinds of outcomes that may foster diversity and the value of public education. “Thus, the ‘problem’ is in great part due to the lack of policies, opportunities, and choices provided by the district and the larger New York City School System” (Roda & Well, 2013, p. 284).

Zimmer et al. (2007) conducted a national longitudinal study of state and local implementation of NCLB, specifically school choice, SES, and student achievement. They found participation in these programs came largely from the population targeted by NCLB

policies and that participation rates for both state and local programs were the highest at the elementary level. They also found the highest participation rates among African Americans in the school choice program and Hispanic population participation rates were higher than white in SES. Students classified as special education or limited English proficient showed higher participation rates in SES than school choice. Students with lower prior achievement levels had higher participation rates in SES than students with lower prior achievement levels eligible for services. Students participating in school choice transferred to higher performing schools and generally to more racially balanced schools. Fusarelli (2007) found that, of those eligible, higher performing students were more likely to take advantage of the school choice option.

Zimmer et al. (2007) found that across seven districts, on average, participation in SES resulted in greater academic gains in reading and math, with students participating for multiple years experiencing the greatest gains. Across six districts, however, there was no statistically significant effect, either positive or negative, found for participants in school choice.

Chakrabarti and Schwartz (2013) conducted a study analyzing the responses of public schools to the Florida Opportunity Scholarship Program (FOSP). FOSP was designed as a sanction that would be applied to schools at which students consistently performed poorly on the state high stakes test. The FOSP sanction allowed for the transfer of students from publicly funded failing schools to private schools on a voucher system, thereby potentially reducing revenues for failing schools. The purpose of the study was to investigate potential adverse incentives for misclassifying students to individual subgroups to garner results and better achievement outcomes.

Using a regression discontinuity approach and an alternate regression-discontinuity strategy, Chakrabarti and Schwartz (2013) analyzed schools that barely avoided the threat of the FOSP sanction with those schools that barely received the classification and associated FOSP sanction. The authors investigated two hypotheses to the research question: “Did the exemptions for certain limited English proficient (LEP) and special education (SPED) students induce schools to classify some weaker students into these categories to remove them from school-grade calculations and artificially boost scores” (pp. 20-21)?

Chakrabarti and Schwartz (2013) found a higher classification rate of students into excluded categories, such as LEP, for schools under the threat of FOSP in the high stakes grade and entry grade. There was no statistically significant finding of classification into the other excluded category of SPED, possibly because that classification carries with it additional threat of student transfer to schools with more robust services and programming, as well as the costs associated with providing services to those classified students remaining in the school’s enrollment. Similar analyses were run for schools barely avoiding the threat of FOSP, to which no such findings applied.

Florida’s system did not include clear incentives, only the threat of sanctions with incentives seen as the absence of sanctions. New York, however, had a system by which school leaders were eligible for monetary rewards for student achievement. Also, the New York system included student scores for LEP and SPED, awarding schools with additional credit for those LEP and SPED classified students’ successes.

Similarly, Haney (2008) found practices invoking the misrepresentation of achievement under NCLB in which the increases in achievement correlated with the increase in grade retentions, especially for minority students, as well as the removal of certain students such as

dropouts from the tested cohort. He concluded that reporting on test results alone has contributed to cases of fraud and administrator ‘push-out’, in which schools force students out of the system in an effort to increase test results

In studying the effects of failure to meet AYP under NCLB, Hemelt (2008) concluded that academic performance suffers in the short run in response to school-wide failure, as compared with subgroup failure. He found that schools that failed to meet AYP targets fostered improvements in short-run student performance and concluded that under NCLB, the scope of failure matters.

Policymakers must use caution when designing systems that include exemptions, special allowances, and/or credits for certain groups of students and the corresponding adverse incentives and unintended consequences. Public reports of poor performance may lead to incoherent stabs at change and/or may demoralize the culture (Forte, 2013). The challenge facing education policymakers is acknowledging and accounting for educating special populations of students who require additional services provided by specially trained educators, especially under circumstances where resources are sparse (Chakrabarti & Schwartz, 2013).

Limitations to sanctions applied for failure to meet AYP. The most noticeable and prominent limitation to the effectiveness of applied sanctions for failure to meet standards under NCLB is the lack of empirical evidence to support the effectiveness of each stage of applied sanctions (school choice, SES, school reconstitution). NCLB law as it was implemented utilized strategies that were not proven to positively impact student achievement (Kim & Sunderman, 2003; Fusarelli, 2007; Malen et al., 2002; Wong, 2013; Forte, 2013; Jimerson, 2005; Hemelt, 2011). “There is little to no evidence that the three most visible

sanctions NCLB imposes on schools in improvement status are effective and some possibility that they may be conceptually unrelated to the notion of enhanced school functionality” (Forte, 2013, p. 82). Malen et al. (2002) also noted the lack of evidence to support the sanction of school reconstitution by noting it as a “prevalent but understudied strategy” (p. 113). “Whether these consequences are actually inducing meaningful changes in schools and students to perform at higher levels remains an open question” (Hemelt, 2011, p. 706).

Another factor limiting the effectiveness of applied sanction is the lack of resources to support implementation. ESEA initially included funding under Title VI to support the capacity building in the SEA, however, it has been removed during reauthorizations (Forte, 2013). One complication that arose with the use of the school choice sanction was ensuring enough receiving schools were amiable and matched to the needs of those receiving the transferred students. The investment by policymakers more fully and directly in the preparation of preservice teachers and the professional development of practicing teachers may strengthen teacher quality and improve the quality of instruction provided to students (Malen et al., 2002). New accountability systems expose the tremendous capacity needs and achievement gap in mid- and small-sized urban communities as illustrated by cases in Michigan, Rhode Island, and New Jersey (Wong, 2013).

Additionally, some researchers worried that school choice could lead to greater ethnic stratification and possibly further segregate student populations in schools (Fusarelli, 2007; Carlson, 2014; Roda & Wells, 2013). Educational leaders and researchers have questioned the practical progression of sanctions, namely the rationale for placing school choice transfer as the front-line sanction, rather than SES. A practical remedy would include switching the order in which the sanctions apply to SES in year one and the school choice transfer option in year two (Fusarelli, 2007). Additional considerations for the improvement of the school

choice and SES sanctions include restricting the school choice option to individual students who are failing in the failing school, and disallowing test scores in calculations for school performance (AYP) of students who transferred to another school under school choice until the student had been under the tutelage of the school for two years; SES providers should contribute one half of one percent of their SES to cover a portion of the administrative costs to speed up turnaround in state test scoring or revamping testing cycles. USDOE requires states collect better data about student use of transfers and SES, including funding for SES in the next reauthorization of ESEA, and more funding and greater flexibility in spending for LEAs in implementing SES (Fusarelli, 2007).

Implications of NCLB for School Improvement

Jaiaini and Whitford (2011) noted that during the George W. Bush administration, accountability as a policy frame occurred more frequently than those of equity or fairness (p. 10). This divergence from the original goal of ESEA signifies an important shift in the paradigm for this law. Using state assessment results and results of the NAEP, the Council of Chief State School Officers (CCSSO) found that student achievement for economically disadvantaged students increased during the time of NCLB implementation; however, they also found student achievement for non- economically disadvantaged students increased during the same time. This suggests that the intent to narrow the achievement gap has not been realized (Allen et. al, 2007). However, Reardon, Greenberg, Kalogrides, Shores, and Valentino (2012) identified a narrowing of the achievement gap between minority-majority relationships, although the patterns do not suggest a strong effect of NCLB on that narrowing of the gap (p. 4). One teacher expressed her concern for the failure of NCLB to improve student learning by decrying the notion that external accountability is a silver bullet.

The enforcers of NCLB policy are not finding new answers to how to teach poor children because they are not looking for them. They thought they knew; they thought by forcing teachers into uniform methods and children into particular textbooks they were being scientific. They thought that threats and punishment would make those poor, lazy children work harder. They thought that by cementing the pole-vaulting bar into the standards that we would all become winning athletes (Allen et. al, 2007, p. 460).

Test-Based Accountability: Cognitive Domain

History of Educational Testing in the US

Academic assessments in the cognitive domain purport to accomplish one of the most complicated tasks in education: to identify learning, i.e. applied cognition, in the brain. As described by the NRC (2001), “assessment is a process of reasoning from evidence. Because one cannot directly perceive students’ mental processes, one must rely on less direct methods to make judgments about what they know” (p. 53). Learning occurs in the innermost confines of the human brain, and it is only by its applied behaviors that we can collect evidence to support a conjecture that learning has occurred. Thorndike (1918) observed that “education is concerned with changes in human beings” (p. 16) “and its effectiveness could be judged by differences in student behavior” (Haertel & Herman, 2005, p. 4).

The design of academic assessments varies, including teacher-created, mass-production by a testing company, or a large-scale assessment system. Results from achievement tests may be used to inform decision-making about students, teachers, school, and programs (Miller et al., 2013; National Research Council, 2001; Salvia & Ysseldyke, 1995). In the cognitive domain, the most prominent designs include diagnostic, formative, summative, and placement (Miller et al., 2013, p. 37), the results of which “are used for a wide array of purposes, ranging from low-stakes diagnosis for instructional purposes to high-stakes such as the award of high school diploma” (Baker & Linn, 2004, p. 50).

The key factor in tests designed to inform classroom instruction that most directly impacts student learning is the actual use of feedback (Haertel & Herman, 2005; Miller et al., 2013). Accordingly, diagnostic and formative assessments are designed to provide information that can directly impact teaching and learning. Alternatively, accountability tests, i.e. “...uses of tests for externally mandated accountability purposes” (Baker & Linn, 2004, p. 51), are described by Miller et al. (2013) as “barometer[s] and lever[s] of reform” (p. 3) (see Chapter 1).

A cursory understanding of the history of assessments and their use in American education systems may be helpful to fully understand the implications and contexts for which the test results may be applied. Where evidence points to early uses of testing for educational purposes as early as 1845 (Shepard, 2008), Ayres (1918) attributed “the real beginning of the scientific measurement of educational products” (p. 3) to “the publication of the Thorndike Scale for the Measurement of Merit in Handwriting in 1910” (Haertel & Herman, 2005, p. 3). Over the next 20 years, an array of tests, including power, speed, intelligence, achievement, and aptitude, were developed and implemented across the nation (Haertel & Herman, 2005). Additionally, school systems began using “norm-referenced” (Haertel & Herman, 2005, p. 5) interpretations of the results and the results were used as an evaluation tool for school success, as early as 1912 and as a tool for ability grouping and “tracking” (p. 5) by 1926.

IQ & Objective-Based Tests. The development of achievement tests led by Thorndike ran concurrently with the development of intelligence tests led by L. M. Terman (Haertel & Herman, 2005; Shepard, 2008). Ralph Tyler influenced the next wave of achievement test development, characterized by “objective-based assessments” (Haertel & Herman, 2005, p. 6), during the 1930s through 1940s, and he later played a key role in the development of the National Assessment of Education Progress (NAEP) (Carr, 2004; Shepard, 2008). Bloom’s

(1956) work advancing behavior objectives influenced test development through the 1970s by “measurement-driven instruction” (Haertel & Herman, 2005, p. 7), in which the interpretation of results by “criterion-reference” was formalized by Glaser (Haertel & Herman, 2005, p. 9). This development coincided with the development and first administration of NAEP, seen by some as a “policy instrument” (Shepard, 2008, p. 32) for its wide use in education policy debates in the national arena.

Minimum-Competency Tests. During the 1970s and 1980s, policymakers “disillusioned” (Shepard, 2008, p. 13) by the consistent achievement gaps, steady decline of test scores, and rising youth unemployment rates (Resnick, 1980), shifted their attention from inputs of school quality to outputs of student learning with a “back-to-basics” (p. 13) mentality achieved through Minimum Competency Tests (MCT) (Haertel & Herman, 2005). “In a single decade (1973-1983), the number of states with some form of MCT requirement went from 2 to 34” (Miller et al., 2013, p. 4). These tests, admittedly, assessed low levels of learning and not higher-order complex thinking skills (Haertel & Herman, 2005). *A Nation at Risk* (1983) “recommended the use of tests as instruments to improve education through their use to a.) certify the student’s credentials, b.) identify the need for remedial intervention, and c.) identify the opportunity for advanced work” (NCEE, 1983, p. 28).

Standards-Based Reform. Emerging from research on the Effective Schools movement of the 1970s and 1980s, the Standards-Based Reform movement spurred the use of performance assessments, which were characterized by their design for students to construct original responses to authentic tasks and the elimination of multiple choice item types (Haertel & Herman, 2005; Miller et al., 2013). Although Standards-Based Reform led to increased alignment with the state curriculum and assessment (Carnoy & Loeb, 2002), there was notable

concern for the ability to build capacity. Some states, though, made strides in standards-based assessment practices. Connecticut, Maine, Maryland, New York, and Vermont locally administered and scored state-developed performance assessments (Conley & Darling-Hamming, 2013). Minnesota, Oregon, Wisconsin, and Wyoming administered locally developed performance assessments that required students to demonstrate proficiency by producing original and authentic work (Conley & Darling-Hammond, 2013).

High-Stakes Accountability. The final wave of educational testing reform may be characterized by the use of high-stakes mandated by NCLB, characterized by serious consequences applied to test results, such as retention in grade level or requirement for graduation.

Test Design

The design and construction of the test determines the ways in which the results may be applied (Haertel & Herman, 2005). As Dunbar (2008) noted, “the *Standards for Educational and Psychological Testing* explicitly recognize that best practice in test development is defined by the assessment context” (p. 266). The NRC report (2001), *Knowing What Students Know*, indicated that “the contrast between classroom and large-scale assessments arises from the different purposes they serve and contexts in which they are used” (p. 8), and that “large-scale, standardized assessments can communicate across time and place, but by so constraining the content and timeliness of the message that they often have limited utility in the classroom” (p. 8). Koretz (2008a) echoed this difficult task of developers by claiming that assessment design always includes “trade-offs” (p. 2; NRC, 2001, p. 8). While the field of test measurement is driven by a complex array of topics such as test specifications, item response theory, standards-setting for cut-scores, technical reports, and interpretive guides (Koretz, 2008a), “it is only a

slight exaggeration to describe the test theory that dominates educational measurement today as the application of 20th century statistics to 19th century psychology” (Mislevy, 1994, p. 19).

Assessment design is dependent upon the purpose for the assessment and the intended use of the results (NRC, 2001). In this vein, Pellegrino (2002) advised, “when we try to design an all-purpose assessment, what we get is something that doesn’t adequately meet any specific purpose” (p. 50). Koretz (2008a) addressed the notion of “trade-offs” in assessment design (p. 2; National Research Council, 2001, p. 8) in which “large-scale, standardized assessments can communicate across time and place, but by so constraining the content and timeliness of the message that they often have limited utility in the classroom” (Koretz, 2008a, p. 8). Standardized accountability tests do not test all skills required by the established standards of the specific test instrument (Conley & Darling-Hammond, 2013; Haertel & Herman, 2005; Rothstein, 2004). “Tests almost always are made up of fewer items than the number actually needed to thoroughly assess the entire domain that is of interest” (Amrein & Berliner, 2002, p. 15). Even the most well designed test is fallible and provides limited amounts of data for professional interpretation on student learning (Koretz, 2008a; Miller et al., 2013). Results are reported in imperfect terms and claims made based on these test scores reflect an inference of scientific measure (Herman, 2004; Miller et al., 2013; Rothstein, 2004). Glasser (1990) asserted, “nothing of high quality, including schoolwork, can be measured by such standard, machine-scored tests” (p. 9).

Accountability tests measure mostly lower levels of knowledge in the cognitive domain (Glasser, 1990; Miller et al., 2013; Rothstein, 2004) and do not reflect higher levels of cognitive processes such as problem solving. Turnipseed and Darling-Hammond (2015) described the detrimental effect of emphasizing low-level cognitive functions as “discouraging critical and creative thinking” (p. 2), citing research on the decreasing levels of measured creativity from

ages 5 to 25. The current design of accountability tests does not foster the development of “metacognitive skills” (Turnipseed & Darling-Hammond, 2015, p. 4), skills attributed to experts in a field and allow learners to think about their own learning to support transfer of learning in solving new problems (Marion & Leather, 2015) which characterize deeper learning as “deep understanding,” or “expert knowledge” (Turnipseed & Darling-Hammond, 2015, p. 5) of content.

Alternatively, formative assessment that is described as a “process” (Popham, 2013, p. 296) is known for its quality to inform teaching and learning according to the “function it serves” (Black & Wiliam, 2004, p. 3). The classification of “formative” to an assessment has been applied when “the evidence is actually used to adapt the teaching work to meet learning needs” (Black, Harrison, Lee, Marshall, & Wiliam, 2003, p. i). This form of assessment is characterized by the provision of feedback to teachers and students to inform teaching and learning (Conley & Darling-Hammond, 2013). Feedback, or “information provided by an agent, i.e., teacher, peer, book, parent, or one’s own experience about aspects of one’s performance or understanding” (Hattie, 2009, p. 174) has been found to result in positive gains on student learning (Black & Wiliam, 1998). Hattie’s 2009 summarization of 23 meta-analyses on the effects of feedback yielded a 0.73 average effect size, analogous to an increase from the 50th to the 77th percentile on a standardized test. Similarly, Black and Wiliam (1998) found the effects of good formative assessment ranging from 0.40 to 0.70 standard deviations, which is similar to an increase from the 50th to the 65th or 75th percentile on a standardized test. Black and Wiliam (1998) were firm in the interpretation of their analyses: “The research reported here shows conclusively that formative assessment does improve student learning” (p. 61). However, Kingston and Nash

(2011) found that when feedback was less constructive for student use for remediation, the effects of feedback dropped to 0.25.

Validity

“Validity is always specific to some particular use or interpretation for a specific population of test takers” (Miller et al., 2013, p. 73). Although the process of test construction involves a complex array of mathematical analyses (Koretz, 2008a), tests are still susceptible to measurement error (Koretz, 2008b). Current state accountability test designs are more closely aligned with the summative evaluation of learning. Researchers have documented the trend of states to use the same test for multiple purposes, thereby invalidating the test for any inappropriate use (McDonnell, 2005; Popham, 2013). For example, states using a summative test design to measure student learning diagnostically would be using an invalid measure to make diagnostic decisions about student learning.

Challenges. Inappropriate teacher behaviors in test preparation and cheating have contributed to validity concerns for testing systems (Koretz, 2004, 2008a, 2008b; Linn, 2008), including narrowing the curriculum to tested content (Herman, 2008; Shepard, 2008). Citing Shepard (1997), Koretz (2008b) noted that some content was not included in tests due to its non-tested status which in turn impacted course enrollment by students (p. 8). Another challenge of validity to state tests includes score inflation, defined as “a gain in scores that substantially overstates the improvement in learning it implies” (Koretz, 2004, p. 99), due, in part, to previously mentioned teacher behaviors. Even the most well designed test is fallible and provides limited amounts of data for professional interpretation on student learning (Koretz, 2008a; Miller et al., 2013).

Unintended Consequences

Furthermore, a series of unintended consequences have contributed to the problem of using accountability tests for high-stakes purposes including: narrowing the curriculum to teach to the test (Amrein & Berliner, 2002, p.17); adversely focusing only on those students who promise the greatest return on investment, i.e. “bubble kids” (Booher-Jennings, 2005, p. 231); over- and under-classification of students receiving special education services (Deming et al., 2016); adverse effects on students’ motivation to learn; and outright cheating by students and adults. Additional behaviors have been noted as ‘gaming the system’: misclassification as SPED; over-diagnosis with ADHD; adjusting discipline policies to restrict student participation in tests; adjusting meal programs to ensure adequate nutrition; and teacher grade level placement according to associate strength(s)/weakness(es) (Figlio & Ladd, 2015). Amrein and Berliner (2002) questioned the validity of test results when teachers “teach to the test” (p. 17). “The harder teachers work to directly prepare students for a high-stakes test, the less likely the test will be valid for the purposes it was intended” (Amrein & Berliner, 2002, p. 17).

Non-Cognitive Domains for Learning

Affective Domain for Learning

A recent empirical study associated with the National Center for Scaling Up Effective Schools (NCSU) found that “instructional quality is not the defining feature of highly effective schools” (Rutledge, Cohen-Vogel, Osborne-Lampkin, & Roberts, 2015, p. 1062). Rutledge et al. (2015) conducted an inductive study of four high schools, two lower- and two higher-performing, and found that the more effective schools incorporated a process they called *Personalization for Academic and Social Emotional Learning* (PASL), which is a “systemic and intentional recognition of and attention to the interdependency of the instructional core and

social emotional activities” (p. 1062). Conceptually, the study was framed around eight components of effective schools (pp. 1063-1065):

1. Quality instruction
2. Rigorous and aligned curriculum
3. Personalized learning connections
4. Culture of learning and professional behavior
5. Connections to external communities
6. Learning-centered leadership
7. Systemic use of data
8. Systemic performance accountability.

Defining a construct. Egalite et al. (2015) define non-cognitive skills as “a set of behaviors, attitudes, and strategies that have been shown to be associated with individual success. It incorporates constructs such as optimism, resilience, adaptability, and conscientiousness” (p. 2).

The Collaborative for Academic, Social, and Emotional Learning (CASEL) at the University of Illinois at Chicago conducts on-going research around SEL and is currently facilitating the Collaborating Districts Initiative, which aims to support the efforts of eight regionally diverse districts in promoting social-emotional learning (SEL) for students. Weissberg and Cascarino (2013) defined SEL as

the process by through which children and adults acquire and effectively apply the knowledge, attitudes, and skills necessary to understand and manage emotions, set and achieve positive goals, feel and show empathy for others, establish and maintain positive relationships, and make responsible decisions (p. 20).

This definition is closer to what some refer to as emotional intelligence, or soft skills, largely propelling by the *Secretary's Commission on Achieving Necessary Skills* (SCANS) report (Department of Labor, 1991), which examined the demands of the workplace to evaluate students' preparedness for entering skills-based employment. SCANS (1991) reported that in addition to foundational skills, competencies for "workplace-know-how" (p. 21) were required for effective job performance. Foundational skills were defined as "basic skills, thinking skills, and personal qualities" (p. 21) while competencies were defined as "resources, interpersonal, information, systems, and technology" (p. 21). A more recent definition of soft skills includes "leadership, teamwork, critical and holistic thinking, logical reasoning, and communication skills" (Brill et al., 2014, p. 175). This term is used across sectors and is associated with employability, retention, and longevity in a respective field.

Another term used to describe non-cognitive ability in the affective learning domain is "executive function" skills, sometimes referred to as self-regulation skills (Zimmerman et al., 1992). These terms capture the constructs of motivation, perseverance, and attention and focus. Luria's (1966) description of executive function described this skill as "anticipation, planning, execution, and self-monitoring" (as cited in Purdy, 2011, p. 78). In 2017, NAEP will begin collecting data on grit, (i.e. perseverance and passion for long-term goals (Duckworth, Peterson, Matthews, & Kelly, 2007, p. 1087) and growth mindset (Dweck, 2006). Moreover, the Partnership for 21st Century Learning, of which Louisiana is a Leadership State, promotes the advancement of 21st Century Skills (National Research Council, 2012) in schools (see Table 2.1).

Table 2.1

21st Century Skills			
Key Subjects and 21 st Century Themes	Life and Career Skills	Learning and Innovation Skills	Information, Media, and Technology Skills
Core Content	Flexibility and	Creativity and	Information Literacy
Global Awareness	Adaptability	Innovation	Media Literacy
Financial, Economic, Business and Entrepreneurial Literacy	Initiative and Self- Direction	Critical Thinking and Problem Solving	ICT (Information, Communications, and Technology) Literacy
Civic Literacy	Social and Cross- Cultural Skills	Communication	
Health Literacy	Productivity and	Collaboration	
Environmental Literacy	Accountability		
	Leadership and Responsibility		

From Partnership for 21st Century Learning http://www.p21.org/storage/documents/docs/P21_framework_0116.pdf

Further still, Conley (2015b) urged scholars in the field to replace the term “noncognitive” with “success skills” when referring to those qualities or skills that require students to use both content knowledge and meta-cognitive learning skills. He proposed “success skills” as the term that should be used “to describe the diverse mix of behaviors and mindsets students need to be effective learners” (blogs.edweek.org).

In 2013, the USDOE permitted the state of California to pilot a new accountability system in nine school districts, in which non-cognitive measures will account for a portion of the overall accountability rating for schools (West, Kraft, Finn, Martin, Duckworth, Gabrieli, & Gabrieli, 2016). These districts are testing out measuring grit and other social-emotional skills such as self-management, growth-mindset, self-efficacy, and social awareness (Fensterwald, 2016).

Paul Reville leads the Education Redesign Lab at Harvard University, which is engaging with a consortium of cities in the By All Means project. This project is aimed at “redesigning

education to restore opportunity” (edredesign.org) through a series of field projects, the first of which pairs health and education services in the community (edredesign.org). The fundamental design of the project includes personalized learning, health and social services, and out-of-school opportunities (edredesign.org), and reflects what some advocate as the “whole-child” approach (Kochhar-Bryant & Heishman, 2010), which accounts for child development across the domains for learning, including school-based health centers, after school enrichment programs, and other supports.

As is often the case, education leaders may act presumptuously in applying a policy or practice on a large scale without the prerequisite frames and supports to implement it effectively. While the promise of including SEL is a worthy endeavor for states and systems in ascertaining school quality, some scholars appeal to education leaders to proceed with caution (West, et al., 2016; West, 2016), while others advise that more research is needed before using such measures for accountability purposes at this time (Egalite et al., 2015; Duckworth & Yeager, 2015). In 2016, Zernike of the *New York Times* ran a special report on testing for social-emotional skills, citing words of caution from some of the top researchers in this area of the time: Angela Duckworth, “I do not think we should be doing this; it is a bad idea”; Camille Farrington, “There are so many ways to do this wrong”; and Martin West, “You think test scores are easy to game...they’re relatively hard to game when you compare them to a self-report survey.”

Measures. The inclusion of this domain as a measurement for student development would account for student habits and behaviors not captured in a measurement tool for cognitive development. To measure aspects of habits and student behavior, researchers can utilize self-report questionnaires, teacher-report questionnaires, performance tasks, attitude surveys, and mental health screeners.

Psychomotor Domain for Learning

In the wake of the passage of ESSA, Connecticut passed a new accountability system that includes, among other indicators, a measure for physical fitness (Burnette, 2016). California assesses its students in grades five, seven, and nine using the FITNESSGRAM[®], a “comprehensive, health-related physical fitness battery” (cde.ca.gov/ta/tg/pf/) that measures aerobic capacity, abdominal strength and endurance, upper body strength and endurance, body composition, trunk extensor strength and flexibility, and flexibility. The results of this test are reported in the school’s accountability report card. Similarly, Kentucky administers the “Practical Living” (education.ky.gov) portion of its state’s program review. And while most states have policy mandating instruction in physical education, few account for this mandate in their state accountability system.

Schneider and Zhang (2013) explored the impact of test-based accountability under NCLB on childhood obesity rates among school-aged children. The study found that lower levels of quality physical activity correlated with higher pressure for academic achievement for high school students. The study also found that state mandates on physical education work to reduce the negative effect of pressure stemming from test-based accountability. Intersection of education and health—where there are health issues in a community, there are likely education deficiencies accompanying the issues.

Basch (2010) noted the negative impact of health-related issues on a student’s motivation to learn and advocates for a public-health strategy that would address the health-related needs to close the achievement gap, especially of low-income minority students living in urban areas. He draws attention to the evidence supporting the claim that “children’s health factors as causal mechanisms through which low socioeconomic status influences academic achievement and

educational attainment ” (p. 6), and identified seven “educationally relevant health disparities” as priorities:

1. vision
2. asthma
3. teen pregnancy
4. aggression and violence
5. physical activity
6. breakfast
7. inattention and hyperactivity (p. 8).

This research suggested a compounding negative effect on academic achievement due to poor development in the psychomotor and affective domains.

Health and wellness facilitated by the local school setting have been systematized by reforms including School Based Health Centers (Kisker & Brown, 1996; Weist, Nabors, Albus, & Bryant, 2006), Communities in Schools (Warren, 2005), and wraparound services (Eber, Sugai, Smith, & Scott, 2002). The Harlem Children’s Zone provides an excellent case study in a school’s systemic approach to wraparound services and supporting the comprehensive needs of its students across the learning domains (hcz.org).

Innovation in State Assessment Systems

In March of 2015, the USDOE awarded New Hampshire’s Department of Education with a waiver to NCLB regulations that included annual locally developed performance assessment measures, called PACE, and reduced the number of tests required of K-12 students to eight. Likewise, California is piloting portfolio designs in lieu of graduation exams in their system using the Stanford Center for Assessment, Learning, and Equity (SCALE) student performance

assessment system (scale.stanford.edu). Other states in the Innovation Lab Network coordinated by the Council for Chief State School Officers, including Maine, Kentucky, Colorado, and Connecticut, are working to develop innovative assessment system designs (CCSSO, 2016). When asked about this opportunity for states to redesign accountability systems under ESSA, Jeff Henig of Teachers College at Columbia University replied:

Looking back to pre-NCLB, we see what we could anticipate as a likely outcome in the future, which is considerable variation in terms of how [states] use greater authority and discretion. Some states were leaders and innovators, some were laggards. They vary in terms of political dynamics, vary in terms of bureaucratic capacity, ... and in terms of what they value....It may take a while for the dust to settle and a new vision for accountability to emerge. But one blueprint for the future may be the past, specifically, the years just before the passage of the NCLB law, which saw a real range of approaches to accountability. (Klein, 2016).

Authentic Assessment for Deeper Learning

Herman and Baker (2009) supported the need for a national system of standards and assessments, but called for a “comparable, but locally adapted course-based evidence of essential competencies” (p. 187); they noted that the current state of accountability policies provided segue for standardization. Authentic assessment incorporates tasks encountered in the real world of any discipline (p. 49), original transfer of knowledge to novel situations, not a contrived scenario as commonly provided for on standardized tests (Kornhaber, 2004). One type of authentic assessment is performance assessment. Marion and Leather (2015) defined performance assessment as

generally multi-step activities ranging from quite unstructured to fairly structured. The key feature of such assessments is that students are asked to produce a product or carry out a performance (e.g., a musical performance) that is scored according to pre-specified criteria, typically contained in a scoring guide or rubric (p. 5).

Performance-Based Assessment

Assessment tasks that measure higher order cognitive processes are learning tasks³ (Miller et al., 2013; Shepard, 2000) and can be useful as a “tool for learning” (Turnipseed & Darling-Hammond, 2015, p. 3). Other types of performance assessment includes observation protocols, portfolio development, and self-report by students (Miller et al., 2013). These types of assessment capitalize on the value of feedback, learning from mistakes, and content mastery (Turnipseed & Darling-Hammond, 2015). Consortia tests, such as the Partnership for the Assessment of Readiness for College and Careers (PARCC) and the Smarter Balanced Assessment Consortia (SBAC), get us closer to more accurately measuring learning with through authentic means (Conley, 2015a; Darling-Hammond, Wilhoit, & Pittenger, 2014); however, as Marion and Leather (2015) described, “once-per-year”, i.e. summative, assessments, are “not enough to drive and support deeper learning” (p. 7). More formative and diagnostic measures are needed to provide actionable data for teachers for impacting instruction.

Need for Advanced System of Test-Based Accountability

It is worth noting that while efforts are underway to development “comprehensive” (Conley & Darling-Hammond, 2013, p. 6) assessment systems, this study is distinguishable by its attention to account for learning across the cognitive, affective, and psychomotor domains. “Comprehensive systems are generally defined as multiple levels of assessment designed to provide information for different users to fulfill different purposes” (Conley & Darling-Hammond, 2013, p. 6). Moreover, a group of twenty assessment experts “advocate for a coordinated system of assessment, in which different tools are used for different purposes—for example, formative and summative, diagnostic versus large-scale reporting” (Conley & Darling-Hammond, 2013, p. 16). The NRC (2001) argued for a more “balanced” assessment system that

³ Task is described as applying learned content to novel situation.

incorporates three principles: ‘coherence, comprehensiveness, and continuity’” (pp. 253-257.

The proposition for this study is to move toward an “advanced” system of accountability, one in which all domains for learning are accounted for in determining school quality labels. Moreover, the proposition is not to increase the amount of time or number of state-mandated tests, but rather to examine current practice and utilize the very best methods of measuring student learning and holding schools accountable for what we expect of our schools for student learning outcomes.

Furthermore, it is suggested that scholars approach the design of test-based accountability system inclusive of social-emotional learning *indicators* as distinguished from a *measurement*. OECD/DAC defines an “indicator” as “A quantitative or qualitative factor or variable that provides a simple and reliable means to measure achievement, to reflect changes connected to an intervention, or to help assess the performance of a development actor” (oecd.org). Similarly, the United States Agency for International Development (USAID) defines the term “indicator” as “A characteristic or dimension used to measure intended changes define in a Results Framework. Performance indicators are used to observe progress and to measure actual results compared to expected results...” (usaidprojectstarter.org). Alternatively, a “measurement” is “the process of obtaining a numerical description of the degree to which an individual possess a particular characteristic. Measurement answers the question ‘How much’” (Miller et al., 2013, p. 28). Thus, an indicator may include qualitative responses, while a measure would include only quantifiable responses.

Linn (2008) also suggested caution in the interpretation of causal effects in education research, while Wong (2013) cautioned researchers and policymakers to acknowledge the evolution of policy in practice and urged the understanding that a new system of accountability

will be dependent upon an appreciation of the interconnectedness of the different levels of the federal government. Although traditionally a role based in the community through the local school board, federal and state accountability policies have taken a more precedential role in local school governance (Carnoy & Loeb, 2002). Cohen and Spillane (1992) identified the relative independence by which states establish policy in education. Because the states depend on the districts for political support and policy execution, in theory, the local district should wield the control over acceptable policy and legislation; however, states have passed reform legislation with relative ease, even when these laws were not supported by local educators (Fuhrman, Clune, & Elmore, 1988). The national and state control of daily school functions do not account for what some studies suggest, that social and cultural influences may have as great an impact on academic achievement as reform policies and instructional guidance (Cohen & Spillane, 1992; Koretz, 1996). Moreover, as McDonnell (2009) noted,

not only do multiple levels of government share authority over public education and responsibility for its funding, but power is also fragmented among institutions within each level. It is intuitively clear that this institutional fragmentation helps explain the significant variation in educational services and quality across states and localities (p. 59).

When policy goals become more outcome-oriented and less process-focused, integral paradigm shifts are imminent and require associated supports for the shifts to take hold. Policies in place for student learning must also accompany supports for implementation and successful integration into practice. The measure of learning in an educational system will be less complete until it can structurally support the complex and organic nature of the learning process across all three domains for learning: cognitive, affective, and psychomotor.

Summary

Establishing structures for school success has proven effective (Pogrow, 2006; Edmonds, 1979). Edmonds (1979) identified school-level factors contributing to the academic achievement of low-income students in urban schools. These factors included creating a culture of high expectations with regular monitoring and supportive leadership, as well as teacher dissatisfaction with complacency and sense of personal responsibility for student success. McCarthy and Still (1993) found a cost-effective intervention, the Accelerated Schools Process, that utilized multiple measures of formative and summative cognitive assessments, while allowing for contextual factors for individualized implementation. In a counter-narrative to the teach-to-the-test disposition, Pogrow (2006) identified positive impacts of the Modularized Continuous Progress approach in which curriculum is modularized based on student competencies and not grade level. He also found increased achievement results correlated with participation in dramatic and musical productions, as well as with immersion in higher order thinking skills, in which questions and answers were provided in the context of small group learning environments to prioritize the process in thinking and learning.

Herman and Baker (2009) identified the need for a national system of standards and assessments, but noted that the current state of accountability policies providing sufficient work-around for standardization. They called for “comparable, but locally adapted course-based evidence of essential competencies” (p. 187). This mention ignites the question of the relevance of accountability policies in an era of common standards and assessments—for states that have adopted the Common Core State Standards and participate in a consortia-based achievement test, what additional purposes are served by state accountability policies? To effectively balance policy development, stability, and change, the alignment of ideas and interests with institutions is critical (McDonnell, 2009). The implementation of policies is dependent on the interpretation of

policies within contextual bounds and struggles for power (McDonnell, 2009; Honig, 2006; Honig, 2009). As accountability systems progress from measures of attainment to growth, perhaps the window for providing more coherence, comprehensiveness, and continuity (Herman & Baker, 2009) is opening in the system of measurement of cognitive development.

CHAPTER 3. METHODOLOGY

The more any quantitative social indicator is used for social decision-making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (Campbell, 1976, p. 49).

Philosophical Underpinnings of Educational Research

Guba's "paradigm dialogue" (Johnston & Christensen, 2014, p. 31) advances the idea that all research is guided by philosophical underpinnings of one's ontology, epistemology, methodology, axiology, and rhetoric. Ontology, the "nature of reality" (Creswell, 2013, p. 20), or "nature of existence" (Crotty, 1998, p. 10), frames one's worldview on the existence of knowledge and origin of truth. Broadly speaking, the term *epistemology* refers to the source of knowledge, *axiology* is the value brought to the research by the researcher, *rhetoric* is the language used to describe the process of study, and *methodology* is the procedure of research (Creswell, 2013, p. 22).

Johnston and Christensen (2014) outlined five general types of education research inquiry: basic, applied, evaluation, action, and orientational (p. 9) with the objective to explore, describe, explain, predict, and/or influence. According to Creswell (2009), "quantitative research is a means for testing objective theories by examining the relationship among variables" and "qualitative research is a means for exploring and understanding the meaning individuals or groups ascribe to a social or human problem" (p. 4). Quantitative research includes experimental, non-experimental, and quasi-experimental designs (ex post facto) involving descriptive or inferential statistical analyses for causal and predictive analytics, whereas qualitative research, generally, is framed around theories of social science or social justice (Creswell, 2013) and includes case study, ethnography, grounded theory, narrative, and phenomenology designs

(Johnston & Christensen, 2014). Qualitative research follows an “inductive” (Johnston & Christensen, 2014, p. 22) approach where knowledge is viewed as created through transformation by social construction through interpretive and participatory experience (Davis, 2004).

Quantitative Research

Quantitative research stems from the metaphysical ontological perspective that there is an ideal world that is ‘out there’, that knowledge is static, and that cognitive development follows a linear trajectory toward an absolute truth. In this perspective, knowledge is viewed as attainable by systematic study (Davis, 2004). Within the traditional linear, hierarchical thinking, there is an embedded assumption of gradual conformity to higher levels of the organization, or development by accumulation (Kuhn, 1962). This worldview, perspective, or ‘paradigm’ (Lincoln & Guba, 2000) is heavily influenced by Descartes’s ideology on the psychophysical dualism introduced the concept of the individual, where objectivity came into consciousness and reason became the primary means by which we come to know (Davis, 2004). In the quantitative tradition, the axiological perspective of the researcher is positioned outside the study. Objective instrumentation is employed to ensure replicability with the purpose of generalizability to broader populations. Researchers have a responsibility to disclaim any potential bias and/or conflicts of interest that may directly or indirectly influence the way in which the researcher interprets the data.

Davis (2004) described two epistemological approaches to underscore the quantitative tradition of inquiry: gnosis (i.e. meaning of life, spiritual knowledge) and episteme (i.e. function of life, practical knowledge), rooted in religion and science. Within gnosis, rhetoric is characterized by understanding the essence and meaning of life through transcendence for deeper

understanding by asking questions around “why” in the areas of mysticism and religion. Methodological designs include vision quests, symbolisms, and hermeneutics, among others. Within episteme, rhetoric is characterized by deductive reasoning according to a scientific method by asking questions around “how” in the areas of rationalism (i.e. analytic philosophy) and empiricism (i.e. analytic science). Methodological designs include but are not limited to experiments and surveys using descriptive and inferential statistics (Davis, 2004).

Qualitative Research

Qualitative research stems from the ontological perspective that knowledge is emergent, socially constructed and transformative, and where truth is discovered as a biological unfolding from within (Davis, 2004). Within this paradigm, knowledge can never be fully attained as there are many truths (Creswell, 2013) bound by the time and place of participants. Values and assumptions shape reality (Erikson, 1986). Where quantitative inquiry pursues absolute truth, qualitative inquiry traces patterns of phenomenon and networks that maintain original diversity of individual parts and complexity within the relationships of those parts (Davis, 2004).

In this tradition, the axiological perspective of the researcher is positioned within the study, in which the researcher is directly a part of the data as the research instrument (S. MacGregor, personal communication, June 13, 2013). There is greater responsibility for the researcher to establish trustworthiness and full disclosure of the researcher in the research (S. MacGregor, personal communication, June 13, 2013). Procedures of research are characterized as inductive, emerging, and shaped by the researcher’s experience in collecting and analyzing the data (Creswell, 2013, p. 22). The purpose of this tradition of inquiry is to study the richest data sources by getting as close to the data as possible, which requires subjective experience. While possible, generalizability is not expressly purposed for this tradition. “The author admits that the

stories voiced represent an interpretation and presentation of the author as much as the subject of the study” (Denzin, 1989, p. 6).

Davis (2004) described two epistemological approaches that underscore the qualitative tradition of inquiry: intersubjectivity (i.e. interpretive, truth understood in terms of social accord) and interobjectivity (i.e. participatory), rooted in humanism and naturalism. Within intersubjectivity, rhetoric is characterized by understanding language as power by asking questions around “what is said and unsaid” in the areas of structuralism and post-structuralism. Methodological designs include semiotics, phenomenology, grounded theory, ethnography, case study, critical theory, race and gender studies, and genealogies. Within interobjectivity, rhetoric is characterized by holistic investigation of humanism and naturalism by asking questions around dependence on relationships, adaptability, and the non-conscious cognitive in the areas of complexity science (i.e. self-organization) and ecology (i.e. the way of being in the world, truth framed in terms of possibilities that arise and lock into place as the universe evolves). Methodological designs include, among others, chaos and complexity theory, string theory, systems theory, rhisomatic theory, green theory, and ecopsychology (Davis, 2004).

The *Handbook of Qualitative Research* (Denzin & Lincoln, 2005) identified eight critical moments in the history of qualitative study: traditional (1900-1950), modernist (1950- 1970), blurred genres (1970-1986), crisis of representation (1986-1990), postmodern (1990-1995), postexperimental inquiry (1995-2000), methodologically contested present (2000-2004), and the fractured future (2005-) in which the term *qualitative research* maintains different meanings for each period during the history. The current stage of qualitative study (fractured future) “confronts the methodological backlash associated with the evidence-based social movement” (Denzin & Lincoln, 2005, p. 3) of scientifically based research.

Qualitative research methods include case study, politics and ethics, participatory inquiry, interviewing, participant observation, visual methods, and interpretive analysis (Denzin & Lincoln, 2005). Creswell (2009) outlined the following strategies of inquiry for the qualitative researcher: phenomenology, grounded theory, ethnography, case study, and narrative in which the researcher employs emerging approaches, asking open-ended questions, and using text or image data collection. The researcher positions himself within the study to collect participant meanings of a context or setting, makes interpretations of the data and validates the accuracy of the findings through collaborative partnerships, and creates an agenda for change or reform.

Grounded theory. Phelps and Hase (2002) identified participatory action research within a qualitative research design as “consistent with the notion of adaptation to environment...in which theory becomes a learning tool for trying out solutions to local problems” (p. 512). Phenomena outlying the norm is not only valued but also embraced as the richness of exception to understanding change processes. Glaser (1965) identified four emergent stages within a constant comparative method of a grounded theory approach: (1) comparing incidents applicable to each category, (2) integrating categories and their properties, (3) delimiting the theory, and (4) writing the theory in a continuous growth model which will generate theoretical properties of the model (p. 439). This approach assumes an awareness of interrelationship between causes, conditions, and consequences (MacGregor, Educational Research lecture, 2013) and views humans as purposive agents—in which people act on the basis of meaning which is defined and redefined through interaction. In this way, generating a theory and conducting social research are part of the same process. The theory that the researcher develops is a “unified theoretical explanation” (Corbin & Strauss, 2007, p. 107) around the “actions, interactions, and social processes of people” (Creswell, 2013, p. 84). This structuralist perspective assumes that reality is

socially constructed and negotiated through our interactions (Davis, 2004); therefore this theory is provisional—an infallible interpretation that is limited to its historical context, and is therefore, in need of continual revision (MacGregor, Educational Research lecture, 2013). “Thick descriptions” (Geertz, 1977, p. 3) can be utilized throughout the process to strengthen the case for validity of the research. These descriptions may include details about the research setting, quotes from participants, and strong action verbs that support “abundant, interconnected details” (Stake, 2010, p. 49).

Lincoln and Guba (1985) used terms such as “credibility,” “authenticity,” “transferability,” “dependability,” and “confirmability” (p. 300) to convey the qualitative approach to external validity and reliability. Qualitative researchers can ensure reliability through “intercoder agreement checks” (Miles & Huberman, 1994, p. 64), an agreement by independent coders to use certain code words to represent particular words and/or phrases in the phenomenon/text under study (Creswell, 2013). Additionally, qualitative researchers strengthen reliability in their research with the development of codebooks for analyzing codes, names, and themes established “as priori” (Miles, Huberman, & Saldana, 2014, p. 27) or as emergent from the research (Miles, Huberman, & Saldana, 2014).

Complexity Science. The emergence of complexity science within the relatively young field of qualitative study has yet to be included in the *Handbook* as such. Davis and Sumara’s (2008) conditions of emergence are the closest resemblance to a method in complexity science. These conditions suggest opportunities for educators to establish such an environment in which emergence is likely to occur within the school setting. The conditions include internal diversity, internal redundancy, neighbor interactions, decentralization of control, randomness, coherence, positive and negative feedback loops, the possibility of dying, sufficient means to preserve

information, rate of flow of information, stability under perturbations, and reproductive instability (Davis & Sumara, 2008).

The current social dynamic of public schooling in America, with its emphasis on standardization and quantitative output, may be recognized, by some, as anti-intellectual with resemblances to antidemocratic ideas in which the process of schooling becoming a process of conformity. “Educational objectives, then, represent the kinds of changes in behavior that an educational institution seeks to bring about in its students” (Tyler, 1949, p. 6). The current emphasis on identifying “needs” as deviating from the norms indicates a rejection of complexity thinking in education. Tyler (1949) identified the importance and value of alignment with learning at school and real-life experiences outside of school, but Dewey (2010) identified the classroom as the most removed environment from authentic experiences—essentially, the worst place to learn.

Complexity science is an enabling, catalytic process that fosters cognitive development (Capra, 1996). Therefore, its possibilities in the classroom as a research setting abound. Complexity science fosters reflection and thoughtfulness, promoting considerations of the possibilities, and as such does not offer “research recipes” or prescribed responses (Kuhn, 2008; Phelps & Hase 2002; Davis & Sumara, 2005).

There is a misalignment with descriptive complexity and education, in that education has an aim to make a difference through goal-oriented criteria (Phelps & Hase, 2002), but complexity maintains that

cognition is more than acquisition of new knowledge: it engages motivation, personalities, learning styles, dispositions and preferences, the whole person. Teaching and learning take place at the intersection of the individual and society, and the outcomes are unpredictable. This is a difficult model for those managers

to entertain who seek certainty, control, predictability, and narrow accountability” (Morrison, 2008, p. 23).

In his work *A Postmodern Perspective on Curriculum*, Doll (1993) identified the possibilities of complexity science within education as a process of exploring the unknown together, underscoring the essential component of relationship within complexity science. Relating one’s knowledge to another with the other’s own ontological and epistemological worldviews incites the synergetic phenomenon of complex adaptive systems. Doll also proposed his vision of a transformative curriculum characterized by the 4 R’s: richness, recursion, relations, and rigor (pp. 174-183).

Possibilities within complexity science for education can be further explored through the notion of the “hidden curriculum,” that which is learned in a learning organization but not explicitly taught, encompassing culture, attitudes, expectations, etc. Additionally, impact of technology on the brain and its role inside, outside, and both concurrently inside and outside the classroom with real-time communication technology provides opportunities for exploration of implications on learning, networks, systems, and knowledge. Possibilities with technology stretch reality through virtual role-playing gaming, real-time audio/video communication, and 3-dimensional technology, including holograms. With technology, what is reality?

Gough (2012) identified triangulation methods of survey instruments in mixed methods approaches as a limitation to the research process and non-explanatory of the phenomena. As with any participatory action research, there are limitations to one’s own experience based on physical composition, age, and lived experience. Phelps and Hase (2002) described action research as unpredictable in that the chain of causes is unrecoverable, therefore not entirely replicable, and the generalizability of action research is relative to the participants or observers

of the study, rather than to the study itself. In the evidence-based era of education research, complexity is a less valuable form of research design due to its unpredictability. According to Capra (1996), “the theory of autopoiesis—the generation of configurations that are constantly new— shows that creativity is a key property of all living systems” (p. 221), generating diversity through reproduction. This theory is in stark contrast to the standardization movement of outcomes-focused test-based accountability.

Another contrast between complexity science and the current state of education in American is in the competitive, combative context of our schooling practices—*A Nation at Risk* was spurred by the space-race and competition with Russia, *No Child Left Behind* invoked incentives and consequences based on performance measured quantitatively, and *Race to the Top* established competition for financial incentives. In this paradigm, education is not viewed as a social process, but an economic commodity.

Furthermore, language is limiting—“Like blinders, the terms we adopt to express ourselves limit the range of our view. The crucial role of language in human evolution was not the ability to exchange ideas, but the increased ability to cooperate” (Capra, 1996, p. 275). Capra (1996) identified networking as the survival-mechanism for life: “In the end, the aggressors always destroy themselves, making way for others who know how to cooperative and get along. Life is much less a competitive struggle for survival than a triumph of cooperation and creativity” (p. 242).

Considerations. While complexity offers the opportunity to develop through failure (Kuhn, 2008), current test-based accountability systems provide low-quality feedback on opportunities for advancement. Doll (1993) asserted that we are in the midst of a paradigm shift, possibly even a mega paradigm shift. However, complexity science cannot, according to Kuhn

(1962), be termed a paradigm, as “...one of the things a scientific community acquires with a paradigm is a criterion for choosing problems that, while the paradigm is taken for granted, can be assumed to have solutions” (p. 37). If we, as a field of researchers, are to embrace complexity science, we must become comfortable with the nonlinear, orderly disorder—in a word, the unknown.

As Industrialization impacted the educational practices of the day, we are in the midst of the impact of Globalization on our society. Ever connected mega-networks have the ability to form, even beyond reality. Memories become alive and the notion of reality becomes obscured. The possibilities of technology have ushered in opportunities for complex adaptive systems to form and we are once again faced with the values of our society, in which our beliefs about what constitutes knowledge are changing (Hendry, Traditions of Inquiry lecture, 2012). What is life? What is knowledge? Living is knowing—an unfolding process of transformation through relating—and knowing is doing is being (Hendry, Traditions of Inquiry lecture, 2012). Complexity science, then, offers education research, not a solution, but an opportunity for a ‘different practice’ (Gough, 2012).

Historically in education research, quantitative research methods have dominated the field. Advantages of quantitative inquiry include the causal inferences drawn from cause/effect relationships and evaluation, focused on providing solutions to problems. Qualitative research studies embrace anomalies of complex phenomena and how natural phenomena of complex systems transform our being (Capra, 1996) and value diversity as “the pattern which connects” (Bateson, 1979, p. 8), aiming not to offer solutions, but the possibilities to further explore solutions (Capra, 1996; Kuhn, 1962). In *An Elusive Science*, Lagemann (2000) traversed the terrain of educational research to its roots and identified the conflict between functionalists (e.g.

John Dewey) and structuralists (e.g. Edward Lee Thorndike), thereby causing divergence, or “bifurcation” (Davis & Sumara, 2006, p. 32), in philosophies of education. In his criticism of the structuralist approach of using teaching manuals in the classroom, Dewey (1980) referenced the impact of industrialization on society.

Through it the face of the earth is making over, even as to its physical forms; political boundaries are wiped out and moved about, as if they were indeed only lines on a paper map; population is hurriedly gathered into cities from the ends of the earth; habits of living are altered with startling abruptness and thoroughness; the search for the truths of nature is infinitely stimulated and facilitated and their application to life made not only practicable, but commercially necessary (p. 6).

Such a criticism of the structuralist approach to learning may be applied to the impact of globalization in today’s modern world. For Dewey, all knowledge was socially constructed through authentic experience with an authentic purpose (1980), which may be limited by the standardized approach to learning with mass-produced materials and recourses. Alternatively, Thorndike (1980) perceived knowledge as an independent activity that could be replicated among different populations and he emphasized statistical analyses of progress.

Both quantitative and qualitative studies have limitations. Quantitative research limits the development of novelty and creativity in the natural environment (Kuhn, 1962), whereas in qualitative studies, knowledge can never be fully known, as it is created in relational, social construction. Datta (1994) advanced the promise for mixed methods designs when he stated “neither the quantitative hook set for the big fish nor the qualitative net scaled for the little fish adequately captures life in most seas. We need a paradigm to help us become scuba divers” (p. 64).

Mixed Methods

Situated within a physical ontology (i.e. qualitative frame), a mixed methods approach to study may be employed for the pragmatist. This approach assumes that strengths from each tradition will total more than the sum of individual parts, an idea known as “complementary strengths” (Johnson & Christensen, 2014, p. 53). This idea has also been referred to as “methodological eclecticism” (Teddlie & Tashakkori, 2011, p. 285), which Hammersley (1996) further delineated.

What is being implied here is a form of methodological eclecticism; indeed, the *combination* of quantitative and qualitative methods is often proposed, on the ground that this promises to cancel out the respective weaknesses of each method (p. 167, italics in original).

The mixed-methods researcher collects data, develops a rationale for mixing, integrates the data at different stages of inquiry, and employs the practices of both qualitative and quantitative research. Johnson, Onwuegbuzie, and Turner (2007) compiled the following working definition of mixed methods research:

Mixed methods research is the type of research in which a researcher or team of researchers combines elements of qualitative and quantitative research approaches (e.g. use of qualitative and quantitative viewpoints, data collection, analysis, inference techniques) for the broad purposes of breadth and depth of understanding and corroboration (p. 123).

Further, mixed methods research is characterized by the following qualities:

1. Methodological eclecticism
2. Paradigm pluralism
3. Emphasis on diversity at all levels of the research enterprise
4. Emphasis on continua rather than a set of dichotomies
5. Iterative, cyclical approach to research

6. Focus on the research question (or research problem) in determining the methods employed within any given study
7. Set of basic “signature” research designs and analytical processes
8. Tendency toward balance and compromise that is implicit within the “third methodological community” (Teddlie & Tashakkori, 2011, p. 287).

For this study, a mixed methods approach was the ideal research strategy for ensuring a thorough and comprehensive understanding of the complexity involved in designing and implementing a test-based accountability system that served the interests of stakeholders.

Study Design: Mixed Methods

This study examined the test-based accountability system in Louisiana under federal mandates and the extent to which the system provided meaningful and actionable data for stakeholders. The intent of this two-phase, sequential mixed methods study was to discover the educational values most prioritized by Louisiana stakeholders and the extent to which current policymakers were willing to pilot an advanced system for test-based accountability, including tests of the higher levels of the cognitive domain, as well as indicators for learning in the affective and psychomotor domains. In the first phase, quantitative research questions addressed the comparison of perceptions about Louisiana’s test-based accountability system and the educational values of stakeholders through the administration of a digital statewide survey. Stakeholders included parent, teacher, teacher/instructional leader, principal, community member, local education agency staff, local education agency superintendent, state education staff, elected official, members of the business community, or members of higher education. Using hierarchical linear modeling, survey responses were correlated along a line of regression

between the dependent and independent variables. Information from this first phase was explored further in a second qualitative phase.

In the second phase, qualitative interviews were used to probe significant comparative relationships of the survey data by exploring aspects of Louisiana's test-based accountability system with lawmakers serving on the House and Senate Education Committees at the Louisiana Legislature. Following up with qualitative research in the second phase allowed better understanding and explanation of the survey results.

Research Questions

Phase I: To what extent does Louisiana's current test-based accountability system deliver results that are valuable to stakeholders?

Phase II: To what extent are policymakers in Louisiana receptive to piloting an advanced system of accountability for Louisiana that includes reported measures in the affective and psychomotor domain to improve student learning outcomes?

Conceptually, this study was framed from a pragmatic worldview (James, 1906) directed at the study of test-based accountability in Louisiana. Methodically, this study followed a descriptive, non-experimental design (Creswell, 2009) of a mixed methods approach to answer the research questions. The researcher used an interpretive lens grounded in Bloom's Taxonomy for learning domains (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956) to more fully understand the extent to which the test-based accountability system in Louisiana delivered valuable information to stakeholders, and the extent to which policymakers were interested in piloting an advanced system of accountability including measures in the affective and psychomotor domain. This concurrent embedded strategy (Creswell, 2009) first utilized a survey approach to examine existing stakeholder perceptions of Louisiana's test-based accountability

system and its alignment with their personal educational values. Concurrently, this study used an inductive, grounded-theory approach to conduct semi-structured interviews with current Louisiana lawmakers on their willingness to pilot an advanced system of accountability that included indicators in the affective and psychomotor domains for learning, as well as cognitive measures of deeper learning (i.e. higher-order thinking skills) in a test-based system of accountability for student learning in Louisiana.

Phase I

The first phase of this confirmatory study (Johnson & Christensen, 2014) consisted of empirical data collection through survey instrumentation to test the null hypothesis, that there was a significant relationship between Louisiana's test-based accountability system and educational values of stakeholders, against the alternative hypothesis was that there is no significant relationship between Louisiana's test-based accountability system and educational values of stakeholders. This design was correlational (i.e. observational) in nature, and non-experimental (i.e. causational). Surveys "provide quantitative or numeric description of trends, attitude, or opinions of a population" (Creswell, 2009, p. 145) in order to make generalizations from sample of a population to the broader population (Creswell, 2009). Feedback was received from five colleagues and trusted advisors on the design of the survey instrument.

Sampling. This study utilized random sampling procedures to collect responses from over 100 respondents. From April 19, 2016 through May 20, 2016, I utilized email and social media communications for a targeted sampling of Louisiana residents. Members of the sample population received invitations to participate via email or social media communication. Where state level associations and/or organizations were available (e.g. Louisiana Association of Educators, Louisiana Association of School Principals, Louisiana Association of

Superintendents, and Louisiana Association of School Boards), state leaders were contacted for assistance in disseminating participation invitations. Representatives of the Louisiana Teacher Leaders and the Louisiana Standards Review Committee were invited to participate. All members of the Louisiana Legislature received an invitation to participate along with local education agency staff members responsible for maintaining the data used for Louisiana's test-based accountability system. Leaders in and representative for Louisiana institutions for higher education, charter management organization, as well as business leaders in chambers of commerce were invited to participate. Education journalists for print and visual media received the invitation to participate with a special request for dissemination via print or social media. The email invitation to participate included a hyperlink to the digital survey instrument, accessible at the following link: https://www.surveymonkey.com/r/LSU_Kahn. Inclusion criterion for this study was residency in Louisiana, and exclusion criterion was non-residency in Louisiana.

Instrumentation. The survey was administered digitally through SurveyMonkey, an online survey development company that provides survey administration solutions through data collection and analysis. This tool increased accessibility for participants to respond via computer-based technology such as desktop/laptop computers, tablet, or smartphone. The survey allowed only one complete survey response per IP address, protection against multiple response bias. Despite this protection, if a participant accessed and responded to the survey on multiple electronic devices (i.e. personal computer and personal smartphone), that participant could weaken the validity of the survey instrument and distort the results. To protect against such bias, the email invitation to participate included a clause requesting each participant to respond to all survey items only once to ensure the validity of the results (see Appendix F). The survey

instrument consisted of twenty-one response items of four distinctive item types: binary, scale, rank, and open response (see Appendix D).

The first section asked participants to identify the role that best applied to them for this survey (parent, teacher, teacher/instructional leader, principal, community member, local education agency staff, local education agency superintendent, state education staff, elected official, member of the business community, or member of higher education). Demographic data was collected to contextualize respondents accordingly and instructions for providing responses were provided for each item type. To strengthen validity of the survey, participants were asked to confirm their residency in the state of Louisiana by selecting one of 69 parishes for their residency. This selection allowed for further analysis of responses by geographic regions of the state. The second section asked participants to rate various aspects of Louisiana's current test-based accountability system. Each item included a section for comments. The third section asked participants a series of questions around factors not currently included in Louisiana's test-based accountability and whether or not these factors should be included; each item included a section for comments.

The fourth section asked participants to rank order goals for education and skills valued in education. Each of these items was replicated from previous studies. The first item asked participants to rank order eight goals of education (basic academic skills, citizenship, critical thinking, emotional health, physical health, preparation for skilled work, social skills and work ethic, and the arts and literature). The second item asked participants to rank order ten skills (creative thinking, goal setting/innovation, interpersonal skills, leadership, listening skills, oral communications, potential career development, problem solving, teamwork, and writing).

Finally, participants were asked general questions around their perceptions of learning and of improving Louisiana's test-based accountability system.

Data Collection and Analysis. After electronic collection and storage in SurveyMonkey, data was exported from SurveyMonkey for import into SPSS predictive analytic software. Using SPSS, I conducted linear regressions of hierarchical linear modeling and analysis of variance (ANOVA). Correlation and regression analyses provided comparative data for determining the significance of the relationship between Louisiana's test-based accountability system and stakeholder values.

I ensured data quality in checking for outliers by reporting standardized residuals. Additionally, I ran descriptive statistics with the Q-Q Plot to identify any significant differences to the mean, which could indicate an outlier. Next, I tested for the assumptions of homogeneity of variance with the Levene's test and the Shapiro-Wilks' test for normality. If these values exceeded .05, equal variance and normality was not assumed and I tested the equality of means with the Brown-Forsythe and Welch tests. I ran descriptive statistics on this data set including the n for each group, the group mean, median, variance, skewness, kurtosis, along with the standard deviation and standard error for each group, including minimum and maximum score ranges. These tests allowed for me to set the criterion level of significance and compute the test statistic and the observed critical value(s). Finally, I tested the null hypothesis with the F-statistic found by running ANOVA and MANOVA tests to determine the predictive relationship for interpreting the results.

Validity and Reliability. Tests for internal consistency were applied, such as the Cronbach's coefficient alpha (Cronbach, 1951). Regarding post hoc tests for reliability, I ran analyses for Tukey HSD post hoc comparison and indicated a significant difference between

groups 1 (low noise) and 3 (high noise). Further, by running a Means Plot, the researcher visually associated the difference between groups 1 and 3, with implications for the study that an increase in noise level is negatively associated with academic performance on a student achievement test.

Phase II

The second phase of this study included a qualitative, semi-structured interview using “purposive sampling” (Creswell, 2013, p. 86) and systematic analytic procedures (Strauss & Corbin, 1998) that followed a constant comparative design (Johnson & Christensen, 2014).

Sampling. The researcher utilized purposive sampling, a non-probability sampling technique, of 2-3 interviewees to reach “saturation” (Creswell, 2013, p. 89), although there was the possibility of using “discriminant sampling” (p. 90) to further validate information gained from the selected interviews. Consideration was given for providing an information-rich sample base for selection by attempting to garner participation from any lawmaker meeting the inclusion criteria of serving on an education committee (see Appendix G). The names and contact information for House and Senate Education Committee members are publically available online at house.louisiana.gov and senate.legis.state.la.us, respectively. Of the sixteen members in the House and seven members in the Senate, all were invited to participate. Interviews were scheduled between May 2, 2016 and May 27, 2016, and lasted between 45 minutes and an hour. As the timing of this study was concurrent with the annual legislative session, it was anticipated that scheduling might prove to be a challenge.

Instrumentation. An interview protocol was designed and utilized to guide the interview process with each selected participant (see Appendix E). As a locally designed tool, the semi-structured interview protocol lacked study on the reliability and validity of the instrument in the

Mental Measurements Yearbook. The protocol included guiding questions aimed at gathering information on Louisiana's test-based accountability system, aspects of state-mandated testing, and interest in including additional indicators for learning in Louisiana's test-based accountability system. The protocol also included probing questions to illicit thoughtful responses and the researcher asked clarifying question, as applicable.

Data Collection and Analysis. Interviews were conducted in the location of the interviewee's choice; the Louisiana legislature interviews were recorded with both electronic and cassette devices. Following the interviews, the audios were transcribed into an electronic file for import into AtlasTI, data analysis, and research software. For data analysis, a codebook was used, including a systematic coding framework. The methodological approach used in vivo coding for themes and may include axial coding (Strauss, 1987) and subcoding (Miles, Huberman, & Saldana, 2014, p. 80) to gain understanding of the voice and listen for the plot. A priori words including 'achievement', 'quality', 'improvement', and 'assessment' were coded using Atlas TI. Quotes that poignantly reflected key themes were noted. Data was analyzed for co-occurrence and analysis of word count.

Trustworthiness. The role of the researcher in this case study focused on an interpretative lens to discover and expose experiential reality as it became rationalized over time inter-rater reliability. Validity of this experience was strengthened by recording and transcribing the interviews, as well as by using a codebook for locating prior codes and emerging themes. Reliability for data interpretation was established by using multiple raters to establish inter-rater reliability.

Ethical Considerations

Due to the non-RCT design and purposive sampling techniques employed for this study, selection bias was assumed for participants. Situated within the social sciences, education research faces unique challenges of ethics, particularly concerning the use of experimental, quantitative research, in which one group inherently does not receive the same services as another, which may have an overall positive or negative effect. Provisions should be considered for augmenting services for interested participants, if participating in a group not receiving services that possibly provides beneficial outcomes.

Engaging complexity in educational research involves researchers in a complex process of marrying complexity habits of thought with a range of aims. It means recognizing that complexity *per se* does not have an ethical intent—it is the researcher who is committed to human betterment (Kuhn, 2008, p. 187).

Due to the nature of qualitative research, there is the possibility of the researcher “interpreting descriptions of descriptions or describing interpretations of interpretations” in which it is the full responsibility of the researcher to “readily concede the difficulty posed in deriving direct causal explanations or predictive proof for complex phenomena within which he is embedded” (Horn, 2008, p. 2). This subjectivity requires the researcher to ask continuously “Did we get it right?” (Stake, 1995, p. 107) to ensure accuracy in data analysis and validity of the research.

Along with approval from an Internal Review Board (IRB), ethical practices associated with qualitative research include

identification of: the researcher, sponsoring institution, purpose of the study, benefits for participating, level and type of participant involvement; indication of how the participants were selected, notation of risk to the participant, guarantee of confidentiality to the participant, assurance that the participant can withdraw at any time, and the provision of names of persons to contact if questions arise (Creswell, 2009, p. 89).

Capra's (1996) assertion that Western society's over-emphasis on self-assertive thinking through social domination, such as patriarchy, imperialism, capitalism, and racism that are exploitative in nature and antiecological, underscores the notion of qualitative research associated with imperialism and colonialism (Denzin & Lincoln, 2005). It is the due responsibility of qualitative researchers to examine, through reflexivity, their own ontological and epistemological assumptions about the research being conducted and to communicate their position within the study to the participants and in the research findings. Reflexivity is the process of reflection on the self as researcher, the inquirer as respondent, the teacher as learner, with self-discovery as a part of the process (Guba & Lincoln, 2005).

Pragmatic challenges anticipated by the researcher included timing of the study. This study was anticipated to coincide with the convening of the Louisiana legislature, during which lawmakers were proposing solutions to a significant budget shortfall. Due to the significance of the budget issue, it was believed that lawmakers might be slow to respond to survey questions and/or requests for interviews.

Anticipated Implications for Future Research

To test the hypothesis that including indicators in the cognitive, affective, and psychomotor domains would contribute to improved student learning outcomes, a question for research subsequent to this study may include: What is the impact of reporting indicators in all three domains for learning on student achievement on the annual state standardized summative test? Pending results, further research studies may include experimental, causal studies of an advanced accountability systems piloted across the state. Upon collection, data collected for measuring and/or reporting student learning across the three domains for learning (cognitive, affective, and psychomotor) could be analyzed for variance using MANOVA procedures to

discover the extent to which achievement in one domain impacts achievement in other domains, and whether a compounding effect is evident.

Future research extending the outcomes of this study may also include investigating the extent to which Louisiana is utilizing flexibility as allowed by ESSA in the design of its accountability system, the impact of incentives on improvement in student achievement (Forte, 2013), the extent to which education policy is having the effect of creating uncontrolled choice (DeBray-Pelolt, 2007), and parent trigger laws (McDonnell, 2012).

Noteworthy. A recent study published by the Louisiana State University (LSU) Public Policy Lab (2016), a division of the Reilly Center for Media and Public Affairs at theanship School of Mass Communication, included information on public opinion of education reforms in public schools. During the month of February 2016, *The Louisiana Survey 2016* collected information via phone from over 1,000 respondents on public opinion about charter schools, the voucher program, Common Core State Standards, school letter grades, amount of testing, funding, safety, and discipline. The study reported a 3% response rate and margin of error +/- 3.1 percentage points. It is worth noting consideration of the LSU (2016) report for triangulation with this dissertation study of advancing test-based accountability systems for improved student learning outcomes.

Axiology

For five years, I was a classroom teacher at both high- and low-performing elementary schools in a district of 43,000 students who attended over 35 schools. I taught fourth grade, the elementary school grade in which the state high-stakes standardized achievement test was administered. I was responsible for communicating high-stakes test results to both parents and

students, and therefore bringing experiential knowledge on the relative impact of high-stakes testing to student achievement.

In 2011, I became the Curriculum Coordinator at my school and was responsible for establishing collaborative partnerships through various school-level committees and designing professional development for local and state agency initiatives. I prepared professional development and support in teaching and learning for classroom teachers, implemented strategies for data-driven decision making, facilitated the after-school enrichment program, and developed summer enrichment learning camps. That same year, the local education agency nominated me to represent the district on the Educator Leader Cadre for the Partnership for the Assessment of Readiness for College and Career, a taskforce established by the state education agency. I partnered with the state agency to serve as a leader of the Communications Committee of the Teacher Advisory Committee. In this capacity, I led the design, implementation, and interpretation of a statewide survey of teachers on the efficacy of communication practices from the state agency, and subsequently presented the results and implications for future practice, directly informing the state's continued transition to higher expectations for student learning. The following year, I was promoted to the Department of Accountability, Research, and Evaluation, a central office administrative position from which I oversaw the administration of the state assessments for the test-based accountability system for all schools in the district.

In 2013, with the support of my supervisor, I applied for and was hired as the Supervisor of Assessment Administration in the Office of Assessment of the state education agency. I led all aspects of training and administration of statewide assessments for over 700,000 students, as well as facilitating professional development and program reports for stakeholders including school board members, parents, district and school leaders, and state legislators. I produced

various reports and analyses informing policy development and communication through press releases, legislative audits, legislative requests, and vendor relations for large-scale statewide assessment systems. I also designed and developed professional development for district superintendents and leaders. Additionally, I initiated a comprehensive plan for professional development on the transition to new standards and assessments, specifically tailored to key changes in the assessment design and shifts in content for the Partnership of the Assessment of Readiness for College and Career (PARCC).

Currently, I hold a full-time faculty position as Visiting Instructor in the Department of Curriculum and Instruction in the College of Education at a state university. In this role, I teach upper division undergraduate pre-service teacher candidates in pedagogical coursework including Developmental Assessment and Research in Early Childhood Education, English Language Arts in the Elementary Classroom, and Classroom Management for Elementary Education. In this capacity I engage learners with philosophical and theoretical frameworks shaping perspectives and opportunities for the practical application of skills. Additionally, I supervise student teachers in early childhood, elementary, and secondary education.

CHAPTER 4. DATA COLLECTION AND ANALYSIS

In the ideal world, we would assess achievement by measuring the ultimate goals of education. (Lindquist, 1951, p. 152)

Phase I

The first research question asked respondents to what extent does Louisiana's current test-based accountability system deliver results that are valuable to stakeholders. To investigate this question, I used survey research methodology with a series of questions that followed a variety of item types (categorical selected response, nominal scale rating, dichotomous, or binary, selected response, ordinal ranking, and constructed response) for descriptive and inferential statistical analyses. Survey data were digitally collected from Thursday, April 28, 2016 through Friday, May 20, 2016. A total of 544 responses were collected with a 74% completion rate representing each of the 11 categorical respondent types (parent, teacher, teacher/instructional leader, principal, community member, local education agency staff, local education agency superintendent, state education agency staff, elected official, member of the business community, member of higher education) (see Table 4.1). Respondents were randomly sampled by targeting communications via email and social media to key leaders of each respective respondent type. Electronic access and a digital device were required for participation. Survey data were collected from a sample of the population of Louisiana residents, not the entire population. Due to this limitation, parametric tests made assumptions about the entire population based on the sample population from which the data were obtained.

Categorical Selected Response

Survey Question 1 (SQ1) required that participants select the category that best described their role for this survey. Understanding that some respondents may have been classified as more

than one category (i.e. parent and teacher), respondent discretion determined the singularly coded value. Due to an oversight in settings, four respondents were able to continue in the survey without responding to this question. Upon notice, the setting was adjusted to require a response for all subsequent respondents.

Table 4.1

SQ1: Please select the category that best describes your role for this survey.

	<i>f</i>	%
Parent	80	14.8
Teacher	228	42.2
Teacher/Instructional Leader	58	10.7
Principal	56	10.4
Community Member	32	5.9
Local Education Agency Staff	45	8.3
Local Education Agency Superintendent	8	1.5
State Education Agency Staff	3	0.6
Elected Official	5	0.9
Member of the Business Community	6	1.11
Member of Higher Education	19	3.5
Total (N)	540	100.0
No response: 4 cases		

Due to instrument design, the survey was susceptible to out-of-state respondents; therefore, Survey Question 2 (SQ2) strengthened the validity of this study by requiring respondents to identify their parish of residence. This allowed for the researcher to exclude any out-of-state respondents. Respondents in this study represented 81% of all Louisiana parishes (see Table 4.2).

Table 4.2

SQ2: Please select your parish of residence.

	<i>f</i>	%		<i>f</i>	%
Acadia	14	2.6	Natchitoches	12	2.2
Allen	4	.7	Orleans	35	6.4
Ascension	6	1.1	Ouachita	13	2.4
Assumption	3	.6	Plaquemines	1	.2
Avoyelles	3	.6	Point Coupee	2	.4
Beauregard	7	1.3	Rapides	6	1.1
Bossier	5	.9	Red River	4	.7
Caddo	2	.4	Richland	2	.4
Calcasieu	40	7.4	Sabine	2	.4
Caldwell	1	.2	St Bernard	2	.4
Cameron	12	2.2	St John the Baptist	1	.2
Catahoula	1	.2	St Landry	13	2.4
Claiborne	2	.4	St Martin	7	1.3
DeSoto	3	.6	St Mary	2	.4
East Baton Rouge	31	5.7	St Tammany	5	.9
Evangeline	4	.7	Tangipahoa	2	.4
Franklin	4	.7	Terrebone	4	.7
Grant	1	.2	Union	1	.2
Iberia	8	1.5	Vermilion	14	2.6
Iberville	2	.4	Vernon	2	.4
Jefferson	9	1.7	Washington	2	.4
Jefferson Davis	4	.7	Webster	2	.4
Lafayette	220	40.4	West Baton Rouge	1	.2
Lafourche	4	.7	West Carroll	2	.4
Lincoln	10	1.8	West Feliciana	2	.4
Livingston	4	.7	Winn	1	.2
			Total (<i>N</i>)	544.0	100.0

Note. All respondents indicated a Louisiana parish of residence. Parishes not represented in the dataset include Bienville, Concordia, East Carroll, East Feliciana, Jackson, LaSalle, Madison, Morehouse, St. Charles, St. Helena, St. James, and Tensas.

Likert-Scale Rating

Survey questions three through six utilized Likert scales for rating perception data of test-based accountability in Louisiana. The nominal categories were converted to ordinal scales (1-5) for each question to be treated as interval data for inferential statistics. This process was challenged and compounded by questions surrounding the range between each category being the same or different.

The independent T-test was not an appropriate inferential statistical test for this data set. T-tests determine whether there is a statistically significant difference between the means of two unrelated groups. T-test tests a null hypothesis (ex. $H_0: \mu = 3$) against an alternative hypothesis (ex. $H_a: \mu \neq 3$). When we test with a T-test, we test the hypothesis that two samples have the same mean. Because the data collected for this study included more than two dependent variables, the T-test was not the most effective test for analyzing statistical significance for the research questions under study.

A more effective test for statistical significance for this study was the one-way analysis of variance (ANOVA). Also known as an omnibus test, ANOVA tests for an overall experimental effect and identifies whether three or more means are the same. ANOVA does not provide specific information about which groups were affected, so post hoc analyses were necessary for more detailed analysis. For each of the Likert-scale rating items, the working null hypothesis (H_0) was that all group means were equal (ex. $\mu = 3$). The alternative hypothesis (H_a) was that all group means are not equal (ex. $\mu \neq 3$). The statistical significance level, or alpha (α), which is the probability of falsely rejecting the null, was tested at the .05 level ($\alpha = .05$), which meant that the probability of falsely rejecting the null hypothesis (Type I error) could happen 5% of the time. Type I errors wrongly reject the null hypothesis, when in fact it is true. A series of parametric tests were applied to test for assumptions of normality, statistical significance, and robustness for reliability and validity.

I tested for the assumptions of homogeneity of variance with the Levene's test. Where values exceeded the $\alpha = .05$, the null hypothesis was accepted. Where values measured below $\alpha = .05$, the null hypothesis was rejected and the alternative was accepted. When the p value, i.e. the ANOVA F statistic, exceeded the acceptable level of significance ($\alpha = .05$) to accept the null

hypothesis (μ = there is no statistical significance between group means), the null hypothesis was accepted and the alternative hypothesis was rejected (μ = there is statistical significance between group means). When the ANOVA F statistic measured below the acceptable level of significance ($\alpha = .05$), the null hypothesis was rejected and the alternative hypothesis was accepted. I checked for data quality by investigating the Q-Q Plot for normal linear regression.

Further analysis through cross tabulations, i.e. joint frequency distribution, allowed for detailed analysis by respondent type. Percentages were reported with the total raw N for each group to provide standardization for interpretation of results. I then applied the Chi Square test of independence to test for association or correlation coefficients of multiple variables. The Chi Square, or Pearson's Chi Square, tests for the expected count if there was no association between variables and against the observed values to determine if those values differ enough to establish statistical significance in variance.

Survey Question 3 (SQ3) asked respondents to rate the extent to which the School Letter Grade (i.e. school performance score) reflected a school's quality in Louisiana. The majority of respondents rated the accuracy of Louisiana's test-based accountability system as below average. The mean and mode ratings for SQ3 was (2- Somewhat Accurate). This data set narrowly passed the test for normality using the Levene statistic (see Table 4.5). Descriptive statistics testing for normality of this data sample indicated this distribution has a positive skew, where the normal distribution was weighted below the mean and the bell-curve included a smaller right tail (see Figure 4.1). Due to the skewness of the results, we knew that most respondents selected a value below the mean (3). The overall extent to which Louisiana stakeholders rated the School Letter Grade as reflecting a schools' quality in Louisiana as below average (1- Least Accurate and 2- Somewhat Average) was 65.9%.

The ANOVA *F* statistic did not measure below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was accepted. (see Table 4.6).

There was no statistically significant variance between groups. Because SQ3 had greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.9). Checking, then, the Likelihood Ratio, I concluded that there is no statistically significant association between respondent type and rating.

Table 4.3

SQ3: Descriptive Statistics

Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

	<i>(N)</i>	<i>Percent</i>
1 Least Accurate	96	19.4
2 Somewhat Accurate	230	46.5
3 Sufficiently Accurate	115	23.2
4 Highly Accurate	43	8.7
5 Most Accurate	11	2.2
Total	495	100.0
No Response: 49 cases		

Table 4.4

SQ3: Descriptive Statistics

Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

	<i>Std. Error of</i>		<i>Std.</i>		
<i>Mean</i>	<i>Mean</i>	<i>Mode</i>	<i>Deviation</i>	<i>Skewness</i>	<i>Kurtosis</i>
2.2788	.04259	2.00	.94754	.665	.209

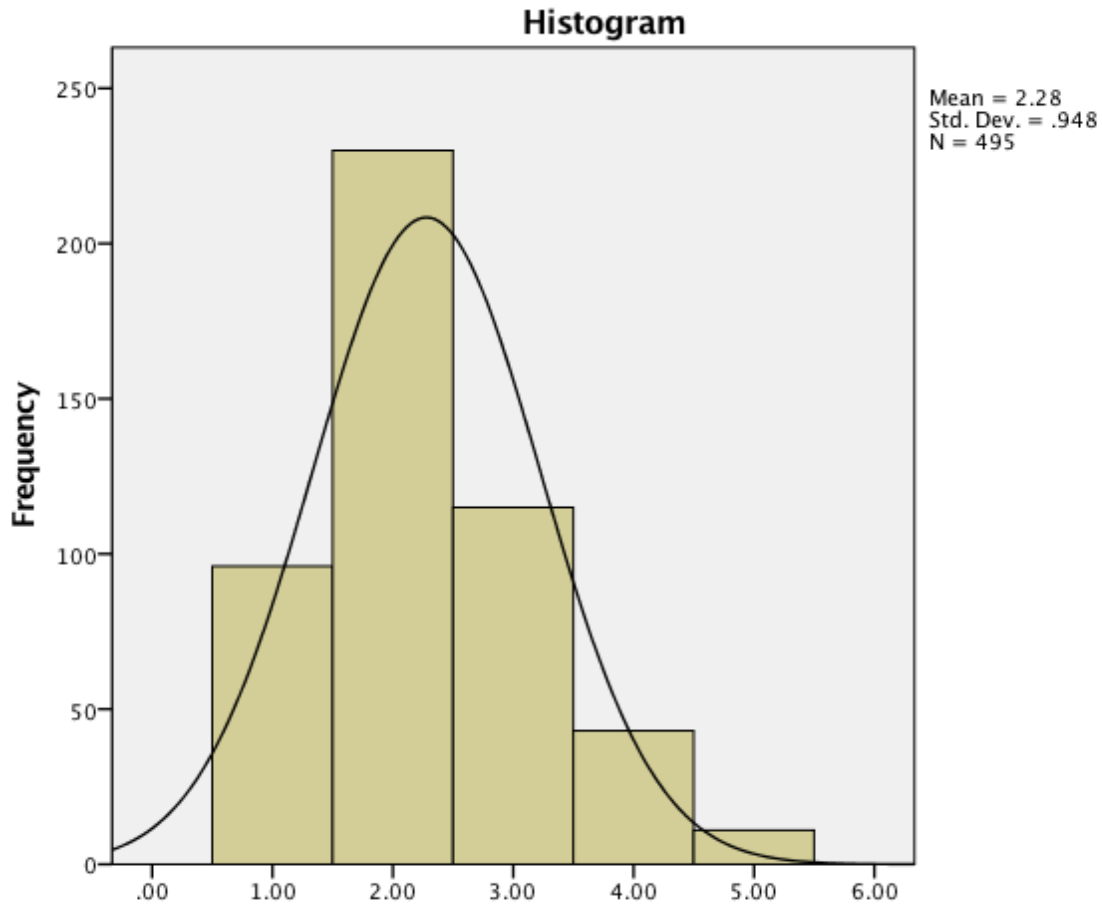


Figure 4.1. Histogram of SQ3: Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

Table 4.5

SQ3: Test of Homogeneity of Variances

Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
1.826	10	481	.054

Table 4.6

SQ3: Oneway ANOVA

Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	7.895	10	.790	.873	.558
Within Groups	434.956	481	.904		
Total	442.852	491			

Table 4.7

SQ3: Robust Tests of Equality of Means

Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

	<i>Statistic^a</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
Welch	1.433	10	33.230	.209

a. Asymptotically F distributed.

Table 4.8

SQ3: Cross Tabulations

Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

	<i>Respondent Type</i>											
	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instructional Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of High Education</i>	<i>Total</i>
1 Least Accurate	23.5	20.7	15.8	16.0	17.9	14.0	25.0	0.0	40.0	16.7	25.0	19.5
2 Somewhat Accurate	45.6	44.7	45.6	36.0	64.3	51.2	62.5	66.7	60.0	50.0	43.8	46.3
3 Sufficiently Accurate	14.7	24.0	24.6	36.0	14.3	23.3	12.5	33.3	0.0	33.3	25.0	23.2
4 Highly Accurate	11.8	7.7	12.3	12.0	3.6	9.3	0.0	0.0	0.0	0.0	6.3	8.7
5 Most Accurate	4.4	2.9	1.8	0.0	0.0	2.3	0.0	0.0	0.0	0.0	0.0	2.2
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(69)	(208)	(57)	(50)	(28)	(43)	(8)	(3)	(5)	(6)	(16)	(492)

No response: 52 cases

Table 4. 9

SQ3: Chi Square Tests

Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a schools' quality in Louisiana?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	27.105 ^a	40	.940
Likelihood Ratio	32.632	40	.790
Linear-by-Linear Association	1.132	1	.287
N of Valid Cases	492		

a. 34 cells (61.8%) have expected count less than 5. The minimum expected count is .07.

Survey Question 4 (SQ4) asked respondents to rate the amount of time students spent on state-required tests in Louisiana. The mean and mode ratings for SQ4 were 4.5 (4- More than Enough Time) and 5 (5- Too Much Time), respectively (see Table 4.10), comprising eighty-seven percent of the total response ratings (see Table 4.11). Descriptive statistics testing for normality of this data sample indicated this distribution had a negative skew, where the normal distribution was weighted above the mean and the bell-curve included a smaller left tail (see Figure 4.2). Due to the skewness of the results, we knew that most respondents selected a value above the mean (3). This data set passed the test for normality using the Levene statistic (see Table 4.12); therefore normality was assumed. Furthermore, robust tests of equality of means could not be performed for SQ4 because at least one group had zero variance.

The ANOVA *F* statistic did not measure below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was accepted. (see Table 4.13). There was no statistically significant variance between the groups of means; therefore, we accepted the null that there is no statistically significant difference between groups. Because SQ4 had greater than 20% expected count less than five, the Pearson's Chi Square assumption had

been violated (see Table 4.15). Checking, then, the Likelihood Ratio, I concluded that there was no statistically significant association between respondent type and rating.

Table 4.10

SQ4: Descriptive Statistics

Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

<i>Mean</i>	<i>Std. Error of Mean</i>	<i>Mode</i>	<i>Std. Deviation</i>	<i>Skewness</i>	<i>Kurtosis</i>
4.4980	.03719	5.00	.83149	(1.779)	2.995

Table 4.11

RQ4: Descriptive Statistics

Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

	<i>(N)</i>	<i>Percent</i>
1 Not Enough Time	5	1.0
2 Somewhat Enough Time	10	2.0
3 Sufficiently Enough Time	50	10.0
4 More Than Enough Time	101	20.2
5 Too Much Time	334	66.8
Total	500	100.0
No Response: 44 cases		

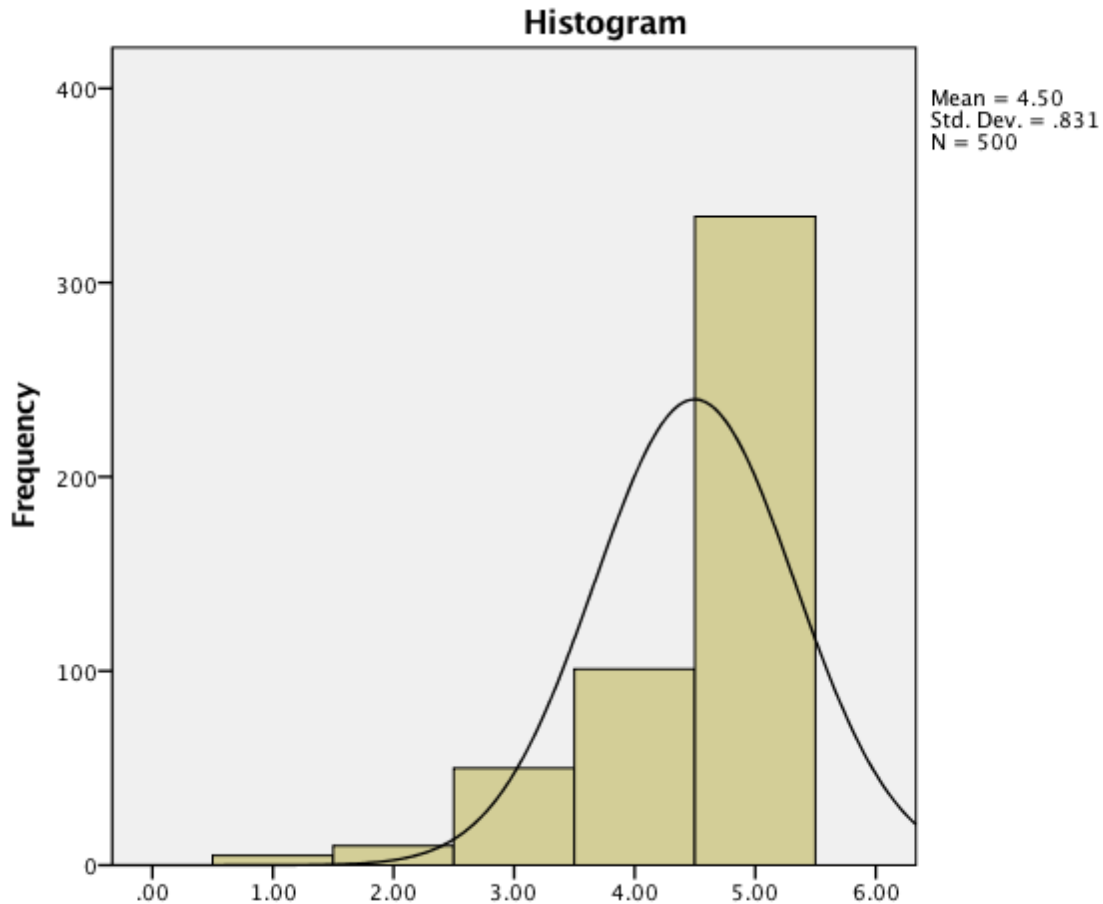


Figure 4.2 Histogram of SQ4: Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

Table 4.12

SQ4: Test of Homogeneity of Variances

Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
2.104	10	485	.023

Table 4.13

SQ4: Oneway ANOVA

Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	8.164	10	.816	1.179	.302
Within Groups	335.828	485	.692		
Total	343.992	495			

Table 4.14

SQ4: Cross Tabulations

Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

Time Spent Testing	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instructional Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of High Education</i>	<i>Total</i>
1 Not enough time	2.9	0.5	0.0	2.0	3.6	0.0	0.0	0.0	0.0	0.0	0.0	1.0
2 Somewhat Enough Time	0.0	3.3	0.0	2.0	3.6	0.0	0.0	0.0	25.0	0.0	0.0	2.0
3 Sufficiently Enough Time	10.1	8.5	15.8	15.7	3.6	14.0	0.0	33.3	0.0	0.0	0.0	10.1
4 More Than Enough Time	14.5	15.2	22.8	27.5	10.7	39.5	62.5	33.3	0.0	0.0	31.3	20.2
5 Too Much Time	72.5	72.5	61.4	52.9	78.6	46.5	37.5	33.3	75.0	100.0	68.8	66.7
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(69)	(211)	(57)	(51)	(28)	(43)	(8)	(3)	(4)	(6)	(16)	(496)

No response: 48 cases

Table 4.15

SQ4: Chi Square Tests

Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	67.907 ^a	40	.004
Likelihood Ratio	65.374	40	.007
Linear-by-Linear Association	.205	1	.651
N of Valid Cases	496		

a. 37 cells (67.3%) have expected count less than 5. The minimum expected count is .03.

Survey Question 5 (SQ5) asked respondents to rate the number of tests students take for state-required tests in Louisiana. The mean and mode ratings for SQ5 were 4.5 (4- More Than Enough Tests) and 5 (5- Too Many Tests), respectively (see Table 4.16). Descriptive statistics testing for normality of this data sample indicated this distribution had a negative skew, where the normal distribution was weighted above the mean and the bell-curve included a smaller left tail (see Figure 4.3). Due to the skewness of the results, we knew that most respondents selected a value above the mean (3). Eighty-nine percent of Louisiana stakeholders rated the number of tests student take for state-mandated tests in Louisiana as more than enough or too may tests (see Table 4.17). This data set passed the test for normality using the Levene statistic (see Table 4.18); therefore, normality was assumed.

The ANOVA *F* statistic measured below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was rejected and the alternative hypothesis was accepted (see Table 4.19). There was statistically significant variance between groups. Robust tests of equality of means could not be performed for RQ5 because at least one group had zero variance. Because SQ5 has greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.20). Checking, then, the Likelihood Ratio, I concluded that there is no statistically significant association between respondent type and rating, though the significance was near acceptance.

Table 4.16

SQ5: Descriptive Statistics

Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

<i>Mean</i>	<i>Std. Error of Mean</i>	<i>Mode</i>	<i>Std. Deviation</i>	<i>Skewness</i>	<i>Kurtosis</i>
4.5480	.03202	5.00	.71605	(1.485)	1.624

Table 4.17

SQ5: Descriptive Statistics

Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

	(N)	Percent
1 Not Enough Tests	1	.2
2 Somewhat Enough Tests	3	.6
3 Sufficiently Enough Tests	51	10.2
4 More Than Enough Tests	111	22.2
5 Too Many Tests	334	66.8
Total	500	100.0

No Response: 44 cases

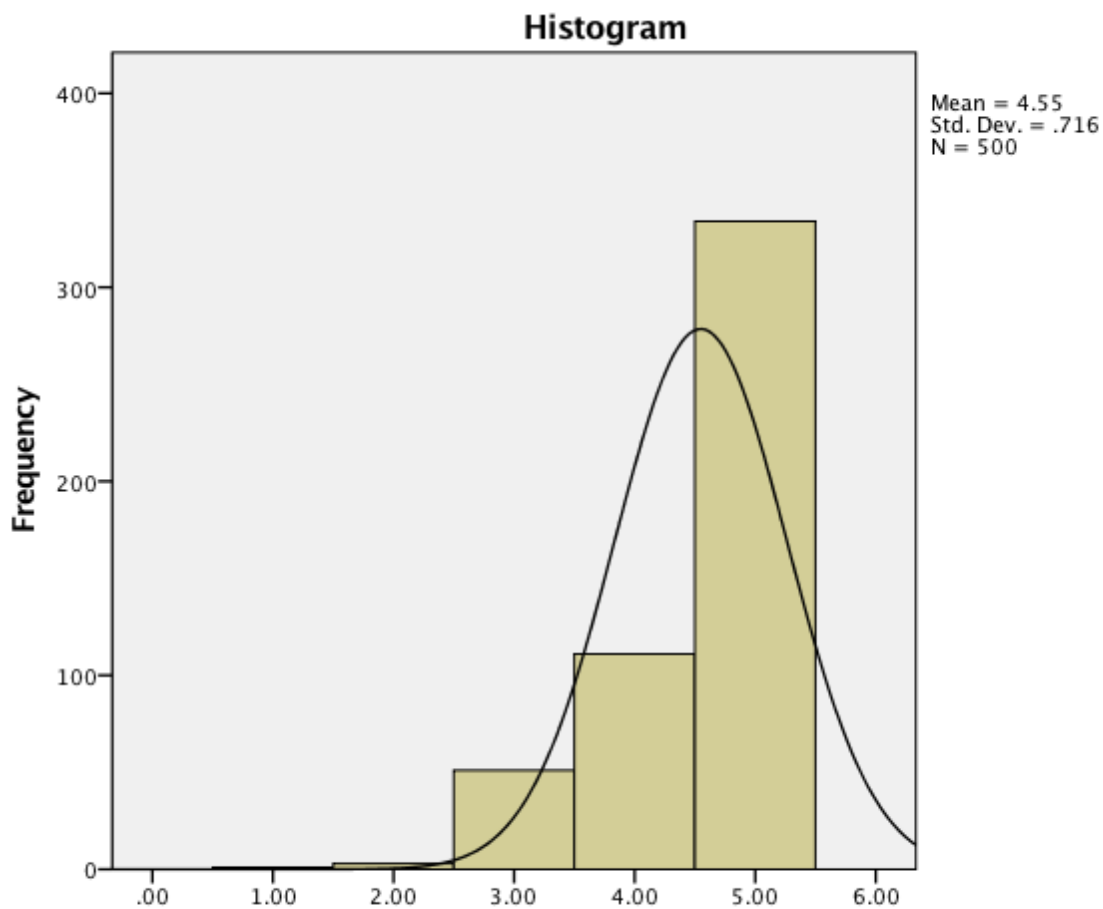


Figure 4.3. Histogram of SQ5: Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

Table 4.18

SQ5: Test of Homogeneity of Variances

Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
4.141	10	485	.000

Table 4.19

SQ5: Oneway ANOVA

Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	9.722	10	.972	1.922	.040
Within Groups	245.302	485	.506		
Total	255.024	495			

Table 4.20

SQ5: Cross Tabulations

Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

	<i>Respondent Type</i>											
	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instructional Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of High Education</i>	<i>Total</i>
Not enough tests	1.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
Somewhat Enough Tests	0.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	20.0	0.0	0.0	0.6
Sufficiently Enough Tests	10.3	6.2	21.1	13.7	10.7	16.3	12.5	33.3	0.0	0.0	0.0	10.3
More Than Enough Tests	19.1	21.3	12.3	31.4	21.4	34.9	37.5	66.7	0.0	0.0	25.0	22.4
Too Many Tests	69.1	71.6	66.7	54.9	67.9	48.8	50.0	0.0	80.0	100.0	75.0	66.5
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(68)	(211)	(57)	(51)	(28)	(43)	(8)	(3)	(5)	(6)	(16)	(496)

No response: 48 cases

Table 4. 21

SQ5: Chi Square Tests

Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	77.297 ^a	40	.000
Likelihood Ratio	54.970	40	.058
Linear-by-Linear Association	.649	1	.420
N of Valid Cases	496		

a. 37 cells (67.3%) have expected count less than 5. The minimum expected count is .01.

Survey Question 6 (SQ6) asked respondents to rate Louisiana's test-based accountability system for K-12 public schools. The mean and mode ratings for RQ5 were 1.7 (2- Okay) and 1 (1- Poor), respectively (see Table 4.22). Eighty-three percent of respondents rated the system as below average, while 13.6% rated the system as average. Only three percent of respondents rated the system above average. Descriptive statistics testing for normality of this data sample indicated this distribution had a positive skew, where the normal distribution was weighted below the mean and the bell-curve included a smaller right tail (see Figure 4.4). Due to the skewness of the results, we knew that most respondents selected a value below the mean (3). This data set failed to pass the test for normality using the Levene statistic (see Table 4.26); therefore, normality was not assumed.

The ANOVA *F* statistic measured below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was rejected and the alternative hypothesis was accepted (see Table 4.27). There was statistically significant variance between groups of means. Because SQ6 has greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.28). Checking, then, the Likelihood Ratio,

I concluded that there was sufficient evidence to support a statistically significant association between respondent type and rating.

Table 4.22

SQ6: Descriptive Statistics

Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

<i>Mean</i>	<i>Std. Error of Mean</i>	<i>Mode</i>	<i>Std. Deviation</i>	<i>Skewness</i>	<i>Kurtosis</i>
1.7315	.03710	1.00	.82872	1.046	.879

Table 4.23

SQ6: Descriptive Statistics

Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

	<i>(N)</i>	<i>Percent</i>
1 Poor	235	47.1
2 Okay	181	36.3
3 Good	68	13.6
4 Great	12	2.4
5 Excellent	3	.6
Total	499	100.0

No Response: 45 cases

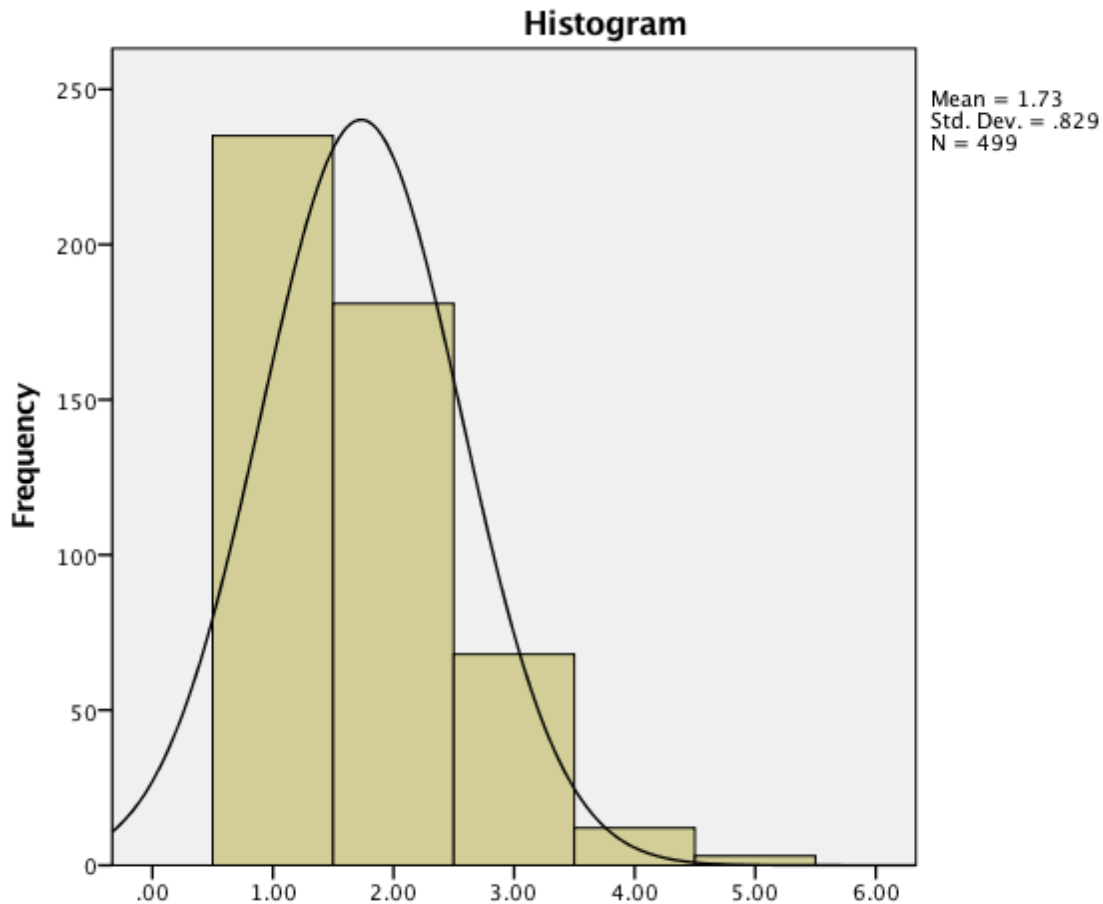


Figure 4.4. Histogram of SQ 6: Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

Table 4.24

SQ6: Test of Homogeneity of Variances

Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
.763	10	484	.665

Table 4.25

SQ6: Oneway ANOVA

Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	25.286	10	2.529	3.879	.000
Within Groups	315.514	484	.652		
Total	340.800	494			

Table 4.26

SQ6: Robust Tests of Equality of Means

Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

	<i>Statistic^a</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
Welch	3.463	10	32.882	.003

a. Asymptotically F distributed.

Table 4.27

SQ6: Percentage Crosstabulations of Overall Quality of Louisiana's Accountability System by Respondent Type

Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

	<i>Respondent Type</i>											
	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instructional Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency Staff</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of High Education</i>	<i>Total</i>
1 Poor	64.7	51.2	52.6	33.3	46.4	14.3	0.0	33.3	60.0	33.3	56.3	47.1
2 Okay	20.6	36.0	33.3	39.2	39.3	52.4	75.0	33.3	40.0	66.7	25.0	36.2
3 Good	10.3	10.9	10.5	23.5	14.3	23.8	25.0	33.3	0.0	0.0	18.8	13.7
4 Great	2.9	1.9	3.5	2.0	0.0	7.1	0.0	0.0	0.0	0.0	0.0	2.4
5Excellent	1.5	0.0	0.0	2.0	0.0	2.4	0.0	0.0	0.0	0.0	0.0	0.6
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(68)	(211)	(57)	(51)	(28)	(42)	(8)	(3)	(5)	(6)	(16)	(495)

No response: 49 cases

Table 4. 28

SQ6: Chi Square Tests

Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	60.251 ^a	40	.021
Likelihood Ratio	67.421	40	.004
Linear-by-Linear Association	7.953	1	.005
N of Valid Cases	495		

a. 35 cells (63.6%) have expected count less than 5. The minimum expected count is .02.

Dichotomous, or Binary, Selected Response

Survey Question 7 (SQ7) asked respondents to identify, either affirmatively or negatively, if schools should be held accountable for providing the nationally recommended amount 60 minutes of quality physical activity every day. RQ7 elicited 79.9% affirmative responses from Louisiana stakeholders (see Table 4.29). This data set passed the test for normality using the Levene statistic (see Table 4.30); therefore, normality was assumed.

The ANOVA *F* statistic measured below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was rejected and the alternative hypothesis was accepted (see Table 4.31). There was statistically significant variance between groups of means. Robust tests of equality of means could not be performed for RQ7 because at least one group had zero variance. Because SQ7 had greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.33). Checking, then, the Likelihood Ratio, I concluded that there was sufficient evidence to support a statistically significant association between respondent type and rating.

Table 4.29

SQ7: Descriptive Statistics

Should schools be held accountable for providing the nationally recommended amount (60 minutes) of quality physical activity every day?

	(N)	Percent
1 Yes	390	79.9
2 No	98	20.1
Total	488	100.0

No response: 56 cases

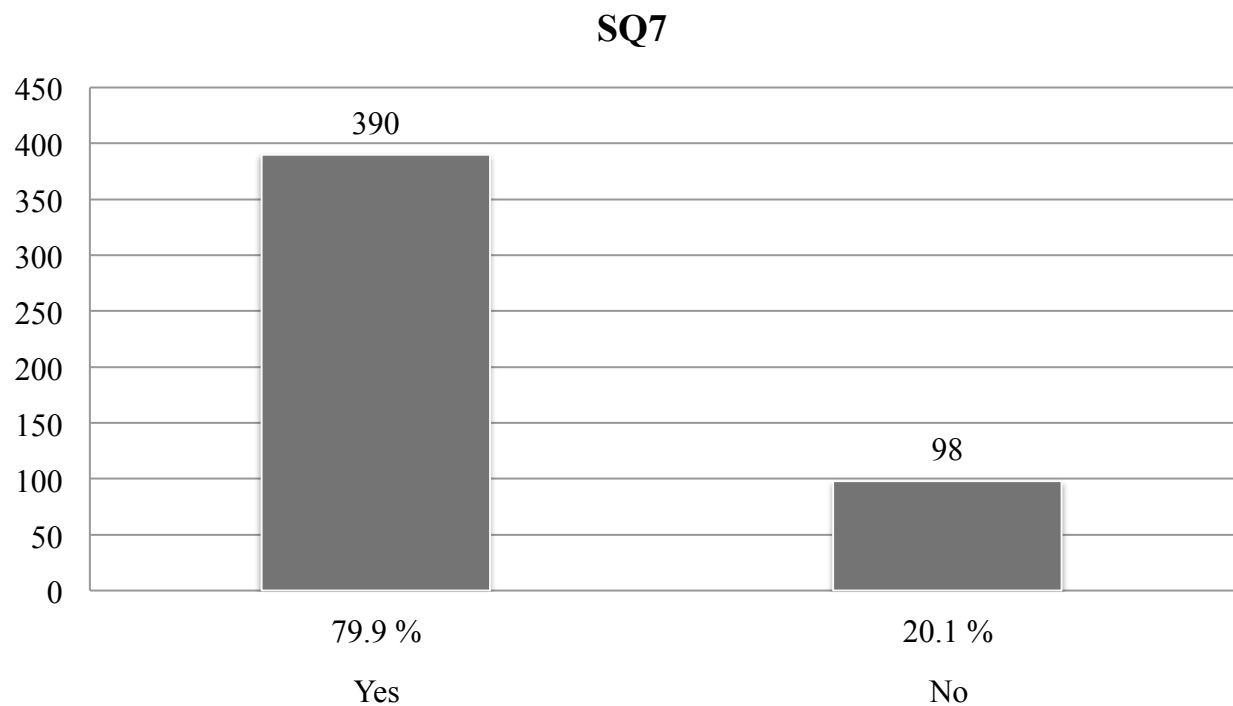


Figure 4.5 Bar graph of SQ7: Should schools be held accountable for providing the nationally recommended amount (60 minutes) of quality physical activity every day?

Table 4.30

SQ7: Test of Homogeneity of Variances

Should schools be held accountable for providing the nationally recommended amount (60 minutes) of quality physical activity every day?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
12.954	10	473	.000

Table 4.31

SQ7: Oneway ANOVA

Should schools be held accountable for providing the nationally recommended amount (60 minutes) of quality physical activity every day?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	6.909	10	.691	4.586	.000
Within Groups	71.249	473	.151		
Total	78.157	483			

Table 4.32

SQ7: Cross Tabulations

Should schools be held accountable for providing the nationally recommended amount (60 minutes) of quality physical activity every day?

	Respondent Type											
	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instr uctional Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of High Education</i>	<i>Total</i>
1 Yes	90.9	83.6	82.1	52.9	92.3	63.4	75.0	66.7	100.0	100.0	73.3	79.8
2 No	9.1	16.4	17.9	47.1	7.7	36.6	25.0	33.3	0.0	0.0	26.7	20.2
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(66)	(207)	(56)	(51)	(26)	(41)	(8)	(3)	(5)	(6)	(15)	(484)

Table 4. 33

SQ7: Chi Square Tests

Should schools be held accountable for providing the nationally recommended amount (60 minutes) of quality physical activity every day?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	42.782 ^a	10	.000
Likelihood Ratio	41.329	10	.000
Linear-by-Linear Association	5.138	1	.023
N of Valid Cases	484		

a. 8 cells (36.4%) have expected count less than 5. The minimum expected count is .61.

Survey Question 8 (SQ8) asked respondents to identify, either affirmatively or negatively, if schools should be held accountable for providing social-emotional learning opportunities for all students. Nearly three-fourths of Louisiana stakeholders responded

affirmatively to SQ8 (see Table 4.34 and Figure 4.6). This data set passed the test for normality using the Levene statistic; therefore, normality was assumed.

The ANOVA F statistic did not measure below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was accepted (see Table 4.36). There was no statistically significant variance between groups of means. Robust tests of equality of means could not be performed for SQ8 because at least one group had zero variance. Because SQ8 had greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.33). Checking, then, the Likelihood Ratio, I concluded that there was no statistically significant association between respondent type and rating.

Table 4.34

SQ8: Descriptive Statistics

Should schools be held accountable for providing social-emotional learning opportunities for all students?

	<i>(N)</i>	<i>Percent</i>
1 Yes	356	73.3
2 No	130	26.7
Total	486	100.0
No response: 58 cases		

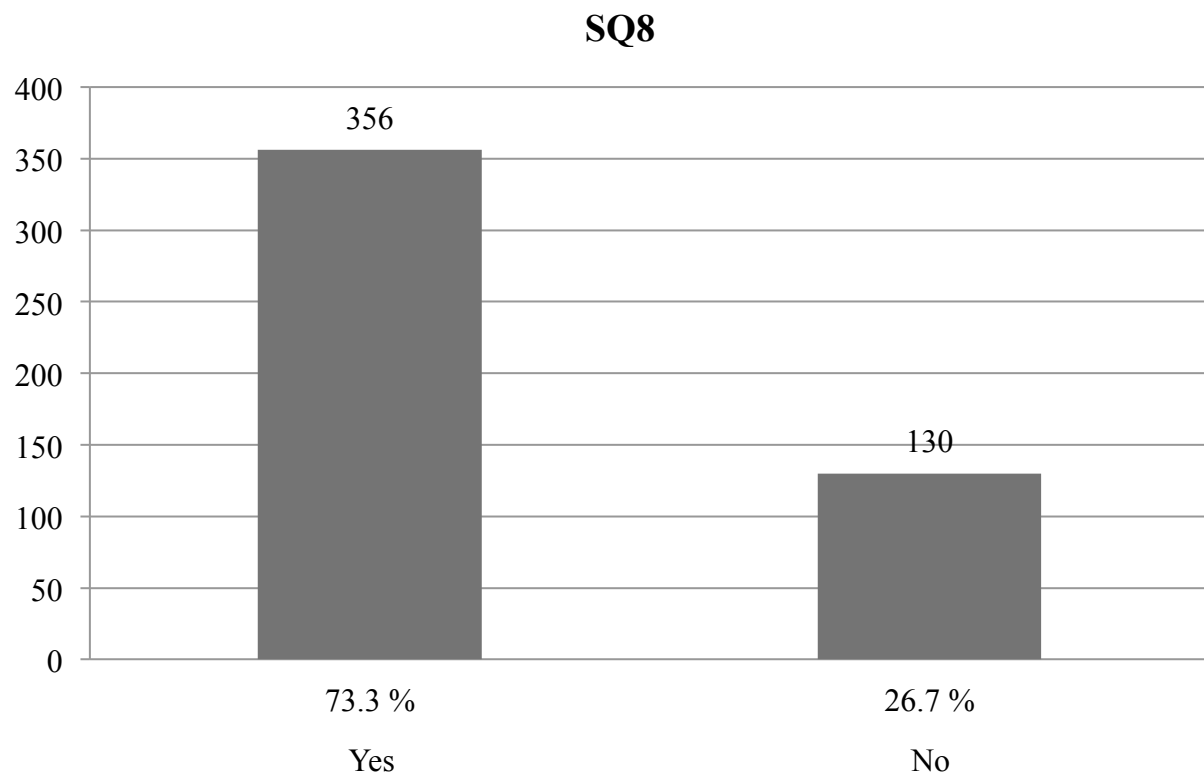


Figure 4.6 Bar graph of SQ8: Should schools be held accountable for providing social-emotional learning opportunities for all students?

Table 4.35

SQ8: Test of Homogeneity of Variances

Should schools be held accountable for providing social-emotional learning opportunities for all students?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
3.767	10	471	.000

Table 4.36

SQ8: Oneway ANOVA

Should schools be held accountable for providing social-emotional learning opportunities for all students?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	1.855	10	.185	.953	.484
Within Groups	91.683	471	.195		
Total	93.537	481			

Table 4.37

SQ8: Percentage Crosstabulations of Overall Quality of Louisiana's Accountability System by Respondent Type

Should schools be held accountable for providing social-emotional learning opportunities for all students?

	Respondent Type											
	Parent	Teacher	Teacher/Instructional Leader	Principal	Community Member	Local Education Agency Staff	Local Education Agency	State Education Agency Staff	Elected Official	Member of Business Community	Member of High Education	Total
1 Yes	75.8	75.6	73.2	62.7	81.5	63.4	75.0	100.0	60.0	83.3	86.7	73.7
2 No	24.2	24.4	26.8	37.3	18.5	36.6	25.0	0.0	40.0	16.7	13.3	26.3
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(66)	(205)	(56)	(51)	(27)	(41)	(8)	(2)	(5)	(6)	(15)	(482)

Table 4.38

SQ8: Chi Square Tests

Should schools be held accountable for providing social-emotional learning opportunities for all students?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	9.557 ^a	10	.480
Likelihood Ratio	9.957	10	.444
Linear-by-Linear Association	.009	1	.927
N of Valid Cases	482		

a. 8 cells (36.4%) have expected count less than 5. The minimum expected count is .53.

Survey Question 9 (SQ9) asked respondents to identify, either affirmatively or negatively, if indicators for non-academic skills (i.e. social, emotional, & physical health) should be included in Louisiana's test-based accountability system. Respondents were nearly split with 43.6% of responses supporting and 56.4% not supporting including indicators for non-academic skills (i.e. social, emotional, & physical health) in Louisiana's test-based accountability system (see Table 4.39 and Figure 4.7). This data set passed the test for normality using the Levene statistic (see Table 4.40); therefore, normality was assumed.

The ANOVA F statistic measured below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was rejected and the alternative hypothesis was accepted (see Table 4.41). There was statistically significant variance between groups of means. Robust tests of equality of means could not be performed for SQ9 because at least one group had zero variance. Because SQ9 had greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.43). Checking, then, the Likelihood Ratio, I concluded that there was sufficient evidence to support a statistically significant association between respondent type and rating.

Table 4.39

SQ9: Descriptive Statistics

Should indicators for non-academic skills (i.e. social, emotional, & physical health) be included in Louisiana's test-based accountability system?

	<i>(N)</i>	<i>Percent</i>
Yes	211	43.6
No	273	56.4
Total	484	100.0
No response: 60 cases		

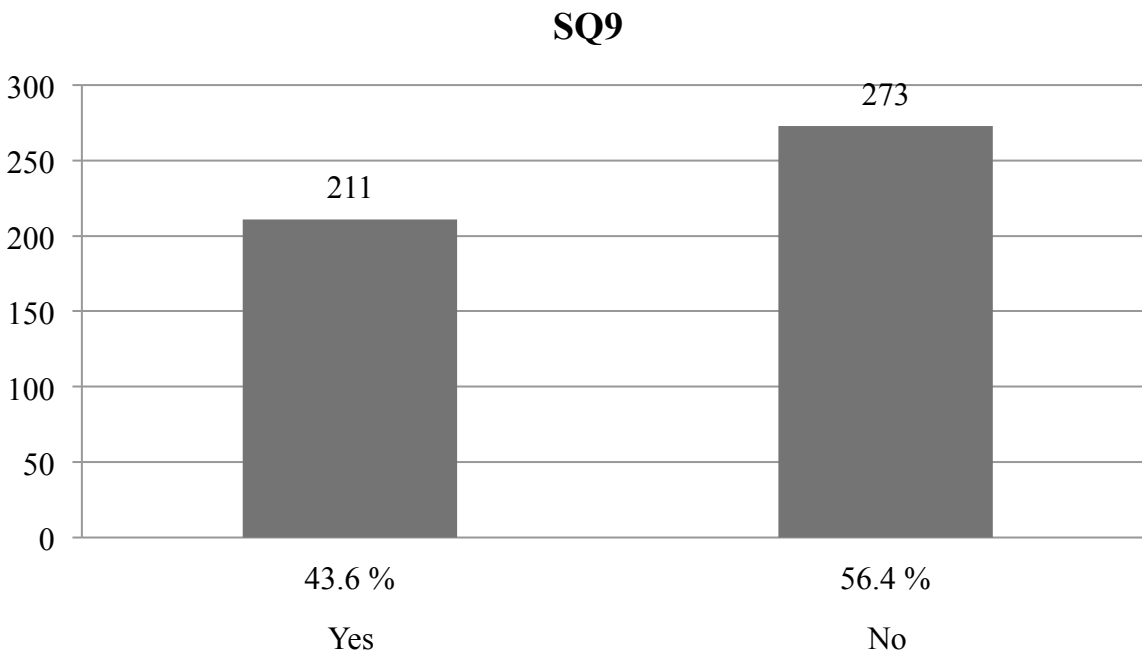


Figure 4.7 Bar graph of SQ 9: Should indicators for non-academic skills (i.e. social, emotional, & physical health) be included in Louisiana's test-based accountability system?

Table 4.40

SQ9: Test of Homogeneity of Variances

Should indicators for non-academic skills (i.e. social, emotional, & physical health) be included in Louisiana's test-based accountability system?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
18.607	10	469	.000

Table 4.41

SQ9: Oneway ANOVA

Should indicators for non-academic skills (i.e. social, emotional, & physical health) be included in Louisiana's test-based accountability system?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	6.264	10	.626	2.632	.004
Within Groups	111.603	469	.652		
Total	117.867	479			

Table 4.42

SQ9: Cross Tabulations

Should indicators for non-academic skills (i.e. social, emotional, & physical health) be included in Louisiana's test-based accountability system?

	Respondent Type											
	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instructional Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency Superintendent</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of Higher Education</i>	<i>Total</i>
1 Yes	54.5	44.9	49.1	27.5	48.1	20.0	50.0	0.0	25.0	80.0	57.1	43.3
2 No	45.5	55.1	50.9	72.5	51.9	80.0	50.0	100.0	75.0	20.0	42.9	56.7
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(66)	(207)	(55)	(51)	(27)	(40)	(8)	(3)	(4)	(5)	(14)	(480)

Table 4.43

SQ9: Chi Square Tests

Should indicators for non-academic skills (i.e. social, emotional, & physical health) be included in Louisiana's test-based accountability system?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	25.510 ^a	10	.004
Likelihood Ratio	27.772	10	.002
Linear-by-Linear Association	1.702	1	.192
N of Valid Cases	480		

a. 8 cells (36.4%) have expected count less than 5. The minimum expected count is 1.30.

Survey Question 10 (SQ10) asked respondents to identify, either affirmatively or negatively, if surveys of students' perception about their school experience should be included in Louisiana's test-based accountability system. Respondents differed by only five percentage points in their overall perception of including student perception data in Louisiana's test-based accountability system (see Table 4.10 and Figure 4.8). One hundred percent of state agency staff and 93.3% of members of higher education responded affirmatively, while the majority of school-based professionals (parents, teachers, teacher/instructional leaders, principals, local education agency staff, and local education agency superintendents), elected officials, and

members of the business community responded negatively. This data set passed the test for normality using the Levene statistic; therefore, normality was assumed.

The ANOVA F statistic measured below the acceptable level of significance ($\alpha = .05$) to accept the null hypothesis; therefore, the null hypothesis was rejected and the alternative hypothesis was accepted (see Table 4.46). There was statistically significant variance between groups of means. Robust tests of equality of means could not be performed for SQ10 because at least one group had zero variance. Because SQ10 had greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.43). Checking, then, the Likelihood Ratio, I concluded that there is sufficient evidence to support a statistically significant association between respondent type and rating.

Table 4.44

SQ10: Descriptive Statistics

Should surveys of students' perception about their school experience be included in Louisiana's test-based accountability system?

	<i>(N)</i>	<i>Percent</i>
1 Yes	230	47.6
2 No	253	52.4
Total	483	100.0
No response: 61 cases		

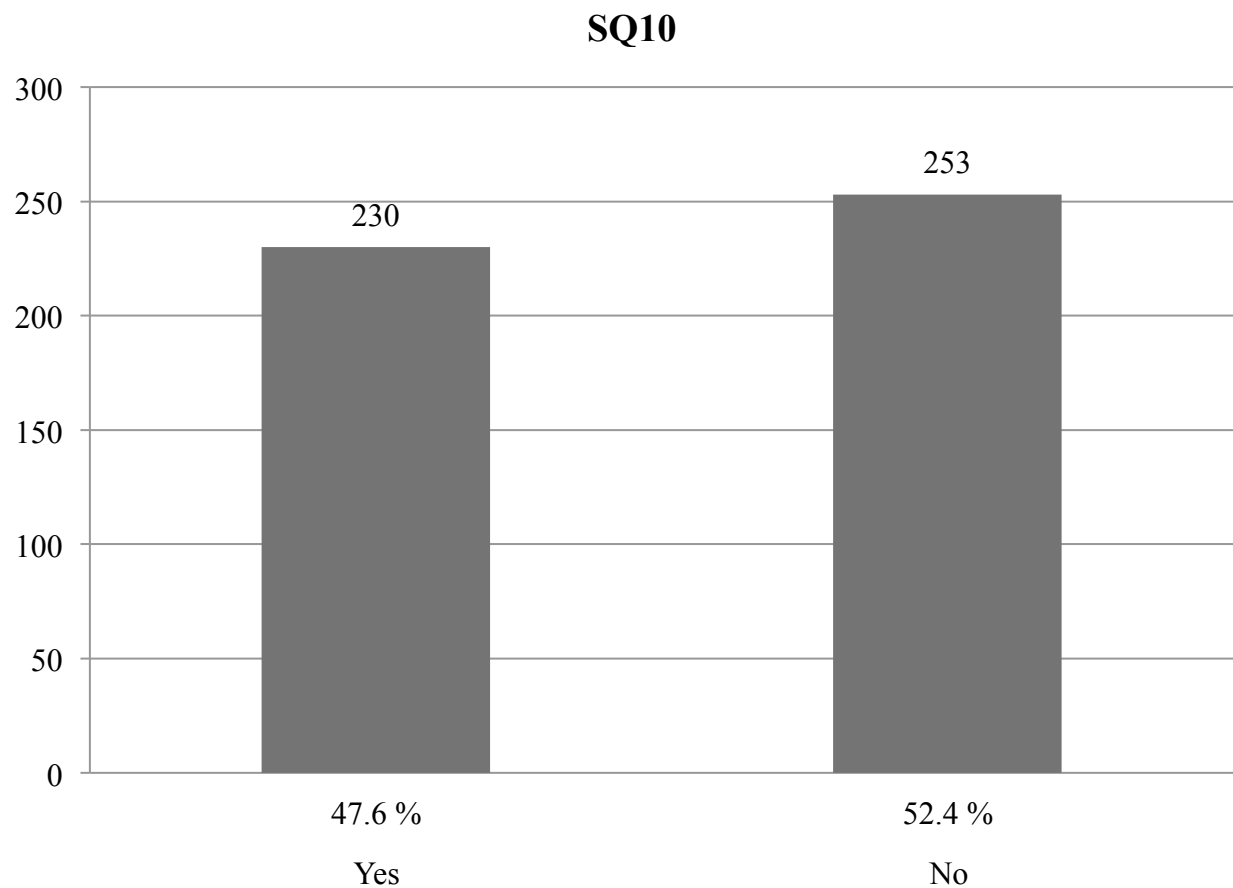


Figure 4.8 Bar graph of SQ10: Should surveys of students' perception about their school experience be included in Louisiana's test-based accountability system?

Table 4.45

SQ10: Test of Homogeneity of Variances

Should surveys of students' perception about their school experience be included in Louisiana's test-based accountability system?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
25.678	10	468	.000

Table 4.46

SQ10: Oneway ANOVA

Should surveys of students' perception about their school experience be included in Louisiana's test-based accountability system?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	8.351	10	.835	3.516	.000
Within Groups	111.169	468	.238		
Total	119.520	478			

Table 4.47

SQ10: Percentage Crosstabulations of Overall Quality of Louisiana's Accountability System by Respondent Type

Should surveys of students' perception about their school experience be included in Louisiana's test-based accountability system?

	<i>Respondent Type</i>											
	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instructional Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of High Education</i>	<i>Total</i>
1 Yes	56.1	49.0	37.0	39.2	66.7	25.0	42.9	100.0	40.0	33.3	93.3	47.8
2 No	43.9	51.0	63.0	60.8	33.3	75.0	57.1	0.0	60.0	66.7	6.7	52.2
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(66)	(206)	(54)	(51)	(27)	(40)	(7)	(2)	(5)	(6)	(15)	(479)

Table 4.48

SQ10: Chi Square Tests

Should surveys of students' perception about their school experience be included in Louisiana's test-based accountability system?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	33.469 ^a	10	.000
Likelihood Ratio	36.958	10	.000
Linear-by-Linear Association	.355	1	.551
N of Valid Cases	479		

a. 8 cells (36.4%) have expected count less than 5. The minimum expected count is .96.

Survey Question 11 (SQ11) asked respondents to identify, either affirmatively or negatively, if students' report card grades should be included in Louisiana's test-based accountability system. Of all survey items, this item was the most evenly split for respondents,

split by only one percentage point (see Table 4.49 and Figure 4.9). Forty-nine percent of Louisiana stakeholders responded affirmatively, while 51% responded negatively. State education agency staff and elected officials were most in opposition, while members of the business community and parents were most evenly split. Members of higher education were mostly in support of incorporating student report card grades in Louisiana's test-based accountability system. This data set passed the test for normality using the Levene statistic; therefore, normality was assumed.

The ANOVA *F statistic* measured below the acceptable level of significance ($\alpha = .05$) to reject the null hypothesis; therefore, the null hypothesis was rejected and the alternative hypothesis was accepted (see Table 4.51). There was statistically significant variance between groups of means. Robust tests of equality of means could not be performed for SQ11 because at least one group had zero variance. Because SQ11 had greater than 20% expected count less than five, the Pearson's Chi Square assumption had been violated (see Table 4.53). Checking, then, the Likelihood Ratio, I concluded that there was sufficient evidence to support a statistically significant association between respondent type and rating.

Table 4.49
SQ11: Descriptive Statistics
Should students' report card grades be included in Louisiana's test-based accountability system?

	(N)	Percent
1 Yes	239	49.4
2 No	245	50.6
Total	484	100.0
No response: 60 cases		

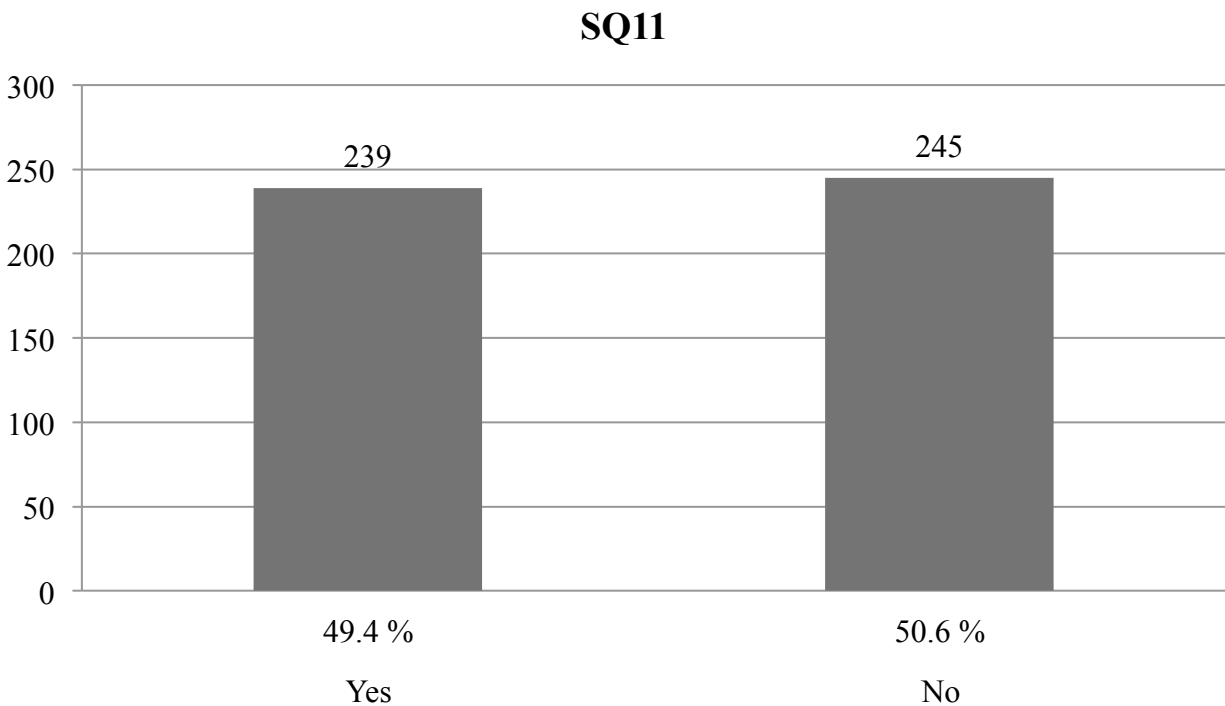


Figure 4.9 Bar graph of SQ11: Should students' report card grades be included in Louisiana's test-based accountability system?

Table 4.50

SQ11: Test of Homogeneity of Variances

Should students' report card grades be included in Louisiana's test-based accountability system?

<i>Levene Statistic</i>	<i>df1</i>	<i>df2</i>	<i>Sig.</i>
12.550	10	470	.000

Table 4.51

SQ11: Oneway ANOVA

Should students' report card grades be included in Louisiana's test-based accountability system?

	<i>Sum of Squares</i>	<i>df</i>	<i>Mean Square</i>	<i>F</i>	<i>Sig.</i>
Between Groups	6.2	10	.621	2.558	.005
Within Groups	114.019	470	.243		
Total	120.225	480			

Table 4.52

SQ11: Cross Tabulations

Should students' report card grades be included in Louisiana's test-based accountability system?

	Respondent Type											
	<i>Parent</i>	<i>Teacher</i>	<i>Teacher/Instr uctural Leader</i>	<i>Principal</i>	<i>Community Member</i>	<i>Local Education Agency Staff</i>	<i>Local Education Agency</i>	<i>State Education Agency Staff</i>	<i>Elected Official</i>	<i>Member of Business Community</i>	<i>Member of High Education</i>	<i>Total</i>
1 Yes	48.5	57.3	41.1	43.1	59.3	27.5	25.0	0.0	20.0	50.0	69.2	49.3
2 No	51.5	42.7	58.9	56.9	40.7	72.5	75.0	100.0	80.0	50.0	30.8	50.7
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
(N)	(66)	(206)	(56)	(51)	(27)	(40)	(8)	(3)	(5)	(6)	(13)	(481)

Table 4.53

SQ11: Chi Square Tests

Should students' report card grades be included in Louisiana's test-based accountability system?

	<i>Value</i>	<i>df</i>	<i>Asymptotic Significance (2-sided)</i>
Pearson Chi-Square	24.828 ^a	10	.006
Likelihood Ratio	26.585	10	.003
Linear-by-Linear Association	2.875	1	.090
N of Valid Cases	481		

a. 8 cells (36.4%) have expected count less than 5. The minimum expected count is 1.48.

Ordinal Ranking

Based on a survey of the general public conducted by Rothstein, Jacobsen, and Wilder (2008), Survey Question 12 (SQ12) asked participants to rank eight goals of education in order of importance. These skills included the arts and literature, basic academic skills, citizenship, critical thinking, emotional health, physical health, preparation for skilled work, and social skills and work ethic. Notably, the top ranked goals of education from 2008 remained positioned similarly in 2015 with Basic Academic Skills (1), Critical Thinking (2), Social Skills and Work Ethic (3). Also notable was the lack of accounting for critical thinking and social skills and work ethic in Louisiana's test-based accountability system. Physical health was ranked least important of all goals of education.

Table 4.54

SQ12: Descriptive Statistics

In order of importance, with 1 being most important and 8 being least important, rank the following goals of education⁴.

	1	2	3	4	5	6	7	8	Rating Average	Response Count
Basic Academic Skills	227	75	45	23	20	14	4	6	2.09	414
Critical Thinking	92	131	64	48	27	27	10	13	2.93	412
Social Skills and Work Ethic	27	65	69	88	67	47	43	15	4.17	421
Citizenship	17	34	66	57	71	55	55	59	4.95	414
Preparation for Skilled Work	23	41	56	63	50	58	61	66	4.97	418
Emotional Health	24	28	31	55	65	71	66	71	5.29	411
The Arts and Literature	12	22	63	49	57	57	69	100	5.48	429
Physical Health	7	23	31	35	58	82	98	79	5.78	413
TOTAL										432

No response: 112

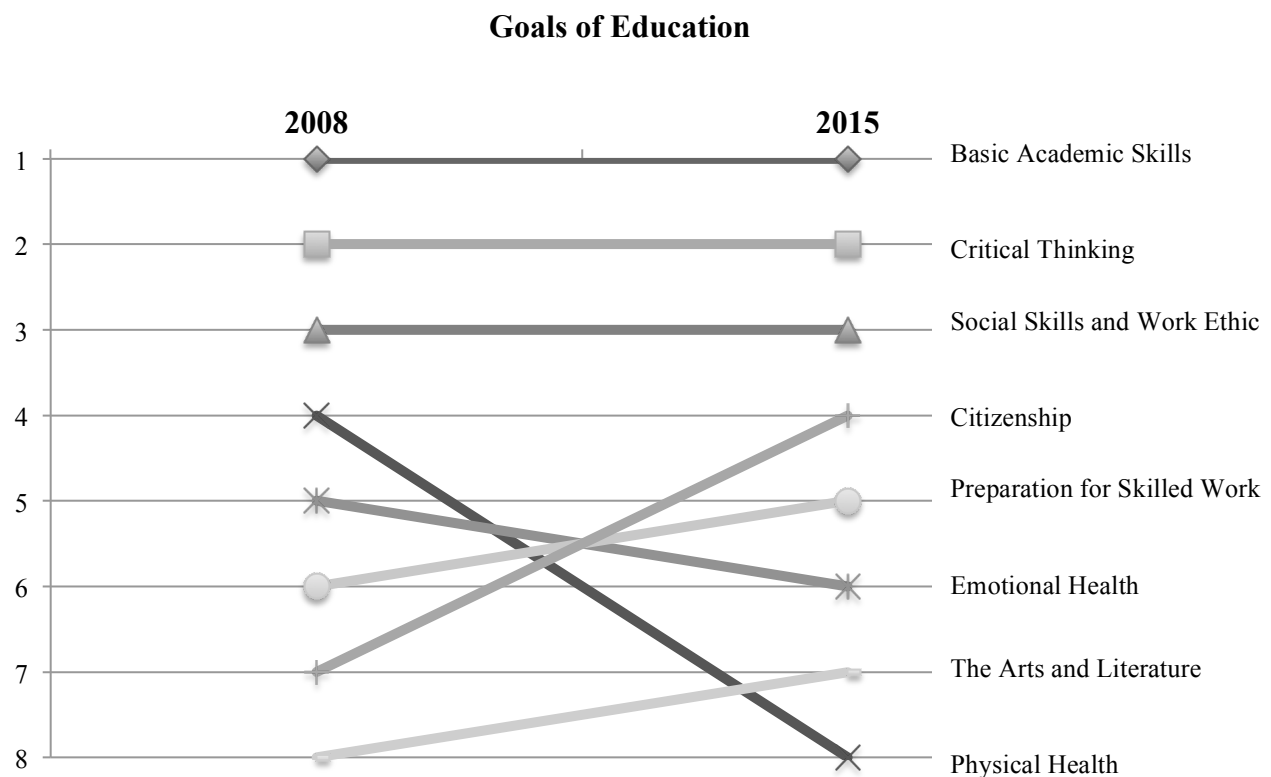


Figure 4.10 Longitudinal plot of SQ12: In order of importance, with 1 being most important and 8 being least important, rank the following goals of education.

⁴ Based on Rothstein, Jacobsen, & Wilder (2008).

Based on the work of Cassel and Kolstad (1990, 1999), Survey Question 14 (SQ14) asked participants to rank 10 skills in order of importance. These skills included creative thinking, goal-setting/motivation, interpersonal skills, leadership, listening skills, oral communications, personal career development, problem solving, teamwork, and writing. Cassel and Kolstad's work included surveys of chief executive officers of Fortune 500 companies, while this survey was not limited to such participants.

In 2015, respondents ranked Problem Solving as the most important skill in education. Surprisingly, Listening Skills, Oral Communications, and Creative Thinking rose in rank to second, third, and fourth most important skills, respectively. These skills are not currently factored into Louisiana's test-based accountability system. Writing rose from the least important skill to the fifth most important skill, with Interpersonal Skills and Teamwork following in the sixth and seventh ranks. Goal Setting/Motivation was ranked eighth most important, Leadership ranked ninth most important, and Potential Career Development fell from sixth most important in 1999 to the least most important skill in 2015. The skills of Writing and Teamwork experienced the most striking changes over time, and warrant further analysis.

It is worth noting the significant difference between survey respondents by year. The 1970 and 1999 surveys included CEO's of Fortune 500 Companies, while the 2015 survey included a random selection of Louisiana stakeholders. Data collection included too few respondents categorized as members of the business community to parse out for replicability, so the sample size included all Louisiana stakeholders. This illustrated the stark contrasts between values of leaders in Fortune 500 businesses with those of Louisiana stakeholders.

Table 4.55

SQ14: Descriptive Statistics

In order of importance, with 1 being most important and 10 being least important, rank the following skills⁵.

	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>	<i>10</i>	<i>Rating Average</i>	<i>Response Count</i>
Problem Solving	80	79	61	71	33	30	26	15	6	4	3.57	405
Listening Skills	77	66	54	54	43	42	31	11	12	8	3.89	398
Oral Communications	42	64	61	60	42	37	41	35	17	1	4.39	400
Creative Thinking	53	41	44	30	42	35	42	39	34	35	5.24	395
Writing	37	42	48	50	54	48	34	30	33	38	5.26	414
Interpersonal Skills	47	35	36	46	37	52	44	43	39	24	5.33	403
Teamwork	21	45	42	45	52	55	47	45	40	17	5.44	414
Goal Setting	37	20	33	24	42	50	55	70	48	17	5.88	396
Leadership	14	9	20	17	34	33	44	68	90	67	7.22	396
Potential Career Development	8	11	14	10	24	18	34	39	69	173	8.1	400
TOTAL												420

No response: 124

⁵ Based on Cassel & Kolstad (1990, 1999).

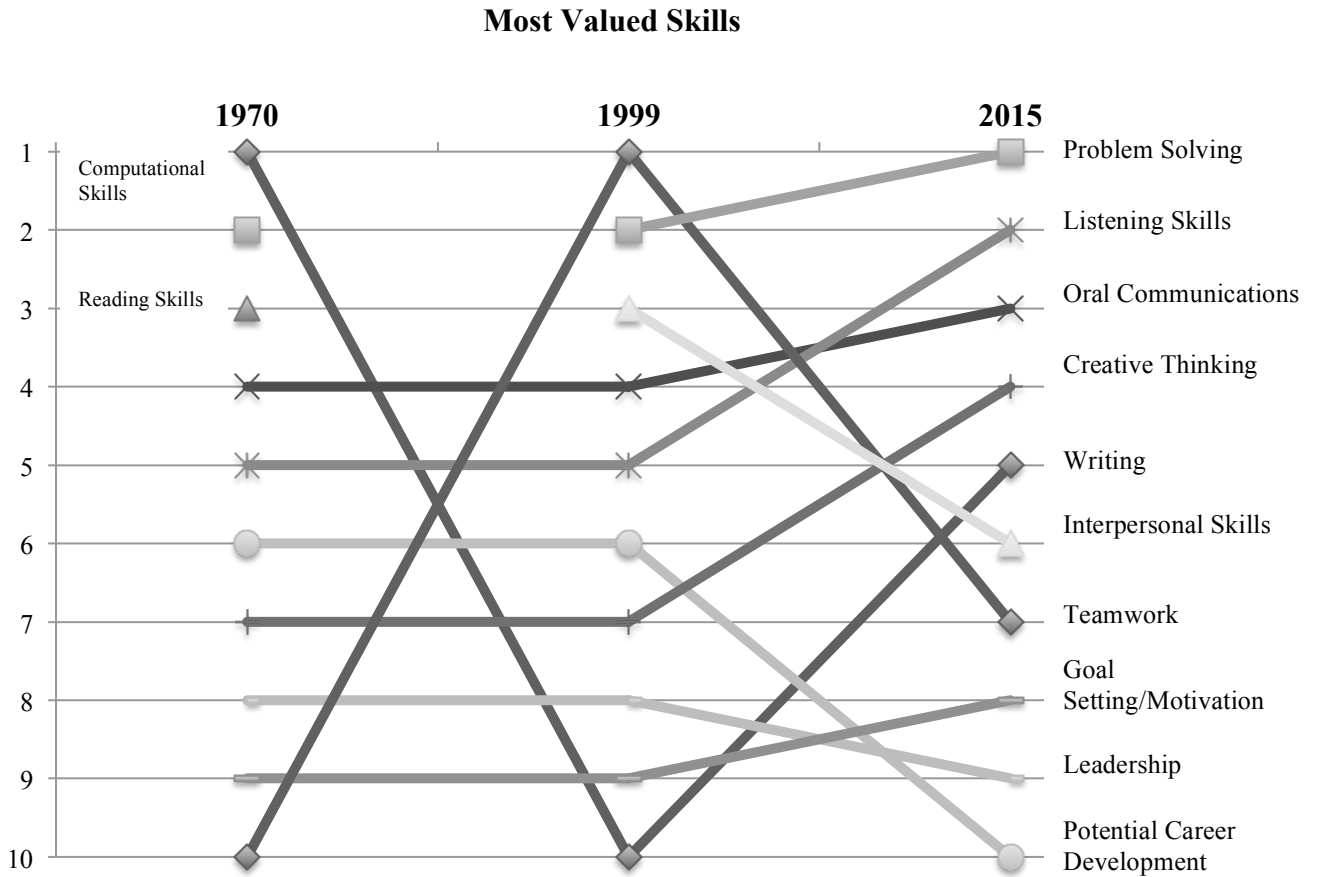


Figure 4.11 Longitudinal plot of SQ14: In order of importance, with 1 being most important and 10 being least important, rank the following skills.

Constructed Response

Survey Questions 3 through 15 afforded respondents with the opportunity to provide constructed responses about their ranking preferences. These comments are included in Appendix E. Timing of this study excluded text-analysis for these items from this study.

Survey questions 16 through 19 asked participants to respond to each question or prompt by writing in the space provided. A total of 960 responses were collected for these items and 3,886 words were analyzed for unitized codes, categorized themes, and word frequencies.

Survey Question 16 (SQ16) asked respondents to identify the top three factors that contribute to student learning outcomes for students in Louisiana. SQ16 received a total of 310 responses including 781 different words (see Appendix E). Textual analysis produced 35 highest-frequency nouns (see Table 4.56) and 33 emergent themes (see Table 4.57). The factors most included were teachers, parents, school, and students.

Textual analysis identified variation in meaning of the theme Teacher Quality. Some respondents used terms such as understanding, caring, and loving to describe teacher quality, while other respondents used terms such as depth of knowledge, quality of training, and skill/ability to describe Teacher Quality. Additionally, textual analysis identified variation in the theme Skill Development. Some respondents included responses indicative of basic skills attainment, such as reading comprehension, while other respondents included terms such as critical thinking and problem solving to describe Skill Development; still others included skills such as interpersonal and communication skills. Three cases included repeated terms for their top three factors; in two cases, the repeated word was poverty and in one case the repeated word was home.

Table 4.56

SQ16: Word Frequencies^a

In your opinion, what top three factors contribute to student learning outcomes for students in Louisiana?

<i>Word</i>	<i>Count</i>	<i>Percent</i>
Teachers	77	2%
Teacher	76	2%
Home	64	2%
Parental	62	2%
Involvement	51	2%
Support	51	2%
School	48	1%
Student	47	1%
Environment	43	1%
Students	40	1%
Quality	38	1%
Motivation	34	1%
Life	32	1%
Learning	31	1%
Curriculum	30	1%
Family	28	1%
Parents	26	1%
Skills	26	1%
Poverty	25	1%
Classroom	22	1%
Parent	21	1%
Ability	19	1%
Schools	19	1%
Discipline	18	1%
Effective	17	1%
Teaching	17	1%
Test	17	1%
Accountability	16	0%
Instruction	16	0%
Resources	15	0%
Education	14	0%
Work	14	0%
Knowledge	13	0%
Status	13	0%
Health	12	0%

Total (N) = 781

a. Table includes highest-frequency nouns.

Table 4.57

SQ16: Themes

In your opinion, what top three factors contribute to student learning outcomes for students in Louisiana?

<i>Theme</i>	<i>Frequency</i>
Teacher Quality	141
Home Environment	100
Parental/Family Involvement	100
Student Work Ethic/Motivation ^a	78
Curriculum and Instruction ^b	72
Skill Development	54
School Culture	45
Socio-Economic Status	28
Student Discipline/Behavior	25
Basic Health (Mental, Physical, Emotional)	24
Resources/Funding	24
Poverty	22
Leadership Quality	18
Value of Education	18
Teacher/Student Relationships	12
Early Learning	12
Community	11
Attendance	10
Student Ability	10
Class Size	9
Access to Opportunity	6
Home/School Connection	6
Readiness to Learn	5
State law & local policies	5
Environment	5
Time	4
Tests	3
Recess/Free Play	3
High Expectations	3
Accountability	2
Communication	2
Peer Influence	2
Politics	2

a. Four cases used the term “accountability” for students. Two cases used the term “confidence” for students.

b. Two cases used the term “standards.”

Survey Question 17 (SQ17) asked respondents to identify, if any, limitations to

Louisiana’s test-based accountability system. SQ17 received a total of 295 responses including

1,127 different words (see Appendix E). Textual analysis produced 37 highest-frequency nouns (see Table 4.58) and 29 emergent themes (see Table 4.59). The most included factors were test, students, teachers, and school.

Tests and testing led both word frequency and theme analysis. The coding for this theme included a variety of concern with testing, such as too much time testing, poor quality of the tests, as well as the impact of testing on the perceived value of education. Two cases indicated concern with test security. They included the length of the testing window for the End-of-Course test system and the susceptibility of items to invalidity due to exposure. One respondent noted the impact of Louisiana's test-based accountability system on race relations. "It drives our best teachers to higher performing schools. It segregates school by driving parents with more resources to live in zones served by A and B schools."

Table 4.58

SQ17: Word Frequencies^a

In your opinion, what, if any, are limitations to Louisiana's test-based accountability system?

<i>Word</i>	<i>Count</i>	<i>Percent</i>
Test	110	2%
Students	84	2%
Tests	41	1%
Teachers	36	1%
Testing	35	1%
School	34	1%
Time	34	1%
Accountability	31	1%
School	30	1%
Learning	26	1%
Student	24	0%
System	20	0%
Children	15	0%
Parents	15	0%
Scores	14	0%
Skills	14	0%
Account	13	0%
Factors	13	0%
Measure	13	0%
Results	13	0%
Score	11	0%
Standardized	11	0%
Education	10	0%
Growth	10	0%
Teaching	10	0%
Accountable	9	0%
Teach	9	0%
Ability	8	0%
Academic	8	0%
Assessment	8	0%
Assessments	8	0%
Standards	8	0%
Individual	7	0%
Life	7	0%
Poverty	7	0%
Stakes	7	0%
Success	7	0%

Total (N) = 1,127

a. Table includes highest-frequency nouns.

Table 4.59

SQ17: Themes

In your opinion, what, if any, are limitations to Louisiana's test-based accountability system?

<i>Theme</i>	<i>Frequency</i>
Testing	71
Limited Scope	35
Out-of-School Factors	26
Instability of System ^a	21
Standardization/Lack of Accounting for Individual Needs and Diversity	20
Negative Impact on Instructional Time	19
Does Not Reflect Student Learning	18
Lack of Real World Applicability	16
Accountability for Schools and Teachers Only ^b	16
Does Not Reflect Teacher or Student Effort	16
Lack of Portraying Whole Child Development	14
Does Not Account for Growth	12
Results Tied to Teacher Evaluation	6
Students Who Do Not Test Well	6
Lack of Teacher Input	4
Does Not Accurately Predict Future Success	4
Susceptibility to Cheating	4
Not All Grades and Subjects Are Accountable	4
Misalignment of Tested Content with State Standards	3
Inadequate Resources/Funding	3
Testing Conditions	3
Difficulty Quantifying Learning	2
All Schools Held to Same Standards	2
Data From System Not Received in a Timely Manner	2
High-Stakes	2
Double Standard for Students and Schools	2
Testing Opt Out	1
Low Cut Scores for Students	1
Societal Impact	1

a. Three cases used the term “not valid,” five cases used the term “not reliable,” and three cases used the term “bias.”

b. Includes needs for accountability for teachers or students.

Survey Question 18 (SQ18) asked respondents to identify what they would do if they were afforded the opportunity to enhance Louisiana's test-based accountability system. SQ18 received a total of 287 responses including 1,225 different words (see Appendix E). Timing of this study excluded analysis and reporting of textual analysis for themes of SQ18.

Table 4.60

SQ18: Word Frequencies^a

If you were afforded the opportunity to enhance Louisiana's test-based accountability system, what would you do?

<i>Word</i>	<i>Count</i>	<i>Percent</i>
Test	115	2%
Students	85	2%
Testing	47	1%
Tests	47	1%
Teachers	40	1%
Accountability	39	1%
School	39	1%
Student	29	1%
Schools	28	1%
System	21	0%
Scores	19	0%
Grade	18	0%
Growth	16	0%
Standards	15	0%
State	15	0%
Teacher	15	0%
Eliminate	14	0%
Skills	14	0%
Assessments	13	0%
Data	13	0%
Measure	13	0%
Accountable	12	0%
Teach	11	0%
Factors	11	0%
Provide	11	0%
Remove	11	0%
Results	11	0%
Assessment	10	0%
Create	10	0%
Education	10	0%
Scrap	10	0%
Teaching	10	0%
Work	10	0%
Act	9	0%
Allow	9	0%
Focus	9	0%
Grades	9	0%

Total (N) 1,225

a. Table includes highest-frequency nouns and verbs.

Survey Question 19 (SQ19) provided respondents with the opportunity to provide any additional feedback. SQ19 received a total of 68 responses including 753 different words (see Appendix E). Due to the timing of this study excluded analysis and reporting of textual analysis for themes of SQ19.

Table 4.61
SQ19: Word Frequencies^a
Any additional feedback?

<i>Word</i>	<i>Count</i>	<i>Percent</i>
Test	23	1%
Students	21	1%
School	20	1%
Teachers	17	1%
Testing	17	1%
Children	14	1%
Schools	14	1%
Teacher	12	1%
Learning	10	1%
Education	9	1%
Kids	9	1%
System	9	1%
Accountability	8	1%
Tests	8	1%
Time	8	1%
Child	7	1%
Student	7	1%
Teach	7	1%
Teaching	6	1%
Classroom	5	1%
College	5	1%
Grade	5	1%
State	5	1%

Total (N) = 753

a. Table includes highest-frequency nouns and verbs.

Phase II

The second research question asked *to what extent policymakers in Louisiana were receptive to piloting an advanced system of accountability for Louisiana that includes reported measures in the affective and psychomotor domain to improve student learning outcomes*. To investigate this question, I used a semi-structured interview protocol with a series of questions and applicable probes. Interviews were conducted between May 2, 2016 and May 27, 2016.

Twenty-three lawmakers currently serving on the state Education Committee in either the House or the Senate met the criteria for participation and were invited to participate. Of the twenty-three lawmakers invited to participate, five responses were received. Two lawmakers agreed to participate. Two lawmakers declined the invitation due to the timing of the study during the second active legislative session, while another declined participation due to former experience with a breach of confidentiality with a previous study similar in nature. Interview data collected included 130 minutes of recorded interview time transcribed into 33 single-spaced pages of transcripts that included 5,826 words.

Participants

Elected officials currently serving on a state education committee in either the House of Representatives or Senate were selected for participation. Although throughout this section, I use variations of the pronoun “he” to refer to each interviewee, the pronouns “he,” “him,” and “his” are used purely as standard reference to the interviewees participating in this study and are not indicative of a participant’s gender or gender identity. Sampling for Interview research followed a purposive design.

Participant A

Upon request of the interviewee, the interview was conducted in a local food establishment. The environment was not the most conducive to the interview process as it produced significant background noise. Participant A had served multiple years on the Education Committee. He was responsive to the interview questions and probes and stayed beyond the requested interview time to discuss additional educational topics of interest. Participant A displayed a strong respect for professional educators, indicating more than once that he was not a professional educator. Thrice he asked for my opinion on specific topics pertinent to the discussion. I gathered a sense of earnestness from him to make changes for improved conditions for education in the state.

More than once, Participant A indicated an interest in practicality over theory—he appeared disgruntled with the system of higher education and teacher preparation programs. His second statement during our interview revealed his frustration with the state Board of Regents. He said:

...she's probably mad at me, because I've been pressing hard, saying "Okay, listen, hell with all the theory, heck with anything. Give them the basics of two years, and put them out as a mentor, put them in a classroom", because too many teachers, once they get into the classroom, they're not prepared...because I'm telling you it infuriates me to lose teachers early, only because they're intimidated, because they're not prepared. And it's not a big deal, it's just that-- don't theories [sic], we'll call them theories, put them out there where the real situation is, and allow them to at least get acclimated, to where when it comes time for them to go in the classroom, you don't chuck them a couple of books, chuck them some keys, and say "Go to it."

When asked the first question which sought to uncover his overall perception of Louisiana's test-based accountability system, he responded, "Well, that's hard to answer. I'm for it. There may be ways we can tweak it and make it better...." When I asked his perception of the impact of test-based accountability on student achievement in Louisiana, he said, "I would hope

that it's improved it. Because I believe that so long as you have a goal, so long as you raise expectations, anybody can meet them. So hopefully it is going where it's supposed to be.”

His perception of the top factors impacting student learning outcomes were quality leadership, school culture, and teacher preparation. When asked about the most challenging issue for policymaking, he responded “all of them”. When asked how he might improve Louisiana’s test-based accountability system, he indicated that the growth factor should be included. When asked about his willingness to pilot an advanced system of test-based accountability that includes indicators in the cognitive, affective, and psychomotor domains he responded, “Why not? I mean, I’ll try anything. I really don’t care. I’m going to try each and everything that I can to try to help.” Participant A demonstrated a particular respect for the profession and for the work of educators. “The profession, it’s just so, so brutal. So brutal to the teachers and so brutal to the kids. It’s just not fair. Just make sure that people have a chance, that’s all I’m for.”

Politics was a strong theme for this participant. At one point in the interview, he asked me to turn off the recorder for an off-the-record description of personal experience with the political process. He indicated the challenge of legislating without clear direction from the Governor.

What is our priority? I mean, is our priority putting new bridge over there, which is a problem. Or is it health or is it Higher Ed or is it K through 12th, I don't know what our priorities are so every day I got to vote. I got to make decisions where the money goes, but you can't be at all things to all people. You've got to decide...

Participant B

At the request of the interviewee, the interview was conducted in a local food establishment near the state Capitol. The environment, again, was not the most conducive to the interview process as it produced significant background noise. Participant B had served fewer

years on the Education Committee than Participant A, but had multiple years serving in government. When asked his perception of Louisiana's test-based accountability system, he responded:

Generally speaking, I think it's good. I think it's positive. It has to be, by its very nature, simple. And it has to reflect an educational purpose. Because at the end of the day, what you're trying to do is transmit some metric of quality to your parent - basic taxpaying parent - who has a child in the system and wants to make an informed decision about their child's future. These would be the system in which they're interacting at the time. So everybody understands an A is good, an F is really bad. You never wanted to bring Fs home and show your dad, in most cases. So people understand that. And I think to someone who is in the education system, it may seem like an oversimplification, but I think it has to be because it has to be succinct.

What I have learned is your basic taxpayer out there is so wrapped up around paying their mortgage, paying their kids' operating expenses, buying the car, paying the car note, paying for insurance, all of life's little expenses as they go along. They don't have a lot of time. Plus, they're worried about losing their job, right? Because the economy's so bad. There's not a lot of time for reflection and analysis of deeply complicated and perhaps, in some cases, contradictory metrics and statistics and detail. They know the school's an A, they know they're doing good. They know it's a D or an F, they got to make some alternative arrangements.

He described an anecdote of a friend of his who didn't pass a licensing exam in a particular field, illustrating the high stakes nature of tests and their limitation on upward mobility.

The pressure built...and did that measure whether he was going to be a good professional. No, not necessary because he was a darn good professional, but he just—but the more failure he experienced, the more pressure he encountered...the pressure was excruciating.

At the same time, he acknowledged the difficulty of measuring attributes not empirically derived.

It would be very difficult to create an assessment that would measure some of those human qualities without opening yourself up to criticism that somehow the testing tool is biased and slanted.

On social-emotional skills he responded, “So it's stuff like that that I don't know why we're not teaching it.” The top three factors impacting student achievement, from his perspective are parent involvement, teacher quality, and students’ basic health. On piloting an advanced system, he replied, “Oh, yeah. No question. I'd agree. I think we should, to change.” When asked the most challenging topic of for policy-making, he responded with “charter schools. Well, it's a complicated process, but the politics behind it are particularly difficult.” However, when I probed to uncover his perceptions about any limitation to the current test-based accountability system, he identified race as a key factor.

Well, the elephant in the room is the racial aspect. No question about it. You know, if you uh, the uh, you can almost, dealing with.

And really, it's typified by-- you should go back and look at the tape of that education committee meeting about two weeks ago. It devolved into this screaming match telling us all about how charter schools are equivalent to the Tuskegee experiments where they exposed those guys to syphilis and then didn't treat them, just watched to see how they-- we had to call in the state police. We got the whole history of slavery, Jim Crow, people screaming and hollering at us up there in the education committee. And all we were trying to do was saying, "Here, here's your schools."

But I think that from the standpoint of assessments and getting buy-in, I think the original question is what's the hardest thing about bringing consensus? And that's really a hard thing because you have members of the committee and members of the legislature that are saying something, and you're saying something...And the public transmitting them back to you. They're all looking at you like, "Why are you telling me this?" You know, like I was somehow the overseer, and they were down there picking cotton. It's that kind of thing, "Why are you telling me?" I don't know how you fold that into your deal, but I think that nobody wants to talk about that. I'm glad this is anonymous.

Table 4.62
Code Frequencies

<i>Code Number</i>	<i>Code</i>	<i>f A</i>	<i>f B</i>	<i>Total</i>
1	Politics	30	4	34
2	Test-Based Accountability Perception	4	12	16
3	Reliance on Experts	12	0	12
4	Ethnicity Factor	0	12	12
5	Respect for the Profession	10	1	11
6	I Don't Know	10	0	10
7	Teacher Quality	6	2	8
8	Leadership Quality	7	0	7
9	Teacher Preparation	7	0	7
10	Charter Schools	2	5	7
11	Policy-Making	4	3	7
12	School Culture	6	0	6
13	Advanced Test-Based Accountability Pilot	1	4	5
14	Number of Tests	4	1	5
15	Students First	5	0	5
16	Difficulty of Teaching	4	0	4
17	Parental Engagement	0	4	4
18	Self-Described Unpopularity	4	0	4
19	Testing Time	3	0	3
20	Vouchers	2	1	3
21	Practicality Over Theory	3	0	3
22	Physical Health	1	2	3
23	Priorities Set by State Government Leadership	3	0	3
24	Complexity of Education	1	1	2
25	Early Childhood	2	0	2
26	Education for Upward Mobility	2	0	2
27	Education Savings Accounts	2	0	2
28	Funding	2	0	2
29	Gifted and Talented	2	0	2
30	Higher Education in Louisiana	1	1	2
31	Test Quality	1	1	2
32	Teachers Unions	1	0	1
33	Advanced Test-Based Accountability Growth Factor	1	0	1
34	Autonomy for School Leaders	1	0	1
35	Complexity of Adjusting from Slave- to Free- Men society	0	1	1
36	Difficulty Legislating	1	0	1
37	Failure of Men in Society	0	1	1
38	Charter Schools in New Orleans	1	0	1
39	Personal Experience as an Educator	0	1	1
40	Praise for State Superintendent of Education	1	0	1

(Table 4.62 continued)

<i>Code Number</i>	<i>Code</i>	<i>f A</i>	<i>f B</i>	<i>Total</i>
41	Quantitative Accountability Measures	0	1	1
42	Quorum Challenge	1	0	1
43	School Improvement Perception	1	0	1
44	Self-Described Ineffectiveness	1	0	1
45	Social Skills Development	0	1	1
46	Impact of Test-Based Accountability on Student Achievement	0	1	1
TOTAL		151	60	211

Co-occurrences

Co-occurrences refer to the context of information. They do not distinguish the meaning of the association, but they can illustrate an association. To better understand the meaning of the association, researchers must further investigate the quotation in which the codes are co-occurring.

Table 4.63
Co-Occurring Code Frequencies

<i>Code Number</i>	<i>Code</i>	<i>f Co-Occurrences</i>
1	Difficulty Legislating	0.93
2	Priorities Set by State Government Leadership	0.92
3	Funding	0.84
4	Policy-Making	0.46
5	I Don't Know	0.32
6	Reliance on Experts	0.29
7	Teacher Preparation	0.27
8	Politics	0.23
9	Ethnicity Factor	0.22
10	Teacher Quality	0.21
11	Students First	0.19
12	Vouchers	0.18
13	Charter Schools	0.14
14	Parental Engagement	0.14
15	Difficulty of Teaching	0.13
16	Test-Based Accountability Perception	0.13

(Table 4.63 continued)

<i>Code Number</i>	<i>Code</i>	<i>f Co-Occurrences</i>
17	Respect for the Profession	0.11
18	Leadership Quality	0.08
19	School Culture	0.08
20	Complexity of Education	0.03
21	Gifted and Talented	0.03
22	Advanced Test-Based Accountability- Growth Factor	0
23	Advanced Test-Based Accountability Pilot	0
24	Autonomy for School Leaders	0
25	Complexity of Slave- to Free- Men Society	0
26	Early Childhood	0
27	Education for Upward Mobility	0
28	Education Savings Account	0
29	Failure of Men in Society	0
30	Higher Education in Louisiana	0
31	Impact of Test-Based Accountability on Student Achievement	0
32	Charter Schools in New Orleans	0
33	Number of Tests	0
34	Personal Experience as an Educator	0
35	Physical Health	0
36	Practicality over Theory	0
37	Praise for State Superintendent of Education	0
38	Quantitative Accountability Measures	0
39	Quorum Challenge	0
40	School Improvement Perception	0
41	Social Skills Development	0
42	Teachers Unions	0
43	Test Quality	0
44	Testing Time	0
45	Self-Described Unpopularity	0
46	Self-Described Ineffectiveness	0

Table 4.64
Co-Occurring Code Power

<i>Code 1</i>	<i>Code 2</i>	<i>Strength of co-occurrence</i>
Difficulty Legislating	Funding	0.50
Difficulty Legislating	Priorities set by State Government Leadership	0.33
Funding	Priorities set by State Government Leadership	0.25
Policy-Making	Priorities set by State Government Leadership	0.25
Teacher Preparation	Teacher Quality	0.15
Ethnicity Factor	Parental Engagement	0.14
Difficulty of Teaching	Students First	0.13
Charter Schools	Vouchers	0.11
Difficulty Legislating	I Don't Know	*0.10
Policy-Making	Self-Described Unpopularity	0.10
Funding	I Don't Know	0.09
I Don't Know	Priorities set by State Government Leadership	0.08
Leadership Quality	School Culture	0.08
Reliance on Experts	Vouchers	0.07
Ethnicity Factor	Policy-Making	0.06
Reliance on Experts	Students First	0.06
Reliance on Experts	Teacher Preparation	0.06
Teacher Preparation	Respect for the Profession	0.06
I Don't Know	Reliance on Experts	0.05
Policy-Making	Politics	0.05
Politics	Respect for the Profession	0.05
Reliance on Experts	Teacher Quality	0.05
Complexity of Education	Politics	*0.03
Charter Schools	Politics	0.03
Gifted and Talented	Politics	*0.03
Politics	Self-Described Unpopularity	*0.03
Ethnicity Factor	Politics	0.02

*Indicates strength of relationship undervalued by weighted average.

Limitations of this Study

This study was limited by the timing of the study. The release of the statewide survey occurred simultaneously with the conclusion of annual statewide testing, a co-occurrence that was unintended. Considering the spike in number of responses, it is possible that those most passionate respondents were more inclined to include their responses and encourage their

associates to participate as well. In that case, snowball sampling may have become a factor unintended by the researcher.

The survey design included limitations. SQ2 should have been the initial question in the survey (parish of residence) and SQ1 (respondent role) should have been subsequent to SQ2. Both SQ1 and SQ2 should have required a response; four participants were able to respond to SQ1 without coding a category before the researcher noticed the setting. The setting was thus adjusted to require a response for respondents thereafter.

The electronic delivery of this survey limited respondents to those with electronic access and digital means for responding. Respondents may have responded via personal computer, tablet, smartphone, or any other device with Internet access and web browser. Those stakeholders without the means for electronic access were unable to participate. This limitation skewed results to only those respondents with the means for access. Including community and school-based focus groups could have strengthened this study to include a wider sample of the population.

Survey items were placed in the instrument to encourage completion rate. The items that required the most mental energy and time for completion, SQ12 and SQ13, were placed near the end of the survey to not dissuade respondents from completion. Therefore, some of the information from SQ3 through SQ 11 may have driven the respondent's ranking, i.e. leading, of SQ12 and SQ13. One way to strengthen the study would have been to have a concurrent survey of only SQ12 and SQ13 to compare results.

One weakness of the survey was the absence of questions around strengths of the test-based accountability system as it is currently designed. SQ17 asked respondents to identify limitations of the system, however, information regarding the strength(s) of the system could

have richly enhanced this study and contributed to advancing the system. It would have been advantageous to learn those factors and aspects of the current system that hold value-add for Louisiana stakeholders.

Terminology included in SQ9 may have been unclear. Respondents may have regarded using the term “indicators” as a “test”, when asking if indicators for non-academic skills should be included in Louisiana’s test-based accountability system terms. Indicators may be naturally observed or empirically measured, however, this distinction was not made clear in the study.

The results of this study were limited to one state in in the Southern region of the country. Context should be considered when applying generalizability to this study.

SQ14 was a type of replicability study, however, the sample population was littered with stakeholders of all varieties, not just CEOs of Fortune 500 Companies. Furthermore, there were not enough respondents from the business community to parse out the business perspective only. Additionally, the item included a clerical error that may have impacted responses. The category, *Personal Career Development*, was unintentionally altered to *Potential Career Development* for this study. There was no measure or accounting for the effect of this change in terminology.

There was a moment in one interview when the interviewee identified his question of “I don’t know why we’re not teaching it [social skills/business etiquette].” I should have probed deeper into his thinking and questioning, however, I reverted back to restating my understanding of his position on the question asked. Finally, reliability of interview analysis was limited to a single coder. Multiple raters would strengthen the reliability of this study.

Future Research

Opportunities to enhance this study include delineating state-mandated testing with district/school/classroom-mandated testing to clarify the quantity of assessments in today’s

classrooms. Allowing a multiple selected response option for SQ1 would allow researchers to further analyze discrete variables that may impact response patterns. A future study may include analysis of test-based accountability by student matriculation to investigate the impact of enrollment patterns on test-based accountability systems.

Opportunities for extending this study include utilizing this data set for analysis of covariance (ANCOVA) for Lafayette Parish stakeholders, in which the N is greater than 100. This could provide more clear analyses of responses by geographic location. Analysis of teacher and principal perception data and constructed responses through multiple T-tests may provide focused data for school leaders. Furthermore, additional analyses may include textual analysis of SQ3 through SQ15 with the interview analyses to create supercodes based on emerging themes.

CHAPTER 5. RESULTS AND CONCLUSIONS

If students are to be well served, accountability must be reciprocal. That is, federal, state, and local education agencies must themselves meet certain standards of delivery while school-based educators and students are expected to meet certain standards of practice and learning (Conley & Darling-Hammond, 2013, p. 35).

Stakeholders in Louisiana identified goals of education and valued skills that exceeded the scope of its current test-based accountability system. Through this study, Louisiana stakeholders made clear their dissatisfaction with the current test-based accountability system due to its limited scope, reliance on test scores, and negative impact on student learning. At the time of this study, Louisiana's Chief School Officer, State Superintendent John White, was hosting public forums and targeting focus groups to gather stakeholder feedback on revamping Louisiana's accountability system according to the newly passed federal legislation, Every Student Succeeds Act (ESSA). Focus groups included educators and education advocates, business leaders, community advocates, and higher education leaders including the following groups: Louisiana Accountability Commission, Superintendents Advisory Council, charter school leaders, Early Childhood Advisory Council, Special Education Advisory Panel, and the Louisiana Teacher Leader and Supervisor Collaborations.

There is a misalignment with the goals and skills valued by stakeholders with the information captured, valued, and incentivized by Louisiana's test-based accountability system. Education leaders in the state should consider opportunities for advancing the system. Lawmakers in Louisiana are amenable to piloting a system that would provide accountability for student learning in the cognitive, affective, and psychomotor domains. Written into the law, ESSA allows for pilot accountability systems. In keeping with this spirit of continuing progress,

expanding knowledge, and better serving the public, states are prime to advance systems of accountability through pilots. In particular, Louisiana now has a storehouse of public opinion data on test-based accountability in Louisiana, along with data collected through public forums and targeted focus groups. It is important, and necessary, that state leaders educate lawmakers about the need for advancing the system of test-based accountability in Louisiana and the range of options allowed for under ESSA.

Challenging the advancement of the state's test-based accountability system is the state's budget crisis, which forced not one but two special legislative sessions in 2016 for Louisiana. The final result of lawmakers' attempts to fill a 65 billion dollar deficit was a 24 million dollar cut to state education aid (<http://theadvocate.com/news/16177656-148/public-schools-take-hit-amid-budget-mess>). This cut is likely to cause education leaders to spend more time and energy deciding how to do more with less, as the standard for being rated an "A" school in Louisiana continues to climb (SBESE, Bulletin 111 §303, December 2015). By 2025, student test scores will have to average "Mastery" to earn the distinction of "A," whereas the current standard is for student test scores to average "Basic". This fiscal challenge may limit district leaders' interest in engaging in additional work toward advancing the test-based accountability system in the state.

For Louisiana to realize the intent of ESSA and to make gains on improving student learning outcomes, considerations for learning in the cognitive, affective, and psychomotor domains should remain at the forefront of policy decisions. Lawmakers should sharpen the focus on closing the achievement gap with an urgency and intensity that each day matters in the lives of our students and in their futures.

REFERENCES

- Allen, J., Altwerger, B., Edelsky, C., Larson, J., Rios-Aguilar, C., Shannon, P., & Yatvine, J. (2007). Focus on policy: Taking a stand on NCLB. *Language Arts*, 84(5), 456-464.
- Amrein, A. L. & Berliner, D. C. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18), 1-74.
- Anderson, L.W. & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of bloom's taxonomy of educational objectives*. New York, NY: Pearson.
- Autor, D., Levy, F., Murnane, R. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, 118(4), 1279-1333.
- Ayres, L. (1918). History and present status of educational measurements. In G. Whipple (Ed.), *The measurement of educational products. The seventeenth yearbook of the National Society for the Study of Education*, Part II (pp. 9-15). Bloomington, IL: Public School Publishing Company.
- Bacon-Blood, L. (2013, November 07). Louisiana students score near bottom on national test. *The Times-Picayune*. Retrieved from www.nola.com
- Baker, E. L (2008). Learning and assessment in an accountability context. In K. E. Ryan & L. A. Shepard (Eds.) *The future of test-based accountability* (pp. 277-292). New York, NY: Routledge.
- Baker, S. (2015). Desegregation, minimum competency testing, and the origins of

- accountability: North Carolina and the nation. *History of Education Quarterly*, 55(1), 33-57.
- Baker, E. & Linn, R. (2004). Validity issues for accountability systems. In S. Fuhrmand & R. Elmore (Eds.), *Redesigning accountability systems for education*, (pp. 47-72). New York, NY: Teacher's College Press.
- Basch, C. (2010). *Equity Matters, Research Review No. 6: Healthier students are better learners: A missing link in school reforms to close the achievement gap*. New York, NY: Teachers College, Columbia University.
- Bateson, G. (1979). *Mind and nature: A necessary unity*. New York, NY: Bantam Books.
- Binet, A. & Simon, T. (1916). *The development of intelligence in children (the Binet-Simon scale)*. Baltimore, Williams, & Wilkins: New Jersey.
- Black, P. Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning—putting it into practice*. Maidenhead, U.K.: Open University Press.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P. & Wiliam, D. (2004). The formative purpose: Assessment must first promote learning. In M. Wilson (Ed.) *Towards coherence between classroom assessment and accountability*, (pp. 20-50). Chicago, IL: The University of Chicago Press.
- Bloom, B. Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956) *Taxonomy of educational objectives, the classification of educational goals – Handbook I: Cognitive domain*. New York, Toronto: Longmans, Green.
- Booher-Jennings, J. (2005). Below the bubble: “Educational Triage” and the Texas education system. *American Education Research Journal*, 42(2), 231-268.

- Brill, R., Gilfoil, D., & Doll, K. (2014). Exploring predictability of instructor ratings using a quantitative tool for evaluating soft skills among MBA students. *American Journal of Business Education*, 7(3), 175-182.
- Brophy, (1988). Educating teachers about managing classrooms and students. *Teaching and Teacher Education*, 4(1), 1-18.
- Brookover, W. & Lezotte, L. (1979). *Changes in school characteristics coincident with changes in student achievement*. East Lansing, MI: Michigan State University, College of Urban Development.
- Brookover, W., Schweitzer, J., Schneider, J., Beady, C., Flood, P., Wisenbaker, J. (1978). Elementary school social climate and school achievement. *American Educational Research Journal*, 15(2), 301-318.
- Brown v. Board of Education, 347 U.S. 483 (1954).
- Buras, K. (2011). Race, charter schools, and conscious capitalism: On the spatial politics of Whiteness as property (and the unconscionable assault on Black New Orleans. *Harvard Educational Review*, 81(2), 296-331.
- Burgess, P. & Kennedy, S. (1998). What gets tested, gets taught; who gets tested gets taught: Curriculum development framework process. Mid-South Regional Resource Center, Lexington, KY
- Burnette, D. (2016). Connecticut approves new school accountability system. Education Week, Retrieved March 4, 2016 from <http://blogs.edweek.org>
- Cadei, E. (2015). The education law everyone wants to fix. *Newsweek*. Retrieved July 3, 2015 from <http://www.newsweek.com>
- Campbell, D. (1976). *Assessing the impact of planned social change*. The Public Affairs

- Center, Dartmouth College, Hanover New Hampshire, USA.
- Capra, F. (1996). *The web of life: A new scientific understanding of living systems*. New York, NY: Anchor Books.
- Carlson, D. (2014). School choice and educational stratification. *Policy Studies Journal*, 42(2), 269-304.
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305-331.
- Carr, P. (2004). Preface. In L. Jones & L. Olkin, (Eds.), *The nation's report card: Evolution and perspectives* (pp.vii-xi). Bloomington, IN: Phi Delta Kappa Educational Foundation.
- Chakrabarti, R. & Schwartz., N. (2013). Unintended consequences of school accountability policies: Evidence from Florida and implications for New York. *Economic Policy Review*, 19(1), 19-43.
- Cohen, D. K., & Spillane, J.P. (1992). Policy and practice: The relations between governance and instruction. *Review of Research in Education*, 18, 3-49.
- Colby, T. (2012). *Some of my best friends are black: The strange story of integration in America*. Penguin Group: New York, NY.
- Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfield, F., & York, R.. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- Conley, D. (2015a). A new era for educational assessment. *Education Policy Analysis Archives*, 23(8), 1-40.

- Conley, D. (2015b). Renaming noncognitive skills to emphasize success. Learning Deeply Blog Series, Retrieved March 16, 2016 from <http://blogs.edweek.org>
- Conley, D. & Darling-Hammond, L. (2013). *Creating systems of assessment for deeper learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Corbin, J. & Strauss, A. (2007). Basics of qualitative research: Techniques and procedures for developing grounded theory (3rd ed.). New York: Sage Publications.
- Council of Chief State School Officers. (2011). *Closing the achievement gap for economically disadvantaged students? Analyzing change since No Child Left Behind using state assessments and the national assessment of educational progress*. Washington, D.C.: Blank, R. K.
- Council of Chief State School Officers. (2016). Innovation Lab Network, Retrieved February 15, 2016 from www.ccsso.org
- Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Creswell, J. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd ed.). Thousand Oaks, CA: Sage.
- Crotty, M. (1998). *The foundations of social research*. London: Sage.
- Cullen, J. & Reback, R. (2006). Tinkering toward utopia: School gaming under a performance accountability system (Working Paper No. 12286). Retrieved from National Bureau of Economic Research website: <http://www.nber.org/papers/w12286>
- Darling-Hammond, L. (2007). *Race, inequality, and educational accountability: The irony of 'No Child Left Behind.'* *Race, Ethnicity, and Education*, 10(3), 245-260.
- Darling-Hammond, L. (2013, April 11). 'Test-and-punish' sabotages quality of children's

- education. Retrieved May 22, 2015 from <https://edpolicy.stanford.edu>
- Darling-Hammond, L., Herman, J., Pellegrino, J., Abedi, J., Aber, L., Baker, E., Bennet, R., Gordon, E., Haertel, E., Hakuta, K., Ho, A., Linn, R., Pearson, D., Popham, J., Resnick, L., Schoenfeld, A., Shavelson, R., Shepard, L., Shulman, L., Steele, C. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2014). *Accountability for college and career readiness: Developing a new paradigm*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- Darling-Hammond, L., & Wise, A. (1985). Beyond standardization: State standards and school improvement [Special issue]. *The Elementary School Journal*, 315-336.
- Datta, L. (1994). *Paradigm wars: A basis for peaceful coexistence and beyond*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Datta, L. (2004). Paradigm wars: A basis for peaceful coexistence and beyond. *New Directions for Program Evaluation*, 61, 53-70.
- Davis, B. (2004). *Inventions of teaching*. New York, NY: Routledge.
- Davis, B. & Sumara, D. (2006). *Complexity and education: Inquiries into learning, teaching, and research*. New York: Routledge.
- Daivs, B. (2008). Complexity and education: Vital simultaneities. In M. Mason (Ed.) *Complexity theory and the philosophy of education*. (pp. 46-61). Malden, MA: Wiley-Blackwell.
- DeBray-Pelot, E.H. (2007). NCLB's transfer policy and court-ordered desegregation: The conflict between two federal mandates in Richmond County, Georgia, and Pinellas County, Florida. *Educational Policy*, 21(5), 717-746.

- DeHouwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Society, Bulletin Review*. DOI: 10.3758/s13423-013-0386-3
- Delandshere, G. (2001). Implicit theories, unexamined assumptions, and the status quo of educational assessments. *Assessment in Education Principles, Policy, & Practice*, 8(2), 113-133.
- Deming, D., Cohodes, S., Jennings, J., & Jencks, C. (2016). When does accountability work? *Education Next*, Winter 2016
- Denzin, N. K. (1989). *The research act* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Denzin, N. K., & Lincoln, Y. S. (2005). Introduction: The discipline and practice of qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (pp. 1–32). Thousand Oaks, CA: Sage.
- Department of Labor (1991). What work requires of school: A SCANS report for American 2000. Secretary's Commission on Achievement Necessary Skills, Washington, D.C.
- Dewey, J. (1899). *The school and society*. Southern Illinois University Press, London, U.K. Arcturus Books Edition
- Doll, W. (1993). A post-modern perspective on curriculum. New York, NY: Teachers College Press.
- Duckworth, A. & Yeager, S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.
- Duckworth, A., Peterson, C., Matthews, M., & Kelly, D. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087-1101.
- Dunbar, S. (2008). Enhanced assessment for school accountability and student achievement. In

- K. Ryan & L. Shepard (Eds.), *The future of test-based educational accountability* (pp. 263-274). New York, NY: Routledge.
- Durlak, J., Dymnicki, A., Taylor, R., Weissberg, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405–432.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House.
- Eber, L., Sugai, G., Smith, C., & Scott, T. (2002). Wraparound and positive behavioral interventions and supports in the schools. *Journal of Emotional and Behavioral Disorders*, 10(3), 171-180.
- Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership*, 37(October 1979), 15-27.
- Egalite, A., Mills, J., & Green, J. (2015). The softer side of learning: Measuring students' non-cognitive skills. *Improving Schools*, The Authors, DOI: 10.1177/1365480215616313
- Elias, M. J., Zins, J.E., Weissberg, R.P., Frey, K.S., Greenberg, M.T., Haynes, N.M., Kessler, R., Schwab-Stone, M.E., & Shriver, T.P. (1997). Promoting social and emotional learning: Guidelines for educators. ASCD.
- Elementary and Secondary Education Act, Pub. L. 114-95 (1965)
- Elmore, R. (2008). *School reform from the inside out*. Cambridge, MA: Harvard Education Press.
- Elmore, R. & McLaughlin, M. (1988). *Steady work: Policy, practice, and the reform of American education*. The RAND Corporation, Report for the National Institute of Education, R-3574-NIE/RC.

- Erickson, F. (1986). Qualitative methods in research on teaching. In M. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 119-161). New York, NY: Macmillan.
- Fadel, C., Bialik, M., & Trilling, B. (2015). Four-dimensional education: The competencies learners need to succeed. Center for Curriculum Redesign, Boston, MA.
- Fasching-Varner, K., Mitchell, R., Martin, L., & Bennett-Haron, K. (2014). Beyond school-to-prison pipeline and toward an educational and penal realism. *Equity & Excellence in Education*, 47(4), 410-429.
- Fensterwald, J. (2016). Rating schools by students' social-emotional skills worth trying, evaluator says. EdSource, Retrieved March 20, 2016 from <http://edsources.org>
- Fenwick T., & Edwards, R. (2010). *Actor network theory in education*. New York, NY: Routledge.
- Ficella, K. & Kitzman, H. (2009). Disparities in academic achievement and health: The intersection of child education and health policy. *Pediatrics*, 123(3), 1073-1080.
- Figlio, D. & Getzler, L. (2007). Accountability, ability and disability: Gaming the system? In T. Gronberg & D. Jansen (Eds.) *Improving school accountability—check-ups or choice: Advances in applied microeconomics 14* (pp. 35-49). Amsterdam: Elsevier.
- Figlio, D. N. & Ladd, H. F. (2015). School accountability and student achievement. In H.F. Ladd & M. E. Goertz (Eds.), *Handbook of research in education finance and policy* (Second., p. 676). New York: Routledge.
- Finnigan, K. & Daly, A. (2012). Mind the gap: Organizational learning and improvement in an underperforming urban system. *American Journal of Education*, 119(1), 41-71.
- Firestone, W. & Herriott, R. (1980). *Images of the school: An exploration of the social*

- organization of elementary, junior, and high schools*. Philadelphia, PA: Research for Better Schools.
- Fishbein, J. M. (1973). The father of behavioral objectives criticizes them: An interview with Ralph Tyler. *Phi Delta Kappan*, 55(1), 55-57.
- Forte, E. (2013). Examining the assumptions underlying the NCLB federal accountability policy on school improvement. *Educational Psychologist*, 45(2), 76-88.
- Fuhrman, S. (2004). Less than meets the eye: Standards, testing and fear of federal control. In N. Epstein (Ed.), *Who's in charge of the schools? The tangled web of education power and policy*. Denver and Washington, DC: Brookings Institution Press.
- Fuhrman, S., Clune, W., & Elmore, R. (1988). Research on educational reform: Lessons on the implementation of policy. *Teachers College Record*, 90, 237-257.
- Fusarelli, L. D. (2007). Restricted choices, limited options: Implementing choice and supplemental educational services in No Child Left Behind. *Educational Policy*, 21(1), 132-154.
- Geertz, C. (1977). *The interpretation of cultures*. New York, NY: Basic Books
- Gerstl-Pepin, C. (2005). *The paradox of poverty*. *Educational Policy*, 20(1), 143-162.
- Glaser, B. (1965). The constant comparative method of qualitative analysis. *Social Problems*, 12(4), 436-445.
- Glasser, W. (1990). *Quality schools*. New York, NY: Harper and Row, Publishers, Inc.
- Goals 2000: Education America Act, Pub. L. No. 1804 (1994).
- Goertz, M. (2005). Implementing the No Child Left Behind Act: Challenges for the states. *Peabody Journal of Education*, 80(2), 73-89.

- Gong, B. (2002). Designing school accountability systems: Towards a framework and process. (pp. 3-37). Washington, DC: Council of Chief State School Officers.
- Gordon Commission on Future Assessment in Education (2013). *A public policy statement*. Princeton, NJ: Educational Testing Service.
- Gough, N. (2012). Complexity, complexity reduction, and ‘methodological borrowing’ in education inquiry. *Complicity: An international journal of complexity and education*. 9(1), 41-56.
- Greenberg, M., Weissberg, R., O’Brien, M., Zins, J., Fredericks, L., Resnik, H., & Elias, M. (2003). Enhancing school-based prevention and youth development through coordinated social, emotional, and academic learning. *American Psychologist*, 58(6/7), 466-474.
- Haertel, E. & Herman, J. (2005). A historical perspective on validity arguments for accountability testing. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing. Yearbook of the National Society for the Study of Education* (pp. 1-34), Vol. 104, Part I. Malden, MA: Blackwell Publishing.
- Hamersley, M. (1996). The relationship between qualitative and quantitative research: Paradigm loyalty versus methodological eclecticism. In J. T. E. Richardson (Ed.), *Handbook of research in psychology and the social sciences*. Leicester: British Psychological Society Books.
- Haney, W. (2008). Evidence on education under NCLB (and how Florida boosted NAEP scores and reduced the race gap). In G. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and school reform* (pp. 91-102). Thousand Oaks, CA: Corwin Press, Inc.

- Hanushek & Raymond (2006). School accountability and student performance. *Federal Reserve Bank of St. Louis Regional Economic Development*, 2(1), 51-61.
- Hattie (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.
- Heifetz, R. & Linksky, M. (2002). *Leadership on the line: staying alive through the dangers of leading*. Boston, MA: Harvard Business School Publishing.
- Hemelt, S. W. (2011). Performance effects of failure to make adequate yearly progress (AYP): Evidence from a regression discontinuity framework. *Economics of Education Review*, 30(2011), 702-723.
- Herman, J. L. (2004). The effects of testing on instruction. In S. H. Fuhrman & R. F. Elmore (Eds.) *Redesigning accountability systems for education* (pp. 141-166). New York, NY: Teachers College Press.
- Herman, J. L. (2008). Accountability and assessment: Is public interest in K-12 education being served? In K. E. Ryan & L. A. Shepard (Eds.) *The future of test-based accountability* (pp. 211-232). New York, NY: Routledge.
- Herman, J. L. & Baker, E. L. (2009). Assessment policy: Making sense of the babel. In G. Sykes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research* (pp. 176-190). (1956)..
- Hess, F. (2002). Reform, resistance,.....retreat? The predictable politics of accountability in Virginia. In D. Ravitch (Ed.), *Brookings papers on educational policy: 2002* (pp. 69-122), Washington, DC: Brookings Institution Press.
- Holmstrom, B. & Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, 7, 24-52.

- Honig, M. (2006a). Complexity and policy implementation: Challenges and opportunities for the field. In M. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 1-24). Albany, NY: State University of New York Press.
- Honig, M. (2006b). Building policy from practice: Implementation as organizational learning. In M. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 125-148). Albany, NY: State University of New York Press.
- Honig, M. (2009). What works in defining “What Works” in educational improvement: Lessons from education policy implementation research, directions for future research. In G. Sykes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research* (pp. 333-347). New York, NY: Routledge.
- Improving America’s Schools Act, Public Law 103-382, (1994).
- Irvine, J. (1988). Urban schools that work: A summary of relevant factors. *Journal of Negro Education*, 57(3), 236-242.
- Jacob, B.A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, 89, 761-796.
- Jaiani, V. & Whitford, V. (2011). Policy windows, public opinion, and policy ideas: The evolution of No Child Left Behind. *Quality Assurance in Education*, 19(1), 8-27.
- Jimerson, L. (2005). Placism in NCLB--How rural children are left behind. *Equity and Excellence in Education*, 38, 211-219.
- Johnson, R. & Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches*. Thousand Oaks, CA: Sage.

- Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research*, 1(2), 112-133.
- Jones, M., Jones, B., & Hargrove, T. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman and Littlefield Publishers.
- Joyce, B., Hersh, R., & McKibbin, M. (1983). *The structure of school improvement*. White Plaines, NY: Longman, Inc.
- Kim, J. & Sunderman, G. (2003). Measuring academic proficiency under the No Child Left Behind Act: Implications for educational equity. *Educational Researcher*, 34(8), 3-13.
- Kim, J. S. & Saunderman, G. L. (2005). Measuring academic proficiency under the No Child Left Behind act: Implications for educational equity. *Educational Researcher*, 34(8), 3-13.
- Kisker, E. & Brown, R. (1996). Do school-based health centers improve adolescents' access to health care, health status, and risk-taking behavior? *Journal of Adolescent Health*, 18(5), 335-343.
- Klein, A. (2016). Path to School accountability taking bold new turns. Retrieved January 20, 2016 from <http://www.edweek.org>
- Klittgard, R. & Hall, G. (1975). Are there unusually effective schools? *The Journal of Human Resources*, 10(1), 90-106.
- Kochhar-Bryant, C. & Heishman, A. (2010). *Effective collaboration for educating the whole child*. Sage Publications, Thousand Oaks: CA.

- Koretz, D. (1996). Using student assessments for educational accountability. In E.A. Hanushek & D.W. Jorgenson (Eds.), *Improving America's schools: The role of incentives* (pp. 171-195). Washington, DC: National Academy Press.
- Koretz, D. (2004). Alignment, high stakes, and the inflation of test scores. In S. Fuhrman & R. Elmore (Eds.), *Redesigning accountability systems for education*, (pp. 99-118). New York, NY: Teacher's College Press.
- Koretz, D. (2008a). *Measuring up: What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Koretz, D. (2008b). Further steps toward the development of an accountability-oriented science of measurement. In K. Ryan & L. Shepard (Eds.), *The future of test-based educational accountability* (pp. 71-92). New York, NY: Routledge.
- Koretz, D. (2008c). The pending reauthorization of NCLB in G. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, and school reform* (9-26). Thousand Oaks, CA: Corwin Press, Inc.
- Koretz, D. M. & Hamilton, L. S. (2003). *Teachers' responses to high-stakes testing and the validity of gains: A pilot study*. CSE Report 610. National Center for Research on Evaluation, Standards, and Student Testing.
- Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy*, 18(1), 45-70.
- Koyama, J. P. (2012). Making failure matter: Enacting no child left behind's standards, accountabilities, and classifications. *Educational Policy*, 26(6), 870-891.

- Kress, S., Zechmann, S., & Schmitten, J. (2011). When performance matters: The past, present, and future of consequential accountability in public education. *Harvard Journal on Legislation*, 48, 185-234.
- Kuhn, T. (1962). *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.
- Lee, J., & Wong, K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4), 797-832.
- Levin, H. M. (1974). A conceptual framework for accountability in education. *The School Review*, 82(3), 363-391.
- Lezotte, L. & Bancroft, B. (1985). School improvement based on effective schools research: A promising approach for economically disadvantaged and minority students. *The Journal of Negro Education*, 54(3), 301-312.
- Lezotte, L. & Passalacqua, J. (1978). Individual school buildings: Accounting for differences in measured pupil performance. *Urban Education*, 13(3), 283-293.
- Lincoln, Y. & Guba, E. (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage.
- Lincoln, Y. & Guba, E. (2000). Paradigmatic controversies, contradictions, and emerging confluences in Denzin, N. and Lincoln, Y. (Eds.). *Handbook of qualitative research*. London: Sage.
- Lindle, J. C. (2009). Assessment policy and politics of information. In G. Sykes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research* (pp. 319-332). New York, NY: Routledge.

- Lindquist, E. F. (1951). Preliminary considerations in objective test construction. In E. F. Lindquist (Ed.) *Educational measurement* (pp. 119- 158). Washington, DC: American Council on Education.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
- Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher*, 32(7), 3-13.
- Linn, R. L. (2008). Educational accountability systems. In K. Ryan & L. Shepard (Eds.), *The future of test-based educational accountability* (pp. 3-24). New York, NY: Routledge.
- Linn, R. L., Baker, E.L. & Betebenner, D.W. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher*, 31(6), 3-16.
- Linn, R., Gronlund, N., & Miller, D. (2013). *Measurement and Assessment in Teaching* (11th ed.), Boston, MA: Pearson Education, Inc.
- Louisiana Department of Education, (2015). Test results. Retrieved February 16, 2016 from <http://www.louisianabelieves.com>
- Louisiana Department of Education (2016). Assessment. Retrieved March 6, 2016 from <http://www.louisianabelieves.com>
- Louisiana District and School Accountability Advisory Commission. (1998, March 18). Recommendations for Louisiana's Public Education Accountability System.
- Louisiana State Board of Elementary and Secondary Education. (December 2015). Bulletin 111: The Louisiana school, district, and state accountability system. Baton Rouge, LA: Author.

Louisiana State University Public Policy Lab, Reilly Center for Media and Public Affairs
at the Manship School of Mass Communication. Retrieved March 4, 2016 from
<http://pprllsu.com>

Mackenzie, D. (1983). Research for school improvement: An appraisal of some recent trends. *Educational Researcher*, 12(8).

Malen, B. (2006). Revisiting Policy Implementation as a Political Phenomenon: The Case of Reconstitution Policies. In M. Honig (Ed), *New directions in education policy implementation: Confronting complexity* (pp. 83-104). Albany, NY: State University of New York Press.

Malen, B., Croninger, R., Muncey, D., & Redmond-Jones, D. (2002). Reconstituting schools: "Testing" the "theory of action". *Educational Evaluation and Policy Analysis*, 24(2), 113-132.

Manna, P. (2006). *School's in: Federalism and the national education agenda*. Washington, DC: Georgetown University Press.

Marshall, C. & Gerstl-Pepin, C. (2005). *Re-framing educational politics for social justice*. Boston: Allyn and Bacon.

McCarthy, J. & Still, S. (1993). Hollibrook Accelerated Elementary School. In J. Murphy & P. Hallinger (Eds.), *Restructuring schooling: Learning from ongoing efforts* (pp. 63-83).

McDermott, K. A. (2007). "Expanding the moral community" or "blaming the victim?" The politics of state education accountability policy. *American Journal of Education*, 44, 77-111.

McDonnell, L. (2005). Assessment and accountability from the policymaker's

- perspective. In J. Herman & E. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement* (pp. 35-54). Malden, MA: Blackwell Publishing.
- McDonnell, L. M. (2009). A political science perspective on education policy analysis. In G. Sykes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research* (pp. 57-70). New York, NY: Routledge.
- McDonnell, L. M. (2012). Educational accountability and policy feedback. *Educational Policy*, 27(2), 170-189.
- McLaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis*, 9(2), 171-178.
- McLaughlin, M. W. (2006). Implementation research in education: Lessons learned, lingering questions, and new opportunities. In M. Honig (Ed.), *New directions in education policy implementation: Confronting complexity* (pp. 209-228). Albany, NY: State University of New York Press.
- Miles, M. & Huberman, A. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks, CA: Sage.
- Miles, M., Huberman, A., & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook*. Thousand Oaks, CA: Sage.
- Miller, M. D., Linn, R. L., & Gronlund, N. E. (2013). *Measurement and assessment in teaching* (11th ed.). Boston, MA: Pearson Education, Inc.
- Mislevy, R. (1994). Evidence and inference in educational assessment. *Psychometrika*, 59(4), 439-483.

- Morrison, K. (2008). Educational philosophy and the challenge of complexity theory. *Educational Philosophy and Theory*, 40(1), 19-34.
- Murnane, R., Willet, J., Duhaldeborde, Y., & Tyler, J. (2000). How important are the cognitive skills of teenagers in predicting subsequent earnings? *Journal of Policy Analysis and Management*, 19(4), 547-568.
- Murphy, J. & Hallinger, P. (1988). Characteristics of instructionally effective school districts. *The Journal of Education Research*, 81(3), 175-181.
- NAE (The Review Committee of the National Academy of Education). (1987). Commentary by the National Academy of Education. In L. Alexander (chair). *The nation's report card. Improving the assessment of student achievement. Report of the study group. With a review of the report by a committee of the national academy of education* Robert Glaser (chairman). Washington, D. C.: Office of Educational Research and Improvement, U. S. Department of Education, and the National Academy of Education.
- National Center on Education and the Economy (2015). Center on International Educational Benchmarking <http://www.ncee.org>
- National Research Council. (1999). *Testing for tracking, promotion, and graduation*. Heubert, J. P. & Hauser, R. M. (Eds.). Washington, D.C.: National Academies Press.
- National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. J. Pellegrino, N. Chudowsky, & R. Glaser, Eds. Washington, DC: The National Academies Press.
- National Research Council (2012). National Research Council. (2012). *Education for life*

- and work: Developing transferable knowledge and skills for the 21st century.*
- Washington, DC: The National Academies Press.
- Nelson, S.W., McGhee, M.W., Meno, L.R., & Slater, C.L. (2007). Fulfilling the promise of education accountability. *Phi Delta Kappan*, 88(9), 702-709.
- No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).
- Northup, T. (2008). *Five hidden mistakes CEOs make: How to unlock the secrets that drive growth and profitability.* Solutions Press.
- Organisation for Economic Cooperation and Development, Programme for International Student Assessment, (2014). *PISA 2012 results in focus: What 15 year olds know and what they can do with what they know.* <http://www.oecd.org>
- Organisation for Economic Cooperation and Development (2016). *Glossary of statistical terms.* Retrieved March 5, 2016 from <http://www.oecd.org>
- Partnership for 21st Century Learning (2016). Retrieved March 5, 2016 from <http://www.p21.org>
- Pederson, P. (2007). What is measured is treasured: The impact of the No Child Left Behind Act on nonassessed subjects. *Clearing House: A Journal of Educational Strategies, Issues, and Ideas*, 80(6), 287-291.
- Pellegrino, J. (2002). Knowing what students know. *Issues in Science and Technology*, 19(2), 48-52.
- Phelps, R. & Hase, S. (2002). Complexity and action research: Exploring the theoretical and methodological connection. *Educational Action Research*, 10(3), 507-524.
- Polikoff, M. (2012). Instructional alignment under No Child Left Behind. *American Journal of Education*, 118(3), 341-368.

- Pogrow, S. (2006). Restructuring high-poverty elementary schools for success: A description of the hi-perform school design, *Phi Delta Kappan*, 223-229.
- Popham, J. (2013). *Classroom assessment: What teachers need to know* (7th ed.). New York: Pearson.
- Purdy, M. (2011). Executive functions: Theory, assessment, and treatment. In M. Kimbarow (Ed.) *Cognitive Communication Disorders* (pp. 83-128), San Diego, CA: Plural Publishing.
- Purkey, S. & Smith, M. (1983). Effective schools: A review. *The Elementary School Journal*, 83(4), 426-452.
- Ramaley, J. (2005). Goals for learning and assessment. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data in accountability testing. Yearbook of the National Society for the Study of Education* (pp. 55-77), Vol. 104, Part I. Malden, MA: Blackwell Publishing.
- Reardon, S., Greenberg, Kalogrides, E., Shores, D., & Valentino, R. (2012) Trends in academic achievement gaps in the era of No Child Left Behind. SREE Spring Conference
- Reber, S. (2007). School desegregation and educational attainment for Blacks. Working Paper 13193, National Bureau of Economic Research: Cambridge, MA.
- Reichardt & P. Swepson (Eds.), *Effective change management using action learning and action research: Concepts, frameworks, processes, applications*. Lismore: Southern University.
- Resnick, D. (1980). Minimum competency testing historically considered. *Review of Research in Education*, 8, 3-29.

- Riedl, B. (2010). Federal spending by the numbers 2010 (Special Report No. 78).
Retrieved from The Heritage Foundation website <http://report.heritage.org/sr0078>
- Roda, A. & Wells, A. (2013). School choice policies and racial segregation: Where white parents' good intentions, anxiety, and privilege collide. *American Journal of Education*, 119(2), 261-293.
- Rosenbaum, D. (2000). "Bush and Gore stake claims to federal role in education", *The New York Times*, August 30, p. A1.
- Rothstein, R. (2004). *Class and schools: Using social, economic, and educational reform to close the black-white achievement gap*. New York, NY: Teachers College Press.
- Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York, NY: Teachers College Press.
- Rowan, B., Bossert, S., & Dwyer, D. (1983). Research on effective schools: A cautionary note. *Educational Researcher*, 12(4), 24-31.
- Rutledge, S., Cohen-Vogel, L., Osborne-Lampkin, L., Roberts, R. (2015). Understanding effective high schools: Evidence for personalization for academic and social emotional learning. *American Education Research Journal*, 52(6), 1060-1092.
- Rutter, M. (1983). School effects on pupil progress: Research findings and policy implications. *Child Development*, 54, 1-29.
- Salazar Perez, M. & Cannella, G. (2011). Disaster capitalism as neoliberal instrument for the construction of early childhood education/care policy: Charter schools in post-Katrina New Orleans. *International Critical Childhood Policy Studies*, 4(1), 47- 68.
- Salvia, J. & Ysseldyke, J. (1995). *Assessment* (6th ed.), Boston, MA: Houghton Mifflin Co.

- Schmidt, W. H. & Maier, A. (2009). Opportunity to learn. In G. Sykes, B. Schneider, & D. Plank (Eds.), *Handbook of education policy research* (pp. 541-559). New York, NY: Routledge.
- Schneider, H. & Zhang, N. (2013). School accountability and youth obesity: Can physical education mandates make a difference? *Education Research International*, 2013, 1-14.
- Senechal, D. (2013). Measure against measure: Responsibility versus accountability in education. *Arts Education Policy Review*, 114, 47-53.
- Shavelson, R., Young, D., Ayala, C., Brandon, P., Furtak, E., Ruiz-Primo, M., Tomita, M., & Yin, Y. (2008). On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment development. *Applied Measured in Education*, 21, 295-314.
- Shepard, L. (2008). A brief history of accountability testing, 1965-2007. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 25-46). New York, NY: Routledge.
- Siegel, H. (2004). What ought to matter in public schooling: Judgment, standards, and responsible accountability. In K. Sirotnik (Ed.) *Holding accountability accountable: What ought to matter in public education*. Teachers College Press, New York, NY.
- Sizemore, B. (1985). Pitfalls and promises of effective schools research. *The Journal of Negro Education*, 54(3), 269-288.
- Sizemore, B., Brossard, C. & Harrigan, B. (1983). *An abashing anomaly: The high achieving predominantly black elementary school*. Pittsburg, PA: National Institute of Education.
- Spillane, J. (2004). *Standards Deviation*. Cambridge, MA: First Harvard University

Press.

Spillane, J. P., Reiser, B. J., & Reimer, T. (2002). Policy implementation and cognition: Reframing and refocusing implementation research. *Review of Educational Research*, 72(3), 387-431.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.

Stake, R. E. (2010). *Qualitative research: Studying how things work*. New York: The Guilford Press

Strauss, (1987). *Qualitative analysis for social scientists*. New York: Cambridge University Press.

Strauss, & Corbin, (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Thousand Oaks, CA: Sage.

Superfine, B. (2005). The politics of accountability: The rise and fall of Goals 2000. *American Journal of Education*, 112(1), 10-43.

Sweeney, J. (1982). Research synthesis on effective school leadership. *Educational Leadership*, 39(5), 346-352.

Teddlie, C., Kirby, P., Stringfield, S. (1989). Effective versus ineffective schools: Observable differences in the classroom. *American Journal of Education*, 97(3), 221-236.

Teddlie, C. & Stringfield, S. (1993) *Schools make a difference*. New York, NY: Teachers College Press.

Teddlie, C. & Stringfield, S. (2007). A history of school effectiveness and improvement research in the USA focusing on the past quarter century. In *Springer International Handbook of Education: International handbook of school effectiveness and improvement*. Dordrecht, The Netherlands: Springer Science+Business Media.

- Teddlie, C. & Tashakkori, A. (2011). Common “Core” characteristics of mixed methods research: A review of critical issues and call for greater convergence. *American Behavioral Science*, 56(6), 774-788.
- Thorndike, E. L. (1918). Fundamental theorems in judging men. *Journal of Applied Psychology*, 2(1), 67-76.
- Tucker, M. S. (2014). *Fixing our national accountability system*. Retrieved from National Center on Education and the Economy website: <http://www.ncee.org>.
- Turnipseed, & Darling-Hammond, L. (2015).
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. Chicago: IL: University of Chicago Press.
- Tyler, R. W. (1972). Why evaluate education? *Compact*, 6(1), 3-4.
- United States Agency for International Development, Project Starter, Indicator Definition and Unit of Analysis Retrieved April 11, 2016 from <http://usaidprojectstarter.org>.
- United States Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, *National Assessment Of Title I Final Report*. Jessup, MD: ED pubs, Education Publications Center, 2007. Print. Volume I: Implementation.
- United States Department of Education. (1995). *The improving America’s schools act of 1994: Reauthorization of the Elementary and Secondary Education Act*. Washington, D.C.
- United States Department of Education, National Center for Education Statistics, Institute

of Education Sciences, National Center for Education Statistics, (2015a). *The condition of education*. (Report number 2015-144). Retrieved from National Center on Education Statistics website: <http://nces.ed.gov>.

United States Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Center for Education Statistics, (2015b). Mapping state proficiency standards onto NAEP scales: Results from the 2013 NAEP reading and mathematics assessments. Retrieved from National Center on Education Statistics website: <http://nces.ed.gov>

United States Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Center for Education Statistics, (2015c). State profile reports. Retrieved from National Center on Education Statistics website: <http://nces.ed.gov>.

United States Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Center for Education Statistics, (2016). State profile reports. Retrieved from National Center on Education Statistics website: <http://nces.ed.gov>.

Wanker, W. & Christie, K. (2005). State implementation of the No Child Left Behind Act. *Peabody Journal of Education*, 80(2), 57-72.

Warren, M. (2005). Communities in schools: A new view of urban education reform. *Harvard Educational Review*, 75(2), 133-173.

Weber, G. (1971). *Inner-city children can be taught to read: Four successful schools*. Washington, DC: Council for Basic Education.

Weissberg, R. & Cascarino, J. (2013). Academic learning + social-emotional learning =

- national priority. *The Kappan*, 95(2), 8-13.
- Weist, M., Nabors, L., Albus, K., & Bryant, T. (2006) School-based health centers. In A. Petti & C. Salguero (Eds.), *Community child and adolescent psychiatry: A manual of clinical practice and consultation* (pp. 149-160). Washington, D.C.: American Psychiatric Publishing, Inc.
- West, M., Kraft, M., Finn, A., Martin, R., Duckworth, A., Gabrieli, C., & Gabrieli, J. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148-170.
- West, M. (2016). Should non-cognitive skills be included in school accountability systems? Retrieved March 30, 2016 from <http://educationnext.org>
- White House, (March 10, 2009). Remarks of the President to the United States Hispanic Chamber of Commerce.
- Willcocks, L. & Lester, S. (1996). Beyond the IT productivity paradox, *European Management Journal*, 14(3), 279-290.
- Witte, J. & Walsh, D. (1990). A systematic test of the effective schools model. *Educational Evaluation and Policy Analysis*, 12(2), 188-212.
- Wong, K. K. (2013). Politics and governance: Evolving systems of school accountability. *Educational Policy*, 27(2), 410-421.
- Yin, R. K. (2003). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.
- Zernike, K. (February 29, 2016). Testing for joy and grit? Schools nationwide push to measure students' emotion skills. New York Times, Retrieved March 17, 2016 from <http://www.nytimes.com>

- Zimmer, R., Gill, B., Razquin, P., Booker, K., Lockwood III, J. (2007). State and local implementation of the No Child Left Behind Act: Volume I--title I school choice, supplementary educational services, and student achievement *National Longitudinal Study of No Child Left Behind (NLS-NCLB)* 1 (1-44). Jessup, MD: RAND.
- Zimmerman, B., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, 29(3), 663-676.
- Zins, J., Bloodworth, M., Weissberg, R., & Walberg, H. (2004). *Building academic success on social and emotional learning: What does the research say?* New York, NY: Teachers College, Columbia University.

APPENDIX A SURVEY INSTRUMENT

Test-Based Accountability for K-12 Public Schools in Louisiana

1. Please select the category that best describes you.

- ☐ Parent
- ☐ Teacher
- ☐ Teacher/Instructional Leader
- ☐ Principal
- ☐ Community Member
- ☐ Local Education Agency Staff
- ☐ Local Education Agency Superintendent
- ☐ State Education Agency Staff
- ☐ Elected Official
- ☐ Member of Business Community
- ☐ Member of Higher Education

2. Please select your parish of residence.

- | | | | |
|-------------------------------------|---|--|---|
| <input type="checkbox"/> Acadia | <input type="checkbox"/> East Baton Rouge | <input type="checkbox"/> Madison | <input type="checkbox"/> St Landry |
| <input type="checkbox"/> Allen | <input type="checkbox"/> East Carroll | <input type="checkbox"/> Morehouse | <input type="checkbox"/> St Martin |
| <input type="checkbox"/> Ascension | <input type="checkbox"/> East Feliciana | <input type="checkbox"/> Natchitoches | <input type="checkbox"/> St Mary |
| <input type="checkbox"/> Assumption | <input type="checkbox"/> Evangeline | <input type="checkbox"/> Orleans | <input type="checkbox"/> St Tammany |
| <input type="checkbox"/> Avoyelles | <input type="checkbox"/> Franklin | <input type="checkbox"/> Ouachita | <input type="checkbox"/> Tangipahoa |
| <input type="checkbox"/> Beauregard | <input type="checkbox"/> Grant | <input type="checkbox"/> Plaquemines | <input type="checkbox"/> Tensas |
| <input type="checkbox"/> Bienville | <input type="checkbox"/> Iberia | <input type="checkbox"/> Point Coupee | <input type="checkbox"/> Terrebone |
| <input type="checkbox"/> Bossier | <input type="checkbox"/> Iberville | <input type="checkbox"/> Rapides | <input type="checkbox"/> Union |
| <input type="checkbox"/> Caddo | <input type="checkbox"/> Jackson | <input type="checkbox"/> Red River | <input type="checkbox"/> Vermilion |
| <input type="checkbox"/> Calcasieu | <input type="checkbox"/> Jefferson | <input type="checkbox"/> Richland | <input type="checkbox"/> Vernon |
| <input type="checkbox"/> Caldwell | <input type="checkbox"/> Jefferson Davis | <input type="checkbox"/> Sabine | <input type="checkbox"/> Washington |
| <input type="checkbox"/> Cameron | <input type="checkbox"/> Lafayette | <input type="checkbox"/> St Bernard | <input type="checkbox"/> Webster |
| <input type="checkbox"/> Catahoula | <input type="checkbox"/> Lafourche | <input type="checkbox"/> St Charles | <input type="checkbox"/> West Baton Rouge |
| <input type="checkbox"/> Claiborne | <input type="checkbox"/> LaSalle | <input type="checkbox"/> St Helena | <input type="checkbox"/> West Carroll |
| <input type="checkbox"/> Concordia | <input type="checkbox"/> Lincoln | <input type="checkbox"/> St James | <input type="checkbox"/> West Feliciana |
| <input type="checkbox"/> DeSoto | <input type="checkbox"/> Livingston | <input type="checkbox"/> St John the Baptist | <input type="checkbox"/> Winn |
| | | | <input type="checkbox"/> Non-resident |

For each of the following, rate each response on a scale of 1 to 5 (1=lowest, 5=highest).

3. Overall, to what extent does the School Letter Grade (i.e. school performance score) reflect a school's quality in Louisiana?

1	2	3	4	5
Least Accurate	Somewhat Accurate	Sufficiently Accurate	Highly Accurate	Most Accurate

Comment(s) _____

4. Overall, how would you rate the AMOUNT OF TIME students spend on state-required tests in Louisiana?

1	2	3	4	5
Not enough time	Somewhat enough time	Sufficiently enough time	More than enough time	Too much time

Comment(s) _____

5. Overall, how would you rate the NUMBER OF TESTS students take for state-required tests in Louisiana?

1	2	3	4	5
Not enough tests	Somewhat enough tests	Sufficiently enough tests	More than enough tests	Too many tests

Comment(s) _____

6. Overall, how would you rate Louisiana's test-based accountability system for K-12 public schools?

1	2	3	4	5
Poor	Okay	Good	Great	Excellent

Comment(s) _____

For each of the following questions, select “yes” or “no.”

7. Should schools be held accountable for providing the nationally recommended amount (60 minutes) of quality physical activity every day?

☐ Yes
☐ No

Comment(s) _____

8. Should schools be held accountable for providing social-emotional learning opportunities for all students?

☐ Yes
☐ No

Comment(s) _____

9. Should indicators for non-academic skills (i.e. social, emotional, & physical health) be included in Louisiana’s test-based accountability system?

☐ Yes
☐ No

Comment(s) _____

10. Should surveys of students’ perception about their school experience be included in Louisiana’s test-based accountability system?

☐ Yes
☐ No

Comment(s) _____

11. Should students’ report card grades be included in Louisiana’s test-based accountability system?

☐ Yes
☐ No

Comment(s) _____

For each of the following, rank each item according to the instructions.

12. In order of importance, with 1 being most important and 8 being least important, rank the following goals of education⁶.

- ___ Basic academic skills
- ___ Citizenship
- ___ Critical thinking
- ___ Emotional health
- ___ Physical health
- ___ Preparation for skilled work
- ___ Social skills and work ethic
- ___ The arts and literature

13. What, if any, additional school-based learning outcomes/qualities would you add to the list in #12? Rank in order of importance.

14. In order of importance, with 1 being most important and 10 being least important, rank the following skills⁷.

- ___ Creative thinking
- ___ Goal setting/motivation
- ___ Interpersonal skills
- ___ Leadership
- ___ Listening Skills
- ___ Oral communications
- ___ Potential career development
- ___ Problem Solving
- ___ Teamwork
- ___ Writing

15. What, if any, additional school-based learning outcomes/qualities would you add to the list in #14? Rank in order of importance.

⁶ Based on Rothstein, Jacobsen, & Wilder (2008)

⁷ Based on Cassel & Kalder (1990, 1999)

For each of the following, respond by writing in the space provided.

16. In your opinion, what top three factors contribute to student learning outcomes for students in Louisiana?

17. In your opinion, what, if any, are limitations to Louisiana's test-based accountability system?

18. If you were afforded the opportunity to enhance Louisiana's test-based accountability system, what would you do?

19. OPTIONAL: Any additional feedback?

APPENDIX B
INTERVIEW PROTOCOL
Preliminary Protocol for Semi-Structured Interviews

Interviewee: _____

Interview Date: _____

1. How many years have you served as an elected representative in the Louisiana legislature?
2. How many years have you served on the Education Committee in the House/Senate?
 - a. During your tenure, what topics stand out to you as most challenging for policy making?
 - b. During your tenure, what topics stand out to you as most challenging for consensus building?
3. Overall, what are your thoughts on Louisiana's test-based accountability system?
4. We're interested in your perception about the testing aspect of Louisiana's test-based accountability system.
 - a. Time: Are students spending sufficient time taking state-mandated tests in Louisiana?
 - i. PROBE: Why do you say that?
 - b. Quantity: Are students taking a sufficient number of tests for state-mandated testing in Louisiana?
 - i. PROBE: Why do you say that?
 - c. Quality: Are tests of sufficient quality to adequately inform students and parents about the students' progress?
 - d. Quality: Are tests of sufficient quality to adequately inform teachers for instructional decision-making?
5. In your opinion, how has test-based accountability impacted student achievement in Louisiana?
 - a. For each response, PROBE: Why do you think that had an impact?
6. In your opinion, what, if any, are limitations to Louisiana's test-based accountability system?
 - a. For each response, PROBE: Why is that a limiting factor?
7. What 3 factors do you think will boost academic achievement for students in Louisiana's test-based accountability system?
8. If you were afforded the opportunity to enhance Louisiana's test-based accountability system, what would you do?
 - a. PROBE: What benefits do you anticipate resulting from those changes?
9. To what extent would you be willing to pilot a system for accountability that incorporated measures for cognitive, affective, and psychomotor development?

APPENDIX C
REQUEST FOR SURVEY

Invitation to Participate

April 19, 2016

Dear Louisiana education stakeholder,

My name is Susan Kahn. I am a Doctoral Candidate of Educational Leadership and Research in the College of Education at Louisiana State University conducting a study on Louisiana's test-based accountability system for K-12 education.

Please consider responding to this brief survey to share your perspective on test-based accountability in Louisiana. This survey should take no more than 10-15 minutes and the response portal will close at 11:59 p.m. on **Friday, May 20, 2016**. The quality of this study will be enhanced by your response. To ensure the validity of results, please respond to all survey items only once.

With Louisiana's implementation of the *Every Student Succeeds Act*, this study is a timely endeavor. It is our hope that the information learned from this study may inform decision-making at the state and local levels. We expect results to be available Summer 2016.

All the best,

-

Susan Kahn, Ed.S.

Doctoral Candidate

School of Education | College of Human Sciences and Education

Louisiana State University

223 Peabody Hall | Baton Rouge, LA | 70803 | 225-578-3202

skahn1@lsu.edu | susankahn.lsu@gmail.com | lsu.edu | lsu.edu/chse/education

APPENDIX D REQUEST FOR INTERVIEW

Invitation to Participate

April 19, 2016

Name (Title)
Street
City, State, Zip
Phone
Email

Dear Representative/Senator [insert],

My name is Susan Kahn. I am a Doctoral Candidate of Educational Leadership and Research in the College of Education at Louisiana State University conducting a two-part study on Louisiana's test-based accountability system for K-12 education. Due to your position as an elected official on the House or Senate Education Committee, you have been selected to participate in the second part of this study as an interviewee.

The interview will consist of a set of 5-10 questions and last approximately 45 minutes to one hour. For this study, all participants will be given pseudonyms and their respective elected office (Senate or House) will remain anonymous. To ensure valid and reliable data analysis, the interview will be recorded and will be conducted, preferably, in your office at the Louisiana legislature.

I understand this request comes during the 2016 active Legislative session. As such, I have arranged for interviews to be conducted during each of the following weeks:

Week 1: May 2-May 6
Week 2: May 9-May 13
Week 3: May 16-May 20
Week 4: May 23-May 27

The quality of this study will be enhanced by your response. Please let me know your availability for participation in this study.

With Louisiana's implementation of the *Every Student Succeeds Act*, this study is a timely endeavor. It is our hope that the information learned from this study may inform decision-making at the state and local levels. We expect results to be available Summer 2016.

All the best,

Susan Kahn, Ed.S.
Doctoral Candidate
School of Education | College of Human Sciences and Education
Louisiana State University
223 Peabody Hall | Baton Rouge, LA | 70803 | 225-578-3202
skahn1@lsu.edu | susankahn.lsu@gmail.com | lsu.edu | lsu.edu/chsc/education

APPENDIX E

CONSENT FORM FOR INTERVIEWS

1. Study Title: *Toward an Advanced System of Accountability for Improved Student Learning Outcomes: A Mixed Methods Analysis of Test-Based Accountability in Louisiana*
2. Performance Site: Phase I: digital survey; Phase II: Louisiana Legislature offices
3. Investigators: The following investigators are available for questions about this study.
M-F, 8:00 a.m. – 4:30 p.m.
Susan Kahn, 337-212-6192
SusanKahn.lsu@gmail.com
4. Purpose of the Study: The purpose of this research study is to examine the extent to which Louisiana's test-based accountability system provides meaningful information to support student learning and the extent to which lawmakers are willing to pilot a more advanced system of accountability in Louisiana.
5. Subject Inclusion: Phase I: Parents, Teachers, Principals, Superintendents, Local Community Members, Members of Institutions of Higher Education and the Business Community; Phase II: Members of the Louisiana Legislature (House and Senate Education Committees)
6. Number of Subjects: Phase I: 50-100; Phase II: 4-6
7. Study Procedures: This study is Phase II of a multi-phase study. In Phase I, the principal investigator (PI) has conducted a survey of education stakeholders on their values in education outcomes and Louisiana's test-based accountability system. In this phase, Phase II, the PI will interview elected lawmakers on their positional values in education outcomes and the extent to which they are willing to pilot a more advanced system of accountability in Louisiana.
8. Benefits: The study may yield valuable information about test-based accountability in Louisiana.
9. Risks: There are no known risks to participating in this study.
10. Right to Refuse: Subjects may choose not to participate or to withdraw from the study at any time without penalty or loss of any benefit to which they might otherwise be entitled.
11. Privacy: Results of the study may be published, but all participants and their institutions will be given pseudonyms to protect each individual's identity.
12. Signature:

The study has been discussed with me and all of my questions have been answered. I may direct additional questions regarding study specifics to the PI. If I have questions about subjects' rights or other concerns, I can contact Robert Mathews, Institutional Review Board (225) 578-8692, irb@lsu.edu, www.lsu.edu/irb.

I agree to participate in the study described above and acknowledge the investigator's obligation to provide me with a signed copy of this consent form.

Subject Signature: _____ Date: _____

Institutional Review Board
Dr. Robert Mathews, Chair
203 B-1 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.6792
irb@lsu.edu
lsu.edu/irb

APPENDIX F
INSTITUTIONAL REVIEW BOARD APPROVAL

IRB Application

Institutional R Board

Tue 4/19/2016 11:11 AM

To: Susan N Kahn <skahn1@lsu.edu>;

Cc: Kenneth J Fasching-Varner <varner@lsu.edu>;

Hi,

The IRB chair reviewed your application, TOWARD AN ADVANCED SYSTEM OF ACCOUNTABILITY FOR IMPROVED STUDENT LEARNING OUTCOMES: A MIXED-METHODS ANALYSIS OF TEST-BASED

ACCOUNTABILITY IN LOUISIANA, and determined IRB approval for this specific application is not needed. There is no manipulation of, nor intervention with, human subjects. Should you subsequently devise a project which does involve the use of human subjects, then IRB review and approval will be needed.

Elizabeth



Elizabeth Cadarette

IRB Coordinator

Office of Research and

Economic Development

Louisiana State University

[130 David Boyd Hall, Baton](#)

[Rouge, LA 70803](#) office

225-578-8692 | fax 225-

578-5983

eamt1@lsu.edu | lsu.edu | www.research.lsu.edu

LSU Research - The Constant Pursuit of Discovery

VITA



Susan Kahn, Ph.D.

susankahnphd@gmail.com

Twitter: @Susan_Kahn

Personal Statement

My objective is to lead educational excellence by impacting student achievement through exemplary educational leadership in policy and practice.

Education

Doctor of Philosophy Educational Leadership and Research	Louisiana State University
Certificate of Education Specialist Educational Leadership	Louisiana State University
Master of Education Educational Leadership	University of Louisiana at Lafayette
Bachelor of Arts Elementary Education	University of Louisiana at Lafayette

Professional Experience

Instructor of Education College of Education Department of Curriculum and Instruction	University of Louisiana at Lafayette, 2014-present
Assessment Administration Office of Assessments	Louisiana Department of Education, 2013-2014
Assessment Coordinator Department of Accountability, Research, and Evaluation	Lafayette Parish School System, 2012-2013
Curriculum Coordinator	Lafayette Parish School System, 2011-2012
Teacher, 4th grade	Lafayette Parish School System, 2006-2009