

Building Evaluation Capacity in Educational Organizations:
A Longitudinal Case Study of One Metropolitan School District

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA

BY

Johnna Rohmer-Hirt

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Jean A. King, PhD
Advisor

July, 2016

Acknowledgements

First and foremost, I would like to acknowledge my Heavenly Father, Lord and Savior, Wonderful Counselor; You are the source of all that I am and all that I have – without Your unconditional love, grace and forgiveness, I would be nothing. Your sacrifice for my sake is unfathomable and undeserved and I will forever worship and praise You in all that I do.

To my parents, who taught me the importance of getting an education as well as keeping my common sense – you gave me wings and allowed me to soar but also provided me a safe and loving place to land when needed. And to my siblings who were among my first teachers, you provided me the foundation and inspiration to continue to fly.

To my husband and children – you are my heart. I love you each so very deeply and I thank God for you always. You have loved and supported me through this journey, walked beside me and encouraged me; brought me coffee and came in just to check to see if I needed anything. This accomplishment is as much each of yours as it is mine. You are the reasons I continue to strive to do more and be a better wife and mother. You each make me so very proud and I am extraordinarily blessed to have you in my life. Also to Cashous Clay and Carshey, my constant writing companions – through long days and sleepless nights, you two were with me the whole time.

To all my extended family and friends, you have supported me and encouraged me, asked me about the progress of my coursework and research and stuck by me as I ebbed and flowed through the process. Thank you for praying me through.

To my department and district co-workers – you, too, have provided tremendous support and encouragement and I am thankful for all that you do; your dedication to the students, families, staff and community members whom we serve is admirable. I could not ask for a better team and I am grateful that each of you are working alongside me on a daily basis. And to those current and past employees who agreed to be a part of this study – my research could not have occurred without you! Your work in forming and shaping our department is commendable and I admire the dedication and foresight you brought to the district. All of our stakeholders have and will continue to benefit because of your contributions.

To Dr. Jean King, thank you so much for being my professor, my mentor and most importantly, my friend. I have learned so much from you and I am eternally grateful. You have afforded me so many experiences and opportunities that would have never been possible without my connection to you. You are an inspiration and someone whom I wish to emulate; I am sincerely honored to know you.

To my committee, I have greatly benefited from your expertise and guidance and I am filled with gratitude for the role you have played in bringing me to the end of this journey. Your willingness and commitment to support those of us entering the world of academia is admirable and definitely do not go unnoticed.

And finally, to my writing group; I am the last of us to finish – it has been a long two years. I will never forget the hours we spent sitting around D’s table trying to figure out the “new” prelim process. You will all forever be a part of this accomplishment. We were in this together and we all made it! We hit submit!

Dedication

I dedicate this dissertation to my family – my husband, Stephen, and my children, Jonathan, Rebekah, and Johannah. May you always know that my most treasured accomplishments are being your wife and mother and all else that I do is because you provide me the aspiration to be the best I can be for each of you.

Jon, Bekah, and Hannah – I am so proud of the people you have become and I cannot wait to see what is next for each of you. You are amazing individuals; uniquely gifted and incredibly wonderful. Having you in my life has made everything else imaginable and nothing seems impossible.

I love each one of you more than words can express...

Abstract

Although evaluation capacity building (ECB) is relatively prevalent in the literature, there is not an extensive empirical body of work related to the topic. The purpose of this study was to add to the existing literature through a 30-year historical longitudinal case study of a large metropolitan school district from the perspective of its leadership. This research examined the longitudinal evidence of the context, processes, and activities that impact ECB within a K-12 educational setting. In addition, it traced the development of an internal evaluation department and the steps to build supports for data use and evaluation in the midst of accountability, using process use as the primary strategy toward ECB.

The study consisted of interviews with 14 district leaders, along with document reviews over the years 1985 to 2015, encompassing significant legislative enactments such as No Child Left Behind (NCLB) and the public accountability era in K-12 education. The district leadership was found to value data use across time, along with community involvement and input, and demonstrated a strong legislative awareness and presence. The leaders who were interviewed also viewed evaluation as mission and vision critical across time. Substantial changes were made over the 30-year time frame in the number and types of staff members in the department, the internal process and activities of the department, how the department collaborated with others across the organization, and the organizational reporting structure. In addition, organizational factors such as the shift from evaluation for mandatory reasons (such as grants and legislative requirements) to evaluation for internal decision making and inquiry occurred

over the 30-year span, along with important changes in technology and data integrity, accessibility, and understandability.

Eight key recommendations to propel ECB emerged from the research: 1) attend to data integrity, accessibility, and understandability – including technology; 2) capitalize on one-time and special funding; 3) recognize that leadership matters and change is disruptive; 4) hire the right people; 5) remember that disposition/personality matters; 6) make ECB an intentional, ongoing focus; 7) work toward consistency, not the new next best thing; and 8) build on previous work instead of completely starting over.

Table of Contents

Acknowledgements	i
Dedication	iii
Abstract.....	iv
List of Tables	ix
List of Figures.....	x
Chapter 1: Introduction and Statement of the Problem	1
Introduction	1
Statement of the Problem	1
History of Legislative Impact on the Development of Evaluation.....	4
Evaluation Capacity Building (ECB).....	11
Purpose of the Study	12
Significance of the Study	14
Definition of Key Terms	16
Overview of Chapters.....	20
Chapter 2: Review of the Literature: Theoretical Underpinnings of ECB	22
Organizational Change and Learning as the Groundwork to ECB	23
Accountability as an Impetus for ECB.....	33
Process Use as a Means to ECB.....	40
Internal Evaluation as a Model for ECB	50
Evaluation Capacity Building	61
Conceptual Framework of ECB	73
Summary	75

Chapter 3: Methods	78
Case Selection and History	80
Data Collection Procedures and Methods	84
Data Analysis	91
Limitations	97
Positionality/Managing Subjectivity	100
Chapter 4: Findings	103
The Beginning: Superintendent Quincy’s and Urtel’s Eras	103
Superintendent Spawling’s Era	105
Interim Superintendent Hadley’s Era	119
Superintendent Osborne’s Era	120
Superintendent Mostenby’s Era	124
Superintendent Cadence’s Era	142
Superintendent Langley’s Era	157
Overview of Departmental Progression	160
Progression of Data Availability	169
Overarching Findings: Document Review	172
Interview Findings: Evaluations Importance to Mission and Vision	180
Chapter 5: Discussion and Implications	182
Discussion of Findings	182
Implications	198
Conceptual Framework of ECB – Revisited	210
Summary	214
Chapter 6: Conclusion.....	215
Research Questions and Answers	215
Implications for Practitioners	223
Implications for Further Research.....	224

Epilogue	226
References	232
Appendix A: Evaluation Capacity in School Districts Interview Instrument	247
Appendix B: Consent Information Form	251
Appendix C: Rubric for Assessing Evaluation Strength	253
Appendix D: State Assessment Expansion Over Time	254
Appendix E: Two-County Program Evaluation Model and Sample Reporting Tool.....	255
Appendix F: Additional Figures from School Board Agenda/Minutes Analysis.....	257
Appendix G: Additional Figures from Two-County Newsletter Analysis.....	258

List of Tables

Table 1: Timeline of Evaluation and Legislative Implications to ECB – Part 1	9
Table 2: Timeline of Evaluation and Legislative Implications to ECB – Part 2	10
Table 3: Timeline of Evaluation and Legislative Implications to ECB: 1985-2015	15
Table 4: Comparison of Four ECB Frameworks	68
Table 5: Matrix of Research Questions, Methods, and Units of Analysis.....	86
Table 6: Method Limitations and Strategies to Address Them	99
Table 7: Timeline of Evaluation and Legislative Implications to ECB with Departmental Progression.....	167
Table 8: Timeline of Evaluation and Legislative Implications to ECB – Departmental Processes, Important Events, and Internal Influences	168
Table 9: Progression of Available Data in Two-County ISD by Data System.....	172
Table 10: Total Annual Average Count of Scheduled Meetings/Events by Activity Type	179
Table 11: Annual Average Percentage of Scheduled Meetings/Events by Activity Type	180
Table 12: Evidence of ECB in Two-County ISD Over Time	219
Table 13: Technology Enhancements and Superintendents’ Data Focus Over Time	220

List of Figures

Figure 1: History of Evaluation and United States Federal Legislation	8
Figure 2: Conceptual Framework of ECB	73
Figure 3: Demographic Changes Over Time in Two-County ISD	82
Figure 4: Annual October 1 st Enrollment from 1985-2015	105
Figure 5: Evaluation Strength in School Board Agendas/Minutes Over Time	173
Figure 6: Overall Percentage Evaluation Connection by Superintendent Era.....	174
Figure 7: Percentage of Evaluation Connection Strength by Superintendent Era.....	174
Figure 8: Percentage of Categorical Evaluation Connection by Superintendent Era	175
Figure 9: Evaluation Strength in Focus on Two-County Publications Over Time.....	176
Figure 10: Five-Year Calendar Activity Percentages	177
Figure 11: Annual Average Calendar Activity Percentages	178
Figure 12: Annual Percentage of Evaluation Activities	178
Figure 13: Number of Years of Service by Superintendent.....	183
Figure 14: Implications for Evaluation Capacity Building in Schools.....	199
Figure 15: Conceptual Framework of ECB Expanded to Include Study Implications...	211

Chapter 1:

Introduction and Statement of the Problem

Introduction

School districts routinely lack the capacity to conduct quality evaluation to keep up with the growing mandates, rising achievement targets, increasing demands to narrow the achievement gap between student groups, and perpetually escalating evaluation needs (Huffman, Thomas, & Lawrenz, 2008; King & Rohmer-Hirt, 2011). In an informal poll following a presentation on evaluation at the February 2012 meeting of the statewide assessment group, an assembly of district personnel with responsibilities for research, evaluation, and assessment, only six of the 41 districts represented had conducted any sort of formal evaluation within the past five years, and only one of them had the capacity to conduct evaluation internally (Minnesota Assessment Group, personal communication, February 17, 2012). Foretelling the emerging challenges, at the turn of the decade, Sonnichsen (2000) forecasted the benefits of the evaluation process to organizations as they confront the dilemmas of performance demands, information overload, and reduced resources.

Statement of the Problem

Educational entities encountered the challenges predicted by Sonnichsen (2000), perpetuated in large part by the No Child Left Behind Act (2001) unprepared, with underestimated apprehension. No Child Left Behind (NCLB) legislation (2001) was a reauthorization of the Elementary and Secondary Education Act (ESEA) that focused on programs for disadvantaged youth, primarily outside of the regular classroom (Mills, 2008). NCLB (2001), however, increased accountability for public school educators by

turning attention to inside the classroom, requiring rigorous academic standards that described what students should know and be able to do at each grade level and more standardized testing than ever before (Fullan, 2016). Put another way, the reauthorization of ESEA to NCLB was “from delivery of service to achieve educational opportunities and equity to standards and outcomes” (Mills, 2008, p. 11). For example, one requirement of NCLB was that schools and districts made adequate yearly progress (AYP) based on students reaching preset targets of achievement with increasingly stringent sanction imposition for not meeting these targets (NCLB, 2001).

NCLB (2001) technically applied to the approximately 57% of public schools in the United States receiving federal Title I funding specifically (www.ed.gov), but has “set a tone throughout educational research communities” and has, in general, affected all schools across the nation (Mills, 2008, p. 10). Guillén-Woods, Kaiser, and Harrington (2008) provided evidence of this sentiment in a case study capturing the impact of NCLB on an after-school program in California. They found that the program had transitioned from collecting and analyzing non-prescribed contextual measurements of socio-emotional outcomes to more rigorous standardized, academic outcomes (Guillén-Woods et al., 2008).

NCLB (2001) was directly preceded by a number of important events. In August of 1981, the National Commission of Excellence in Education (NCEE) was formed through authority of legislation and given a charge to, among other things, determine the quality of teaching and learning at all levels (K-16) in the nation (NCEE, 1983). Next, the Improving America’s Schools Act (IASA), which was a reauthorization of ESEA preceding NCLB, along with the Goals 2000: Educate America Act (both of 1994), were

enacted to take a step forward in monitoring learning outcomes for all students as opposed to only disadvantaged students or students considered at risk of academic failure (Jorgensen & Hoffman, 2003). The signing into law of these two pieces of legislation mandated three critical components: (a) content and academic standards; (b) aligned assessments measuring progress toward the standards in the areas of reading and math, given once in each grade band – grades 3-5, grades 6-9, and grades 10-12; and (c) accountability systems for detecting schools not succeeding in helping all students reach the defined academic standards (Fullan, 2016). NCLB (2001) extended mandated testing to all grades 3-8 and one administration in high school in the areas of reading and math and added the administration of a science test once in each of the previously defined grade bands along with sanctions for schools receiving Title I funding for not reaching academic targets.

As stated above, a significant factor instigating authorization of the NCLB Act was the report *A Nation at Risk* (NCEE, 1983). One of the main findings exposed in *A Nation at Risk* (NCEE, 1983) was the discrepancy between the performance of American students and that of students in other nations around the world, especially in the areas of math and science. Other nations, such as Canada, England, Singapore, and Norway, enacted accountability legislation, many prior to the authorization of NCLB, in response to the growing milieu of global comparison emerging from this report and others like it, intensifying the necessity for educational reform in the United States to compete (Gregory & Clarke, 2003; McEwen, 1995; Møller, 2009; Webb, Vulliamy, Häkkinen, & Hämäläinen, 1998).

From the compulsion of educational reform emerged the importance of schools to gain the ability to measure the effectiveness of the programs, curriculum, and instructional methods they were employing (Mills, 2008). McEwen (1995) presaged the inevitable accountability rush, recognizing that in times of “diminishing resources and competing priorities, accountability for results will increase” (p. 9). Russ-Eft and Preskill (2009), referring to the shift in the federal government toward accountability, reflected that “increased emphasis on government program accountability and organizations’ efforts to be lean, efficient, global, and more competitive led to more urgent calls for evaluation” (pp. 41-42).

History of Legislative Impact on the Development of Evaluation

Evaluation’s development and U.S. legislation have a related and intertwining history, with legislative acts affecting the development of program evaluation through the years. In the early 20th century, scholars such as E.L. Thorndike were instigating advances in educational testing and measurement, strengthening the standardized testing progression, and leaders such as Frederick W. Taylor were conducting analyses of time and motion to optimize employee skills forming the basis for organizational change movements (Fitzpatrick, Sanders, & Worthen, 2004; Russ-Eft & Preskill, 2009).

By the 1930’s and 1940’s, many governmental services and agencies were established as a result of the response to the Great Depression, including Roosevelt’s New Deal (Fitzpatrick et al., 2004; Russ-Eft & Preskill, 2009). Paralleling this legislation, Ralph Tyler was in the midst of conducting his famous Eight-Year Study on the differences of traditional and progressive high schools on student outcomes, which became an impetus for a shift from norm-referenced testing, which had become

commonplace in schools, to criterion-referenced testing (Fitzpatrick et al., 2004; Russ-Eft & Preskill, 2009). In addition, the Eight-Year Study transformed thinking around measuring student progress throughout their coursework, demonstrating the benefits and effectiveness of employing questionnaires, observations, examination of student work and performance-based assessments, to augment testing (Tyler, 1986; Tyler, 2013).

In 1957, the United States government enacted the National Defense Education Act, focusing on curriculum development in the nation, propagated by Sputnik and the Russian space program (Jorgensen & Hoffman, 2003). At the same time, privately supported evaluations using educational assessments were being conducted in schools, mainly by university researchers (Fitzpatrick et al., 2004; Russ-Eft & Preskill, 2009).

By the mid-1960's, President's Johnson's War on Poverty and Great Society legislation formed the basis for the Elementary and Secondary Education Act (ESEA), championed by Robert Kennedy, focusing on compensatory education for disadvantaged children (Jorgensen & Hoffman, 2003). This legislation embedded an evaluation component into federal allocation of monies, beginning the area of public accountability, and the extensive growth of evaluation practices (Fitzpatrick et al., 2004; Russ-Eft & Preskill, 2009).

In the 1970's and 1980's, the evaluation field continued to expand and practices such as internal evaluation emerged. Other concepts such as organizational learning and evaluation use – using the results of evaluation toward specific organizational purposes as well as the process of evaluation as a means to teach evaluation – gained momentum (Torres & Preskill, 2001). (See Figure 1.) With continued attention on international comparisons, the U.S. Department of Education formed the National Commission of

Excellence in Education to determine the state of education in the United States, culminating with the report entitled *A Nation at Risk: The Imperative for Educational Reform*, discussed above. At the same time, funding for evaluation and governmental programs greatly decreased under President Reagan (Russ-Eft & Preskill, 2009).

Legislatively in the 1990's, ESEA was reauthorized as the Improving America's Schools Act (IASA), and in conjunction with the Goals 2000: Educate America Act, governmental focus shifted to quality education for all students, not just disadvantaged students, with standards and aligned testing requirements for all public schools receiving federal Title I dollars (Jorgensen & Hoffman, 2003; Mills, 2008). Related to evaluation, training programs strengthened, moving evaluation into all organizational areas, associations formed in the previous decades grew and connected internationally, and the practice of participant involvement in the evaluation process increased (Fitzpatrick et al., 2004; Russ-Eft & Preskill, 2009).

The 21st century brought new legislation that significantly changed the educational landscape. The ESEA was again reauthorized, this time as the No Child Left Behind Act (NCLB), explicated above. Additionally, the America Recovery and Reinvestment Act (ARRA) provided one-time monies to districts and schools to innovate changes by saving and creating jobs and reforming education by focusing on college-and-career readiness, data systems, teacher effectiveness, and intervention and support for low-performing schools (ARRA, 2009). Building evaluation capacity within organizations grew in intensity, becoming more central to the field (Fitzpatrick et al., 2004; Russ-Eft & Preskill, 2009).

In December 2015, NCLB was again reauthorized into the Every Student Succeeds Act (ESSA). The ESSA (2015) lessened federal control and turned direct oversight to state-level entities while maintaining the testing requirements and the mandate to intervene with schools deemed as needing assistance. Accountability shifted from a primary focus on proficiency to an incorporation of growth and closing the achievement gap between student groups. Graduation rates and college-and-career readiness were also integrated factors along with attention to early childhood education. The student groups measured also expanded to incorporate youth experiencing homelessness, children of military families, and children from other countries with little to no formal education. Also required was a measurement non-related to testing such as student engagement, attendance, or discipline, left up to each state's choosing (ESSA, 2015).

Below in Figure 1 is the visual overview of the full evaluation and legislative history using gradation to depict growing intensity of evaluation components. Following the figure, Tables 1 and 2 are zoomed in views of Figure 1, broken into the progression from Ancient China through the 1950s and the 1960s through 2010s, respectively.

Timeline of Evaluation and Legislative Implications to Evaluation Capacity Building

Year	1800s	1900s	1910s	1920s	1930s	1940s	1950s	1960s	1970s	1980s	1990s	2000s	2010s
Early Evaluation 2000-2001 E.C.	Chinese Dynasties Exams to determine professions	Horace Mann: Written essay Exams in place of oral exams	E.L. Thorndike: Investigated educational testing and measurement	Frederick W. Taylor, Time & Motion Study to optimize employee skills	Use of norm- referenced tests to measure comprehense	Ralph Tyler: Eight-year study on differences in learning between progressive high schools and traditional schools on student outcomes	Tyler's work propagated a shift from norm- referenced to criterion- referenced testing	Embedding of an evaluation component into federal education policy the area of accountability and the growth of evaluation practices	Michael Scriven: Evaluation to determine merit or worth Emergence of numerous tests and writing on evaluation: Campbell, Scriven, Stake, Stufflebeam, Whitley, Patton	Michael Patton: Evaluation for Utilization Focused Evaluation gains prominence Emergence of evaluation testing programs, a growing variety of models of evaluation, professional organizations and numerous journals - Program Evaluation - Standards created	Greater emergence of evaluation in all organizational areas Increased focus on stakeholder involvement in the evaluation process Concerted effort to connect internationally around evaluation	Increased acceptance of mixed methodology in conducting evaluations Increased emphasis on insurgence of internal evaluation units within organizations	Increased reliance on evaluation and evaluative thinking as a way of operation within organizations
Evaluation Developments	<p>Development and Growth of Standardized Testing</p> <p>Understanding and Development of Organizational Change</p> <p>Evaluation as a Growing Field</p> <p>Surfacing of Organizational Learning</p> <p>Emergence of Internal Evaluation Defining and Refining of Process Use</p> <p>Evaluation Capacity Building</p>												
	Scientific & other Goal teachers: Use of questioning in the learning process												
U.S. Legislation													
			Establishment of many governmental agencies and agencies as a result of the Great Depression and Roosevelt's New Deal	National Defense Education Act (1958) in response to Sputnik, focusing on curriculum development	Johnson's War on Poverty, B. Grant Carter's Elementary and Secondary Education Act (1965) focused on compensatory education to disadvantaged children	National Commission on Excellence in Education (1983) regarding the state of education in the United States Focus on evaluation and governmental programs to greatly decreased under Reagan	Reauthorization of ESEA to Improve America's Schools Act (1994) Goals 2000: Educate America Act (1994) Twentieth Anniversary legislation focused on quality education for all students with standards and aligned testing requirements	Reauthorization of ESEA to No Child Left Behind (2002) Greater amount of testing required and sanctions imposed for not making adequate yearly progress	America Reinvests and Reinvestment Act (2009) Race to the Top Bill (DEB - Not yet enacted) Every Student Succeeds Act (12/2015) Focused on linking the achievement gap and additional measures of student outcomes such as growth and college readiness - more accountability to states				

Adapted based on Argenson & Ruffman (2008); Fitzpatrick, Sanders, & Worthen (2004); and Bane-EP & Pineda (2008)

Figure 1. History of evaluation and United States federal legislation

Table 1. *Timeline of Evaluation and Legislative Implications to ECB – Part 1*

	Early Evaluation	Modern Day Evaluation						
	2000 - 300 B.C.	1800's	1900's	1910's	1920's	1930's	1940's	1950's
Evaluation Developments	Chinese Dynasties: Exams to determine professions	Horace Mann: Written essay Exams in place of oral exams	E.L. Thorndike: Instigated advances in educational testing and measurement	Frederick W. Taylor: Time & motion analysis to optimize employee skills	Use of norm-referenced testing in schools became commonplace	Ralph Tyler: Eight-year study on differences between traditional and progressive high schools on student outcomes	Tyler's work propagated a shift from norm-referenced to criterion-referenced testing	Privately supported evaluations using educational assessment conducted mainly by university researchers
	Socrates & other Greek teachers: Use of questioning in the learning process	Development and Growth of Standardized Testing Understanding and Development of Organizational Change						
U.S. Legislation						Establishment of many governmental services and agencies as a result of the Great Depression and Roosevelt's New Deal		National Defense Education Act (1958) in response to Sputnik, focusing on curriculum development

Adapted based on Jorgensen and Hoffman (2003); Fitzpatrick, Sanders, and Worthen (2004); and Russ-Eft and Preskill (2009)

Table 2. *Timeline of Evaluation and Legislative Implications to ECB – Part 2*

	Modern Evaluation					
	1960's	1970's	1980's	1990's	2000's	2010's
Evaluation Developments	Embedding of an evaluation component into federal allocation of monies, beginning the era of accountability and the growth of evaluation practices	Michael Scriven: Evaluation to determine merit or worth Emergence of numerous texts and writings on evaluation: Campbell, Suchman, Stake, Stufflebeam, Wholey, Patton	Michael Patton: Evaluation for more than making judgments; Utilization-Focused Evaluation gains prominence Emergence of evaluation training programs, a growing variety of models of evaluation, professional organizations and numerous journals - Program Evaluation Standards created	Greater emergence of evaluation in all organizational arenas Increased focus on stakeholder involvement in the evaluation process Concerted effort to connect internationally around evaluation	Increased acceptance of mixed methodology in conducting evaluations Insurgence of internal evaluation units within organizations	Increased reliance on evaluation and evaluative thinking as a way of operation within organizations
	Era of Public Accountability					
	Organizational Change	Emergence of Internal Evaluation		Evaluation Capacity Building		
	Evaluation as a Growing Field	Defining and Refining of Process Use				
	Surfacing of Organizational Learning					
U.S. Legislation	Johnson's War on Poverty & Great Society Legislation: Elementary and Secondary Education Act (1965) focused on compensatory education to disadvantaged children	National Commission on Excellence in Education: A Nation at Risk (1983) depicting the state of education in the United States Funding for evaluation and governmental programs is greatly decreased under Reagan	Reauthorization of ESEA to Improving America's Schools Act (1994) Goals 2000: Educate America Act (1994) Together this legislation focused on quality education for ALL students with standards and aligned testing requirements	Reauthorization of ESEA to No Child Left Behind (2002) Greater amount of testing required and sanctions imposed for not making adequate yearly progress	America Recovery and Reinvestment Act (2009) Race to the Top Bill (2013 – not enacted) Every Student Succeeds Act (12/2015) Focused on closing the achievement gap and other student outcomes (growth and college readiness) - reduces federal control, grants	

Adapted based on Jorgensen and Hoffman (2003); Fitzpatrick, Sanders, and Worthen (2004); and Russ-Eft and Preskill (2009)

Evaluation Capacity Building (ECB)

The changing educational environment catalyzed an imperative need to cultivate and expand organizations' evaluation capability. Acknowledging this growing need, promulgated in large part by increased accountability through NCLB and subsequent reauthorizations of ESEA, Russ-Eft and Preskill (2009) contend the following:

Evaluation is no longer something that should only be done by those outside of organizations; indeed . . . all of us have responsibilities for asking evaluative questions, seeking information to answer those questions, making decisions based on the answers, and taking action to implement the recommendations. (p. 46)

Their statement is an endorsement of the need for evaluation capacity building.

Evaluation Capacity Building (ECB) is a critical issue in education because, given the complexities of proliferating student needs, expanding skills and abilities required for post-secondary success, the capability to compete in a global society, increased accountability, declining resources, and rapidly changing environments, school districts need the ability to conceptualize, design, and conduct quality evaluations to ensure effective and efficient programs and operations. Furthermore, districts need to be able to determine that these programs are implemented with fidelity and are achieving the desired results. Quality evaluations, in this sense, are gauged against the Program Evaluation Standards of utility, feasibility, propriety, accuracy, and evaluation accountability (Yarbrough, Shula, Hopson, & Caruthers, 2011).

Baizerman, Compton, and Stockdill (2002a) define ECB as a:

. . . context-dependent, intentional system of guided processes and practices for bringing about and sustaining a state of affairs in which quality program evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations/programs/sites. (p. 109)

Others provide similar definitions (Cousins, Goh, Clark & Lee, 2004; King, 2007; Preskill & Boyle, 2008). In their major points about ECB, Stockdill, Baizerman, and Compton (2002) state that ECB and evaluation practice differ. Concurring with this sentiment, Huffman et al. (2008) make the distinction that ECB is different than merely conducting evaluation due to the intentionality given to building evaluation capacity within an organization. They continue, clarifying that ECB goes beyond professional development in that it purposefully attempts to “extend the development of individuals to affect future work of the organization” (Huffman et al., 2008, p. 359).

Although the issue is of utmost importance, there are currently no proven models or existing research on frameworks for organizations to follow (Preskill & Boyle, 2008). Scholars have provided frameworks related to ECB (Baizerman et al., 2002a; Cousins et al., 2004; García-Iriarte, Suarez-Balcazar, Taylor-Ritzler, & Luna, 2011; King & Volkov, 2005; Preskill & Boyle, 2008; Stockdill et al., 2002), yet the contextual factors affecting organizational capacity to conduct quality evaluation related to the frameworks have not been extensively studied. The empirical experiences that educational entities undergo to embed evaluative thinking into everyday processes is another area that has not been studied.

Purpose of the Study

The purpose of this study was to examine the process and effectiveness of building evaluation capacity within school districts to address the circumstances described above. Given the growing culture of accountability partnered with the declining resources available to meet the increasing demands of rigorous academic proficiency and growth for all students, schools and districts need to be able to determine what is working

and what is not. The desire to prepare every student for success in whatever his or her chosen path entails is an equally driving force for school personnel. School districts are seeking expertise and creating internal positions and departments to address these needs, however, guidance and direction as to how to build and sustain evaluation capacity are lacking.

Brandon and Higa (2004) characterize the literature on ECB as a “meager body of work” (p. 140), and Stockdill et al. (2002) say it is limited. Given the passage of time since Stockdill et al. (2002) and Brandon and Higa (2004) conducted their studies, one might imagine that the body of literature may have sufficiently expanded. Negating that supposition, however, Labin, Duffy, Meyers, Wandersman, and Lesesne (2012) articulate that there is still a need to study ECB with “much room for growth and refinement” (p. 23). Preskill and Boyle (2008) also call for a more “robust knowledge base” of ECB, adding that ECB is “an area ripe for exploration,” especially through empirical research (p. 457), an appeal mimicked by others (Amo & Cousins, 2007; Brandon & Higa, 2004). Baizerman et al. (2002a) articulate a need for “the real, actual orientations and related practices of everyday ECB work” (p. 113) only possible through reflective practice of ECB practitioners.

From an empirical standpoint, the literature is relatively silent on whether educational leaders believe that evaluation is important to the mission and vision of the school district, however, Cousins, Goh, & Clark (2006) say that mission and vision clarity are “key strategic building blocks found in learning organizations” (p. 158). In addition, researchers have called for further study on the factors that contribute to organizational evaluation capacity, especially related to process use as a strategy

(Balthasar, 2008; Cousins et al., 2004; Cousins et al., 2006; Harnar & Preskill, 2007; Robinson & Cousins, 2004). With these things in mind, the following are research questions that were explored in this study. The overarching question was: What has been the experience of one school district in building evaluation capacity? The study addressed three sub-questions:

- 1) What are the perceptions of district leaders as to the importance of evaluation to the mission and vision of the district?
- 2) What is the evidence of evaluation capacity within the organization overtime?
- 3) What contextual events expedite or delay evaluation capacity building?

Significance of the Study

This study provides longitudinal evidence of the context, processes, and activities that impact evaluation capacity building within K-12 educational settings. It is different than other studies that have been done to date in that it spanned a thirty-year history in one suburban school district, drawing from the documented experiences of two formally-trained evaluators at different points during the thirty years, along with extensive records of activities and events across that time period. The study traced the evidence of evaluative activities over time as well as the development and integration of an internal evaluation unit and the steps taken to build structures to support evaluation and its use throughout the school district, grounded in the development of evaluation and the federal legislation related to education throughout the thirty-year interval from 1985 – 2015. (See Table 3.)

Table 3. *Timeline of Evaluation and Legislative Implications to Evaluation Capacity Building from 1985 – 2015*

Timeline of Evaluation and Legislative Implications to Evaluation Capacity Building						
Timeline	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010 - 2015
Evaluation Developments	Era of Public Accountability					
	Emergence of Internal Evaluation Defining and Refining of Process Use Surfacing of Organizational Learning			Evaluation Capacity Building		
U.S. Legislation	National Commission on Excellence in Education: A Nation at Risk (1983) depicting the state of education in the United States Funding for evaluation and governmental programs is greatly decreased under Reagan	Reauthorization of ESEA to Improving America's Schools Act (1994) Goals 2000: Educate America Act (1994) Together this legislation focused on quality education for ALL students with standards and aligned testing requirements	Reauthorization of ESEA to No Child Left Behind (2002) Greater amount of testing required and sanctions imposed for not making adequate yearly progress	America Recovery and Reinvestment Act (2009) Race to the Top Bill (2013 - Not yet enacted) Every Student Succeeds Act (12/2015) Focused on closing the achievement gap and additional measures of student outcomes such as growth and college readiness - reduces federal control, grants more autonomy to states		

Adapted from Jorgensen and Hoffman (2003); Fitzpatrick, Sanders, and Worthen (2004); and Russ-Eft and Preskill (2009)

Baizerman, Compton, and Stockdill (2002b), related to building evaluation capacity, acknowledge that “practices are difficult to name and there is little description of how these are carried out; more research is needed” (p. 102). In addition, Huffman et al. (2008) emphasize that there is a need for the field to extend ECB methods beyond individual professional development or training and to focus on collaborative methods. This study contributes to the knowledge base of evaluation by identifying and naming the practices used for ECB, including collaborative approaches, incorporating reflective practice, and by filling some of the gaps in literature described above.

Related to the field of education, Huffman et al. (2008) found that “within the context of accountability, educators seek methods of using data for decision making and they realize that increasing their capacity for conducting evaluation and using evaluation data are ways to achieve that goal” (p. 363). At this time the field of education is

somewhat overwhelmed with the implications of public accountability, dwindling resources, the growing needs of individual students, and the desire to do whatever it takes to meet the ever-increasing demands associated with the role of ‘educator.’ Based on the findings of Huffman et al. (2008), ECB is one way to assist in this realm. This study benefits the knowledge base in education by providing school districts empirical research on the steps, strategies, and activities that can propel an organization toward ECB.

Definition of Key Terms

Accountability. Accountability is defined by the business dictionary online as the obligation of an individual or organization to account for its activities, accept responsibility for them, and to disclose the results in a transparent manner (www.businessdictionary.com). The literature pertaining to accountability in education defines it as “the process of evaluating school performance on the basis of student performance measures” (Hanushek, Machin, & Woessmann, 2011).

The compounding of these definitions captures the essence of the term in the literature related to educational accountability. The resulting definition – “accountability is the process of evaluating school performance on the basis of student performance and the obligation of an individual or organization to account for its activities, accept responsibility for them, and to disclose the results in a transparent manner” – is the definition best representing accountability within this paper.

Evaluation. Michael Scriven in 1967 defined evaluation as “judging the worth or merit of something” (Fitzpatrick et al., 2004, p. 5). Fitzpatrick et al. (2004) go on to define evaluation for themselves as “the identification, clarification, and application of defensible criteria to determine an evaluation object’s value (worth or merit) in relation to

those criteria” (p. 5). In the simplest terms, Michael Quinn Patton states that evaluation is “. . . said to answer three questions: What? So what? and Now what?” (Patton, 2008, p. 5). He formally expands the definition of program evaluation, however, beyond that of judgment, describing it as “the systematic collection of information about the activities, characteristics, and results of programs to make judgments about the program, improve or further develop program effectiveness, inform decisions about future programming, and/or increase understanding” (Patton, 2008, p. 39). Evaluation in this paper will represent the broader definition, submitted by Patton, moving beyond judgment to purposes of program improvement, informing decisions, and expanding understanding.

Evaluation Capacity Building. Baizerman et al. (2002a) define ECB as a “context-dependent, intentional system of guided processes and practices for bringing about and sustaining a state of affairs in which quality program evaluation and its appropriate uses are ordinary and ongoing practices within and/or between one or more organizations/programs/sites” (p. 109). In other words, ECB is the systematic and intentional development, encouragement, and support of evaluation and evaluative thinking throughout the organization, along with the leadership commitment and dedicated resources to employ and sustain these efforts. The definition offered by Baizerman et al. (2002a), in conjunction with the restatement of this definition, is the meaning of ECB put forth in this paper. As to differences between ECB and Patton’s (2008) *process use*, Preskill and Boyle (2008) explain that “one distinguishing characteristic of ECB, and one that differentiates it from *process use* is the inclusion of clearly identified objectives” (italics added, p. 450).

Evaluative Thinking. Vo (2013) developed a working definition for evaluative thinking in her dissertation research. She defined it as:

A particular kind of critical thinking and problem-solving approach that is germane to the evaluation field. It is the process by which one marshals evaluative data and evidence to construct arguments that allow one to arrive at contextualized value judgments in a transparent fashion. (p.107)

In this study, I have expanded this definition beyond arriving at value judgments. Evaluative thinking in this study is defined by the researcher as an internalized need for evaluation – “What? So what? Now what?” (Patton, 2008, p. 5) -- as a way of being and of conceptualizing one’s work, i.e., consistently asking questions about program, policy, and practice efficiencies and seeking ways to systematically determine if they are producing the desired results. In addition, evaluative thinking is striving to be a proficient consumer of research, data, and information so as to be able to critically review and interpret them. Finally, it is having the ability to identify when, how, and where to gain support and consultation within an organizational context.

Internal Evaluation. Sonnichsen (2000) defines internal evaluation in this way; “[i]nternal evaluation is a systematic process for collecting and analyzing reliable performance data and interpreting results to inform and clarify issues that will illuminate and facilitate the decision-making process within organizations” (p. 3). He goes on to state that internal evaluators are “fundamentally problem solvers and change agents” (Sonnichsen, 2000, p. 3). According to Sonnichsen (2000), “practicing internal evaluation in organizations is the art of merging science with management, applying research principles and techniques...to determine effectiveness, efficiency, economy, and performance” (p. 3).

Organizational Change. From an organizational standpoint, Quattrone and Harper (2001) profess that “defining change is problematic” (p. 403), as it has multiple aspects and implications. They explain that change is an intentional choice to move in a rational direction based on information and knowledge, however, they also contend that movement in organizations is forced based on pressure and isomorphism (Quattrone & Hopper, 2001). Change can also be viewed as a “range of skills and knowledge that is altered rather than a specific action,” and the alteration is “not just substitution but could also include strengthening existing skills” (Weick & Quinn, 1999, p. 377). In addition, change encompasses “cognitive restructuring in which words are redefined to mean something other than had been assumed, concepts are interpreted more broadly, or new standards of judgment and evaluation are learned” (Weick & Quinn, 1999, p. 372). Tsoukas and Chia (2002) paint the picture of change in organizations to be a “reweaving of actors’ webs of beliefs and habits of action to accommodate new experiences obtained through interactions” (p. 567).

The definition of organizational change used in this paper is a culmination of ideas from Quattrone and Harper (2001), Weick and Quinn (1999), and Tsoukas and Chia (2002). Change within organizations is both intentionally chosen and compelled by outside forces, is both acquiring new skills and meanings and strengthening knowledge that already exists, and is collaborative and individual acquisition obtained through interactions.

Organizational Learning. Torres and Preskill (2001) propose that organizational learning is “a continuous process of growth and improvement that (a) uses information or feedback about both process and outcomes to make changes; (b) is integrated with work

activities, and within the organization's infrastructure; and (c) invokes the alignment of values, attitudes, and perceptions among organizational members" (p. 388). They go on to explain that as organizational learning relates to evaluation, it involves:

- Establishing a balance between accountability and learning roles for evaluation;
- Integrating the evaluation function and evaluator role within the organization;
- Developing frameworks for relating findings about particular programs and initiatives to broader organizational goals;
- Sustaining a spirit of ongoing inquiry which calls for learning incrementally and iteratively over time;
- Providing time for reflection, examination of underlying assumptions, and dialog among evaluators, program staff, and organizational leaders; and Reconsidering traditional evaluator roles and the skills evaluators need (Torres & Preskill, 2001, p. 388)

The definition presented by Torres and Preskill (2001) encompasses aspects related to accountability, ECB, and internal evaluation.

Process Use. Patton (2008) defines process use as "individual changes in thinking, attitudes, and behavior, and program or organizational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process" (p. 155). The definition stated above assumes participation of stakeholders throughout the evaluation process. Process use, at its simplest, is learning evaluation by doing it.

Overview of Chapters

Chapter 2 of this paper provides an overview of the literature related to organizational change, organizational learning, accountability, process use, internal evaluation, and evaluation capacity building (ECB) along with a conceptual map illustrating connections among these bodies of writing. Chapter 3 describes the research methodology used in this study to examine ECB practice in the context of a school

district. In this chapter, the chosen case is described, along with a brief description of the methods used to collect and analyze the data. In Chapter 4, the findings of the study are presented in a chronology of events told in story form, culminating with a summary of findings from the different methods of data collection. In Chapter 5, a discussion of the research is provided, explicating overarching themes and implications of the findings with ties made to existing literature. Finally, in Chapter 6, answers to the research questions are discussed along with suggestions for further research related to ECB within educational settings. The chapter ends with conclusions, reflections and future hopes.

Throughout this paper, pseudonyms are used for the names of the organization and the participants in the study. The setting is described as a metropolitan area school district in a Midwestern state. By making this choice, the anonymity of the participants and the school district will be maintained as much as possible.

Chapter 2:
Review of the Literature: Theoretical Underpinnings
of Evaluation Capacity Building

The foundation of ECB is supported and upheld by other theories and concepts such as organizational change/organizational learning, accountability, and process use and is strengthened by the role of internal evaluation. The following literature review examines each of these components more fully and explores specific premises related to accountability and process use as underpinnings for ECB.

Crossan, Lane, and White (1999) explicate a framework proposition process built on the declaration of key premises that serves as the basis for the structure of this review. Crossan et al. (1999) write that “a framework defines the territory and takes us a step closer to a theory” (p. 523). They go on to explain that the process includes “identifying the phenomenon of interest, stating key premises or assumptions underlying the framework, and describing the relationship among the elements of the framework” (Crossan et al., 1999, p. 523). The phenomenon that will be explored in this chapter is evaluation capacity building. Key premises will be discussed later in the chapter in the sections related to accountability and process use, along with the increased effectiveness of the ECB process through internal evaluation. Prior to establishing the premises, however, organizational change and organizational learning will be examined.

Literature was selected for review by conducting searches of ‘evaluation capacity building,’ ‘process use,’ ‘accountability,’ ‘internal evaluation,’ and ‘organizational learning’ in two primary databases (Google Scholar and JStor), as well as e-journals such as *New Directions for Evaluation*, the *American Journal of Evaluation*, *Evaluation*, and

Evaluation and Program Planning. Further searches were done based on links to related articles provided in each search. In addition, reference lists of selected articles were cross-checked for potential discovery of other related research until reference lists and searches produced extensively replicated sources. Moreover, relevant literature from coursework was included as well.

Organizational Change and Learning as the Groundwork to Evaluation Capacity Building

Examining organizational change. The French writer Francois de la Rochefoucauld (1613-1680) is credited with saying, “The only thing constant in life is change” (en.thinkexist.com, 4.21.2013). Weick and Quinn (1999) assert that the “acceptance of change as a constant” (p. 375) is an initial step to understanding organizational change. One would be hard pressed to find an individual who could not relate a story of change in his or her life or work place, some positively based, others with negative outcomes. The following section of this paper is a discussion of organizational change and theories on organizational change and strategies contributing to successful organizational change. The literature related to organizational change comes from many organizational contexts yet are applicable to educational entities, as well.

Theories about organizational change. Organizational change is commonly referenced in research literature as both transformational and episodic, encountering either inertial forces within the organization, or adaptive and continuous (March, 1981; Hannan & Freeman, 1984; Weick & Quinn, 1999; Tsoukas & Chia, 2002; Plowman et al., 2007). Additional scholarly analysis of change found that “radical change can be the result of a pattern[s] of small, micro level changes that occur over time” (Plowman et al.,

2007, p. 516). Plowman et al. (2007) expand the contributions described above with the idea that change can be continuous and radical. Their research categorizes change through combinations of convergent or radical descriptions of change paired with continuous or episodic classification, all with either intentional or accidental components. They determined that “ongoing interactions of amplifiers, organizational conditions, and small changes enable local adaptations to accumulate into a pattern that is radical and continuous” (Plowman et al., 2007, p. 538). Change that is continuous and episodic, as well as convergent and radical, exemplifies what is occurring in education today.

In general, it is easier to change appearance, procedures, and policies that are closer to the edges of the organization than those at the organizational core. Furthermore, organizations are prone to protect routines and the rules used to switch between routines (Hannan & Freeman, 1984). The primary cause of this instinctive inactivity is related to inertial forces within the organization. Even so, the potential impetuses for more extensive change or for stronger inertial forces are desired reliability and accountability, imposed either internally or externally (Hannan & Freeman, 1984) or a construed imbalance between the organization and its environment (Schneider, Brief, & Guzzo, 1996; Weick & Quinn, 1999; Quattrone & Hopper, 2001; Tsoukas & Chia, 2002; Plowman et al., 2007). Although inertia implies that organizations typically change at a slow rate provoked by an intentional overt act, it does not imply that they will not or cannot change (Hannan & Freeman, 1984). These conditions are true in American education, with far greater intensity since the release of *A Nation at Risk* (NCEE, 1983) spurred public outcry for greater accountability and evidence of effectiveness in the nation’s schools.

Schneider et al. (1996) contend that the only way to truly approach change is through “total organizational change” (p. 6), i.e., radical, organization-wide transformation in virtually every aspect of the company’s culture and climate. Radical change of this type can also be labeled episodic change (Weick & Quinn, 1999). Episodic change, as described by Weick and Quinn (1999), is characterized as “infrequent, discontinuous, intentional” change spurred by a “misalignment between inertial structure and perceived environmental demands” (p. 365). “Episodic change requires both equilibrium breaking and transitioning to a newly created equilibrium and is mostly associated with planned, intentional change” (Weick & Quinn, 1999, p. 371). Leaders in the legislature over the last two decades seemingly agree with Schneider et al. (1996) as demonstrated by widespread demand for sweeping educational reform, characterized in laws such as IASA (1994), Goals 2000 (1994), NCLB (2001), ARRA (2009) and ESSA (2015).

Hannan and Freeman (1984) caution that in undergoing ecologically responsive change, however, timing should be of primary consideration especially when the change is of great magnitude. Frequent, ongoing changes, particularly significant radical changes “often threaten [organizational] legitimacy” (Hannan & Freeman, 1984, p. 149). Repeated attempts of episodic change in response to environmental changes can be very detrimental as the potential of the inability to keep up with the changing environment exists (Hannan & Freeman, 1984). Hannah and Freeman’s (1994) assertion is evident in educational entities today. Districts are increasingly unable to keep up with governmental changes, including intensifying unfunded mandates, mounting reliance on testing, and relentless requirements of reporting (King & Rohmer-Hirt, 2011).

An opposing perspective claims that “organizations are continually changing, routinely, easily, and responsively, but change within organizations cannot be arbitrarily controlled” (March, 1981, p. 563). Hannan and Freeman (1984) concur with this view, as well, stating that organizational change is, for the most part, spontaneous and serendipitous. Tsoukas and Chia (2002) put it another way; “organizations are sites of continuously evolving human action” (p. 567). Weick and Quinn add to the discussion, declaring that “from a distance, when observers examine the flow of events that constitute organizing, they see what looks like repetitive action, routine, and inertia dotted with occasional episodes of revolutionary change. But a view from closer in suggests ongoing adaption and adjustment” (Weick & Quinn, 1999, p. 362). The perspective of ongoing, continuous change is also a reality in education today. Districts and schools within districts seemingly are not changing from an outside perspective, yet they regularly write and monitor school improvement plans, which are only truly evident to those within the organization (King & Rohmer-Hirt, 2011).

Weick and Quinn (1999) name this type of alteration continuous change depicted by “ongoing, evolving, cumulative change” (p. 375). They go on to explain that “the distinctive quality of continuous change is the idea that small continuous adjustments, created simultaneously across units, can cumulate and create substantial change” (Weick & Quinn, 1999, p. 375). Weick and Quinn also describe a process relating to continuous change that they deem as organizational improvisation, alleging that “variable inputs to self-organizing groups of actors induce continuing modification of work practices and ways of relating” (Weick & Quinn, 1999, p. 375). They argue that “organizations produce continuous change by means of repeated acts of improvisation involving

simultaneous composition and execution, repeated acts of translation that convert ideas into useful artifacts that fit purposes at hand, or repeated acts of learning that enlarge, strengthen, or shrink the repertoire of responses” (Weick & Quinn, 1999, p. 377). Finally, Weick and Quinn maintain that continuous change differs from the other types of change in terms of “continuity and scale” (Weick & Quinn, 1999, p. 378). The type of change described by Weick and Quinn (1999) is a principal type of change underlying ECB.

Strategies for successful organizational change. Identifying and measuring successful and meaningful change is very difficult because of the “fundamental ways in which changes are transformed by the processes of change [itself]” (March, 1981, p. 569). Even though the success of change is difficult to measure, research, and experience, Schneider et al. (1996) allege that “organizations as we know them are the people in them; if the people do not change, there is no organizational change” (p. 1). Research also reveals that the way to gain new information and to change one’s opinion is through conversational dialogue with others who have differing opinions or ways of looking at things (Weick & Quinn, 1999, referencing Schein, 1996). Grounded in this exposition, Judge, Thoresen, Pucik, and Welbourne (1999) write that the “success of change efforts lay in the abilities and motivation of individuals within the organization” (p. 117).

Research shows that organizational leaders are of utmost importance for worthwhile change (Fullan, 2016). Kotter (1996) emphasizes this by stating that, “only leadership can get change to stick by anchoring it in the very culture of an organization” (p. 30). Leaders in organizations must first undergo personal changes and sort out their ideas about what is important and the direction to head in before change is attempted for others in the organization (Weick & Quinn, 1999). Plowman et al. (2007) conclude based

on their research that “leaders serve as ‘sensegivers,’ giving meaning to the changes that were unfolding rather than creating and directing the changes Their use of language was skillful because it gave meaning to emergent changes and helped draw attention to the pattern that was forming” (p. 538).

Organizational processes and structure also play a role in productive change. Research conducted by Weick and Quinn (1999) suggests that “successful firms had well-defined managerial responsibilities and clear project priorities while also allowing the design processes to be highly flexible, improvisational, and continuously changing,” along with “richly connected communication systems” (p. 371).

Expanding upon individual change, March (1981) states that “organizations are also transformed in the process. [Organizational transformation occurs as] organizations develop and redefine goals while making decisions and adapting to environmental pressures” (p. 570). March (1981) goes on to say that “small signals are amplified into larger ones, and the general implication is that routine adaptive processes have consequences that cannot be understood without linking them to an environment that is simultaneously, and endogenously, changing” (p. 571).

Change, however, has to be more than superficial changes, changes in words, text, or structure in an organization. There is no such thing as a “silver bullet” change that will really make a difference (Schneider et al., 1996, p. 7). In fact, Weick and Quinn (1999) summarize the work of Bunker and Alban (1992) by affirming that

. . . gathering data from the environment and sharing it widely, [employing] real-time analysis and decision making, learning about the whole organization, [utilizing a] mixed model of driving change involving both senior management and the organization, and [being] participant centered... (p.374)

are ways to increase the likelihood of success in the alteration process. Weick and Quinn (1999) continue, imparting that

. . . depictions of successful revolutions, however, tend to downplay the degree to which earlier sequences of incremental change made them possible. This oversight is serious because people tend to attribute the success of revolution to its break with the past and its vision of the future, whereas that success may actually lie in its connection with the past and its retrospective rewriting of what earlier micro-changes meant. (p. 379)

Findings related to participant involvement are aligned with the work of Schneider et al. (1996), as well.

The probability of actually having the change take root is enhanced when people feel their work is challenging, when they can participate in decisions regarding how the change will be achieved, and when their interpersonal relationships are characterized by mutual trust. Only by attending to the full range of these concerns will the psychology or feel of the organization be altered and the changes take root. (Schneider et al., 1996, p. 4)

Collinson, Cook and Conley (2010) state that “in the practical world or schools and school systems, organizational learning provides a sustainable avenue for change and an opportunity for continuous renewal from within” (p. 109).

Examining organizational learning. Organizational learning can be thought of as a subset of organizational change. Schein (1996) contends that “all forms of learning and change start with some form of dissatisfaction or frustration generated by data that disconfirm our expectations or hopes” (p. 29). He goes on to explain that either the information can be ignored and rationalized, hence the situation remains status quo, or it can be accepted, prompting change (Schein, 1996). Fullan and Miles (1992) add to the connection between learning and change affirming that “all change involves learning [so] conditions that support learning must be part and parcel of any change effort” (p. 749). In

the case of a learning organization, the specific change that occurs can be in effectiveness (Huber, 1991; Kim, 1993), cognition or action (Crossan et al., 1999; Huber, 1991; Kim, 1993; Senge, 1990), or a strategic renewal (Crossan et al., 1999). Since Senge (1990) first introduced the seminal five dimensions of a learning organization, the literature began to examine this facet of organizational change.

Delving into the realm of organizational learning soon exposed other quandaries. Many researchers began to consider types of learning such as *survival* or *adaptive learning* versus *generative learning* (Senge, 1990), *exploitation* versus *exploration* (March, 1991), or *feedback* versus *feed forward* (Crossan et al., 1999). The concept of *adaptive learning/ exploitation/ feedback* is characterized by refining what is already known or improving practices and processes that already exist within the organization, and is seemingly best aligned with continuous change. The concept of *generative learning/ exploration/ feed forward* is characterized by investigating new ideas and innovations that could be used to improve or reform the organization. This type of learning is seemingly most closely aligned with episodic change. A balance of both types of organizational learning is needed to for an organization to remain healthy and viable (March, 1991).

In addition to the type of change that occurs in organizational learning, researchers also considered the organizational levels at which learning occurs (Crossan et al., 1999; Kim, 1993; Senge, 1990; Torres, 1994). Learning that occurs at the individual, group, and organizational levels is what March (1991) refers to as a “nested system” (p. 72). March (1991) goes on to explain that learning within the system is a reciprocal process that he calls “mutual learning” (p. 73), contending in broad terms that:

Organizations store knowledge in their procedures, norms, rules, and forms. They accumulate such knowledge over time, learning from their members. At the same time, individuals in an organization are socialized to organizational beliefs. (p. 73)

Senge (1990) explains that “organizations learn only through individuals who learn.

Individual learning does not guarantee organizational learning. But without it no organizational learning occurs” (p. 129). Torres (1994) states this concept plainly, saying, “what does not happen first on an individual level cannot happen on an organizational level” (p. 329).

Evaluation capacity building as a specific type of organization learning is relevant at multiple levels, as well. King and Volkov (2005) emphasize that ECB pertains to an organizational scope, not merely a focus on improving individual evaluation skills.

Taylor-Ritzler, Suárez-Balcazar, García-Iriate, Henry, and Balcazar (2013), underscoring the importance of the organization as a whole to the capacity building process, state that

ECB practices are [initially] intended to develop individual knowledge, skills, and attitudes related to evaluation among the organization’s staff. In turn, organizational learning capacity, which includes leadership, culture, systems and structures, and communication, is assumed to facilitate and/or hinder the transfer of individual learning into organizational processes and practices. (p. 192)

Summarizing the relationship between individual and organization related to mutuality, Huffman et al. (2008) assert that

. . . the ideal state is that in which the growth in evaluation capacity of the individuals has helped to expand the organizational capacity by pushing the organization forward. As the capacity of individuals grows, it expands the capacity of the organization as a whole. Likewise, as the organization expands its capacity, more support is provided the individuals to expand and grow in evaluation capacity. In the end, this process allows both the individuals and the organization to expand evaluation capacity in such a way that they are both able to effectively engage in evaluation in the hope of responding to the

demands of stakeholders for accountability information and to improve programs. (pp. 367-368)

Organizational learning and evaluation. Bickel, Millett, and Nelson (2002) assert that “building evaluation into [an organization’s] practice is a critical tool...in becoming [a] learning organization” (p. 1). They follow this up by insisting that “the internal evaluator is an integral component of the learning architecture of the organization” (Bickel et al., 2002, p. 6). Regarding an organizational learning approach to evaluation, Torres and Preskill (2001) describe the role of evaluator as incorporating that of change agent, yet recognize that not every evaluation or evaluator will be appropriate for this type of approach.

In a case study of an Edison Project school in California, Sutherland (2004) equates *evaluative practices* and *data use*, stating that “evaluators have made a compelling case for the link between evaluative practices (e.g., data use) and organizational learning” (p. 278). Cousins et al. (2006), through a review of literature, indicate a “link between ECB and organizational learning” (p. 161), as well.

Further defining evaluation’s link to organizational learning, Torres and Preskill (2001) acknowledge that learning organizations engage in more participatory processes of evaluation. They outline the stages toward using an organizational learning approach to evaluation as progressing from a more traditional approach to more developed stages where participation in evaluation, the questioning of an organization’s core functions, assumptions, and beliefs, and reflecting on practice (the participants’ own and that of the organization) are applied, routine, and “even part of [employees’] ongoing work” (Torres & Preskill, 2001, p. 392) – a concept aligned to the underpinnings of ECB. The evaluation approach progression described by Torres and Preskill (2001) is in reference

to the “extent to which [the evaluation] is participatory and designed to maximize reflection and dialogue about the evaluation findings and where appropriate, near- or long-term action among its stakeholders” (p. 390). Huffman et al. (2008), connecting to the aspects of participatory evaluation described by Torres and Preskill, make an association specifically to evaluation capacity building, stating that “ECB should be a real-world, hands-on process of learning that takes place in both a social and cultural context” (p. 360).

Suárez-Herrera, Springett, and Kagan (2009) build on the link of organization learning to participatory evaluation as well as to ECB, stating that participatory evaluation “facilitates the development of a holistic process of intentional change” (p. 321). They do, however, issue a caution against emphasizing just instrumentation in an evaluation, as well as only examining superficial changes such as uncovering and correcting inefficiencies – *how* things are done, exposing that this could actually lead to an opposition to intentional change (Suárez-Herrera et al., 2009). Instead, they advocate for deeper inquiry through the evaluation, probing into more consequential aspects such as organizational processes, goals, strategies, and assumptions – *why* things are done (Suárez-Herrera et al., 2009), eliciting the need for meaningful participation, not feigned participation.

Accountability as an Impetus for Evaluation Capacity Building

Møller (2009) explains that “in the English language, it is possible to lexically distinguish between *accountability* and *responsibility*, although accountability to some extent has replaced responsibility. In Norwegian both of these concepts refer to the same meaning, which is responsibility” (italics in the original, p. 1). McEwen (1995)

characterized accountability as “*who* is responsible for *what* and to *whom*?” (emphasis in the original, p. 3). She goes on to say that “although the questions are straightforward, the answers are not, and accountability systems are complex” (McEwen, 1995, p. 3).

Researchers agree that accountability mandates have created increased attention to evaluation and data use, raised awareness and intentionality in research-based practice, and intensified public reporting of results (Møller, 2009; Nelson & Eddy, 2008; Rodosky & Muñoz, 2009; Schmitt & Whitsett, 2008; Strahan, Cooper, & Ward, 2001; Sutherland, 2004). At the same time, these same mandates have drastically narrowed the focus of evaluation, amplified reliance on quantitative measures, created an environment of compliance, and monopolized the time of district evaluation personnel in effort to support the growing requirements (Eddy & Berry, 2008; Harmon, 2006; King & Rohmer-Hirt, 2011; Mabry, 2008; Schmitt & Whitsett, 2008).

Preskill and Boyle (2008) indicate that there is usually a catalyst to begin ECB in an organization, many times related to accountability, external pressures, and reaction to governmental policy making or planning. No Child Left Behind (NCLB) Act (2001) legislation, enacted in 2002, brought reform in the educational setting to a whole new level. Rodosky and Muñoz (2009), through a case study in Jefferson County Public Schools in Louisville, Kentucky, found that the external mandates of the NCLB Act (2001) greatly influenced their work within the district, putting systems and processes in place to help meet the legislative accountability requirements of the law.

In the nearly fifteen years since NCLB’s inception, many school systems have restructured, creating systems, processes, and sometimes internal departments to manage the requirements of state and federal accountability initiatives, along with other

evaluation and research activities to help improve educational effectiveness and ultimately build evaluation capacity within the district (King & Rohmer-Hirt, 2011; Nelson & Eddy, 2008). Russ-Eft and Preskill (2009) contend that “evaluators have increasingly been working with internal members to help them develop evaluation knowledge, skills, and beliefs” (p. 45), as well.

Despite restructuring efforts, due to responsiveness to NCLB legislation, many districts have confirmed that the responsibilities of research, evaluation, and assessment have become intertwined and overlapped, compromising the merits of evaluation purpose, design, and practice by forcing evaluation to be relegated to align to mandated accountability requirements due to limited time and resources (Eddy & Berry, 2008; Sutherland, 2004). Two of the four pillars of NCLB Act (2001), *scientific-based research* and *accountability*, have affected educational evaluation to a greater extent than the others (Eddy & Berry, 2008; Mills, 2008). To examine accountability as an effective means of ECB, three underlying premises must be considered.

Accountability and high quality evaluation. There is a common belief amongst many legislators that scientifically-based research, mandated by NCLB accountability legislation, leads to high quality program evaluations which, in turn, fosters effective programming and increased student achievement within the organization (Schmitt & Whitsett, 2008). Scientifically-based research, defined by NCLB (2001) means:

... research that involves the application of rigorous, systematic, and objective procedures to obtain reliable and valid knowledge relevant to education activities and programs and includes research that: (i) employs systematic, empirical methods that draw on observation or experiment; (ii) ...test the stated hypotheses; (iii) relies on ... reliable and valid data; (iv) is evaluated using experimental or quasi-experimental ...with a preference for random-assignment experiments. (NCLB, 2001, pp. 1964-1965)

Within the Department of Education (DOE) Federal Registry (2005), further clarification specific to evaluation stated that “evaluation methods using an experimental design are best for determining project effectiveness” (DOE, 2005, p. 3586). After explicating the acceptable methods and the acceptable conditions, the DOE (2005) lists capability of demonstrating causal relationships as the bottom line for method selection.

Assumptions embedded in this premise are that scientifically-based research exemplifies high quality evaluation and that evaluators must possess the skills necessary to conduct valid scientifically-based research, implying in this context, the use of rigorous quantitative methodology. Substantiating these assumptions, Schmitt and Whitsett (2008), drawing on their experience from a case study of the Austin (Texas) Public School District, assert that evaluators must gain the skills and knowledge to conduct rigorous studies employing experimental or quasi-experimental designs. They go on to say, for example, that evaluators “must also understand and use sophisticated analytical techniques, such as multilevel modeling ... that can remove variance in outcomes due to preexisting group differences” inherent in school programs (Schmitt & Whitsett, 2008, p. 54). Eddy and Berry (2008) reiterate and expand this truth in light of NCLB, positing that “district evaluators require a necessary level of expertise both to conduct and to be consumers of scientific-based research....[and] a similar level of expertise [needs to] extend to teachers” (p. 97). Regardless of NCLB preferences and evaluators’ commitment and skill, Schmitt and Whitsett (2008), understanding the realities of public education, point out the difficulties of conducting this type of study in an educational setting, due, in part, to the mobility of students and teachers within and

between systems, causing threats to the validity of evaluations because of attrition and contamination.

Accountability for increased evaluation. In addition to prompting high quality evaluation defined as experimental or quasi-experimental in design, another assertion related to accountability is that increased external accountability, as imposed by NCLB legislation, will increase evaluation in educational agencies, which, in turn, could potentially build evaluation capacity. For example, the No Child Left Behind (NCLB) Act (2001) dedicates seventeen sections to evaluation, listing such items as the amount or percentage of budgets that can be dedicated to evaluation; language related to what should be evaluated such as policies, programs and state/local educational agencies themselves; how evaluations should be conducted methodologically; and to whom evaluation findings should be disseminated.

Providing an overview of numerous case studies and examples, Eddy and Berry (2008) highlight the strong connection between accountability and evaluation in their findings. Rodosky and Muñoz (2009) corroborate that evaluation is a significant part of the accountability focus within their setting in Kentucky, as well. Møller (2009) also substantiates the idea of increased evaluation as a product of increased accountability, however, cautions about the mandates instigating compliance instead of capacity building.

Furthering the advice on compliance, Harmon (2006) confirms that organizations are most effective when they move past mere compliance with external mandates to become learning organizations with evaluation capacity. Snibbe (2006) adds to the admonition, urging organizations to turn “evaluation into an opportunity for learning,

rather than an opportunity for judging” (p. 45). Nelson and Eddy (2008) take this concept further, suggesting that schools should transition from compliance to capacity building for increased success.

Part of the evaluation process outlined in the NCLB (2001) legislation is the dissemination of evaluation findings. The Department of Accountability, Research, and Planning in Jefferson County School District is organized around outcomes, bearing the responsibility for creating multiple reports to communicate these results to various stakeholders (Rodosky & Muñoz, 2009). Møller (2009), writing from the perspective of accountability in Norway, shares similar experiences related to communication, stating that “schools are increasingly perceived as the unit of measurement, clearly implying new expectations of public reporting” (p. 37). In England, data from tests “serve as very public and powerful labeling tools” (Gregory & Clarke, 2003, p. 69). McEwen (1995), speaking of Canadian accountability, maintains that evaluation and evaluation capacity building programs are examples of ways to remain focused on the performance and academic success of students, which is a primary function of public school districts.

Data use and evaluation capacity. There is no question that data use and evaluation are connected, but a more explicit conception held by some is that using and understanding data can be equated to conducting evaluation; therefore, increased reliance on data, including and beyond data that is established through NCLB, is believed to lead to increased evaluation capacity. Over and above NCLB, evaluators are committed to multiple measures of performance to inform decision making (Nelson & Eddy, 2008; Strahan et al., 2001). Schmitt and Whitsett (2008) extend the argument, stating that program evaluators must “design and implement evaluation projects to monitor various

programs throughout the school year, but they must also collect, analyze, interpret, and describe a multitude of additional data” (p. 48), claiming that in their experience NCLB has expanded, yet focused, the work of evaluators within a school district.

Data use following NCLB, ranging from perceptual data to achievement data, for example, has permeated the educational system, has impacted district and school level improvement planning, prompted comprehensive assessment systems, and influenced decision making at all levels of the organization (King & Rohmer-Hirt, 2011; Nelson & Eddy, 2008). Based on case study research, Nelson and Eddy (2008) contend that as a requirement of NCLB, teachers must not only be highly qualified, but must also be effective, thus requiring the need for increased training in data use and understanding. They go on to assert that steps can be taken to increase data-informed decision making through the analysis of student work and professional practice. They conclude that creating the culture of data reliance is a long process and that not all data are equally valued by all stakeholders (Nelson & Eddy, 2008).

Creating a culture of data use and inquiry can be accomplished through evaluation (Snibbe, 2006). Toward ECB, Nelson and Eddy (2008) state that the “current system requires that educators in all capacities participate in evaluative activities at the school site more than they have in the past” (p. 44), adding that these efforts have grown to extend far beyond external accountability and disaggregated assessment data mandated by NCLB. Aligned to moving beyond NCLB-type mandates toward organizational improvement and sustainability, Sutherland (2004) explains further that:

. . . the synergistic relationship between state, district, and school levels combine to set the stage, in this case for mandated external requirements that schools must use data in promoting a culture of continuous improvement. . . Extrinsic motivators and intrinsic

motivation...are necessary for such accountability demands to become synonymous with organizational improvement...It is only when accountability for improvement becomes synonymous with organizational improvement that a culture of data use for continuous improvement will find intrinsically derived sustainability. (p. 278, 279)

Strengths and weaknesses of accountability for evaluation capacity building.

The strengths of accountability as an approach to ECB include an increased focus on data accessibility, use, and understanding throughout educational agencies (Nelson & Eddy, 2008). In addition, opportunities for increased evaluation expertise and research on evaluation have emerged and are acknowledged as fundamental (Eddy & Berry, 2008). Eddy and Berry (2008) recognize that the weaknesses associated with accountability as an approach to ECB center on practicality and resource limitations, directly violating the *feasibility* standards from Joint Committee's Program Evaluation Standards in the sense that a primary condition for feasible evaluations, as defined, is possessing and dedicating adequate resources for conducting evaluation (Yarbrough et al., 2011). The feasibility of conducting the necessary research and evaluation to support ECB in the current educational environment given the unrealistic mandates of time and resources devoted to accountability is of great question (Eddy & Berry, 2008). Furthermore, the reliance on single-source test score data instead of multiple measures as the filter for prioritization of student needs is problematic (Eddy & Berry, 2008). In consideration of resource limitations, the scope spans far beyond finances, including time, personnel, and expertise as under- or un-available commodities (Eddy & Berry, 2008).

Process Use as a Means to Evaluation Capacity Building

The purpose and application of evaluation have been debated for years, with many evaluators resting on the idea that use is an essential component (Amo & Cousins,

2007; Balthasar, 2008; Forss, Rebien, & Carlsson, 2002; Patton, 2008; Preskill & Caracelli, 1997; Preskill & Torres, 2000; Shula & Cousins, 1997). The conclusion that use is important in evaluation is validated through the Joint Committee on Educational Evaluation Standards for Educational Evaluation's decision to list utility as the first program evaluation standard category (Yarbrough et al., 2011). Although evaluation use has been of continuing discussion, Patton (2008) added common language to the conversation of evaluation use by distinguishing use of findings from use of process, contributing terminology and definition (Amo & Cousins, 2007; Harnar & Preskill, 2007).

One theoretical approach to ECB is through conducting evaluation and more specifically, using the process as a teaching tool, known as *process use* (Patton, 2008).

Patton (2008) defines process use as:

. . . individual changes in thinking, attitudes, and behavior, and program or organizational changes in procedures and culture that occur among those involved in evaluation as a result of the learning that occurs during the evaluation process. Evidence of process use is represented by the following kind of statement after an evaluation: "The impact on our program came not just from the findings but from going through the thinking process that the evaluation required." (quotation marks in original, p. 155)

In a study conducted by Forss et al. (2002), participants expressed a sense of value in going through the process of the evaluation despite unsurprising findings, which illustrates the type of statement found in Patton's definition.

Learning by doing, which is the essence of process use, has "attracted considerable attention" (Shulha & Cousins, 1997, p. 199). In a survey taken by members of the Evaluation Use Topical Interest Group (TIG) of the American Evaluation Association (AEA), 47% of all respondents felt that evaluation enhanced individual

learning, 48% felt that group learning was enhanced through evaluation, and 33% stated that evaluation develops systematic inquiry skills (Preskill & Caracelli, 1997). As early as the turn of the 21st century, Preskill and Torres (2000) wrote that evaluation for the sake of gathering information was no longer sufficient without a learning component embedded. Cousins et al. (2006) concurred, contending that it is by actually engaging in evaluation that participants come to understand the benefits. Indeed, related to the nature of adult learning, in regards to instructional strategies that increase retention of learning new knowledge or skills, engaging in or doing the activity produced a 75% retention rate, exceeded only by actually teaching the concept or skill to others (Silberman, 2006). Toward making a lasting difference, Patton (2007) asserts that “all the hard work of facilitating an evaluation process may yield more enduring outcomes ... than only findings (as important as they may be), for their relevance diminishes rapidly” (p. 111). Patton’s declaration creates a natural connection and progression to ECB.

Regarding process use as a means to ECB, King (2007) proclaimed that “process use and ECB may well be a marriage made in heaven” (p. 46). In a survey taken by members of the AEA questioning what process use looks like, 18% of the 481 respondents described process use as evaluation that is part of the daily routine; learning and change that occurs in individuals, groups and organizations; or specifically named ECB (Harnar & Preskill, 2007). Suggesting further alignment, Labin et al. (2012) published findings from a broad-based research synthesis of ECB literature indicating that 67% of the 61 cases reviewed mentioned use of involvement in an evaluation as a strategy of ECB. Looking at process use from an organizational perspective, Cousins et al. (2004) stated that:

[T]hrough doing evaluation and developing the capacity to do it, organizations become more adroit in constructing shared representations of knowledge and structures, predisposed to generate new knowledge, inclined to capture and interpret external information, and apt to question basic assumptions about the organization, its goals, and strategies for achieving them. Such consequences, in the parlance of evaluation utilization, represent group or organization level “process use.” (p. 101)

Process use as a means to ECB is undergirded by the following three premises: (a) building evaluation capacity through process use requires a long-term commitment; (b) process use to build evaluation capacity has to be an intentional endeavor to be effective; and (c) process use requires participants to be engaged in the evaluation process.

Building capacity through process use is a long-term commitment. Building evaluation capacity through process use implies that the evaluator is present and connected to the organization over a long period of time to allow for learning the evaluation process to occur (Preskill & Torres, 2000; Stockdill et al., 2002). ECB requires evaluators to engage long-term in the environment with which they are involved, developing relationships with participants, and creating experiential learning opportunities related to evaluation (King, 2002). Stockdill et al. (2002) assert that “sustaining ECB is a project that is never complete” (p. 8), proposing that the process of ECB requires developing an internal evaluation unit. The never-ending process of ECB is reiterated by Baizerman et al. (2002b). In addition, Baizerman et al. (2002b) stress the importance of context, emphasizing that “ECB is highly context-dependent; formed by practitioners’ judgment and based in site realities” (p. 104), a statement that further accentuates support for internal evaluators for this work.

In a case study of a large suburban school district in Minnesota, King (2002) describes her work as an internal evaluator, affirming that process use, conducting

evaluation to teach evaluation, can be used as a means of ECB. Taut (2007) describes ECB as an iterative cycle of generating and testing new evaluation approaches, requiring support and follow-up by the evaluator to sustain high-quality application of evaluation. In an action research project in a large international development organization, Taut (2007) elucidates connections to internal evaluation through her description of conducting research as an insider, allowing her to adjust to the needs and realities of the people in this setting, as well as adapting to the changes in context, implicating a long-term relationship with the organization.

ECB technically could be approached by hiring outside evaluators to consult with an organization to conduct quality evaluations, employing process use to assist in this effort. There are distinct advantages of this model, such as perceived objectivity, prospective willingness to disclose contentious findings or criticize the organization, and a potentially broader range of evaluator skills and experiences (Clifford & Sherman, 1983; Conley-Tyler, 2005; Love, 1991; Nevo, 1994; Sonnichsen, 2002). Many organizations, however, lack the resources to hire external evaluators, especially for the length of time needed to build adequate capacity. In addition, organizations find that there is a steeper learning curve related to culture and climate when the evaluator is external and discover that the outside evaluator may not obtain valid data if participants are reluctant to disclose authentic or unfavorable information (Conley-Tyler, 2005; Forss, Kruse, Taut, & Tenden, 2006; King, 2002; King & Rohmer-Hirt, 2011; Nevo, 1994). As described by McDonald, Rogers, and Kefford (2003) upon the conclusion of a five-year case study of ECB in a government agency, “it is clear that while external evaluators might appropriately conduct particular evaluations, and be a resource, the building of an

organization's evaluation capacity must be managed from within the organization" (p. 11).

ECB through internal evaluation depicts several strengths, such as insider knowledge, availability, opportunities for influence, and an organizational investment (Clifford & Sherman, 1983; Conley-Tyler, 2005; Love, 1991; Sonnichsen, 2002). Preskill and Torres (2000) point out that an advantage of an internal evaluator in facilitating learning is the ability to connect program or departmental concerns to organizational factors, as well as having greater accessibility to information. In addition, a survey taken by members of AEA's Evaluation Use TIG found that internal evaluators were more likely than external evaluators to view the purpose of evaluation as a facilitation of organizational learning (Preskill & Caracelli, 1997). Through statistical analysis of a broad-based study of all 300 or so governmental evaluations in Switzerland between 1999 and 2002, Balthasar (2008) found a negative relationship between distance connecting the evaluator and evaluatees in regard to process use. In other words, the greater the distance in the relationship of the evaluator and evaluatees, the less effective process use efforts were perceived (Balthasar, 2008). From an internal vantage point, evaluators have a greater likelihood of forming long-term relationships, thus fostering process use (Preskill & Torres, 2000).

Building capacity through process use takes intentional planning and action.

King (2002) contends that the evaluator must "purposely engage people in the ways that would teach them about the evaluation process" (p. 69). In a later article, King (2007) defines ECB as the "purposeful means of building an organization's capacity to conduct and use evaluations in the long run" (p. 45). 'Purposeful' in both contexts suggests

intentionality. Aligning to the definition of ECB, Baizerman et al. (2002b) argue that “practitioners take on bifocal or dual vision, attending to both evaluation studies and their use and to organizational structure, culture, and process, which makes these studies and their use possible” (p. 104), emphasizing that ECB is intentional work. Describing lessons learned from a study on process use, Preskill, Zuckerman, and Matthews (2003) affirmed that “perhaps [the] most important finding, ‘though terribly obvious in hindsight, is that process use should be *intentional*’” (italics in the original, p. 438). Forss et al. (2006) similarly state that “capacity building does not occur automatically [either]” (p. 139). Responding to a survey about process use, 39% of AEA members, who also referenced engaging stakeholders in evaluations within the survey, concur, describing process use as “intentionally employing various learning processes” (Harnar & Preskill, 2007, p. 33).

Preskill and Boyle (2008) infer that intentionality on the part of the evaluator leads to learning evaluation; and programs become more effective when encompassed in evaluative thinking. This logical progression is explicit in their Multidisciplinary Model of ECB (Preskill & Boyle, 2008). Preskill (2008) more deeply describes this attention to learning, saying that a learning perspective within an evaluation context means:

. . . being intentional and purposeful about what it is we want people to learn from and about evaluation processes and findings and situating our practice within a learning paradigm. By default then, we do not just assume that engaging people in evaluation will lead to the kinds of learning outcomes we desire. Instead, we must consciously guide stakeholders and others along a more direct path that leads to their learning from and about evaluation theory and practice. (p. 130)

Learning during participation in evaluations can occur unintentionally, or incidentally, but it is not as effective, and therefore does not easily support ECB.

Incidental learning, as opposed to intentional learning, did not foster organizational growth or competitiveness in the 41 Manitoba, Canada schools represented by survey respondents (Cousins et al., 2006). Intentional learning, in contrast, was found to have a significant impact on extent, direction, and delivery of organizational learning strategies (Cousins et al., 2006).

Patton (2007) advises, however, that intentionality for the evaluator does not necessarily constitute intentionality for the participant. Findings in the study conducted by Preskill et al. (2003) depict the value of clarifying the intention of learning from the evaluation to the participants, stating that if participants were asked to “reflect on their learning throughout the evaluation process, it would likely increase their attention to learning and potentially increase the amount of learning they would derive from the evaluation experience” (p. 440).

Building capacity through process use requires stakeholder participation.

ECB most effectively requires a commitment to a collaborative or participant-oriented approach to evaluation, which for this paper, will be used interchangeably. Through a broad-based research synthesis, Labin et al. (2012) found that of the cases that reported theory-guided ECB strategies, two-thirds explicitly mentioned a type of collaborative evaluation process. Suárez-Herrera et al. (2009) declare that “it is well known that one of the features supposed to enhance process use of evaluation is the involvement of stakeholders in many aspects of the evaluative process” (p. 326). Robinson and Cousins (2004) aver that through substantial and long-term participation in evaluation, process use is likely to occur. They go on to explain that process use is demonstrated through acquired and developed knowledge and skills aligned to systematic and evaluative

thinking (Robinson & Cousins, 2004). When explicating participatory evaluation, broadly defined as “collaboration between evaluators and non-evaluators” (p. 88) in the evaluation process, Cousins and Whitmore (1998) observe that “utilization is often associated at least as much with the process of doing the evaluation as with the findings themselves” (p. 89). In a later study on ECB, Cousins et al. (2004) confirm that “by virtue of their proximity to the evaluation, stakeholders may develop in ways that are quite independent of the findings or substantive knowledge emerging from the inquiry”, referring to process use aspects of evaluations (p. 106). Zapico-Goñi (2007) affirmed this outlook, declaring that participatory evaluation could be an effective means of instigating learning and change within stakeholders.

Studies have found a link between perceptions of participation in evaluation and participant learning. Findings from a survey of the Evaluation Use TIG of the AEA showed that 72% of respondents agreed that involving multiple stakeholders in evaluations increases process use (Preskill & Caracelli, 1997). Brandon and Higa (2004), using a pre- and post-project questionnaire model, found statistically significant differences in stakeholders’ perceptions of their evaluation capabilities, as well as their confidence in themselves as evaluators, after participating in evaluation activities. Results from a later survey of the AEA membership showed 39% of respondents related stakeholder involvement in evaluations with process use (Harnar & Preskill, 2007). Suárez-Herrera et al. (2009) uphold a connection between a participant-oriented approach to evaluation and ECB, stating that “as every participant becomes an evaluator in participatory evaluation practice, we could also consider these roles as a set of functions,

skills, and abilities that the stakeholder must develop in order to perform appropriate evaluations” (p. 329).

Case study research also substantiates and strengthens evidence of a connection between participation in evaluations and participant learning when internal evaluators are present. Subsequent to conducting a longitudinal case study, Robinson and Cousins (2004) posited that internal participatory evaluation has the potential to lead to “deep-rooted organizational effects” (p. 18). García-Iriarte et al. (2011), in a case study examining process use as a means to ECB through the work of the internal evaluator actively engaging participants in the evaluation process, found that staff experienced increased knowledge and skills related to evaluation, as well as committed to using the data from the evaluation. Experiencing similar results in an internal evaluator role, King (2002), upon engaging stakeholders in every aspect of the evaluation process, professed that “everyone became an evaluator to some extent, working to make sense of his or her own practice” (p. 76).

Strengths and weaknesses of process use for evaluation capacity building.

The weaknesses of process use for evaluation capacity building include the need to engage in long-term commitments, work intentionally and purposefully to plan learning opportunities, and provide opportunities for participants to engage in the process. All of these things require more time and money to conduct evaluations (Cousins et al., 2006; King, 2002). On the contrary, the strengths of process use for ECB are the same. By engaging long-term in an organization, relationships are formed over time, which increases accessibility and understanding of culture and context on the part of the evaluator (King & Rohmer-Hirt, 2011). Also, intentionality increases the likelihood of

success (Baizerman et al., 2002b; King, 2007). Finally, by committing to participation by stakeholders in all aspects of the evaluation process, there is a greater possibility that they will learn to conduct high quality evaluation independently in the future (King, 2002; King, 2007; Patton, 2008). Fetterman (2003) highlights process use as a means of ECB this way:

Process use becomes a tool with which program staff members and participants build capacity. They internalize the logic of evaluation by conducting evaluation on a daily basis. As they use evaluation it becomes a part of the normal planning and management of [their work]. (p. 49)

Internal Evaluation as a Model for Evaluation Capacity Building

With escalating accountability and declining financial resources plaguing public education, the need for increased evaluation is eminent. The broadening divide between performance levels of varying student groups, in the midst of growing diversity in our school systems, is further amplifying the necessity to determine program effectiveness. Related to diminishing funds and greater scrutiny of instructional practices, educational organizations must improve efficiency, again, reinforcing the need for evaluation yet adding the component of frugality. Increasingly, "the benefits of the evaluation process are being recognized as useful to organizations as they confront the dilemma of performance demands, information overload, and reduced resources" (Sonnichsen, 2000, p. 33).

In an effort to address the above mentioned conditions, internal evaluation is a strengthening emergent option, developing into a more acceptable viability in many settings. Years ago Love (1983) argued that "in times of economic turbulence, the internal evaluation process is essential for survival, because it provides the information

crucial for program improvement, accountability, and planned change under adverse circumstances” (p. 5). As Sonnichsen (2000) states it, “[T]he demand for accountability and requirement for high performance in organizations, combined with the increased understanding of the value of internal evaluation as a useful strategy to enhance performance, will have major, positive implications for the practice of internal evaluation” (p. 48).

As further support, Sonnichsen (2000) provides this insight:

Insider environments and the potential allegiances formed by these conditions require new models of evaluation, specifically crafted to function in the insider climate; models that maintain the robust dimensions of traditional science, yet are adapted for use within organizations to support the demands of dynamic and rapidly evolving circumstances. (p. 23)

Plainly stated, “ECB is far easier for an internal ECB practitioner” (Baizerman et al., 2002b, p. 102). In addition, having a dedicated staff member within an organization to design and implement ECB activities and secure resources to support these activities is critical for effectiveness of the process (Baizerman et al., 2002a).

Referring to educational settings, Stufflebeam (1997) explained that:

External evaluation services, while important, [can]not fully serve educational organizations because there [are] far too few evaluation centers, companies and consultants to address the evaluation needs of about 16,000 school districts, 50 state education departments, and hundreds of other educational organizations. Organizations [need] an ongoing process of internal evaluation to help the staff to constantly learn from experiences and improve practices and regularly to report accomplishments to sponsors and other external audiences. (p. 3)

Sonnichsen (2000) firmly asserts that “the ultimate objective is to build evaluation capacity in the organization to an acceptance level where evaluation is perceived as an indispensable component in the structural, administrative, and operational configuration

of the organization. Institutionalization is accomplished when evaluation activity is legitimized as a meaningful pursuit and the results routinely used in the decision-making process" (p.18).

Evaluation and internal evaluation. In the North American Encarta English Dictionary, the word *evaluation* is defined as “the assessment of value: the act of considering or examining something in order to judge its value, quality, importance, extent, or condition” (July, 2013). The Merriam-Webster Dictionary puts it this way: “to determine the significance, worth, or condition of [something] usually by careful appraisal and study” (July, 2013). As human beings, we are continuously evaluating our surroundings and the events that occur around us. Worthen and Sanders (1987), from an academic perspective, defined evaluation as “the determination of a thing’s value” (p. 22). They go on to provide a greater explanation of evaluation as it relates to education.

In education, [evaluation] is the formal determination of the quality, effectiveness, or value of a program, product, project, process, objective, or curriculum. Evaluation uses inquiry and judgment methods, including: (1) determining standards for judging quality and deciding whether those standards should be relative or absolute; (2) collecting relevant information; and (3) applying the standards to determine quality. Evaluation can apply to either current or proposed enterprises. (Worthen & Sanders, 1987, pp. 22 – 23)

Evaluation is a logical informant in the arena of accountability, yet, for evaluation to be used as an accountability tool, four components must exist: (a) the evaluation must be conducted, (b) the evaluation must be rigorous throughout its process, (c) the results must be shared with staff to enable reflection and necessary changes, and (d) the results must be shared with other stakeholders, including the public, to allow for judgments and decisions to be made in this broader realm (Hoefler, 2000).

Educational accountability, according to Nevo (2001), can be “traced back to the long tradition of controlling schools by means of external evaluation” (p. 96). Ideally, even beyond accountability, as Mathison (1991) puts it, “program evaluation is [or should be] part of the general information-process activities employed to monitor and improve the functioning of the organization” (p. 162). Evaluation can provide development and outcome information as well as proactive advantages. As Sonnichsen (2000) puts it, "viewing organizational activities through the analytic lens of evaluation is a useful approach to prevent complacency and to detect and solve problems before they become unmanageable" (p. 48).

The combination of continuous improvement and the mandates of external accountability, in an environment of declining resources, provides a strong basis for internal evaluation. Expanding this notion, Love (1991) states that “developing internal evaluation resources is an investment... As a corporate resource the internal evaluator can assist in obtaining the information required by senior managers, funding bodies or high echelons of government for their planning and accountability purposes, as well as meet the information needs of the organization itself” (p. 5).

Internal evaluation is characterized by evaluation conducted by employees inside the organization. Sonnichsen (2000) explains it this way: "Internal evaluators are employees of their organizations and have career commitments to the organization, continuous contact with program operations and personnel, and a greater understanding and appreciation of the organizational culture and personalities. Their vantage point in the organization aids in developing an institutional memory and in analyzing the dynamics surrounding issues" (p. 30). In contrast, Sonnichsen (2000) goes on to say that

"external evaluators are commonly defined as contractors from outside the organization who conduct specific ad hoc evaluations specified in their contracts" (p. 30).

On a superficial level, internal evaluation seems similar to other forms of evaluation, but as Love (1998) points out, there are four significant differences from other applied research activities. First, the relationships between the evaluator and the other stakeholders within the organization are considerably different, as are the goals, sponsors, and audiences of the evaluation. In the case of internal evaluation, the organization itself is the evaluation sponsor and the audiences of the evaluation results are organizational members at varying levels. Second, the focus on organizational development and planned change, encompassing the structure of the organization, the methods used in conjunction with work performance, and the service delivery provide another difference. Third, the aspect of managing change and organizational learning as components of the process presents variation. Fourth, the concept of evaluation as an integral part of managerial decision making offers yet another difference.

Internal evaluation – characteristics of effectiveness. At its most effective, internal evaluation plays a key role in the strategic planning and decision-making processes of the organization (Clifford & Sherman, 1983; Kyriakides & Campbell, 2004; Love, 1991; Mathison, 1991; Sonnichsen, 2000). Mathison (1991) draws the connection to the work of Stufflebeam and others pertaining to the CIPP (context, input, process, and product) model, saying that it “exemplified decision-making models for internal evaluation” (p. 161). As such, internal evaluators must possess many different kinds of skills and play many roles within the organization. Clifford and Sherman (1983) characterize internal evaluation as an “interdisciplinary venture” (p. 28), requiring the

evaluator to have a wide range of skills, as well as the ability to acquire new skills quickly, needing both “technical and analytical skills, along with interpersonal and organizational skills, recognizing where and when each of these skill sets should be applied” (p. 28).

Among other things, the evaluator must be perceptive and able to read the environment and individuals (Baizerman et al., 2002a). Speaking of the criticality of skills needed in effectively facilitating ECB activities, Preskill et al. (2003) assert that “if the facilitator treats individuals with respect, provides opportunities to engage in genuine dialogue and reflection, ensures that all voices are heard (no one feels silenced or marginalized), comes across as open-minded and trustworthy, and is not overly controlling, then it is likely that the group’s learning will be greater” (p. 430).

Love (1991) illustrates that when skills for internal evaluators are at their best, they are exemplified by their performance of five key roles: supporting, diagnosing, consulting, informing, and linking. *Supporting* involves providing clarifying information about organizational mission, vision, and goals, as well as descriptive information about programs and services provided by the organization. *Diagnosing* entails anticipating, identifying, and describing areas of concern within the organization and collecting relevant data to create viable, solvent options. *Consulting* requires providing meaningful information for creative and effective problem solving along with coaching staff members to appropriately interpret and utilize the information obtained. *Informing* encompasses ongoing communication and interaction between the evaluator and organizational staff members, especially management. *Linking* implies using the information in a way that not only supports the determination of solutions, but also

anticipates implications of current decisions on the organization's long-term goals and strategic planning (Love, 1991). Clifford and Sherman (1983), in relation to the aspects contained in the role of linking, affirm that organizations demonstrate a "long-term commitment to change through the enhancement of the quality of decision making in the organization" (p. 23).

Sonnichsen (2000) enhances Love's role conception by adding:

. . . the evolution of quality information and knowledge as strategic enabling mechanisms in organizations, indispensable to product development and high-quality service delivery, creates opportunities for internal evaluators. Internal evaluation offers organizations the opportunity to incorporate into their management structure an independent, systemic review function that moves horizontally and vertically throughout the organization's structure collecting, processing, and disseminating information about program functions, policy issues, problem areas, and performance. (Sonnichsen, 2000, p. 12)

As in real estate, the three most important factors for internal evaluation are location, location, location. Location in this connotation has propinquity and structural implications. Much of the literature around successful internal evaluation asserts that the evaluator should report to the highest level of management in the organization. (Clifford & Sherman, 1983; Love, 1983; Sonnichsen, 2000). In conjunction, the internal evaluation office should be in close proximity to the top ranking officials within the organization, as well. Love (1983) elucidates that to fully have the ability to impact decision making, be linked to strategic planning and management activities, and act as advisor or consultant to senior and program management, the internal evaluator has to be able to access and make use of formal and informal channels of communication.

The internal evaluator's success is also dependent upon the perception of worth in the organization, customarily linked with the reporting and organizational structure and

inclusion in key committees and processes affecting the organization, not necessarily to serve as a decision maker in these situations, but to supply information and bring the evaluation lens clearly to the table, acting as advisor for technical and behavioral aspects as they relate to evaluation (Love, 1983; Sonnichsen, 2000). Clifford and Sherman (1983) add to the explanation by saying that to promote change, the internal evaluator needs professional and technical credibility and access to the top, supplying executives with useful information.

Strengths of internal evaluation. There are many strengths of an internal evaluation model. Internal evaluators possess insider knowledge of the culture, context, and content of the organization. In other words, "internal evaluation is about organizations: knowing their culture, learning their behavior, understanding their functions, questioning their assumptions, investigating their problems, researching their policy issues, detecting their flaws, and celebrating their accomplishments" (Sonnichsen, 2000, p. 20). Torres (1991) speaks to the benefit of understanding organizational context and goals, saying that it provides insight and understanding of organizational leaders, allowing the evaluator to design evaluations that attend to these factors, while also potentially helping the evaluator to acknowledge and work through potential stakeholder conflicts that could sabotage evaluative activity. Further, internal evaluators can check their perceptions and understanding of the organizational context more readily with other members.

Another strength of internal evaluation is the availability of the evaluator to support on-going evaluation needs and to collect the associated data. As an employee of the organization, internal evaluators are on site to better anticipate, identify, and interpret

management evaluation needs. The fact that internal evaluators are part of the organization helps to strengthen relationships and build trust between the organization members and the evaluator while lessening anxiety toward evaluation, enabling authentic data collection efforts. Availability within the organization gives the evaluator the knowledge to select evaluation strategies that are most appropriate to the questions asked and the methodologies that will be most accepted and best support the inquiry, as well (Love, 1998; Mathison, 1991). Torres (1991) further asserts that “organization members are likely to find greater meaning in findings grounded within the organizational context, therefore, more likely to take action on the findings, implementing changes in a more timely manner” (p. 192). Put differently, “the organizational context is crucial for internal evaluation since it affects both the evaluation methodology and the utilization of evaluation results” (Love, 1998, p. 149).

Opportunity for timely and on-going communication, including promoting evaluation in general and encouraging the use of evaluation findings throughout the organization, is another advantage of internal evaluation. Sonnichsen (2000) calls the promotion of evaluation *advocacy evaluation*. Chelimsky (1994) introduces the concept of an *evaluation broker*, and Winberg (1991) insists that evaluation must be *marketed*. Winberg (1991) affirms that marketing should occur for evaluations that are “planned, continuing, and completed... and reports should be ‘sold’ at all levels of the organization” (p. 118). In support of this notion, Sonnichsen (2000) maintains that “use of evaluation products is far from automatic and will depend, to a great extent, on the salesmanship of evaluators” (p. 17). Adding further explanation, Chelimsky (1994), in describing her work at GAO, speaks to the importance of

. . . maintaining continuous dialogues with [stakeholders] on progress and decisions, orienting the ongoing work toward customers' information needs, [and] producing work that stood up under scrutiny. The unit had to learn how [the organization] operated, to grasp the role that evaluation (or information in general) could play in those operations, and to develop evaluative products that took both these points into account. (p. 498)

The evaluator's investment in the organization is also a significant advantage of internal evaluation. Employment connects the evaluator to the organization, providing a basis and desire for betterment in the workplace. In other words, internal evaluators have a long-term commitment with the organization that establishes a vested interest in performance, enables participation in organizational planning, lends to credibility, and helps to build positive relationships (Love, 1998; Mathison, 1991). Furthermore, internal evaluators, benefiting from an on-going presence in the organization, can document and track evaluative changes overtime (Torres, 1991). In addition, "the continuing experience and feedback – in short, the institutional memory – of internal evaluators makes the opportunity to study, understand, and interpret multiple contexts (e.g., political, economical [sic], historical, and physical) even greater" (Torres, 1991, p. 192).

Weaknesses attributed to internal evaluation. Weaknesses attributed to internal evaluation include the evaluator's lack of specialized skills and expertise, the perceived absence of objectivity, the potentially diminished willingness to criticize, and the insufficiency of an outside lens or accountability (Conley-Tyler, 2005; Love, 1998; Scriven, 1991; Sonnichsen, 2000). As part of this appraisal, internal evaluators are often critiqued for not possessing the expertise necessary to design and conduct rigorous evaluations (Hoefler, 2000; Torres, 1991). In a study related to schools conducting self-evaluation in particular, deficiencies were attributed to all parts of the evaluation process,

including adequate forms of research, devising appropriate questions, incorporating core structures, selecting or constructing valid measurement instruments, analyzing data, and drawing legitimate inferences (Blok, Slegers, & Karsten, 2008). In fact, Sonnichsen (2000) asserts that “occasionally, the technical incompetence of evaluators will be cited as justification for rejection of evaluation results” (p. 12).

A prevalent notion that plagues internal evaluation is the issue of perceived objectivity. Objectivity is credited as an impartiality or detachment to the situation or program. An objective evaluator is thought to be unbiased in his or her approach to the object being evaluated. Love (1998) and others speak to the absurdity of the reality of objectivity, as everyone is influenced by the events and circumstances they have encountered, and these views and opinions are brought into everything we do. That said, however, the perception of objectivity is a real phenomenon. Perceived objectivity typically is more readily attributed to the external evaluators, especially in the presence of contentious or litigious situations (Conley-Tyler, 2005; Love, 1998).

Torres (1991), referencing internal evaluators, exposes the “delicate, often unclear counterpoise between (a) being identified as an evaluator and consultant, (b) maintaining some degree of external objectivity, and (c) understanding the program and its constituents enough to be contextually oriented” (p. 193). She goes on to say “of any one aspect of the program, the evaluator as consultant-mediator may be said to be ‘in it,’ but not ‘of it’” (Torres, 1991, p. 193). This declaration speaks to the evaluator’s need for autonomy, to be *in the water* and not lose the ability to see, perceive, and make inferences *about the water*. Speaking to the need to balance relationships with independence, Torres (1991) maintains that evaluators must “exercise the issue-raising

approach... judiciously. Doing so is central to maintaining credibility and effective long-term relationships with both management and constituents” (p. 195). Adding clarification, Sonnichsen (2000) points out that "internal evaluation, however, is a management tool designed to furnish unbiased information that improves decisions, not a mechanism for decision justification" (p. 21).

Clifford and Sherman (1983) discuss the potential managerial manipulation of the witting evaluator and/or the data collected in the evaluation to satisfy organizational leadership or to advance personal interests. Then again, they go on to say that external evaluators can fall into this same trap in an effort to secure future contracts. In contrast, relating to internal evaluators, they also state that this duplicity is not as common as may be thought for the following reasons (Clifford & Sherman, 1983, p. 40):

- Many times, there are many people with multiple perspectives involved
- Others would detect and identify suspicious activity or analysis of data
- Manipulation is not easy and very counterproductive from an evaluation-use perspective
- Most internal evaluation is formative or developmental in nature, therefore manipulation is illogical
- Many involved have ownership of the evaluation process and accountability for results

Evaluation Capacity Building (ECB)

Compton (2009) explicates that “ECB is a type of organizational change” (p. 66). Taylor-Ritzler et al. (2013) provide direction to the concept, explaining that “ECB efforts should focus on organizational leadership, a learning culture, and developing appropriate and adequate resources devoted to evaluation” (p. 202). They expound upon the importance of these components, stating that “even when individual staff members have the knowledge and motivation to engage in evaluation activities such as mainstreaming and use, these activities are less likely to occur if their organization does not provide the

leadership, support, resources, and necessary learning climate” (Taylor-Ritzler et al., 2013, p. 200).

Baizerman et al. (2002b) refer to a “state of affairs” (p. 104) within the organization, incorporating the concept of the necessary learning climate, as well as “all that is seemingly necessary in a particular organization for evaluation work to be done and used” (Stockdill et al., 2002, p. 8). ECB strives to reside in a supportive site culture where evaluation is commonplace and “how things should be done around here” (Baizerman et al., 2002a, p.110, 112). Preskill and Boyle (2008) advise, however, that every organization is contextually different and may have differing reasons for commissioning ECB. Understanding why an organization is seeking to build evaluation capacity is important and can help identify the strategies and approaches that are most effective in the given circumstances. They go on to explain that “the selection of strategies should at least take into account the participants’ characteristics; available organizational resources; relevant evaluation, learning, and organizational change theories; and desired learning objectives and expected outcomes” (Preskill & Boyle, 2008, p. 448).

Explicit in Preskill and Boyle’s (2008) statement is the fact that ECB should have clearly defined objectives. Huffman et al. (2008) clarify that objectives are essential at both the organizational and individual levels, contending that ECB goes beyond professional development of individuals with the intention to affect the organization long-term. Further, Baizerman et al. (2002a) maintain that it is important for ECB to have an identifiable process with activities seen as part of the larger organizational processes aligned to strategic planning, goal setting and improvement efforts. Compton (2009) adds

specifics to the discussion based on findings from a case study of his work with the American Cancer Society, declaring that “development of intra- and inter-organizational structures was an ongoing process managed for both the short run (completing discrete studies) and the longer run (developing and maintaining collaborative relationships and infrastructures to increase organizational capacity for evaluation)” (p. 63).

While building the skills of individuals within the organization and the organization itself through structures and supports, it is also important to build the demand for evaluation and its use within the organization (King & Volkov, 2005; McDonald et al., 2003; Stockdill et al., 2002). King and Volkov (2005) put it this way: “In the ECB process, there is a strong need to develop a receptive culture in which demand for and effective use of evaluation output can grow” (p. 14). Consciously building demand for evaluation and its use is what Sonnichsen (2000) refers to as evaluation *advocacy*, as described earlier. *Advocacy* entails actively and continuously working to solicit leadership support and resources devoted to ECB, along with endlessly cultivating an understanding of the importance of evaluation throughout the organization. Stockdill et al. (2002) diligently point out, however, that demand must be tied to a rational purpose and is not an end in itself.

Still, even with declarations of the necessary components of effective ECB, “little is known about what it takes to successfully sustain the ECB process” (Stockdill et al., 2002, p. 23). Logically speaking, however, Preskill and Boyle (2008) insist that if individuals are unaware or unwilling to apply their knowledge, skills and beliefs about evaluation to future and ongoing evaluation practice, or do not have the opportunity to do so, sustainability is virtually unattainable. For this reason, the ECB practitioner must be

explicit about expectations of *transfer of learning* throughout the ECB process (Preskill & Boyle, 2008).

Frameworks and models for ECB. The ECB body of work contains a few frameworks, models, and tools that have been developed through a review of the literature or through case study research to assist practitioners with building evaluation capacity. The following is an overview of some of the structures.

Baizerman et al. (2002a) created a framework of ECB consisting of three structural elements: overall process; actual practice; and occupational orientation and practitioners' roles. The three structural elements in the Baizerman et al. (2002a) framework focus through the lens of the ECB evaluator. Within their framework, they identify “core ECB themes, basic ECB concepts, relevant ECB knowledge, and basic ECB skill competencies” related to and grouped by the three structural elements (Baizerman et al., 2002a, p.114). *Core ECB themes* are the underlying principles inherent in the framework, such as “ECB is a context-dependent practice, ECB work is never complete, and every opportunity now must be understood in relation to larger ECB processes and to envisioning how they influence the longer-term opportunities” (Baizerman et al., 2002a, pp. 114-115). *Basic ECB concepts* are the conceptual underpinnings of ECB, such as “emergence, intentionality, and ongoing and sustaining” (Baizerman et al., 2002a, p. 114-115). *Relevant ECB knowledge* outlines primary information the ECB practitioner must understand, for example, “how to work collaboratively, how to work in highly complex organizations, and quality evaluation, as specified by the Joint Standards” (Baizerman et al., 2002a, pp. 114-115). *Basic ECB skill competencies* depict key abilities the ECB practitioner must possess, for instance, “teach

evaluation process and uses, work within and across organizational structures and cultures, and work reflectively and politically” (Baizerman et al., 2002a, pp. 114-115). The framework also provides a “practical checklist for site assessment” (Baizerman et al., 2002a, p. 111) of ECB indicators, aligned with the same three structural elements as the framework itself.

King and Volkov (2005) established a framework based on experiences in three different organizations – a science museum, a school district, and a community center – as well as “the theories and practices” of “many intersecting fields” (p. 12). In their framework, three major categories are identified: organizational context, structures, and resources (King & Volkov, 2005). The major categories in the King and Volkov (2005) framework focus through the lens of the organization, in contrast to the Baizerman et al. (2002a) framework.

Organizational context encompasses both internal and external components. Externally, the organization is affected by legislative “mandates and other requirements stemming from required accountability measures” (King & Volkov, 2005, p. 13). External components also encompass the location of the organization in “time and place” and an “external environment supportive of change” (King & Volkov, 2005, p. 13). King and Volkov (2005) contend that “internal organizational context is key to determining feasibility of ECB” (p. 13). Internal components include leadership support, a person or people enthusiastic about evaluation, organizational “interest in and demand for evaluation” (p. 13), an environment willing to undergo change, and culture and ability of data-driven decision making (King & Volkov, 2005).

Included in the structural category are things like an ECB plan, organizational infrastructure to support evaluation and peer learning, which includes collaboration, and an organizational culture and expectation of doing and using evaluation (King & Volkov, 2005). Other structural components are time for reflection, an effective communication system, and an identified and understood way to solicit feedback pertaining to decision making (King & Volkov, 2005).

The resource category identifies the necessity of dedicated funds, people, and technological tools to effectively conduct evaluation, teach the processes, and work toward sustainability (King & Volkov, 2005). Also included are materials and sources to provide adequate professional development, coaching, and support (King & Volkov, 2005).

Taylor-Powell and Boyd (2008) generated an evaluation capacity building framework similar to the work of King and Volkov (2005) in many ways, including the focus of the main components on the organization. Taylor-Powell and Boyd's (2008) framework outlines the main components as professional development, resources and supports, and organizational environment. These categories are a slight variation from the King and Volkov (2005) framework, however, many of the underlying concepts are very similar, just aligned differently in some cases.

The component of professional development encompasses areas such as "training, technical assistance, collaborative evaluation practices, mentoring and coaching, and communities of practice" (Taylor-Powell & Boyd, 2008, p. 58). The resources and supports component contains the areas of "evaluation and ECB expertise, evaluation materials, evaluation champions, organizational assets, financing, technology, and time"

(Taylor-Powell & Boyd, 2008, p. 58). Organizational environment includes “leadership, demand, incentives, structures, and policy and procedures” (Taylor-Powell & Boyd, 2008, p. 58).

The Preskill and Boyle (2008) Multidisciplinary Model of Evaluation Capacity Building incorporates an ECB “initiating, planning, designing, and implementing” section (p. 444) and an ECB “sustaining” section (p. 446). The first section, devoted to the ECB plan, is framed by the espoused ECB goal of development of “evaluation knowledge, skills, and attitudes” affected by “motivations, assumptions, and expectations” (Preskill & Boyle, 2008, p. 444). This section also describes ten strategies that could be incorporated into the design of the process, implemented, and evaluated for effectiveness for individuals and the organization as a whole (Preskill & Boyle, 2008). The second section identifies the “processes, practices, policies and resources” needed for sustainability (Preskill & Boyle, 2008, p. 446). This section promotes such things as an “evaluation framework, shared evaluation beliefs and commitment, and an integrated knowledge management evaluation system,” along with an evaluation plan, procedures, and dedicated resources (Preskill & Boyle, 2008, p. 445). (See Table 4 for framework comparisons.)

Table 4. Comparison of Four ECB Frameworks

Category	Description	Baizerman et al. (2002a)	King & Volkov (2005)	Taylor-Powell & Boyd (2008)	Preskill & Boyle (2008)
Structures/ Systems	Process in place and strategies /plan for ECB	X	X	X	X
	Dedicated person(s)/ oversight group	X	X		X
	Identifiable structures in place	X	X		X
	Resources dedicated/aligned (including finances, time, technology, etc.)	X	X	X	X
	ECB unit appropriately located in organization	X			
	Demand responsive	X	X	X	X
	Trained ECB practitioner	X	X	X	X
	“Outward” (external) awareness / collaborative opportunities	X	X	X	X
	Understanding of core evaluation principles	X	X		X
	Commitment to quality evaluation (as per Program Evaluation Standards)	X	X		X
	Policy integration		X	X	X
Culture/ Org. Context/ Org. Environment	ECB recognized as legitimate	X	X		
	ECB explicit throughout organization	X	X	X	X
	ECB embedded in org. – common, positive understanding	X	X		X
	Inclusive / transparent / participatory	X	X		X
	Multidisciplinary practices	X	X		X
Practices	Context-based	X	X	X	X
	Continuous learning about evaluation	X	X	X	X
	Leadership support	X	X	X	X
	Evaluation “champions”	X	X	X	
	Organizational support for change		X		X
	Data-driven/data-informed decision making	X	X		X
	Participatory evaluation approach	X	X	X	X
	Data collection & record-keeping system		X	X	X
	Peer learning structures / collaboration	X	X	X	X
	Feedback / communication mechanisms		X	X	X
	Incentives		X	X	
Evaluation of ECB efforts		X		X	
Resources	Evaluation materials		X	X	X
Professional Development	Training	X	X	X	X
	Technical assistance		X	X	X
	Mentoring / Coaching	X	X	X	X

Adapted from Baizerman et al. (2002a), King and Volkov (2005), Taylor-Powell and Boyd (2008), and Preskill and Boyle (2008).

Variations of ECB. There are three main variations or extensions of ECB prevalent in the literature, as well: (1) institutionalizing evaluation, (2) mainstreaming evaluation, and (3) managing evaluation. In some cases, the terms are used almost interchangeably. For instance, Stockdill et al. (2002), in describing the process of making ECB “*ordinary practices*” (italics in the original, p. 9) within an organization, declare that “once [ECB is] institutionalized as an ordinary ‘way we do things around here,’ we can say that evaluation and its uses have been mainstreamed” (p. 9).

Stufflebeam (1997) exemplifies institutionalizing evaluation as creating and maintaining internal evaluation units and structures to support and inform a systematic process of evaluation. He explains that this systematic process of evaluation would “inform the organization’s on-going decision process, help assure its accountability to funders and constituents, and complement and cross-check external evaluations of the organization’s contributions” (Stufflebeam, 1997, p. 4). Stufflebeam (1997) promotes a “unified” (p. 5) stance to both the “concept of evaluation” (p. 5) and the “evaluation model” (p. 14) and relates this work to his *CIPP* model (Context, Inputs, Process, Products). Some of the notional underpinnings for the *concept of evaluation* advocated by Stufflebeam (1997) as indicators of unification include the following:

- Attending to all levels and components of the organization
- Attending to customer/client experience – promoting a broad range of constituents
- Promoting continuity and flexibility
- Identifying and using context-driven approaches
- Experiencing leadership and staff support
- Decision-making impact across all levels of the organization

- Supporting improvement (formative evaluation) and accountability (summative evaluation)

The items listed in the *evaluation model* (Stufflebeam, 1997) mostly align with the components of quality evaluation as indicated by the Program Evaluation Standards (Yarbrough et al., 2011).

Mainstreaming evaluation is described by Sanders (2002) as the “process of making evaluation an integral part of an organization’s everyday operations” (p. 254). He goes on to explain that “the hallmarks of mainstreaming are an organization whose culture values evaluation, whose practices make evaluation continuous, and whose history is of ongoing use of evaluations to improve the organization’s effectiveness” (Sanders, 2002, pp. 255-256). He emphasizes that mainstreaming epitomizes evaluation as an organizational value embedded into the organizational culture instead of being marginalized (Sanders, 2002). Sanders (2002) compares the concepts of institutionalization, evaluation capacity building, and mainstreaming. He describes institutionalizing as committing to evaluation through “tangible evidence” (p. 254) and policy, capacity building as “improving evaluation quality and use” (p. 254), and mainstreaming as “moving evaluation to the forefront of organizational thinking and behavior” (Sanders, 2002, p. 254). He further explains that institutionalization “does not necessarily lead to evaluation use or to valuing evaluation throughout the organization” (Sanders, 2002, p. 254). Additionally, he points out a potential overlap between capacity building and mainstreaming, stating that ECB may, in fact, incorporate the qualities of mainstreaming but it is not a required element, implying just cause to categorize them differently (Sanders, 2002). Finally, culminating the essence of mainstreaming, Sanders

(2002) expounds that “evaluation is seen as a means for learning, changing, and becoming more effective in moving toward the vision for the organization” (p. 256), explaining that organizations that have mainstreamed evaluation characteristically use evaluation to “move toward goals (short term), mission (long term), and vision (ultimate ideal)” (p. 256).

Managing evaluation and ECB. Baizerman and Compton (2009a) expand beyond ECB to a leadership/oversight role, explicitly describing the act of *managing evaluation*. Managing evaluation, or more specifically *effective managing*, is conceptually defined as

. . . the *everyday mundane action* necessary in each *organizational context and moment* to make possible one or more evaluation studies, the work of evaluators, and the collective working of an evaluation unit for the purpose of using quality evaluation for program improvement, accountability, or evaluation capacity building, among other initiatives. (p. 13, emphasis in the original)

In another piece, Baizerman and Compton (2009b) depict managing evaluation as a form of expertise: “professional work grounded in evaluation studies, but essentially it is about how to do organizational work” (p. 80), adding that it is difficult to describe, but easier to experience or show. They partially illustrate the practice of expertise as the ability to manage across levels, i.e., above to supervisors, across to other departments within the organization, and below to subordinates (Baizerman & Compton, 2009b). Baizerman and Compton (2009b) further express the need to manage across entities and types of work.

Managing also encompasses multiple levels of oversight. Baizerman and Compton (2009b) portray a need to manage studies by “building demand, doing studies and using studies” (p. 81); managing workers by “supervising and developing workers as

workers” (p. 81) or, in other words, developing skills and knowledge; and managing units by “building demand for evaluation and for [the] unit and sustaining the unit by doing its everyday work” (p. 81).

Compton (2009) delves into the differences of managing evaluation and managing evaluation capacity building. He asserts that “ECB is a structure as well as a related, joined process, and managing them is a qualitatively different practice than managing a single study, a set of evaluations, a group of evaluators, or an evaluation unit” (Compton, 2009, p. 57). Compton (2009) further distinguishes between the two, saying a program evaluation manager supports the evaluator, whereas the ECB manager supports organizational development. Compton (2009) adds to this notion, affirming that “the ECB manager must be the agent who organizes, leads, and works to sustain development of action systems that advocate for and support evaluation” (p. 62), thus aligning his views to the research on intentionality and internal evaluation.

Conceptual Framework of ECB through Accountability, Process Use, and Internal Evaluation

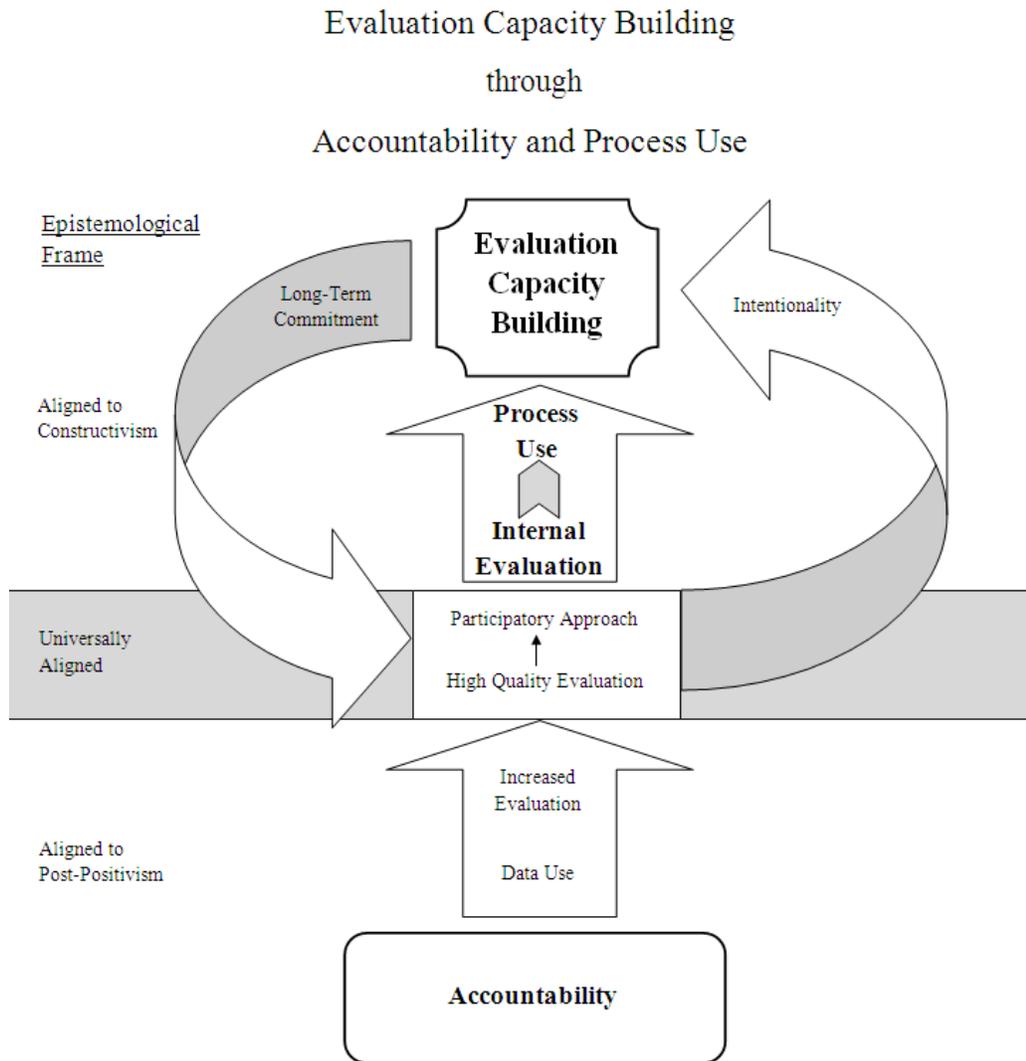


Figure 2. Conceptual framework of ECB.

The conceptual framework (Figure 2) depicts ECB through the underlying premises of accountability and process use, employing an internal evaluation model, as described above. It also incorporates the epistemological frame aligned to different aspects of the model, as well as factors found in the literature to support ECB efforts.

Accountability, as prescribed by legislation such as NCLB, is hoped to lead to data use and increased evaluation within public school districts. Accountability, epistemologically, is aligned with the post-positivist worldview. Post-positivism is characterized by objectivity, a scientific approach to research, quantitative causation, and generalizability (Chelimsky, 1997; Crotty, 1998). Accountability supplies one of the reasons *why* behind schools districts' needs to conduct high quality evaluation.

High quality evaluation is a desired outcome of accountability legislation, school districts and presumably, every evaluator, but means different things in varying contexts. For legislators, high quality evaluation refers to evaluation designed from a post-positivist perspective, incorporating the “gold standard” design of random controlled trials or, at the very least, a quasi-experimental design. (NCLB, 2001). For evaluators, high quality evaluation is measured against the program evaluation standards of utility, feasibility, propriety, accuracy and evaluation accountability (Yarbrough et al., 2011). High quality evaluation is aligned universally with all epistemological frames. There is not one type of theorist who values it over another; different researchers simply approach evaluation differently based on their methodological and epistemological beliefs.

An example of one approach to high quality evaluation is collaborative or participatory evaluation. Participatory evaluation is a necessity for process use, as if there is not stakeholder participation, stakeholder learning from the process is virtually impossible. Process use is grounded in the idea that individuals construct knowledge based on experiences, known as constructivism which is more qualitative in nature (Crotty, 1998). The constructivist perspective is characterized by “interpreting, understanding and making sense of experiences often within a social context. Learners

are not passive...but rather active participants in the construction of their own knowledge and the use of that knowledge in their work” (Preskill & Torres, 2000). Process use is most effectively facilitated through internal evaluation (Preskill & Boyle, 2008; Sonnichsen, 2000).

Through intentionality in planning and conducting high quality evaluations and a long-term commitment to the charge, evaluation capacity can continue to build, which again is most successfully accomplished through an internal evaluation model (King, 2002, 2007; Baizerman et al., 2002b). ECB, too, is established from a constructivist paradigm. Huffman et al. (2008) exemplify this, affirming that for stakeholders to develop in-depth understanding of the concepts related to evaluation, they have to engage in evaluation in a social and cultural context in order to co-construct understanding with others, thus building organizational evaluation capacity.

Summary

Evaluation capacity building is one of the goals of many evaluators (Baizerman et al., 2002a; Cousins et al., 2004; García-Iriarte et al., 2011; King & Volkov, 2005; Preskill & Boyle, 2008; Stockdill et al., 2002), but actually accomplishing it is extremely difficult. ECB at its core is organizational change, rooted in organizational learning, which has to overcome the inertial forces within an organization to grow. Even so, given the internal and external pressures of current demands, financial constraints, and overall complexities inherent in today’s educational institutions, building organizational capacity to conduct quality evaluation and think evaluatively is as important as ever.

Evaluation capacity can be approached through external accountability mandates and/or through using the evaluation process to teach evaluation. Underlying each of these

approaches are several premises. Capacity building through accountability assumes that conducting scientifically-based research leads to increased knowledge and ability to conduct high quality program evaluations. Also inherent in this approach is that increased external accountability will increase evaluation, building capacity within the organization. Likewise, is the notion that using data will increase evaluation relevance, thus increasing evaluation capacity?

Capacity building through process use postulates a number of underpinnings, as well. First is a supposition that ECB takes a long period of time, requiring an extensive commitment. Next, building capacity is an intentional process. Finally, ECB necessitates stakeholder involvement in the evaluation process for learning to occur.

Evaluation capacity building can be accomplished through either external or internal means, both with associated strengths and weaknesses. Yet, even if the financial provisions were available, external evaluation is not always an effectual or viable option. Baizerman et al. (2002b) allege that “ECB work is more craft than art or science” (p. 103) and is likely “far easier for an internal evaluator” (p. 102). Carden and Earl (2007) simply state that an internal evaluator can promote “methodological development and processes of evaluative thinking that balance the opportunity to learn and the need for accountability” (p. 63), combining both strategies into successful ECB. Huffman et al. (2008) add to this aspiration, projecting that “ideally, an organization has the capacity to engage in evaluation for both external accountability purposes and for the sake of building capacity to improve programs and better serve constituents” (p. 358). The necessity to better serve our students and improve programs to more effectively meet

their ever-increasing educational needs makes ECB and all that underscores it a nonnegotiable imperative for school districts everywhere.

Chapter 3:

Methods

Patton (2007) maintains that “the most basic wisdom in evaluation is that you begin by assessing the situation, figure out what information is needed, determine relevant questions and then select methods to answer those questions” (p. 109). In an attempt to follow Patton’s guidance, through a review of the literature, I have *assessed the situation* and *figured out what information is needed* and have posed *relevant questions* (Patton, 2007). Chapter 3 addresses Patton’s (2007) last basic step, where the *selected methods to answer those questions* are described.

The purpose of this study was to examine the process and effectiveness of building evaluation capacity within a large suburban Midwestern school district by examining the events, activities, processes and outcomes of that district over a number of years. The mounting culture of accountability coupled with the instability of funding create a less than adequate foundation on which to build programming to ensure post-secondary success for all students, and therefore schools and districts require focus and direction with ever-increasing intensity. The added pressure for educational entities to be all things to all people further exacerbates the need for information about program effectiveness. To address these concerns, schools and districts need to build evaluation capacity to provide the necessary guidance to direct programming and make decisions. This study strived to add to the empirical data on ECB practice by conducting a systematic, longitudinal, embedded, single case study (Yin, 2009), historically spanning 30 years, from 1985 – 2015, in a large suburban Midwestern school district.

The research questions that guided this study were as follows:

An overarching question: What has been the experience of one school district in building evaluation capacity?

Three sub-questions:

- 1) What are the perceptions of district leaders as to the importance of evaluation to the mission and vision of the district?
- 2) What is the evidence of evaluation capacity within the organization over time?
- 3) What contextual events expedite or delay evaluation capacity building?

The overarching question explored the historical journey of one school district in building evaluation capacity to potentially inform other educational entities with respect to their path to evaluation capacity. Sub-question one elicited district leaders' perceptions of the importance of evaluation related to the mission and vision of school districts, which served as a basis of support for conducting evaluation and building evaluation capacity within the organization. Sub-question two was designed specifically to empirically gauge educators' perspective of the magnitude of evaluation capacity throughout the organization from various leadership roles aligned to research, evaluation, and testing activities within the school district of study. Sub-question three addressed the varying contextual implications that supported or hindered evaluation capacity building, adding to the empirical literature by studying a broad scope of time to not only measure what initially occurred to affect evaluation capacity building, but also what affected the maintenance and sustainability of evaluation capacity within an educational organization. The findings and discussion that stemmed from these questions strengthen the ECB and educational change literature.

Case Selection and History

The Two-County Independent School District (ISD) was selected for this study for many reasons. It is the largest district in the state and has a positive reputation as a leader related to instructional innovations. The district has a notable presence at the state level, from former school board members serving in the state legislature to numerous district employees serving on committees for the state department of education. The district is large enough that it has a department dedicated to research, evaluation and assessment activities and is known for having a commitment to evaluation. In addition, I have a personal connection to Two-County ISD, living in the community and working for the district which affords me access to information at a greater level.

The Two-County Independent School District (ISD) was established in 1952 in a Midwestern metropolitan region after an overwhelmingly positive community vote to consolidate numerous smaller districts in the area. The school district currently serves 13 communities, educating almost 37,000 students in grades K-12, reaching its peak size to date in 2004-2005 and gradually declining since then. Two-County district also has thriving early education and 18-21 year old and adult basic education programs, substantially increasing the total number of students served by the district.

Since its creation, eight different superintendents, including an interim superintendent, have led Two-County ISD. Historically, the district has been relatively stable from a leadership standpoint, however, the last eight to ten years have seen momentous leadership turnover with over 60% of the school and central leadership new to their roles, including two superintendent transitions and numerous associate superintendent switches during that timeframe. Shifts in leadership have occurred due to

retirements, attrition, illnesses, resource allocation adjustments, and, in small part, resignations. Organizational memory has diminished greatly with these changes, as well.

In addition to leadership turnover, there have been substantial organizational structure changes with the most significant modifications occurring in the central office and support staff configurations. Budget cuts, superintendent changes, school board input, and perceived or real inefficiencies have prompted extensive and ongoing adjustments to the cabinet structure and numerous central departments over that same eight- to ten-year period.

The district currently operates five traditional high schools, six middle schools, and twenty-four elementary schools, along with two alternative high schools, one alternative middle school, and numerous specialty sites for students with unique needs, such as special education and behavioral, mental health, or chemical dependency needs.

District demographics. Demographically, Two-County ISD has undergone changes, as well. Figure 3 shows that over the past eighteen years, the district has grown from approximately 7% students of color to 25%, from 15% of students qualifying for free or reduced-priced services to approximately 35%, and from nearly 0% students qualifying for language acquisition services (Limited English Proficient – LEP) to approximately 7%. During that time, the percentage of students qualifying for special education services has remained relatively constant, between 10 – 14%, however, the significance of student needs has increased with many students exhibiting multiple diagnoses.

In some circumstances, the district has maintained or grown very gradually, such as in the percentage of students qualifying for special education services and the

percentage of students qualifying for English language acquisition services. In other cases, the district has advanced faster than the system can change, which has been the case with the increase in students of color; the staff composition has not transitioned quickly enough to be representative of the percentage of students of different races, creating an area of strife for the district.

Another demographic area of extreme growth is the percentage of students qualifying for free- or reduced-priced services. The district saw a dramatic surge, coinciding with economic downturns in the country beginning in the 2008-2009 school year, and the percentage has since remained relatively steady over the past three or four years.

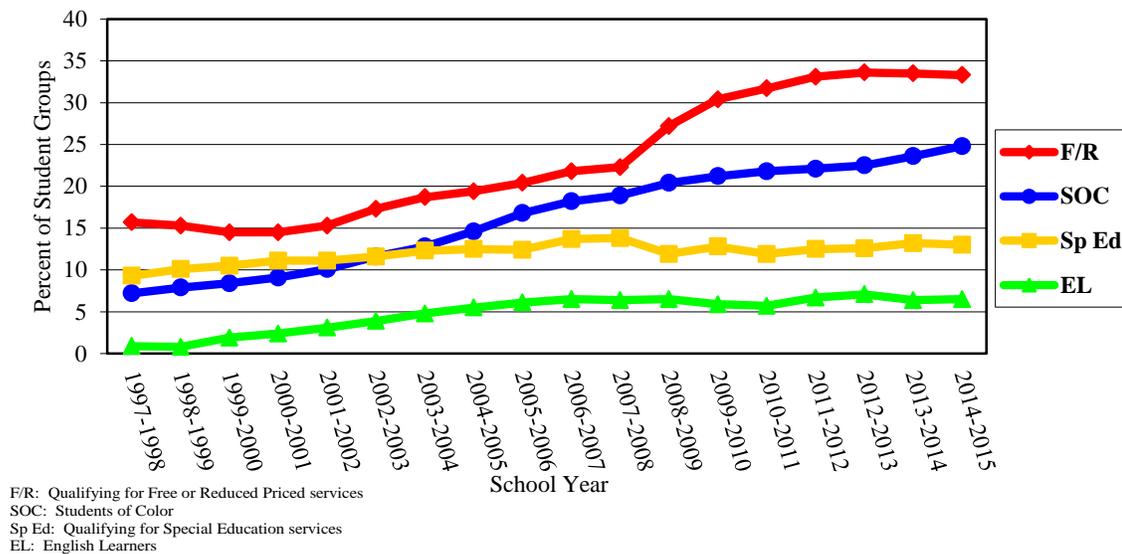


Figure 3. *Demographic changes over time in Two-County Independent School District.*

District structure. Structurally, the district is currently governed by a six-member school board and is led by a superintendent assisted by three associate superintendents and seven other cabinet members. Each associate superintendent oversees one level of schools (elementary, middle school, or high school) as well as a variety of central departments and special programs such as Curriculum & Instruction,

Special Education, Student Services, Supplemental Programs, and Alternative Educational Programs. The Research, Evaluation and Testing department was aligned to one of the associate superintendents until the 2014-2015 school year and now reports directly to the superintendent. The operations side of the district, including Early Childhood and Adult Basic Education programming through Community Education, is overseen by the remaining cabinet members, creating a relatively flat leadership structure. There have been multiple changes in organizational structure throughout district history, initiated in most part by either budgetary constraints or leadership changes, again, amplified in frequency over the past eight to ten years, as described above.

Although there is some autonomy at the school level, central office staff collaboratively develop most comprehensive programming and activities for use across the system. Two-County ISD has common curriculum, representative committee structures, and attempts to maintain consistency in practice and operation across schools and levels to the degree possible.

Two-County has traditionally been forward-thinking in regards to accountability and equal opportunities for students from classroom to classroom, as well, instituting assurance of learning assessments well before the state in which it is located. The district has also played an integral role in the development of statewide educational practices and assessments, serving as a resource to the state and providing input and guidance on numerous state committees.

Data Collection Procedures and Methods

Case study research. Many researchers call for longitudinal case studies to determine effects over time, ECB outcomes, and potential sustainability (Amo & Cousins, 2007; Baizerman et al., 2002a; Cousins et al., 2006; García-Iriarte et al., 2011; Shula & Cousins, 1997). More directly, Robinson and Cousins (2004) assert that longitudinal case studies are a “preferred mode of inquiry” in this domain (p. 3).

Responding to the call of researchers in the field, this study was designed as a longitudinal embedded single case study (Yin, 2009) examined from a historical perspective (Merriam, 2001). Yin (2009) posits that case study research is used to “contribute to our knowledge of individual, group, organizational, social, political and related phenomena” and “allows investigators to retain the holistic and meaningful characteristics of real-life events” (p. 4). He goes on to explain that case study research is most appropriately used with *how* or *why* questions, when control of behavioral events is not required, and when the focus is on contemporary events (Yin, 2009). Case study research was appropriate for this inquiry because the topic of investigation was current, did not occur in a controlled environment, and centered on *how* evaluation capacity is built in an educational setting and *why* it is or isn’t sustained. Yin (2009) also states that a unique strength of case study research is “its ability to deal with a full variety of evidence— documents, artifacts, interviews and observations” (p. 11), which also aligned to this study.

Merriam (2001) asserts that case study research is “a particularly suitable design if you are interested in process.” She also suggests that case study research can be described as having three distinctive traits; they “can be characterized as being

particularistic, descriptive and heuristic” (Merriam, 2001, p. 29). Merriam (2001) adds further definition, stating that *particularistic* refers to a “focus on a particular situation, event, program or phenomenon” (p. 29), *descriptive* is in reference to the “rich, thick description of the phenomenon under study” (p. 29), and *heuristic* “means that case studies illuminate the reader’s understanding of the phenomenon under study” (p. 30).

Based on these features, case study research was appropriate for this study and, specifically, a single case was examined. Yin (2009) states that a rationale for using a single case over multiple cases exists when conducting a longitudinal case; studying two or more different points in time and investigating how certain conditions changed over time. Merriam (2001) connects longitudinal case studies to historic case studies stating “the key to historic case studies... is the notion of investigating the phenomenon over a period of time (p. 35). She goes on to explain that “the researcher still presents a holistic description and analysis of a specific phenomenon (the case) but presents it from a historical perspective” (Merriam, 2001, p. 35).

Embedded case studies are studies that incorporate more than one unit of analysis (Yin, 2009), which also pertained to this research. Yin (2009) explains that certain case studies lend themselves to mixed methods research, suggesting another meaning to the term ‘embedded.’ Some case studies contain other embedded research methods. For instance, in some investigations, “embedded case studies rely on more holistic data collection strategies for studying the main case but then call upon surveys or other ... techniques to collect data about the embedded unit(s) of analysis” (Yin, 2009, p. 63).

The following is a description of methods that were used to answer the previously mentioned research questions, presented from a mixed-methods epistemological frame.

Table 5 outlines the four questions that guided the longitudinal, embedded, single case study research design, the methods that were employed to collect data, and the aligned unit of analysis for each question. Narratives of these methods are provided below.

Table 5. *Matrix of Research Questions, Methods, and Units of Analysis*

	Overarching Research Question: What has been the experience of one school district in building evaluation capacity?	Research Question #1: What are perceptions of district leaders as to the importance of evaluation to the mission and vision of the district?	Research Question #2: What is the evidence of evaluation capacity within the organization overtime?	Research Question #3: What contextual events expedite or delay evaluation capacity building?
Methods	Individual & Group Interviews, Participant-Observations, Document/Archival Records Review			
Unit of Analysis	Individual, Organizational	Individual	Organizational	Individual, Organizational

Individual interviews. Individual interviews were conducted with people identified as being associated with research, evaluation, and assessment responsibilities at some point over the timeframe of study to inform each of the research questions. Individuals were chosen based on their oversight or association with the current department charged with research, evaluation, and assessment-related processes. Fourteen people were interviewed in person or over the phone per the convenience of the interviewee. In addition, archival employment records from the timeframe of the study were reviewed to ensure that employees fitting the above description were not inadvertently overlooked; however, there were six individuals fitting the above description who were unavailable for interviews. All timeframes across the study had at least two interviewees represented and three people who also led evaluative activities that worked very closely with these individuals were available for interviews, representing perspectives of their work together.

The interview items were created, pilot tested, and reviewed through a think aloud process to increase understandability and construct validity (Dillman, Smyth, & Christian, 2009). The initial interviews lasted an average of 84 minutes, ranging in length from 42 minutes to 127 minutes with a series of 17 questions, some containing multiple components. Answers were both audio recorded using two different audio devices and recorded through notes taken by the researcher, which were enhanced through the writing device with video-capture of the notes, linked to the audio that was occurring simultaneously. Second interviews were conducted with participants either individually or in groups to provide reaction and additional input to the study's findings to increase internal validity. Merriam (2001) refers to this process as "member checks" and describes it as a means of enhancing internal validity (p. 204).

At the completion of each interview, reflections were documented related to surprises, salient thoughts, overarching themes, connections to other interview respondents or other data collected, and additional thoughts or questions that arose. Audio recordings were sent to an independent company to be transcribed. Upon return, transcripts were reviewed for accuracy and completeness, and then read in their entirety to identify connections. Responses were coded by question and transferred to a spreadsheet. A secondary spreadsheet was created with one row per respondent and one column per question for analysis purposes.

The questions related to the organizational context and district priorities during the respondent's time of employment; the person's specific role and association with research, evaluation, and testing responsibilities; his or her perception of the importance of research, evaluation, and testing activities relative to the mission and vision of the

district; the state of the district relative to data availability, accessibility, understandability and use, along with degrees of evaluation use during his or her tenure; perceptions of the capabilities within the district to use a defined range of evaluation strategies identified within the review of the literature; departmental goals and vision during their oversight; departmental staff availability and capacity to conduct evaluation; collaboration with other departments; changes that occurred to the department during their tenure; and internal, external and legislative impacts to the direction of the department charged with research, evaluation, and testing activities. The interview questionnaire included Likert-type items pertaining to the relationship of evaluation to the mission and vision of the district; the number of staff capable of using varying evaluation strategies; the type of data available to staff; the presence of a dedicated department and trained staff aligned to research, evaluation, and assessment; and demographic items for analysis purposes, but interviewees were encouraged to elaborate upon their responses to these structured options. (See Appendix A.) Consent was obtained and confidentiality maintained. (See Appendix B.)

Group interviews. Group interviews (or second individual interviews in some cases) were conducted with the same individuals who participated in the individual interviews to review the findings, determine if important information was missed, and provide participants the opportunity to add anything additional that was remembered following the initial interview or as a result of group discussions and review of the overarching study findings, the captured events, and the district history. This step was included to increase the validity and reliability of the findings (Merriam, 2001).

Participant observations. Participant observation is a “special mode of observation in which [the researcher] is not merely a passive observer. Instead, [the researcher] assumes a variety of roles within a case study situation and may actually participate in the events being studied” (Yin, 2009, p. 111). Given that the researcher in this case study is an employee of the district being studied with duties assigned to research, evaluation, and assessment, participant observations were conducted to respond to the research questions. Participant observations included meetings and events related to planning, conducting, or training others in the processes of evaluation and evaluation-related activities, as well as meetings and events to share evaluation or evaluation-related findings. Systematic notes were taken related to the events including dates and times, participants, activities, direct observations, perceived or interpreted observations, and efforts made to build evaluation capacity, along with the reception of these efforts by participants. In addition, agendas, presentation materials, notes from other participants and other documentation were also reviewed as part of the research process.

Document review. Documentation is a common and appropriate form of evidence collection for nearly every case study conducted (Yin, 2009). Documentary information is found in a variety of correspondence, agendas and minutes from meetings and gatherings, reports, other formal studies or reviews, or public information appearing in the print media. School board minutes and agendas from January, 1985 through December, 2015 were reviewed, along with district newsletters ranging from Autumn, 1984 through Fall, 2015, various web pages obtained through the computer-based search engine entitled *Wayback Machine* that facilitates archival searches of websites based on the uniform resource locator (url), which was the address of the district website, in this

case, and numerous boxes of files, reports, personal notes, and documentation about departmental activities over the years examined, totaling 1,121 entries captured in Excel.

Other documents that were reviewed in conjunction with this study included the number and type of scheduled calendar meetings/events, agendas, and minutes from evaluation planning meetings; training sessions and presentations; Power Point presentations created relating to evaluation; evaluation training materials; documents related to previous studies conducted within the district; memos, emails, and other correspondences obtained relating to research, evaluation and assessment activities; and personal notes and documentation related to evaluation collected from past and present employees. Documents were reviewed from the range of the timeframe to determine efforts that occurred as well as potential long-term effects of those efforts. Review of the documentation aided in addressing all four research questions.

Archival records. Archival records consist of files or records maintained typically by the organization (Yin, 2009). They can take the form of census or statistical data, budgets, employment records, service records, or survey data, for example (Yin, 2009). Archival records were reviewed to determine employees associated with the current department charged with activities and process related to research, evaluation, and assessment. In addition, organizational charts from the timeframe of the study were reviewed to assess changes in organizational structure over time. Also, reports related to prior evaluations and community surveys, summaries of budget cuts, organizational historic narratives, special reports, legislative required reports and community task force reports from across the time period were reviewed.

Data Analysis

Yin (2009) expounds three principles to maximize the benefits from the sources of evidence and “help deal with the problems of establishing the construct validity and reliability” (p. 114) that are many times associated with case studies. The three principles are to “use multiple sources of evidence, create a case study database, and maintain a chain of evidence” (Yin, 2009, p. 114, p. 118, p. 122). This case study incorporated all three principles, which not only assisted with establishing validity and reliability, but also aided the analysis of the data.

Internal validity from Yin’s (2009) definition did not directly apply to this case as the study was not seeking to establish causal relationships. Merriam (2001), however, talks about internal validity from the standpoint of “how research findings match reality” (p.201). She goes on to state that “one of the basic assumptions underlying qualitative research is that reality is holistic, multidimensional, and ever-changing; it is not a single, fixed, objective phenomenon waiting to be discovered, observed, and measured as in quantitative research” (p. 202), and, therefore, determining validity in the same way is unsuitable. Merriam (2001) continues, explaining that in qualitative research, since the researcher is the “primary instrument of collection and analysis...interpretations of reality are accessed directly through their observations and interviews” (p. 203). In this way, she argues that internal validity can thus be viewed as a strength of qualitative research (Merriam, 2001). Even so, Merriam (2001) offers six strategies to enhance internal validity that were employed as part of this study. They are:

- Triangulation of data – using multiple sources and/or methods to collect data
- Member checks – having participants verify interpretations
- Long-term observations – gathering data over a period of time
- Peer examination – asking colleagues to comment on the findings as they emerge

- Participatory or collaborative modes of research – involving participants
- Researcher’s biases – clarifying researcher’s assumptions, worldview and theoretical orientation. (Merriam, 2001, p. 205)

Related to external validity, the goal of case study research was to “expand and generalize theories (analytic generalization) and not to enumerate frequencies (statistical generalization)” (Yin, 2009, p. 15). The findings of this case study are potentially transferable or applicable to settings similar to the one studied, but, by definition, are not generalizable in a quantitative sense.

In terms of reliability, Merriam (2001) offers suggestions aligned with Yin’s (2009) three principles. She builds on Lincoln and Guba’s (1985) concept of *dependability* and *consistency* as opposed to results replication in explaining reliability in qualitative studies (Merriam, 2001). Merriam (2001) explicates that “the question then is not whether findings will be found again but whether the results are consistent with the data collected” (p.206), offering a few practices to increase dependability within the results. These practices include “explaining the assumptions and theory behind the study, the basis for selecting the participants, and the social context from which data were collected; triangulation of data in terms of using multiple methods of collection and analysis; and keeping an audit trail – providing an explanation of how you arrived at the results” (Merriam, 2001, pp. 206-207).

Given Yin’s (2009) first principle stated above, as well as Merriam’s (2001) first component of enhancing internal validity, and that the researcher was committed to a mixed methodology epistemological frame, the data from each method were ultimately considered and analyzed collectively to ensure that the case study’s findings were supported by the “convergence of information from different sources, not quantitative or

qualitative data alone” (Yin, 2009, p. 110) and that each research question was answered thoroughly. The strategies used in analyzing the data from this study were as follows.

In general, individual and group interview data, participant observation data, and document and archival records reviews were organized systematically, coded, and stored both electronically and in paper format. They were read and re-read, immediately following collection, throughout the data collection process, and after all data had been collected, looking for evidence of or connection to organizational change, accountability, evaluation activity, process use and/or ECB. Each was coded and reviewed multiple times to ensure consistent coding throughout the study process. The data were analyzed and re-analyzed for converging themes across all data collection methods and the coding and analysis were periodically reviewed by a doctoral advisor from a different university and discussed with an individual who recently earned a doctor of philosophy degree from this sponsoring university to increase validity and reliability of the study.

Specifically, the transcribed interviews were transferred to database software, organized with one respondent per row and one question per column to support item level analysis across participants. Specific questions were coded by certain characteristics and analyzed across time. For instance, the item inquiring about district culture, soliciting three words to describe the district at the respondent’s time of evaluation activity oversight, were coded as positive, neutral, or negative and arranged chronologically by employment dates. Also, evaluation activities and technological capabilities were chronicled over time and analyzed for changes and advancements. Overarching themes across interviews were identified, and events were acknowledged and arranged by superintendent tenure.

As noted above, a number of different documents and archival records were reviewed as part of this research study. Departmental documents, notes, files, presentations, meeting agendas and minutes, planning and training records, data collection tools, spreadsheets, and other artifacts were reviewed, logged, and arranged with a row per record. Column headers were similar to those in the interview file and included folder/document context, material timeframe/dates, concepts/content, salient concepts, surprises, overarching themes, meaningful quotations, comparisons with other data/information, and additional questions/information needed. Records were analyzed across rows for commonalities and emergent themes, as well as identification of significant activities or events that stood out in the information.

School board agendas /minutes along with corresponding artifacts from the meetings from January, 1985 through December, 2015 were reviewed. Categories documented were document source, document type, date, topic/purpose, evidence of evaluation connection, connection strength (none, weak, moderate or strong), audience/participants, and additional comments/notes. Connection strength was determined based on a rubric developed based on evaluation activity progression. Agendas/minutes were coded as strong if there was evidence of formal evaluation activities including such things as identifying criteria and standards, systematic data collection or formally-designed research/evaluation studies. Documents that contained evidence of formal/systematic data collection only (not the full evaluation process) or displayed more in-depth analysis of data, informal evaluation or simple action research were coded with moderate strength. Agendas/minutes that showed evaluative activity limited to testing, only stating criteria for decision making (with no other elements of an

evaluation process), simple data analysis, or other simplistic levels of evaluative activity were coded with weak connection. Documents that contained none of these things were coded as none. (See Appendix C.) Evaluation strength was then coded zero to four aligning to the categories of none, weak, moderate, and strong, respectively, displayed graphically in chronological order, and categorized based on corresponding superintendent tenure. Important events were coded in the spreadsheet as information was being recorded and then again when records were being reviewed.

District newsletters that were sent to all community residents were reviewed from the time period of Autumn, 1984 through Fall, 2015 following the same process as the school board agendas/minutes and added to the same spreadsheet for analysis. These records were also coded numerically by evaluation connections as described above and were displayed graphically in the same manner. Website information for the district and the state during the 30-year period was reviewed based on availability by using a search engine called *Wayback Machine*, which allows the user to access archived web pages by entering the uniform resource locator (url) of the sites of interest, which in this case were the district website and the state department of education website.

In addition, a calendar analysis was conducted of the researcher's scheduled calendar meetings and events. A three-month comparison sample (September – November, 2007) from when the researcher first assumed the responsibilities of director of the department and then a more extensive five-year time frame (February, 2011 – December, 2015) were coded and analyzed for frequency of events related to evaluation activity in the areas of testing, evaluation, research, and general data accessibility and use, herein called *evaluation-type events*. Events related to district-wide testing or testing

support were placed in the testing category. Events related to formal evaluation such as meetings with evaluation leadership teams, planning evaluation studies, discussing evaluation findings, etc. were coded as evaluation. Events related to research conducted in the district mainly aligned to research requests for employees seeking advanced degrees, colleges and universities looking for subjects/sites for research partnership, or national agencies requesting participation in research were coded in this category. All meetings related to data collection, technology tools to capture or display data such as the warehouse, data interpretation meetings related to non-testing or non-formal evaluation events, and any other general data support (analysis planning, data request discussions with the requesters to clarify data needs for upcoming decision points, etc.) These evaluation-type events were quantified and calculated as a total of all calendar events and represented as a percentage of all events, both in aggregate and by event type as listed above. The count of each evaluation-type event was also calculated as a percentage of the total number of evaluation-type events and compared over time. Graphic representation of these activities was created and compared across the study period for changes and evidence of increased evaluation activity.

A fundamental feature of case study research is the “ability to trace changes over time” (Yin, 2009, p. 145). In this case study, a timeline of chronological events was compiled as an analysis strategy. As part the timeline, the employment tenures of staff associated with research, evaluation, and assessment responsibilities (participants of individual and group interviews) and departmental changes and important events stemming from interviews, documents, and archival records were chronicled along with internal influences affecting the organization and internal processes and structural

developments. (The timeline is included in Chapter 4.) In addition, evaluation implications and education-related governmental legislation was traced throughout the same time period. Accounting for this chronology allowed for examination of the interrelatedness of these events. Results from this study are potentially transferable to other settings and contexts that resonate with the one studied.

Limitations

Case study research and the methods chosen for this case pose certain limitations. One of the greatest criticisms of case study research is a shift from initially articulated research questions to different questions altogether (Yin, 2009). Yin (2009) suggests that “one way to increase the sensitivity to such slippage is to have a set of subunits. Thus, an embedded design can serve as an important device for focusing a case study inquiry” (p. 52). This study follows Yin’s (2009) supposition by attending not only to the overall case, but also to the subunits of the tenure of each superintendent within the timeframe as well as the individual participants who were selected.

Merriam (2001) also provides additional information on the limitations of case studies, pointing out the one of the concerns is the need for both the researcher and reader to be aware of potential bias in the final product, as well as the “inherently political nature of case study evaluations” (p. 42), offering that the study may illuminate discrepancies in people’s perspectives in what they believe and what is disclosed about the studied phenomenon.

Yin (2009) asserts that “a major strength of case study data collection is the opportunity to use many different sources of evidence” (pp. 114-115). Many different sources of evidence, however, create a major limitation – managing an abundance of

data. To address the magnitude of data, systematic records were kept, data were sorted and categorized in terms of importance to the case study, and case study databases to organize the data were created and maintained (Yin, 2009), thus increasing reliability of the study.

There was another limitation related to this study that I would like to again highlight. Whereas fourteen leaders of Two-County ISD from 1985 – 2015 were interviewed, there were key three individuals who would have certainly been assets to the information pertaining to ECB and evaluative activity development in the district. The first was Mr. Spawling, superintendent from 1975 – 1990, who was unable to be reached for participation. The second was Dr. Wheaton, associate superintendent of secondary schools, who directly supervised or supported five of the departmental leads, but she had retired, moved away, and was unwilling to participate in this study. The third was Dr. Villard, associate superintendent of elementary schools, who, although she never had direct supervision of the department leader, definitely engaged in evaluative activities and led this work to a large extent within the elementary level, sometimes in conjunction with the research, evaluation, and testing department and sometimes in tandem. Dr. Villard was unable to participate due to health issues. It is certain that all of these individuals could have added detail, insight and perspective not necessarily shared by others.

This limitation was mitigated by interviewing leaders reporting directly to these individuals, who worked very closely with them, in an effort to gain perspective and context from the time periods represented these people. Furthermore, there were at least two individuals interviewed across all time frames of the study. The school board

agendas/minutes and the district community newsletter that were a part of the document review also spanned the entire time frame of the study.

Other limitations of this case study include concerns related to chosen methods.

The following table outlines a variety of potential limitations associated with the methods employed in this study, along with ways they were overcome or lessened. The limitations noted are adapted from Yin’s (2009) table of strengths and weaknesses of varying sources of evidence or methods (p. 102).

Table 6. *Method Limitations and Strategies to Address Them*

Method	Limitations	Strategies to Combat Limitations
Interviews (individual and group)	<ul style="list-style-type: none"> • Willingness and availability of interviewees • Response bias • Articulation and recall inaccuracies • Reflexivity – responding with what is believed to be desired 	<ul style="list-style-type: none"> • Researcher’s position in the organization provided more ready access and flexibility in meeting interviewees’ schedules/needs • Development of precise, well-written measures allowed interviewees to provide novel responses to the questions • Conceptualization of interviews was viewed as “verbal reports” (Yin, 2009, p. 108), not absolute truths • Comparison of information across other interviews and data sources were made • Items were developed, piloted and administered following Tailored Design Method (Dillman et al., 2009)
Participant Observations	<ul style="list-style-type: none"> • Selectivity – difficulty of broad coverage without a team • Reflexivity – affecting the procedure of events due to observations • Bias due to participant observer’s potential manipulation of events • Participant observer support of organization 	<ul style="list-style-type: none"> • Awareness of the possibility of not observing full scope of events was maintained • Input of other participants to validate observations was sought • Researcher actively sought out researcher subjectivity for awareness and disclosure (Peshkin, 1988) • Researcher training and experience were intentionally employed

Method	Limitations	Strategies to Combat Limitations
Document / Archival Records Review	<ul style="list-style-type: none"> • Availability & accessibility – document/record retrievability • Author reporting bias – may contain “unmitigated truths” (Yin, 2009, p. 105) • Interpretation bias – recorded for purpose other than case study 	<ul style="list-style-type: none"> • Knowledge and relationship with personnel in organization provided greater likelihood of accessing documents and records by knowing whom to solicit • Long-term employment in the district provided historical knowledge, as well as direct and secondary access to documents and records • Researcher used multiple sources of data to corroborate findings • Researcher worked to determine purpose for document/record to help avoid misinterpretation

Adapted from Yin (2009), p. 102.

By attending to the identified limitations, the potential effects were mitigated. This case study had unique circumstances, however, so as a researcher, I needed to be diligent to be aware of and manage subjectivity, aligning with Merriam’s (2001) espoused limitation described above.

Positionality/Managing Subjectivity

Falling under the Program Evaluation Standards (Yarbrough et al., 2011) of Transparency and Disclosure, I must let the reader know that I am a current employee of the Two-County School District. I have been under contract in the district since January, 1995 and have held a multitude of positions. I began my career in the district as a math teacher, department leader, and district-wide curriculum writer. Along the way, I became an instructional coach, a school-level administrator, an analyst at both the school and district levels, and eventually accepted the directorship of research, evaluation, and testing, a role that I currently hold. As part of my duties in this position, I oversee student assessment data, the development of the data warehouse and other data structures, district-wide student testing systems, program evaluation, and general data collection,

analysis, and management. Along with my service to the district, my doctoral faculty advisor also has ties to the Two-County School District, having served in the departmental oversight role just fewer than ten years prior to my tenure in the department, as part of a university/public school district collaborative exchange for approximately a two-year period.

Having such a close relationship and intricate knowledge of the organization within which my research was conducted could have posed a significant concern if not managed appropriately. Different epistemological frames address this differently. Phenomenology from an interpretivist frame, for instance, talks about intentionally or consciously ‘bracketing’ the “researcher’s own knowledge and suppositions... so as not to taint the data” (Crotty, 1998, p. 83). In doing so, the researcher tries, to the best of his or her ability, to see things and experience things in a new way, uninhibited by previous beliefs or experiences. From a constructionist view, meaning is constructed or ascribed an object or situation based on our experiences, and there are many ways to construct and assign meaning with no ‘true’ interpretation (Crotty, 1998). Positivists, on the contrary, contend that there is only one truth which is ‘given’ and must be discovered by the researcher. A hypothesis is posited and then tested through direct, scientific observation in a controlled setting (Crotty, 1998). The approach I used to address my subjectivity was most closely aligned to the phenomenological approach and is described below.

To manage my subjectivity, I employed an approach outlined by Peshkin (1988). Peshkin (1988) begins by stating that “when [researchers’] subjectivity remains unconscious, they insinuate rather than knowingly clarify their personal stakes” (p. 17). He goes on to assert that “researchers...should systematically identify their subjectivity

throughout the course of their research,” likening subjectivity to a “garment that cannot be removed, insistently present in both the research and non-research aspects of our life” (Peshkin, 1988, p. 17). Following Peshkin’s (1988) technique, I deliberately paid attention to “the warm and the cool spots, the emergence of positive and negative feelings, the experiences I wanted more of or wanted to avoid and when I [felt] moved to act in roles beyond those necessary to fulfill my research needs” (p. 18). When I encountered such circumstances, I systematically documented the time, place, participants, and circumstances, along with my specific thoughts, feelings, and reactions related to the event. This documentation became part of my analysis as I identified potential themes in my personal *subjective I’s* (Peshkin, 1988) so as to be cognizant of the potential impact to aspects of the research process.

As another safeguard to subjectivity, Yin (2009) insists that “the most important advantage presented by using multiple sources of evidence is the development of *converging lines of inquiry*, a process of triangulation and corroboration” (pp. 115-116, emphasis in original). Gathering numerous sources of evidence attended to construct validity by seeking to provide the same findings through multiple measures and thus lessened the potential for any one finding to skew or invalidate the study results.

Chapter 4:

Findings

Two-County Independent School District (ISD) is a metropolitan school district in the Midwest; it is the largest district in its state of residence and is among the top 160 largest schools districts in the nation (k12.niche.com/rankings, retrieved May 25, 2016). It currently serves approximately 37,000 students from 13 municipalities enrolled in kindergarten through twelfth grade. It did not begin that way, however. State legislation catalyzed the creation of Two-County ISD, serving approximately 4000 students at the district's inception. From the beginning, however, regardless of the size of the district, staff have overwhelmingly deemed *the people* its greatest asset.

The Beginning: Superintendents Quincy's and Urtel's Eras (July, 1943 – June, 1964 and July, 1964 – June, 1975, respectively)

In 1947, the state legislature prompted communities to elect “survey committees to study the programs, facilities and finances of school districts.” After four years of intensive study, one of the primary discoveries of the selected survey committee was that the small, one-room school houses speckled across the county were over-crowded and limited in what they could offer students educationally. Many community residents and families of students attending school in the small school house settings had independently concluded this, as well. Some of the one-room schools closed their doors, sending students to the county's public schools on a tuition basis prior to the completion of the committee's work.

In a concluding report to the citizens and legislature in 1951, the committee resolved the following:

The great changes that have taken place during the hundred years that have passed since people first settled in this area have brought tremendous changes in the kinds of lives we have to live today. Life in our times is so complex that it requires a great deal of knowledge to understand and extensive training to earn a living in it. A comprehensive school program adapted to modern needs requires considerable investment in school building facilities. The county school survey committee believes that in the area designated as District B there will be sufficient financial resources to provide modern school facilities and to maintain them without undue burden on the taxpayers of the area. (Two-County ISD, *Schools for Our Times* report contained in a 100th Anniversary School Celebration Presentation, 2004)

To that end, the committee recommended consolidating the rural schools into a new school district. The consolidation area included all or parts of twenty-six small districts in one county and five districts in an adjacent county, hence aptly naming the proposed district, Two-County Independent School District. On April 8, 1952, the community voted overwhelmingly to support the consolidation as proposed.

There exists a folktale about the size of the district. It is said that Mr. Quincy climbed to the top of the water tower and claimed all that he could see as Two-County district. Now, Mr. Quincy could only see well out of one of his eyes. As the story goes, "if Mr. Quincy had two good eyes, the district would be a lot larger" (Hadley, interview, June 6, 2016).

In the early years of district, enrollment grew rapidly, and by the 1960's and early 1970's, one or more new schools were being built each year along with additions to existing facilities. Twenty-four schools, including schools at the elementary, middle or junior high and high school levels, were built between 1960 and 1976 to keep up with enrollment escalations. During this period, the district also purchased existing facilities, acquiring a building in the mid-1960's and opening a technical college as part of the

district’s educational offerings in 1967. Throughout the next thirty to forty years, enrollment ebbed and flowed slightly between years with an overall trend of growth, reaching a peak of just over 40,000 students in the 2004-2005 school year, when the district took over the top enrollment ranking in the state. Figure 4 depicts the annual October 1st enrollment over the period of study.

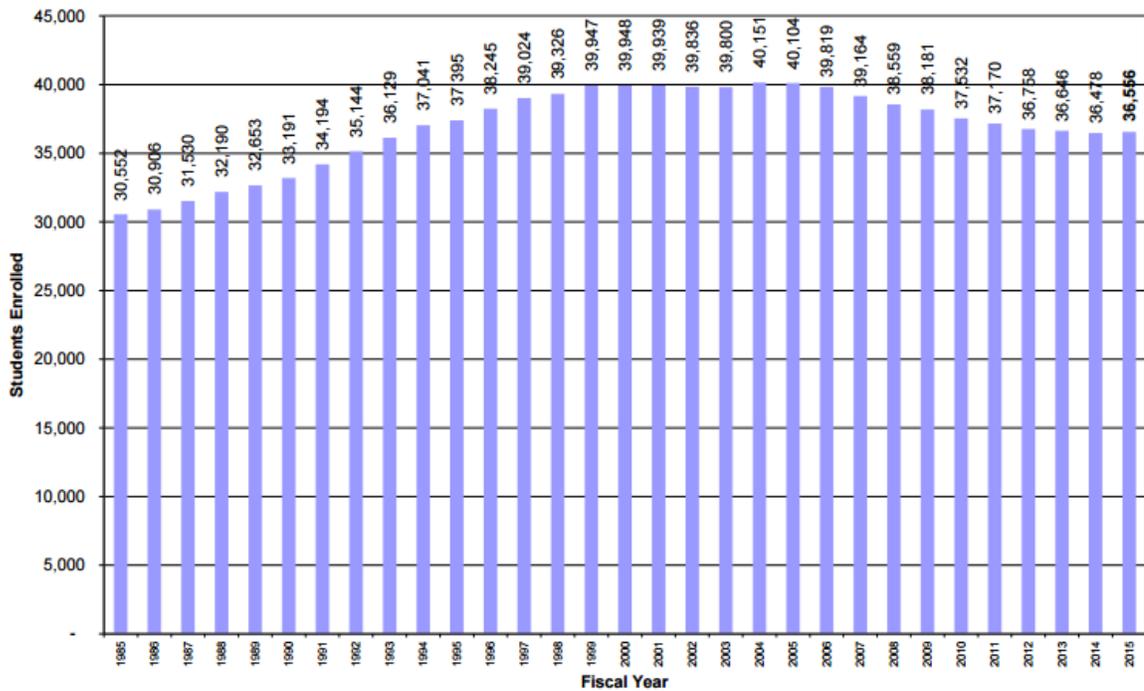


Figure 4. Annual October 1st enrollment from 1985 – 2015.

The intense growth in the district in the early years of formation, provoking the need to build and buy an expansive number of facilities, along with the low tax-base/corporation-poor make-up of the district and the process for state fund allocation, instigated a continuing pattern of fiscal insufficiency that plagues the district even still.

Superintendent Spawling’s Era (July, 1975 – July, 1990)

In 1975, as the rapid building era was reaching its end, Two-County ISD hired a new superintendent, the third superintendent who had served since the district’s creation.

Mr. Spawling was a young man, accepting his second superintendent position at the age of thirty-seven; his first was in a very small outstate district at the age of thirty-two, giving him the distinction of being the youngest superintendent in the state at the time. Almost immediately, Mr. Spawling faced controversy in his superintendency, moving the then scattered central office staff into one building, centrally located in the district boundaries in 1977. The plan and facility intended to co-locate staff and create a central office were viewed as an unnecessary luxury, and the building, although modest in appearance, was deemed the “Taj Mahal” (Gillian, interview, July 10, 2015). While contentious, bringing all central staff together, as opposed to spread out in numerous sites across the district, proved extremely advantageous.

Mr. Spawling, through interactions at the superintendent level, had learned of other districts in the area and around the nation creating positions related to evaluation and assessment. These positions were in response to federal legislation such as the National Defense Education Act of 1958, the beginning of Head Start, President Johnson’s War on Poverty and Great Societies Legislation, and the Elementary and Secondary Education Act (ESEA) of 1965, requiring evaluation as part of federal funding requirements. ESEA spurred the creation of regional laboratories of educational research and development across the country and prompted the federally-identified *Great Cities* (recognized based in large part on enrollment size) to create model research, evaluation and data units as part of their central education staff. In an effort to position the district in line with these other districts, the year following his entrance into the district, Mr. Spawling posted a position for a director of a newly-formed Department of Planning and Evaluation.

At the same time, a young scholar with extensive experience, training, and passion in the areas of research and evaluation was seeking a transition from one of these internal evaluation and data units located in a federally identified *Great City* in an extreme northern portion of the state. He applied for the position in Two-County ISD that was, in essence, the exact job he was currently performing, bringing with him familiarity with evaluation and data units in two separate *Great Cities* (the position he was vacating and one in another state where he was previously employed). He also possessed academic training and involvement with state and federal evaluation programs, criterion-referenced and norm-referenced large scale testing systems, and an earned doctorate of philosophy in the area of educational administration with a focus in research.

Mr. Spawling offered Dr. Mostenby the position, and the strong impetus for evaluative activities was initiated in Two-County ISD, including continued connection and involvement with state and federal offices, as well as service on various governmental advisory committees. Dr. Mostenby, as director of the Department of Planning and Evaluation, reported directly to the superintendent and was aligned with the business and finance operations of the district following the rule of thought at that time that evaluation units should be separate from the programs they were evaluating and if they were reporting to the department or entity that they were evaluating, the perceived objectivity was compromised.

In 1976, the state decided that there was a need to collect information from districts about the curricular process and its impact on student outcomes. Dr. Mostenby, being extremely connected to the state and federal governmental agencies because of prior employment and ongoing evaluation consulting work for them, was instrumental in

creating the Planning, Evaluation and Reporting system (PER), which was a precursor to the Systems Accountability Reporting system, legislated as part of the No Child Left Behind (2002) requirements. This annual report contained evidence of curricular studies and standardized testing results (which were mainly norm-referenced at the time – a test of basic skills and a district developed writing test). Even though the range of assessments was sparse and the sampling of students only represented a few grade levels across the system, it was used as an outcome measure of district effectiveness. As part of the PER requirements, curricular areas had to be studied in cyclical increments, with certain content areas studied each year. As part of these studies, citizens were included in each content review. These studies were extremely thorough, including literature reviews of the latest applicable research, data analysis of extant data as well as new data from surveys and group interviews, and material reviews based on identified criteria and standards.

With the rapid growth in the district, as well as shifts in where populations were living within the district, boundary changes were a constant conversation, and shifting attendance areas became very much the norm. Again, the district sought the input of the community and convened study committees to determine the best courses of action to recommend to the school board. These committees were run very similarly to the content study committees, but instead of a curricular area, they focused on boundaries and trying to project school enrollment implications aligned to facility sizes. At one point in the early 1980's, the district decided to do an in-depth analysis of the geographic locations of students living within the district-wide attendance area. Up to this point, the only

information available was from old census information blocked in large areas. Dr.

Mostenby describes the project in this way:

So what we did is we mapped the district and literally filled the floor of the central office board room with papers that had the images of all the neighborhoods drawn to scale with the lots drawn to scale with the house addresses on them. Then we assigned the geo-codes to those. That became the basis for identifying one of the first systems, demographic computer-based systems. Even then we could see the great potential of being able to assign other data sets to those locators. (Mostenby, interview, August 10, 2015)

The experience, along with related activities in previous positions, solidified Dr. Mostenby's awareness and understanding that demographic data were extremely important to assessment and program evaluation.

Also in the early 1980's, there was a growing focus on educational outcomes and public scrutiny of educational systems, spotlighted by the federal government's appointment of the National Commission on Excellence in Education and the resulting report entitled *A Nation at Risk*. The report was not very complimentary, but was probably what education needed to shake up the status quo, according to Mr. Hadley, the director of elementary education at the time. An organization for business members in the state, coinciding with the attention on education, began to articulate its concern for educational outcomes and accountability. One of the leaders of this movement conducted a study that looked at a number of outcome measures common to many state school districts and attempted to link them to the organization's accepted universal performance indicators. He concluded that the state was not matching up to its capacity (Mostenby, interview, August 10, 2015). In response to this assertion, a group of professionals – Dr. Mostenby and his colleagues in other districts around the metropolitan area – wrote a counter paper, saying that from both testing and statistical perspectives, using assessment

results in this manner was inappropriate. This was the first time that a group of individuals in positions focusing on evaluation and testing came together, and it launched an

. . . informal group of people who were involved in testing, assessment, and program evaluation and were speaking in some unity to the [State] Senate and the House and trying to temper aggressive use of testing in a direction that wasn't going to be fruitful or legitimate. (Mostenby, interview, August 10, 2015)

Dr. Mostenby went on to reflect that “sometimes the policy need outstretches the capacity of testing [but the legislatures] just ignore the capacity of testing and move straight ahead, and you wind up with a challenge” (Mostenby, interview, August 10, 2015). The group continued to meet regularly, shared experiences, and supported each other’s work through collaborative conversations.

During these conversations, the director of another district’s department of evaluation and assessment shared his work pertaining to outcome-based testing on a continuous basis, independent of and in-between administrations of norm-referenced testing. Dr. Mostenby acknowledged that in the same timeframe, the state legislature was struggling with a way to monitor educational effectiveness and was disparaging of norm-referenced testing because it was not a precise enough measure and contained bias toward some groups. In addition, the state, districts and schools did not have data available to guide instruction and ensure effective programming for all students.

Considering the development of a criterion-referenced testing system in one district and the fact that the state desired this type of monitoring for all districts, many in the state proposed adopting this model statewide. However, since the created assessments were curriculum-dependent and the state did not have a mandated curriculum, legislators

instead mandated either using a state-supported system of monitoring or having individual districts develop their own comparable criterion-referenced outcome measures.

In conjunction with the increased attention on education and student outcomes, Two-County ISD staff participated in numerous evaluation and improvement activities as well as creating annual goals for the school board and superintendent. For example, the district applied for and received a corporate-sponsored grant to participate in the *Effective Schools Program*. As part of this program, each school in the district created annual improvement plans, incorporating an evaluation of the school's performance based on a variety of data including teacher performance measures, strategies for improvement, timelines for implementation, and evidence of effectiveness and progress.

The district also joined an association that provided external evaluation of content areas in a cyclical schedule similar to the state-mandated curriculum studies. These external evaluations provided information related to program quality and recommendations based on observation and document review conducted by educational content experts. Aligned to this type of accreditation evaluation, the district also conducted intensive studies of its vocational technical institute. These studies occurred on a recurrent basis as mandated by the state's post-secondary system. The studies and resulting reports incorporated input from and involvement of a broad range of stakeholders and were presented to the school board upon completion.

Another evaluation activity in practice during this time was graduation follow-up studies, as mandated by legislation. The district would annually survey the current year's graduates as well as the graduating class from five year's previous to determine what individuals were doing as well as their perceptions of preparedness to enter a post-

secondary educational institute, the workforce, the military, or whatever their other path of pursuit. This information was shared with the school board and district personnel in an effort to improve programming.

In the 1980's, Two-County ISD purchased its first computers. Dr. Mostenby, at the time, had been doing computing on extraordinarily sophisticated calculators, about the size of a notebook, which could accept a small two inch by half inch paper containing as many as two hundred pairs of data. This offered a tremendous change in analytic power over previous tools even though it was not computer-based. Dr. Mostenby, not having the same confidence in the new tools as others in the district, viewed the computers as more of a liability risk from an insurance and security standpoint than an asset.

In 1981, the district teachers went on strike because a contract settlement could not be reached. This dispute created great upheaval amongst staff, pitting teacher against teacher and teachers against other bargaining units. For staff who had a dual contract, working part time as a teacher and part time as a central administrator consultant, it was especially difficult. Trust and collegiality between employees was threatened and, in some cases, eroded and seemingly lingered as a filter through which teaching staff viewed central employees.

As time went on, Dr. Mostenby became more and more integrated with budgeting, finance, and program measures in an effort to match data to different program components in a structured way to monitor fiscal efficiencies and program effectiveness. This work was not new to him as he had gained experience in these areas as part of his first *Great City* evaluation unit assignment. The connection to finance also existed in his

second assignment through organizational alignment to the business director to maintain evaluation objectivity. Through his experiences, Mostenby came to discover that there were three non-instructional factors that were paramount to managing education; enrollment, finances, and personnel costs. If one of these elements changed, the others, and all other district operations, would change, as well.

In Two-County ISD, he had direct oversight of student testing and assessment and all the information systems within the district, but, with a greater emphasis on budgeting and finance, time available to dedicate to student assessment and its implications for practice diminished. Even so, he was well-respected, with a positive demeanor and other leaders in the organization, including the instructional leadership, would seek out his insight and advice, viewing him as a valuable collaborator.

In 1986, Two-County ISD 11 formed a partnership with the state's land grant university through their Center for Educational Policy Studies. The study was an evaluation of courses offered at the high school level, marking the first formal collaborative evaluation with an institute of higher education for the district.

Also, in 1986, Dr. Mostenby, in direct response to the legislative mandate and his shortage of time to dedicate to student assessment, envisioned the Student Assessment Department, which would be under his supervision, and posted a coordinator position. Dr. Eickner, who at the time had not yet finished her doctorate, saw the posting in a newspaper in the Chicago area where she was residing. She had heard that the Two-County school district was progressive and her parents lived in the area of the district, so she decided to apply. Dr. Eickner came with a background in teaching high school English, had a master's degree in history, and was currently working as an administrator

in gifted education. She had experience with assessment in terms of identifying, selecting, and evaluating students for gifted programming, but student assessment was not her area of study. Although others who applied for the position had degrees and experience in research and evaluation and Dr. Eickner was just completing her doctoral degree in educational administration, she

. . . was astonished when [she] was picked. . . but what [the leaders] wanted required more of an organizational, political role because [the district] had no department at all – nothing. [My] job was to set it up and to create district-wide assessments in key areas and start providing data back to schools that would be useable to them. (Eickner, interview, July 8, 2015)

Dr. Mostenby, reflecting on Dr. Eickner’s performance, stated that she “brought an articulation with teachers that allowed her to speak the language unusually well, and [she had] a good feeling for what it was – she had a great feeling for what would work with program improvement, related to instruction and teaching and away from the assessment.” He went on to say that “she was hands and heads over my performance” (Mostenby, interview, August 10, 2015). Following one particular presentation, the staff even applauded. Dr. Mostenby asked in disbelief if the crowd had just clapped for the assessment person, adding that that had never happened before. Remembering Dr. Eickner’s ability to explain data and evaluative findings, Ms. Gillian, professed, “[she] was just like the most wonderful gift to have because I felt like I could understand it! [Then] I could share that information more with teachers” (Gillian, interview, July 10, 2015). Dr. Eickner acknowledged strength in having a formidable content background, allowing her to lead the effort of developing content assessments while gaining the respect of the teachers who were enlisted to assist in the process.

The district structure included administrative-level content specialists at the central level, who were assigned to oversee the content departments, organize, and provide staff development opportunities, and, with the innovative district initiative, collaborate on the development of content assessments. Dr. Eickner identified that everyone seemed to know about norm-referenced testing, but knew nothing about criterion-referenced testing. The other concept at play was a matter of trust. Teachers had the impression that they had to be the ones to write the tests for them to be accepted and deemed to have any quality, even though they did not have any training to do so, which added complexity and opportunity: complexity in organizing multiple teams of teachers and keeping them engaged in the work, and opportunity to teach them to write quality assessments and providing them basis for using the data that were generated given that there was not a widespread, comprehensive culture of using assessment data within Two-County ISD at this time. Dr. Eickner recalled that “for most of the teachers, [giving the assessments] was compliance, not enthusiasm” at that time (Eickner, interview, July 8, 2015).

Dr. Eickner worked diligently with different sectors of the curriculum department to develop, pilot, and fully administer content assessments in three-year cycles to provide instructional information to teachers, administrators, and central office staff. After the three-year administrative cycle for an assessment, the group would reconvene to examine the item analysis for the test, making necessary adjustments at that time. As the “*new kid on the block*”, overseeing a brand new department and reporting to the new director of finance, she worked hard to build relationships and inroads with the well-established curriculum team of consultants.

Shortly after Dr. Eickner was hired, a new director of curriculum was hired. He had a strong belief that, since she was working with classroom assessments and collaborating with the curriculum team on these efforts, Dr. Eickner should be within his oversight. This created intense contention between the curriculum department leadership and Dr. Eickner, and consequently, she was not included in any meetings with them as a collaborative group; her work with the consultants occurred on an individual basis, driven by their request to begin the process for their content area, which mainly aligned to the curriculum study cycle along with clear-cut directives from the superintendent, Mr. Spawling, and Dr. Mostenby that it had to be done.

Mr. Spawling, in the late 1980's, went to a national conference with other superintendents and school board members and attended a session about a district in another state that was in litigation. The lawsuit was founded on the circumstances where parents of a group of students who were denied their diplomas were suing the district because the district, in their view, should be held accountable to ensure at least a basic education for all students and the students should not be held liable for the district's failure. He returned with the notion that Two-County school district needed to safeguard itself from a similar situation and posed the challenge to Dr. Mostenby, who, in turn, presented it to Dr. Eickner, who seldom interacted with Mr. Spawling directly.

Dr. Eickner recalled that they thought it “would be a really good idea to set a floor on a high school diploma and to say that [students] have to at least have *this* because those are the skills that are really necessary to survive out in the world” (Eickner, interview, July 8, 2015). She said that the idea was “not to get really fancy about it because there were many jobs that did not require higher algebra or anything like that”;

they really intended to “just go for what [students] need to exist in the world and make sure all of the kids have that” (Eickner, interview, July 8, 2015).

This grounding philosophy was the basis for a district-wide assessment, administered to every student beginning in 8th grade, to ensure mastery of basic competencies in the areas of mathematics and reading. The tests for each subject were developed in sections by content strands, but the entire test was given the first time a student took it. For students who did not pass the entire assessment the first time (meaning they achieved 70% or higher on each section), only the portions that were below the 70% threshold were re-assessed and repeated in other forms until they were passed. Review and remediation sessions were offered intermittently between scheduled administrations, and diplomas were withheld for students not meeting the required benchmarks. Some of the only times Dr. Eickner interacted with the school board were to deliver counts of students not passing the test and, thus, not earning a diploma, as part of an annual comprehensive student assessment report. Upon Dr. Eickner’s departure from the district to work at the state level, this assessment served as the pre-cursor to the first statewide high-stakes test for students to measure the basic understandings deemed necessary for competent living and was paired with performance assessments across grades that were meant to measure high standards of performance.

Over time Mr. Spawling developed a combative relationship with the state legislature and governor over educational vision and funding disagreements. He did not approve of the direction to push post-secondary options and open enrollment for high school students and was not shy about stating that publically, as well as overtly antagonizing state officials, “following [the governor] around the nation saying what a

bad idea that was” (Cadence, interview, July 2, 2015). He was overheard telling Dr. Mostenby, who was director of finance at the time, that he was going to keep spending, even to the point of bankruptcy, in an effort to make a point with the governing body. One legislator professed to the district’s legislative liaison, “You’re never getting anywhere as long as he’s superintendent” (Cadence, interview, July 2, 2015).

Not sharing Spawling’s new-found disregard for the legislature, Mostenby expounds:

You have to pay attention to it. You have to be there. It helps to be deeply involved on a trusted level with the people who are making the governmental policy statements. The department [of education] and the staff, the legislature and the legislators and their staff, the educational groups and bodies, the parent groups and bodies – absolutely critical. The more you can interpret and articulate to the parents in your community or district, the stronger your base is going to be to talk to the policy makers in saying yes or no [to the bill proposals]. Really important. Tough thing to do. (Mostenby, interview, August 10, 2015)

In March, 1990, there was another threat of a strike, and teachers were liberally sharing their discontent with the community. This information made it back to Mr. Spawling, who promptly wrote a six-page letter all in upper case, which was given to all staff in hard copy format. This action was not well-received by staff. In June of 1990, Mr. Spawling left the district and an interim superintendent was named.

Another serious threat of strike ensued in 1991, with negotiators reaching an agreement just before midnight the day before a strike vote was scheduled. A ten-year pattern of strike threats began with pressures of striking resurfacing in 2000 and again in 2010.

Interim-Superintendent Hadley's Era (July, 1990 – January, 1991)

In July, 1990, interim superintendent Hadley, who was previously an elementary associate superintendent, took the helm in Two-County ISD. He served the district in this capacity for seven months, through January, 1991. Under Mr. Hadley's direction, the first random-sampled survey facilitated by an outside professional vendor was conducted to solicit input from the community regarding their thoughts on a levy referendum the school board was considering. This was another period of rapid growth in the district, and needs were outpacing revenue, prompting the district to go to the taxpayers for help. Speaking of a need for clear, united messaging to the community, Mostenby attested that "everyone has to be on board with the needs of the district and the integration of those needs. You cannot have people fighting over the pieces of the pie when you're petitioning the public or going to the school board to propose a program reduction and/or adjusted program funding" (Mostenby, interview, August 10, 2015).

By this time, computers were being purchased and used superficially, but none of them were networked or connected in any way. In the student assessment office and in most schools, scanning machines tallied test results recorded on special computer forms. There were approximately 120 criterion-referenced classroom assessments that were administered across the district, and the assessment department had scanning them down to a science, returning the results to schools in two days. Dr. Eickner knew that the data would only be used if the department got the tests back to the teachers immediately, so they worked for an extremely quick turnaround.

The assessment department also had a few computers, but Dr. Eickner disagreed with the leader of technology as to how to expand technology and, specifically, which

platform to purchase. The director of technology was an Apple platform supporter, and Dr. Eickner recognized that a PC platform better supported large-scale database work, which caused tension and competition for resources. Even though both the technology director and Dr. Eickner reported to Dr. Mostenby, the forum to discuss this situation or to ever collaborate together was non-existent. Eventually, the district went the direction of the technology director and purchased Apple computers for every classroom in an effort to have access for all staff members. Dr. Eickner independently pursued computers with the alternative platform in support of the department's work. In addition to trying to increase effectiveness internally, Dr. Eickner also worked with external vendors to develop software to more efficiently run large batches of classroom assessments.

Superintendent Osborne's Era (February, 1991 – July, 1995)

In February, 1991, Mr. Osborne joined the district as superintendent. Culminating his first year, Mr. Osborne issued massive cuts in central office personnel, dismissing all administrative content consultants in the form of a letter they all received one morning in February, 1992. They were given the directive to find positions back in the classroom, and, if they chose, they could re-apply for the fraction of remaining replacement positions, classified as teachers on special assignment (ToSAs). It became the feeling of district personnel that Mr. Osborne was hired with the understanding that he would reduce the number of central office administrators and thus work to repair relations with the teaching staff. Others hypothesized that Mr. Osborne thought the district was too big and was trying to break it into sections to give it a smaller feel and run it more like a *ma and pop shop*, lacking clear systems and documentation of process, instead of a large metropolitan school district.

Mr. Osborne's philosophy was that the ToSA positions were not going to be aligned with content expertise, but were to serve as process support, leaving the content expertise to the classroom teachers in each subject area. Following this same philosophy, he realigned elementary and secondary leadership into *clusters* affiliated with geographical areas of the district instead of levels. Each cluster was assigned a central leader and a couple of ToSAs to provide process support as requested by building staff.

Each cluster also was given the autonomy to determine its academic area of focus, which in some cases meant ToSAs trained in math, for instance, might be supporting the cluster in writing improvement. Principals from all levels within the cluster met together, and meetings were not held by level. This provided the opportunity to have vertical collaborative conversations between principals whose schools were *feeders* for other schools, but left a void in collaboration by level.

In 1992, the state legislature passed a law creating rigorous, results-oriented graduation rules for all public school students modeled after the assurance of basic learning assessments employed by Two-County School District, developed under Dr. Eickner. The legislation outlined a basic standards assessment to ensure that students possessed the minimal skills necessary to successfully live and work following graduation and high standards oriented toward excellence, as measured by performance assessments, providing students the opportunity to demonstrate mastery of high-level skills. Dr. Eickner, recruited by state officials, left the district for the assessment position at the state department, and another person with a teaching background was hired in her place. Dr. Dooren was hired as the coordinator of student assessment in July, 1992. The prevailing philosophy related to assessment had shifted to a belief that the work should be

aligned to curriculum and instruction, so Dr. Dooren's supervision was shifted to the person serving as the associate superintendent overseeing the curriculum department. This facilitated much more collaboration with the new curriculum ToSAs, providing a channel to support their work around assessment more closely.

Dr. Dooren attended and completed her doctorate in educational administration while serving as coordinator of student assessment, continuing a pattern of degree completion begun with Dr. Eickner. The majority of the effort within the department under the oversight of Dr. Dooren was the continuation of classroom assessments, the assurance of learning tests, and broad-scale support of the curriculum study process. Dr. Dooren also continued annual assessment reporting to the school board with minimal interaction with them beyond that. As part of her role, she attended meetings with employees from other district with similar job duties as her own, which provided an avenue for collaboration and learning from each other. Dr. Dooren recalled conversations about data being collected differently by "shifting from equal access to equal outcomes...which changes the whole program" (Dooren, interview, July 7, 2015).

Besides the annual assessment presentation, evaluative activities for the most part ceased during Mr. Osborne's tenure, with a few exceptions. In 1992, the district agreed to implement the first statewide student survey dealing with students' perceptions of school activities; school culture and climate; health and wellness behaviors such as food consumption and exercise; risky behaviors such as drinking and substance use; dating experiences; and out-of-school activities. This survey was administered to students in grades six, nine and twelve and was scheduled to be given every three years.

In 1993, the district participated in the implementation of a state-sponsored graduation rule pilot that was an off-shoot of 1992 state legislation on standards development, enacted to take a first step toward defining outcomes for all students in the state. Dr. Eickner through her work in the state Department of Education developed a system of performance packages based on standards aligned to various courses in ten different areas, including reading and math but expanding far beyond. Each student was required to pass a series of courses, demonstrating upon completion of the series, competency of the full range of standards that were embedded into the courses. Competency was measured through successful completion of all aligned courses and had to be completed to graduate. Dr. Eickner was seeking districts to agree to pilot the packages, and a grant of \$80,000 was available to offset the costs of the pilot. Two-County ISD agreed to participate, which provided an opportunity to be on the forefront of the learning curve that was certain to accompany this state mandate. After the pilot, in 1994, a presentation was made to the school board.

Other activities that occurred during Mr. Osborne's superintendency included a district-assembled task force to study facilities and scheduling and numerous program improvement grants that were submitted to the state department of education. Mr. Osborne resigned in July, 1995, at which time Dr. Mostenby was appointed superintendent. Ms. Gillian reflected on Dr. Mostenby's appointment to the superintendent's position after Mr. Osborne's departure.

[The school board] felt they needed some stability. Osborne had come in; he'd made all this huge change, really alienated a lot of people, and then he left. It was clear he was here as a stepping stone to a wealthy district. I really felt I knew most about what was happening in the curriculum and assessment departments because we would always eat lunch together and talk about all this stuff. And then once that was

blown away by Osborne, I never after that time felt I was quite as much in touch. (Gillian, interview, July 10, 2015)

Superintendent Mostenby's Era (July, 1995 – December, 2008)

By the time Dr. Mostenby was superintendent, the assessment department was intricately embedded into curriculum. Two new very dynamic and competent curriculum leaders, Dr. Villard and Dr. Wheaton, were hired into the district and were serving as cluster leaders, but also assumed oversight of the elementary and secondary levels, one aligned to each level. With numerous other responsibilities, Dr. Mostenby delegated the assessment responsibilities to these curriculum leaders, however, reinstated regular reports to the school board regarding data use, assessment, and evaluation activities.

In addition to continued curriculum studies, external evaluations were conducted related to students needing accommodations, but not qualifying for special education (504 plans) and average daily attendance studies. The district joined an organization from a division of the state land grant university dedicated to applied research and educational improvement and another state academic excellence organization, as well as applied to participate in grants related to the Improving America's Schools Act (1994).

Other new legislation impelled changes related to the district's technical college. In 1995, the state legislature passed a law assembling all public technical and community colleges and state colleges and universities into a new state system, thus relegating district ownership of the college to the state. As consolation for this decree, the district received support to establish the first of its kind secondary technical program for high school juniors and seniors, physically attached to the technical college to facilitate collaboration and partnering across programs.

In 1996, the state basic skills test, mandated as part of the 1992 standards legislation, was implemented for the first time to eighth grade students in the areas of reading and math. Students had to pass each assessment with 75% correct to qualify for their diploma. If students did not pass on the first administration, they had to continue taking the assessment annually until they did pass, or they would be denied their diploma until a passing score was achieved. This mandate, along with record-keeping related to successfully demonstrating completion of the high-level standards aligned to a wide range of courses and the performance assessment requirements as the other part of the same legislation, proved extremely difficult and taxing for districts, among them Two-County. Further adding to the strain on districts to create processes and practices to manage the extensive record keeping and expanding state testing, a basic writing test was implemented statewide in 1997, along with state comprehensive assessments in grades three and five in the content areas of reading and math. The addition of these assessments began an almost annual process of expanding and changing state-required assessments that continued through the duration of this study. (See Appendix D.)

The cluster system, established by Mr. Osborne, had been mostly abandoned, and the central office was again aligned primarily to grade level bands under the supervision of Dr. Villard and Dr. Wheaton. The number of staff in the district office was not reinstated, however, and the unofficial district motto of *doing more with less* was in full force. People were empowered to do what needed to be done, and sometimes that meant something outside the person's job description if he or she had the skill, ability, expertise, or sometimes even willingness to do it, adding to, and strengthening by some accounts,

the *ma and pop* feel of the district office. Dr. Florence, reflecting on leadership changes in terms of how things operated, explained:

We ran a certain way with Spawling because of who he is. Osborne – we had a totally different structure because of the type of person he is. Then Mostenby came in and because of his personality, it changed dramatically again, even though he kept a lot of Osborne’s structures in place. But Mostenby very much delegated, very much. (Florence, interview, July 8, 2015)

In the same way, cabinet members, who were convened for the purpose of advising the superintendent on key decisions, did not serve as an overarching decision-making body; instead, each cabinet member was authorized to make necessary decisions for his or her sector of oversight and did not need to formally gain cabinet input or approval before moving forward. Often cabinet members conferred with other cabinet members and leaders, but each was sanctioned and trusted to make his or her own decisions.

With the completion of the graduation standards pilot, Two-County ISD recognized the need to continue the system of training and support for teachers and other stakeholders developed through the pilot. In addition, there was recognition that the district needed to be working toward increasing the capacity to gather, understand, interpret, and use data to make decisions related to curriculum and instruction, which was overtly communicated as the expectation by Dr. Mostenby and the school board. Supporting staff in how to use data to make decisions, improve instruction and programs, and monitor district priorities were added to the focus of the department. In an explanation as to why the move to using data for anyone who did not inherently see, Dr. Mostenby argued that “without [data] you have nothing. If it isn’t meaningful data you have nothing. You’re kidding yourself; you’re floating; you’re working on gut alone, and

that's not going to work. It's gone, too! Those days are gone" (Mostenby, interview, August 10, 2015).

The school year 1998-1999 was a difficult year in the student assessment department. Dr. Dooren left the district at the conclusion of the previous year for a position in a neighboring district, and the coordinator role was not successfully replaced. The departmental testing functions were still required, so the Curriculum Director oversaw the work of the three secretaries in the department who stepped up to facilitate and maintain the district-wide testing obligations successfully.

In the area of technology, Two-County ISD was making great progress under Dr. Mostenby's leadership. Steps were being taken to improve the district's network capabilities and increase technology. In 1999, a new position was created and an Information Systems Coordinator was hired to work closely with a principal on special assignment appointed to assist with implementation of a new student information system purchased that same year. The acquisition of the student information system prompted the start of conversations about who had access to what data – the whole notion of data security and data privacy. In addition, conversations and training were critical around accurate entry of data and data checking to insure integrity in the data.

With forward strides occurring in technology, the ability of the district to capture and house student data and the intensifying public outcry for educational accountability, the picture became clear that the district needed to move forward with data availability, data accessibility, data understanding and interpretation and building evaluation capacity. In an effort to address these growing needs, Dr. Villard and Dr. Wheaton devised a plan.

The two district leaders were aware of a governmental employee exchange program that allowed two public organizations to enter into an agreement whereby one agency would pay the salary of the other agency's employee in exchange for that employee to be on loan to the funding entity. Dr. Wheaton had control over one-time funds that were distributed from the state to assist districts in implementing the graduation rule and decided this would be a good way to use the money effectively.

Given the past interaction between the district and the state land grant university and central employees' personal experiences with professors at the university through course-taking, the district leaders approached a professor in the area of evaluation studies about coming to the district full-time as part of an exchange. After careful consideration of the associated pros and cons (contemplating the interesting aspect of a professor "putting her money where her mouth is") and input and approval from the university, Dr. Kincade agreed to join the district in August, 1999 for a two-year term as coordinator of research and evaluation. The plan also included a provision to allow Dr. Kincade to focus on research and evaluation activities; one of the ToSAs, Dr. Anthony, would join the department and assume direct responsibility of the ongoing student-assessment-related activities. Dr. Kincade and Dr. Anthony would be supervised by the curriculum director on paper, but in actuality, the secondary associate superintendent would direct their work. As with the previous two coordinators of the student assessment department, both of these new department members had a background of teaching in the K-12 public education system; Dr. Kincade with beginnings as a seventh grade English teacher and Dr. Anthony as a middle school technology education and math teacher. Kincade believes that in her work with Two-County ISD, "[her] my personality [was] an important factor. I

could think of some professors ...that if you sent them in [to a school district] it wouldn't work. I'm a school person. I like people. I'm silly. These were [her] people; [she] was one of them whereas a lot of people from the university wouldn't have had that" (Kincade, interview, July 28, 2015).

One of Dr. Kincade's first orders of business was to analyze and carry forward the work of the former graduation rule pilot as well as to design an implementation study, honoring the funding stream supporting this endeavor that was allocated for this purpose. Beyond that task, Dr. Kincade developed broader goals for the department for year one of her exchange:

- Continue to support the district testing program
- Establish a system structure for evaluation and research
- Create a school profile/template process that supports the school improvement plan process and graduation rule implementation
- Collect data for selected district initiatives
- Create active links to the state university and other institutes of higher education
- Market the Assessment Department both internally and externally

In January, 2000, Dr. Kincade was included on the school board agenda, and, for the first time, a departmental plan with specific evaluation activities and goals was presented.

As novel as this was, Dr. Kincade's interaction with the school board and even with Dr. Mostenby was extremely limited; she spent her time working mostly at the central curricular level and not at all at the policy level. Dr. Kincade puts it this way; the research and evaluation department "did not get to the superintendent at all. I met with him once the whole time I was there, which I think is an astonishing thing." She went on to say, "I did not meet much with the principals [either]. I occasionally went to principal meetings when there was a specific topic" (Kincade, interview, July 28, 2015).

Indications that the department could and should emerge as a department in its own right, separate from the curriculum department, were beginning to surface. Dr. Kincade was charged with determining a name for the evolving department, but to no avail, as none of the suggestions were accepted. In addition to naming the department, Dr. Kincade took initial steps in determining foreseeable needs in the department with the imminent expansion of testing and accountability. She, along with Dr. Anthony, drafted a proposal to add staff to the department to support the growing responsibilities, beginning with a request for an additional secretary.

Technologically, computers were present, the network was formed, and information was being collected, but data were still exceptionally hard to retrieve. When data were needed, a request would be made to the information systems department secretary, and after a week or two, depending on how busy she was, data would be available. Scanning sheets were still widely used by schools and the student assessment department.

Even so, Dr. Kincade and her team made tremendous progress in planning and conducting evaluations of various programs using a participatory approach as well as engaging in a wide variety of other evaluative activities, all with the deliberate intent of building evaluation capacity in the district. The departmental goals for year two of Kincade's employment were as follows:

- Build the system's capacity for evaluation/research
 - Conduct annual "research" projects on issues important to the district
 - Support school level program evaluation by building staff
 - Facilitate and foster classroom level action research by teachers
- Conduct studies on selected district initiatives
 - Self-study support
 - Graduation standards implementation plan
 - Collaborative study of Title I program effectiveness

- Analysis of district K-12 math achievement
- Exploratory study of multi-age groupings in the district
- Participatory study of talent development
- Evaluation of diversity and district’s desegregation plan
- Examine and manage the district’s testing program
- Provide routine assessment department services and support
- Broaden the assessment department’s scope and viability

Collaboration with external evaluators expanded in the district during this timeframe, due in part to the increased focus on evaluation under Dr. Kincade and to her connection to a host of trained evaluators through the university evaluation studies program and the broader evaluation community. Data collection by means of surveys and focus groups also increased, again, related to Dr. Kincade’s expertise, experience, and professional interests and commitments. Kincade reflected:

I never felt we were very well evaluated ourselves, which is ironic, I know. But we were just scrambling to [keep up]. We worked really hard; everybody in that department worked hard and played hard, too. I mean, that’s what was so fun about it. I do think that it was planting the seeds and it took different leaders, technological development, time passing, [and] legislative mandates. (Kincade, interview, July 28, 2015)

In August, 2001, Dr. Kincade “handed over the baton of evaluation” to Dr. Anthony upon the end of the exchange and her departure from full-time employment with Two-County school district. The accountability age in public education had fully matured, and testing was materializing as the main player in the evaluation, research, and assessment arena, which caused the *baton of evaluation* to be almost fully dropped from conception as the district focus necessarily swung toward testing. Two-County ISD, in an effort to poise itself for the impending accountability movement, was shifting toward new forms of standardized assessment to screen students for intervention or extension,

monitor progress, guide instruction, and predict performance on the expanding state tests scheduled to be mandated.

Dr. Anthony agreed to stay aligned to the reverted student assessment department in a ToSA role, overseeing the operations of the department with the assistance of a newly created assessment coordinator role and two secretarial positions. He continued the work of supporting the schools with school improvement plans, understandable and assessable data to inform their plans, and interpretation of state standardized assessment data and district-generated assessment data to gauge and monitor effectiveness of action steps included in the plan. Dr. Anthony served on varying committees through the state department of education and attended meetings with others with job duties aligned to research, evaluation, and assessment across the state in an effort to stay connected to the happenings in the evaluation world, including the inevitable enactment of a new law applied to education in the country.

Prior to the election of George W. Bush into the presidency, Mostenby describes an interaction he had with a reporter.

The reporter asked if I had any concerns about the candidates' platforms for education. I naively said something that I truly believed at that point, that with the [reauthorization of the] Elementary and Secondary Education Act they can affect the resources available but it's almost impossible for them to affect the policies at the local school level. Then came No Child Left Behind and I was wrong. In fact, the policies, the practices and the focus of every individual school district in the country [was impacted] because it introduced [system] requirements or requirements for performance, for good or bad. (Mostenby, interview, August 10, 2015)

In January, 2003, the No Child Behind Act (2002) was brought to the school board as the nation's new accountability law and, with it, many mandates and implications for K-12 education. As a result of the overwhelming need to keep up with

expanding understanding and guidance of these requirements emerging from internally-formed district committees and externally-formed state committees, along with the rapidly-changing state academic standards and the obligation to align curriculum to them, the district abandoned the curricular study process to focus on managing accountability and testing requirements.

Dr. Anthony stayed in the departmental leadership role, supporting increased district-wide testing and providing whatever assistance was possible to as many schools as possible through mid-2003, when a principal on special assignment, Ms. Jordasen, was appointed as the head of the department. Dr. Wheaton called Ms. Jordasen into her office one day and told her “we are really concerned with all these tests and how we are going to manage this” (Jordasen, interview, July 23, 2015). Dr. Wheaton went on to explain that she needed Ms. Jordasen to aid in this endeavor.

Ms. Jordasen had a background in teaching and school-level administration, had credibility with her peers, and was well-known in the district for the work she was then doing to create processes and efficiencies in various systems related to emerging technological tools such as the student information system. Recalling feeling tentative about serving directly in a central role, Ms. Jordasen reflected the perception in schools about the district office at the time saying, “There was always so much change over there. So you finally got it, then there might be a restructure, or there might be someone else’s program or thing that became the more favored one or whatever” (Jordasen, interviews, July 23, 2015).

The general feeling of district leadership was that the department needed an administrative position to firmly establish it as a separate, capable, and viable

department, give it a name, and build on the planning and insight of an expanded departmental structure started by Dr. Kincade, bringing the departmental model closer to full fruition. In addition, things in the district related to increasing external accountability and testing mandates were changing so quickly that processes were not formalized, but instead were more fluid and devised as situations arose.

Given the circumstances and the identified needs, Dr. Wheaton believed that Ms. Jordasen was the person who could facilitate and manage these changes; she was skilled in the areas of organization and process, and she was available, given that she was already on special assignment at the district level. Charging Ms. Jordasen with inaugurating the enhanced department and a director-level presence fit with the district's go-to empowerment practices otherwise commonly known as "fill the role you're capable of filling because it needs to be done" (Inettow, interview, July 16, 2015). As part of Ms. Jordasen's role to learn about and to manage the expanding mandated testing program, she joined state committees as a representative of Two-County ISD.

Along with increased testing, the district launched a number of stakeholder perception surveys, building on the work of Dr. Kincade. In the early- to mid-2000s, an anti-bullying survey was created and administered to students to more closely monitor bullying behaviors in the district. Questions of this nature were asked as part of the state-administered student survey, but data were desired on a more frequent basis than every three years (which was the incremental administration of the state survey) as well as to validate the data of each by providing another source.

Parent and community satisfaction surveys were also developed, along with reestablishing the community survey conducted in the 1990's. The parent survey was

developed in conjunction with the state land grant university and was instituted and approved by the school board. The parent survey was administered annually on a voluntary basis and by phone by an outside agency to a random-sampled quarter-sized segment of the district. The community survey was also administered through random sampling of the district as a whole.

External evaluation remained a supplement to internal evaluation activities, and outside evaluation units were employed on an ongoing basis. For instance, the district contracted with an evaluation unit associated with the state land grant university to conduct an evaluation of the work funded through a state-issued compensatory pilot grant. Another example re-engaged Dr. Kincade and a colleague with the district when they were hired to evaluate the implementation of an elementary instructional model proposed to decrease the workload for fourth and fifth grade elementary teachers. In an illustrative example of process use as a means of building evaluation capacity, Dr. Florence recalled that participating in this project with Dr. Kincade was professionally transformative saying:

[Participating in the evaluation] brought the whole focus of program evaluation into our day-to-day work and our willingness to respond to data that we see, really not even just the data, but the information that was gathered from the analysis of the data, and really take a look at the work that we were doing. I think in elementary, at least, we continued to do that for some time, but that really changed our work. (Florence, interview, July 8, 2015)

Dr. Kincade remained a close collaborator, coach, and consultant with the district throughout time, working on numerous evaluation-related projects, even involving one of her university classes on a large reading evaluation project for the district benefiting both

Two-County ISD by having a large team of evaluation interns of sorts and the students by providing real-world evaluation experience.

By 2004, the district had investigated, selected, and implemented the beginning phases of a data warehouse that would provide access to a wide range of student data all in one system, to applicable stakeholders, including parents. Data that were now available included demographic data, achievement data from local, state and national assessments, marks and course history, enrollment data, attendance, student plans such as health plans and individual learning plans, and program data such as interventions and special courses in which students were enrolled. Attention to access, understandability, and application was intentionally placed on the teacher role first, to most significantly affect student achievement. Access to data and training for other stakeholders with the overarching objective of data accessibility and understanding for all grew from there. In addition, greater attention was paid to role-based security measures to ensure data privacy and how to leverage data, such as occupational codes and employee assignment records, to automate staff access as much as possible. Questions about what data to display and who should be able to see which types of data became part of the ongoing conversation, along with again, ways to ensure integrity in the data and protect student and family information. For instance, students qualifying for free-or-reduced-priced services are sensitive data that may be limited to certain users.

Mostenby, prophesying about the continued development of technology to support all captured data, advised that “the change we’re in now is the merging of these extraordinary data systems into unified systems that allow for stability and continuity over time, longitudinal studies. [In response to this development], there has to be an

articulation by all of us, [central departments, school personnel, parents], of where technology is going to evolve [so it moves] into the support system that we need – we have to anticipate it; we cannot just let it happen” (Mostenby, interview, August 10, 2015). Mostenby also cautioned that educators and stakeholders in the educational world needed to be the drivers of the next technological tools to ensure they meet our needs – the needs educators envision – instead of letting industry, as the non-experts, decide for us; all in the context of keeping the interests and needs of every individual student in mind.

As part of the training efforts and support for the data needs of school sites with the transition to a full-fledged accountability system and an inundation of new data, part-time, school-level achievement analyst positions were created, first at the high schools, then at the middle and elementary schools. At the high school level, the analysts were housed at the site and supervised by the school principal, yet supported by the student assessment department. At the middle and elementary schools, centrally-assigned analysts supporting multiple sites, were housed all together in a central location, and were supervised by the formally-named student assessment department now entitled research, evaluation and testing (instead of assessment because the curriculum department, formally named curriculum, instruction and assessment, did not want to give up or share the *assessment* name or its connection to the teaching and learning cycle).

In addition to the school-level achievement analysts, Ms. Jordasen worked with district leadership to instate a technology specialist and a district-level analyst into the department. Given the professed direction of the state to move toward an online format of testing and the increased software used in the department, a technology specialist was

regarded as a necessary addition. Likewise, with the increased amount of data available and the training needed for staff to most effectively use these data, the need for predictive modeling and the demand for ways to identify students with deficits along with specificity around pinpointing what those deficiencies were, and the general necessity for overall data support for school leadership, the achievement analysts, and central staff, a district-level analyst was warranted and approved, as well. Along with seeking new positions, Ms. Jordasen worked diligently to have the three secretary positions reclassified to reflect the advanced level of skill necessary to effectively fulfill their job functions within the department, but without success. These positions had become training positions for higher-paying, less-demanding jobs in schools or other departments.

Soon after the departmental foundation of roles were all established and filled, dissention ensued. Budgets and resources were limited and tight, as always, and needs were disproportionately high. Schools started looking at central staff and deciphering how they might acquire more in-house support. In the same way, central office departments started looking at other central departments to determine how they might gain direct access to staff. In essence, high schools were questioning the need for an expanded central research, evaluation, and assessment department; middle schools wanted what high schools had in the form of supervision of their own achievement analysts and not having them deployed by the district; and the elementary curriculum department wanted oversight of the elementary achievement analysts to directly determine their function without the need to collaborate with another department.

In the midst of this controversy, in fact the day after Ms. Norhakah began as the district analyst in November, 2006, Ms. Jordasen retired and Ms. Norhakah was solicited

to take the interim directorship under the supervision of Dr. Wheaton, associate superintendent of secondary schools. In the spring of 2007, Ms. Norhakah was named director of research, evaluation, and assessment. Continuing the pattern of K-12 teaching capability, Ms. Norhakah brought experience as a classroom math and science teacher serving at elementary, middle school, and high school; an instructional coach; a school-level administrator; and an achievement analyst into the role.

Reflecting on the people who had lead the department over the years, Dr. Mostenby conjectured that “[he didn’t] think anyone in that office has ever lost the focus of who their true person was to serve or how to do it” (Mostenby, interview, August 10, 2015). Also like her many of her predecessors, Ms. Norhakah served on numerous state-level committees and advisory boards and maintained strong connections with the state assessment group comprised of colleagues from other districts, the state university, corporate data users, and data support personnel.

The discord between and within schools and central departments was quickly dispelled with explanation and collaboration, as well as demonstrated value of the department through services and support provided, except for the dispute with the elementary curriculum department. Despite substantial research and rationale as to why the elementary analysts should stay connected to the research, evaluation, and assessment department and aligned with the other achievement analysts, Dr. Wheaton, the secondary associate superintendent, decided it was not worth the fight and conceded supervision of the elementary analysts to the elementary curriculum director under Dr. Villard, creating an estrangement between departments and the analysts at the elementary and secondary levels that lasted as long as the analyst positions, which were eventually victim to further

budget cuts. Dr. Florence, director of elementary curriculum at the time, put the impact of staffing into words, saying “I think money has been a huge thing – I mean staffing possibilities (and limitations) has been a huge factor in the way this whole thing has evolved, [referring to evaluation capacity building within the district]” (Florence, interviews, July 8, 2015).

Part of the charge given Ms. Norhakah through the posted director job description was to create a process for research requests and conducting research within the district. Another aspect of the job expectation was to examine and evaluate the survey administrations in the district and to develop a process for data collection support. The need for an updated survey tool was identified as part of the survey analysis so in 2007, the existing system was upgraded and processes for both data collection support and managing research requests were founded, communicated, and implemented. In addition, more direct training and support related to goal setting, types of data sets available, interpretation, and appropriate use of different data sets and school improvement planning and monitoring were instituted.

The research, evaluation, and assessment department, after analyzing the district survey condition, identified unmet needs based on data from the state student survey and state mandates for garnering feedback from graduates about their educational experience. In response to those needs, the department developed two annual surveys: a ninth grade transition survey, and a senior exit survey. These data provided contextual information to high schools and the district that supplemented the demographic, achievement, and program data as part of ongoing self-review and improvement processes.

Administering additional surveys created another dilemma, however. The department was already over-extended in terms of available staff compared to work responsibilities. Ms. Norhakah, recognizing the discrepancies in necessary departmental size, presented a convincing case for an added staff member, and an educational data coordinator was posted and hired. This person assumed responsibility for all survey input, coordination, and distribution, along with initial reporting of results among other duties related to supporting and capturing district-wide common assessment data.

Also in 2007, at the state level, an education council convened by the governor released a report, outlining academic success and systemic improvement indicators that were meant to serve as evidence of student preparedness for global competition in the 21st century with a focus on readiness for both post-secondary and workforce opportunities. Whereas this plan was not legislated and, therefore, not widely adopted, it was a catalyst for more in-depth attention to college and career readiness, renewed interest in students' course-taking patterns, and reenergized attentiveness toward increased graduation rates.

This report was a precursor to 2013 state legislation enacted intended to create the *world's best workforce* as determined by five prescribed indicators of success: kindergarten readiness, third grade literacy, college and career readiness, achievement gap minimization, and graduation rates. As part of the legislation, an annual public hearing and interaction with the school board, along with a summary report to the state were mandated, increasing the collaborative ties between the research, evaluation, and testing department, many other departments across the organization, and school personnel in an effort to bring together a coordinated instructional plan for improvement of these indicators.

The district endured a watershed moment in 2008 when over sixty percent of the district school-level and central leadership turned over. Principals at all school levels, central office directors, all three associate superintendents (elementary, secondary and central operational oversight), and the superintendent were part of the transition, leaving the district's organizational memory almost non-existent. In December 2008, Dr. Mostenby retired, and Mr. Cadence, the former associate superintendent with central operational oversight, came out of retirement and accepted the superintendent position.

Superintendent Cadence's Era (January, 2009 – June, 2014)

Reflecting on the importance of leadership and the experiences of his predecessors, Mr. Cadence shared:

Well, certainly the superintendent changes [things]. I mean, what Spawling established really saw us through a period of immense growth and [he] established, I thought, a real strong staff in curriculum and research, evaluation and testing. Then, it got largely dismantled by Osborne. And then, I think Mostenby, having been part of the administration with Spawling, really wanted the days of Eickner and [some of the curriculum administrative consultants] back but he couldn't do it. I mean, you just didn't have the money and you didn't have the [school] board support to pull it off – to add a lot of administration. (Cadence, interview, July 12, 2015)

Even given his awareness of the importance of his role and a reflective nature, Superintendent Cadence's tenure was riddled with dysfunction and overwhelming challenges from the start; Nonetheless, progress was made in many arenas by the end of his employment.

With the retirement of all three associates the year before, three new, first-time central office administrators accepted their positions – one with elementary oversight, one with secondary oversight, and one with K-12 central departmental oversight, all with promising potential, having served as school principals for years prior and all, seemingly,

extremely capable leaders. Within the first few months of their service, two of the three associate superintendents (the one aligned to secondary and the one aligned to K-12), along with Mr. Cadence, jumped into the candidates' pool for the open superintendent's position upon Dr. Mostenby's announced retirement in early fall, 2008. Tensions grew among the associates when only one of them, the K-12 associate superintendent, surfaced as a co-finalist with Mr. Cadence, producing a downward-spiraling whirlpool of animosity and opposition amongst the associate superintendents that persisted until one of them left the district two years later. The acrimony that existed at the associate level permeated the entire central office, creating a difficult, even hostile, work environment for that two-year period. Remnants of the strong silos around levels remained, however. Statements like *from a high school perspective* or *from an elementary perspective* became commonplace, and staff were very careful to not overstep their realm of oversight. Dr. Bennett explained that during this time frame, "the [research, evaluation and testing] department actually did a lot of workarounds ... to try to move things in a positive direction because leadership wasn't always receptive to having research [and] evaluation. [They were] seen as a threat in some cases as opposed to a partner and a help" (Bennett, interview, July 2, 2015).

The exit of the secondary associate superintendent opened a revolving door of central leadership changes at the associate and director levels, spinning for nearly all of Mr. Cadence's time as superintendent. In addition, frequent budget reductions and ongoing restructuring of the organizational chart, especially pertaining to top leadership, became almost routine, adding to the tension and anxiety throughout the central office. The elementary/secondary/ K-12 central-departmental-associate-superintendent model

was shifted to an elementary/ middle school/high school-associate-superintendent model with each assigned specific central departments to supervise. The next shift included an addition of a chief academic officer, joining the existing three associate superintendent positions assigned to levels. Many of these positions were held by multiple people joining and leaving the organization over the years of Mr. Cadence's district oversight.

Even as new positions were being added, central office ToSA positions were being reassigned or eliminated following the school boards mantra of keeping the cuts away from the classroom, which in actuality meant not cutting teachers in the classroom but central support staff and programs could definitely fall victim in the process of reductions. For the research, evaluation and testing department and other central departments alike, this meant that many times, highly-trained staff would be reassigned to other roles or eliminated altogether. Florence acknowledged the impact of leadership changes, expounding on the notion that change is disruptive to evaluation and ECB efforts. "It's been kind of hit-and-miss because there's ... changes in leadership, changes to focus, and that makes it really hard to maintain focus on a program" (Florence, interview, July 8, 2015).

Out of adversity, however, sometimes stems opportunity. The departmental leaders who were supervised by the K-12 associate superintendent, Dr. Bennett, were prohibited from interacting with the secondary curriculum director, any of the secondary ToSAs, or the secondary principals (outside of scheduled principal meetings which Ms. Norhakah was still allowed to attend), without seeking permission and being granted approval from the secondary associate superintendent. These circumstances afforded Ms. Norhakah the gifts of time by freeing up calendar events, by eliminating meeting

obligations, increasing accessibility, and the convenience to work more closely with other central departments that were supervised by Dr. Bennett to plant evaluative-thinking seeds and work to grow evaluation capacity. One such person was the recently-hired director of special education, who was given the directive upon her employment to find a way to evaluate special education programming.

In 2009, the federal government allocated lump sum payments to districts based on criteria related to district demographics. Two-County's intentionally decided to try to make lasting impacts as opposed to quick-fix, stop-gaps like some other districts around the state and nation were doing. Capitalizing on the one-time federal funds provided through the American Recovery and Reinvestment Act (2009) and the fact that Ms. Norhakah had entered a doctoral program at the state land grant university in the area of evaluation studies, the district contracted the services of Dr. Kincade for a two-year period to serve as collaborator and coach for developing the district's philosophy of program evaluation, a formal cyclical model for program evaluation (See Appendix E.) and planned approach for evaluation capacity building, including an intentional learning component as every interaction with stakeholders as part of the evaluation process. Partnering with special education was the perfect opportunity to implement the development of this process, beginning an ongoing cycle of program evaluation that continues to be active in some of their programs even today. (Other programs were beleaguered by leadership turnover; yet in some of these the partnerships have been rebuilt and progress made and in others, the process will need to begin again.)

In addition to obtaining the assistance of Dr. Kincade, funds were dedicated to add two staff members to the research, evaluation, and assessment department (ToSAs

aligned to evaluation support) and expanding data visualization and reporting through data dashboard displays for principals and central administrators. Program evaluation capacity and efficient data accessibility were needed more than ever, as the nation and state were in the midst of a recession, and resources continued to diminish in large part due to declining enrollment and stagnant state funding, which created the urgency for the district to ensure that activities and programming produced meaningful outcomes for students and families. Bennett verbalized her thoughts on the connection between limited resources and evaluation saying:

What [ultimately] brought forward the importance of really being sure that what you were investing in with limited resources was really going to give you the best outcomes for kids...was no money coupled with accountability, accountability, accountability for all student groups. That leveraged some of the need for program evaluation. It also forced the hand around the operation side [of the organization] but because the operation side was not being measured, I would offer, not a lot of things went into that in terms of evaluation [at that time].
(Bennett, interviews, July 2, 2015)

With the funding incentive to surge forward in terms of evaluation capacity came the understanding that the dollars would be exhausted after two years and the ToSA positions, along with the benefit of working closely with Dr. Kincade again, would vanish, intensifying the desire to make a difference through the funding. The initial outcomes of this partnership resulted in the first formal plan for evaluation to the school board since that of Dr. Kincade in 2000 and began a frequent, ongoing interaction between Ms. Norhakah (representing the research, evaluation, and assessment department) and the school board.

Gillian, in speaking about another area of impact for the ARRA dollars acknowledges:

The ARRA money was important, along with the strategic investments because part of those strategic investments we made were supported with stimulus dollars. So I think those were important in giving us the thought that we could do a little more than just keep supporting – we could actually begin improving rather than just maintaining. (Gillian, interviews, July 10, 2015)

The strategic investments were aligned to initiatives identified as catalysts of systems improvement based on data at the cabinet level of the organization.

With the strategic investments, strategic partnerships with external evaluators continued, especially related to contentious or political programs. For instance, as part of the strategic investments, funds were aligned to improving the K-12 mathematics program. Underlying this seemingly innocent line item, however, was the fact that there was disagreement between the school board and some elementary leaders as to which program was most advantageous for students. To gain multiple perspectives and greater perceived objectivity, three outside evaluators were hired to perform an evaluation, one of elementary mathematics, one of secondary mathematics and one of K-12 mathematics.

Beyond the immediate dysfunction in the central office, one of the challenges facing Two-County ISD was declining enrollment, impacted by changing demographics in the communities within the district boundaries. What were once vital homes of young families with school-aged children were now homes of aging adults with no children living there. Acknowledging the harsh reality of what faced the district, Cadence bluntly said, “Money data has always probably driven this joint, [along with] the added enrollment data, because it’s at the basis of everything” (Cadence, interview, July 2, 2015). Even though the district routinely sought and acquired grants to supplement funding, budget declines were a consistent part of the landscape in Two-County ISD.

The district enacted the familiar study committee structure that had been used across the years, but instead of studying curriculum content, the committee was charged with identifying and studying issues facing the district. Although the structure had been used before, the committee was different in that community members served as chair-people, with district personnel among the membership, and an internal district leadership contingent serving as support to the process. Another difference was that the meetings were used as a forum to bring in experts on certain topics, sometimes internal to the district, sometimes external, possibly from other state or private industries, to inform the committee members about the topic at hand. The committee culminated in recommendations to the board related to their charge. This committee was the first of its kind, but brought forth two others in the series.

Budding from the findings and recommendations from the first committee, a second such committee was formed to craft recommendations to the board about the process to *right size* district facilities by closing schools in order to match the decrease in enrollment that had been occurring throughout the previous years. Again, an internal leadership team worked in tandem with the committee, sharing pertinent gathered information and independently identifying solutions with the intent of working toward consistency in conclusions across stakeholder groups. The third committee in this series centered on identifying long-range solutions for the future of the district, but focused mainly on budget implications from the ongoing enrollment decline instead of programming. School closings were not the only outcome of declining enrollment; central office reductions continued to ensue.

Perhaps because of the continual turmoil that plagued Mr. Cadence's tenure as superintendent or perhaps despite it, a primary tenet of his superintendency was to evolve the district from a *ma and pop*, casual-running district to a district operated more like a large corporation, with increased process and practice documentation, formal strategic and long-range planning and investing, identified performance indicators for the district as a whole as well as for central departments, and regular updates and formal reports to the school board based on data pertaining to these indicators.

Although always a data user himself, Mr. Cadence recognized that the time of being able to get by with not using data were long part of the past for all stakeholders in the system and everyone, not just central office, needed to overtly use data for decision making. Mr. Cadence explained,

I think the biggest difference is internal staff now doing the actual research here; [gathering the data, analyzing it, creating meaningful reports]. That's a huge change from anything I recall. And so those kinds of things are milestones within the district. In terms of evaluation and stuff, in the past it was kind of off in the corner in many, many ways. It was almost like you'd go see the doctor. You'd go see the doctor and you'd get an evaluation of how you're doing. Then, you wouldn't see the doctor for a year. I think that's a fair analogy.

And so now, it's like having the Mayo Clinic at your side. You not only get the one doctor, you get this team of people that you have access to and then all this data, immediately at your fingertips. You go in the morning and get your blood work done; I can read it online by the afternoon and it says exactly that you're tracking these various elements – and it's almost instantaneous. What happens in the morning can give you the information that lets you say, 'we need to do this or that' by the afternoon. I think education has done exactly the same thing but that's in contrast to taking an annual visit to the doctor, which we did for all of at least two of those decades – that's all we did. (Cadence, interview, July 2, 2015)

Data became even more central to conversations, processes, and decisions throughout the district (although not all stakeholders recognized or trusted that decisions were made based on data or that the “right” data were used). Soon after the scrutiny began, Mr. Cadence came to work very closely with the research, evaluation, and assessment department and relied on defensible data as a foundational fortress on which to stand in the times of calm and in times of turbulence.

Data, more than ever before, were the tools sought to dispel misconceptions and inaccuracies arising from the challenges or uninformed constituencies, both internally and externally positioned. The alignment between the school board and district leadership of their beliefs and expectations around data use served as a great impetus for a stronger culture of evaluative activities and the transcending use of data through the system. Cadence expounded the need for just-in-time data, adding context and rationale for district actions and communications to a wide range of stakeholders. Knowing the research, evaluation and testing department had the capability to support the superintendent’s data requests was invaluable; Cadence stressed, “[It’s] very, very important; having [Ms. Norhakah] say, ‘I can give you that data – we can get that’ – it’s just huge” (Cadence, interview, July 2, 2015).

Another action of Mr. Cadence was to transition the cabinet from an advisory body to a collective decision-making team. In doing so, the distinct delineation between the operational side of the organization and the instructional side of the organization, as well as the divides among associates aligned to school levels, slowly started to diminish. Along with this, all central departments began to contemplate and articulate how their

departments and roles supported the mission and vision of the district – conceptualizing potential connections from their work to educational outcomes for students.

One of the cabinet decisions was to hire a consultant to work with central leaders and the school board on effective governance and management structure, interaction, and processes of operation. The consultant introduced tools for processing and reporting information, among them a matrix of criteria and standards, organized in ranges representing a continuum of performance levels. The school board and central leadership, through a collaborative process, determined the indicators of a successful district as well as the corresponding levels of performance associated with a need to intervene in the identified area, a need to closely monitor the identified area (as it was in jeopardy of unacceptable performance), a level of acceptable performance in the identified area, a level progressing toward vision in the identified area, and a performance at a level of vision in the identified area.

Ms. Norhakah and her team facilitated the finalization of the scorecard and began an annual population of the tool, presented to the school board each year with supporting documentation and visualization. This presentation soon developed into a formal annual achievement report to the school board aligned to the identified key performance indicators in the scorecard. In subsequent years, the report continued to evolve, incorporating both achievement and perception data from a wide variety of measures as a means of district-level progress monitoring aligned to a philosophy of continuous improvement. Ms. Norhakah, in an effort to align district reporting tools, applied the scorecard matrix concept to the program evaluation model and incorporated a similar

reporting tool capturing the stakeholder-generated criteria and standards, with ranges of performance, for each program engaged in the evaluation cycle.

The scorecard process illuminated the need for additional data collection tools to measure identified indicators of effectiveness that were not currently being monitored. Two of these examples were the creation of a student engagement survey and an employee engagement survey. These surveys were developed internally by the research, evaluation, and assessment department based on a review of existing research, shared with a wide spectrum of stakeholders for input, piloted, implemented and administered annually.

Mr. Cadence also wanted each department to develop scorecards of key performance indicators related to the functions of the division. This provided the opportunity for Ms. Norhakah to build inroads of support within each central department, creating further opportunities for collaboration and evaluation capacity building, throughout the organization, beyond the natural and traditional connections with the curriculum department.

Another process implemented by Mr. Cadence was the directive that he himself, all cabinet members, principals, and eventually directors, would annually set and monitor goals related to the district strategic plan, school board initiatives, individual professional objectives, and other identified key performance indicators. As part of this professional growth and accountability structure, Mr. Cadence worked with the school board to determine his own goals for the year, then worked with the cabinet members on their goals, and instructed the cabinet members to work with the leadership in each of their areas of supervision to set their goals. Ideally, individuals' goals aligned to the

supervisors' goals to create systemic alignment and improvement. As part of this process, Ms. Norhakah consulted with the superintendent and many cabinet members on effective goal writing and appropriate data for monitoring and measuring the goals, as well as with the school board to prepare a formal report depicting the superintendent's performance relative to the goals that were set.

An ongoing avenue for collaboration and interaction was formed among Ms. Norhakah, the superintendent, and the members of the school board during this time to provide means to meet each party's data needs. By working closely with the superintendent and the school board, Ms. Norhakah could anticipate the types of information that would be helpful in making decisions and building contextual understanding and capacity to think evaluatively and use data effectively. Many transformative reports were created from this connection, as well as a deeper understanding of and appreciation for the value of evaluation and data. Cadence, in speaking directly about Two-County's evaluation efforts and experiences suggests this:

It seems to me that [research, evaluation and testing] at one time only was involved with the curriculum – only assessment – didn't really evaluate anything else; didn't stray out of K-12...all within content areas; all within state testing; a lot of it motivated by 'how are we comparing to everybody else?' – so little innovation, if any, little emerging issues – in other words, established issues that you've got to react to. What I liked about [Ms. Norhakah's] work on the achievement gaps, certainly early college, some of the other things is that those weren't even defined issues by somebody else. They were issues that we defined as important to what we're trying to market as far as this district goes and react to. But who knew those were going to be sweet spots? So I think that is a huge change from just what I would say, the difference being reactive for about 25 years of the three decades [of this study] and then being proactive in about the last five...I think [the other superintendents before me] were outwardly-driven rather than inwardly-driven. (Cadence, interview, July 2, 2015)

Outwardly-driven was in reference to reacting to external mandates, legislative accountability, and public pressures. *Inwardly-driven*, in contrast, referenced proactively seeking data and information related to organizationally identified areas of need, spurred by questions arising from data analysis and evaluation deemed important and necessary by central and building leadership.

Cadence also recognized the added pressures on educational evaluators as expectations grew, expounding that “the accountability in all these roles has really ratcheted up, where you could be average or good and do pretty well [in the past]. And now, you’ve got to be really good to do these positions. And if you [aren’t], you don’t make it. You might make a year or two, but that’s all you’re going to make. You can’t keep people that aren’t functioning at high levels” (Cadence, interview, July 2, 2015).

To increase opportunity for discussion and collaboration with the school board on improvement strategies, key issues, and important district and community topics, Mr. Cadence and his cabinet devised a plan that was approved by the school board to change the format of one of the two regular monthly school board meetings to a work session, keeping the governing members and district leadership more informed and aligned pertaining to the actions of the district. In addition, a structure of sub-committees about specific topics such as curriculum department initiatives, scorecard revision review, labor/management concerns, legislative priorities, etc., where portions of the school board were appointed to different committees, was instated, allowing for deeper study and insight for the affiliated school board representatives on the various topics.

The decision to restore curriculum studies after their long hiatus sprang from these meetings, and a specific process was detailed. The process included stakeholder

involvement, including data collection for input and materials review similar to before, however, albeit more formally documented, overall, the process was not as in-depth as the first generation studies, but did include more frequent communication and interaction with the school board throughout the study.

In 2010-2011, the district again experienced cataclysmic circumstances in somewhat of a perfect storm scenario. As part of the transformation into a contemporary organization, the district began a review of existing policies. One of the policies relating to curricular expectations was extremely out-of-date, which prompted a revision for the school board. The first and subsequent revisions were worded poorly, causing misinterpretation amongst staff, families, and community members, which spurred tremendous outcry, amplified by the contextual environment in which it was brought to the school board. The policy was entitled the *neutrality* policy in its original form. It next became known as the *controversial topics* policy and was finalized with the title of the *safe and respectful learning environments* policy.

Separate from the policy review, tragedy struck the district in the form of student suicides. The students and families directly part of the heartbreaking events, as well as others involved in and affected by the happenings, were immersed in feelings of oppression, bullying, and non-acceptance, feelings that continued over a period of a few years. These events created a lens through which the revised policy was viewed. What ensued was national and even international revulsion and disgust from a linking of these events, resulting in terrible misunderstandings and assumptions. The upheaval made unimaginable and unbearable circumstances catastrophically more divisive and

adversarial, culminating in a lawsuit and an eventual consent decree between the district and the federal government related to bullying and harassment education and prevention.

Again, out of these abhorrent circumstances, value could come. A leadership team that was already formed to assess and analyze the culture and climate of schools and the district, especially in regards to the presence of bullying and harassment, intensified its focus. A parallel community task force was formed to examine these aspects through a broader array of stakeholder perceptions including students, parents, staff and community members. The task force mirrored the depth and level of involvement of the new community study committee model used to study district issues, facility use, and future directional focus alternatives; an uncommon model amongst school districts (Gillian, interview, July 10, 2015). In addition, the program evaluation model was applied to help develop participant-generated indicators of effective anti-bullying and anti-harassment programming for staff, students, families, and community members, again, providing the platform for evaluation capacity building.

In April, 2014, Mr. Cadence implemented a change in the cabinet structure, creating and adding top management level positions on the operational side of the organization to the cabinet while shifting other department leaders off. At the same time, the central office was significantly downsized, returning numerous ToSAs to classroom positions and leaving a skeleton crew to support data use and evaluation from the central level. In June, 2014, Mr. Cadence resigned as superintendent, and Mr. Langley was appointed his replacement.

Superintendent Langley's Era (July, 2014 – present)

Upon first glance, Mr. Langley's first year as superintendent seemed to follow the pattern of turmoil that beleaguered Mr. Cadence's time in the role. The organizational changes made just prior Mr. Cadence's departure left the central office in chaos in many ways with undefined roles for new and existing positions, far fewer central staff to assist schools and accomplish the necessary tasks of each department, and a general lack of clarity as to how things were going to work. Feelings of frustration and being overwhelmed were communal, yet nobody knew where to turn for support. Along with the disorientation at the central level, Superintendent Langley, within the first six months in the role, was faced with the sudden, unexpected death of a long-time school board member that sent shockwaves throughout the community.

Frustration and tension ran high. Staff conflict ensued, and friction was consistently present, somewhat reminiscent of the start of the previous superintendent transition, yet there was a sense of hope and vitality in the district's new leader. Many people in the district and community knew Mr. Langley from prior interactions with him as a student, teacher, school administrator, and community member in the district. Questions and concerns about how the district would operate under Mr. Langley were non-existent or quickly dismissed as staff, students, families, and community members had opportunities to interact with him, instilling a feeling that the district would be all right and the central office discord would be acknowledged and addressed.

Mr. Langley professed and displayed an understanding and appreciation for data and evaluation and quickly sought information through which to understand context, ground decisions, and inform strategic priorities. Mr. Langley was committed to aligning

district reporting and communication, working toward presenting and “messaging” things in a way that was meaningful and understandable to all stakeholders, including parents and community members. Langley explained, “Since the [department] sees so much data, they end up informing change as a natural byproduct. I referenced information from the department, I think, ten times in my back to school welcome – about hard data, about our system, parent perspectives, student perspectives, staff engagement, staff perspective” (Langley, interview, August 6, 2015).

Langley worked to clarify the alignment among the mission, vision, district strategic priorities, and organizational goals, identifying and mitigating any perceived disconnections and inconsistencies. He set forth to develop visuals to elucidate the overarching continuous improvement and decision-making process through a model that included explicit interactions with the school board to provide progress updates and gather input.

Mr. Langley held and professed the firm belief that the changes and improvements made were because “the data are telling us it’s a need, and therefore we want to change; it is not because some external force or someone in a higher level of authority is making us change” (Langley, interview, August 6, 2015). Ms. Norhakah maintained and even strengthened the link among the superintendent, school board members, and the research, evaluation, and assessment department as a part of and supporting these ongoing planning and development conversations, advocating for and demonstrating how evaluation services could be of great benefit to the overall organization.

Providing a depiction of the research, evaluation and testing department staff,

Langley offers:

They're responsive ... they're also constantly partnering on things like employee engagement or community surveys. Depending on the area that the district needs data, they're working with other departments to analyze the data, to determine good questions that will be statistically reliable – to analyze data and provide information. (Langley, interview, August 6, 2015)

Opportunities for expanded internal support for program evaluation continued to arise, stretching into collaboration with nearly all central departments providing an avenue to anticipate and assess staff data needs and to develop reports and processes to meet those needs. In addition, new state legislation aligned with creating the best workforce possible through effective programing and instructional systems to ensure each student's success brought further occasion to interact with other departments and provide leadership for evaluation, goal setting, and progress monitoring activities.

Ongoing and frequent reports to the school board continued and were augmented with individual data requests made by school board members to answer community questions, provide understanding or context for an upcoming or emergent district issue, or support background knowledge and implications for school board decisions. Ms.

Gillian described the research, evaluation and testing department as:

A huge change agent for the district because it has provided the information that has underpinned so many of the changes that have occurred. And I think the reason we've been so successful is because of those changes. I think that brought us from, you know, a pretty good district to a really good district given our demographics and our funding. (Gillian, interview, July 10, 2015)

In July, 2015, Mr. Langley made adjustments to the organizational structure once again, this time to address the issues created by the previous year's restructuring. Some of

the positions that were added that did not go as anticipated were cut, and positions that had been cut resulting in unintended outcomes were reinstated. Another important change for the work of evaluation capacity building and strengthening evaluative thinking throughout the organization was that Ms. Norhakah began reporting directly to the superintendent. As part of the shift, she met weekly with the Mr. Langley to discuss upcoming data and evaluation needs and how the research, evaluation, and assessment department could most effectively support the whole organization. Weekly meetings provided the forum to bring important information to the attention of the superintendent while also providing him the necessary data to most effectively perform his role. Increased awareness and visibility of the importance of evaluation and its necessity throughout the organization occurred through this structural change and the relationship built with district personnel. Recognition of the vital role that the research, evaluation, and assessment department played in the district and the significance of this work to all staff came through the approval of a second district analyst position in December, 2015, even in the midst of other budget cuts.

Overview of Departmental Progression

Two-County Independent School District (ISD) experienced many changes over the thirty-year period of study from 1985 – 2015. The department charged with the activities related to research, evaluation, and assessment was established and expanded over this timeframe. The name and reporting structure of the department changed, as did the types of activities on which it focused. Explanations of the primary findings of this study are below.

The department charged with evaluative activities was officially formed in 1986 under the name of the assessment department with the primary focus of student assessments. Activities related to evaluation prior to 1986 and beyond 1986 with the exception of the student assessment activities were conducted and managed through the department of planning and evaluation, which transitioned into the department of finance and planning. At this time, content-area studies were conducted through curriculum and only supported minimally through the assessment department. The assessment department was managed by the student assessment coordinator, who reported to the director of finance and planning. The department also contained 3 clerical staff, added in stages shortly after its inception. The student assessment department, at this time, had limited interaction with school staff, only meeting occasionally with individual principals by request to look at emerging district-wide common assessment data.

In 1992, the department became known as the student assessment department under the leadership of the coordinator of student assessment. This position reported to the associate superintendent of secondary schools and continued to focus primarily on student assessments and supporting content-area studies conducted through the curriculum department, expanding into the beginnings of support for using data to inform decisions. Three clerical roles remained in the department, as well. Interaction with school staff remained limited and was mainly with individual principals upon their request.

For the 1998-1999 school year, the student assessment department had no direct supervisory oversight. In 1999, the department, under the same name, gained a coordinator of research and evaluation position as oversight along with a student

assessment facilitator who was a teacher on special assignment (TOSA) role. The coordinator and the facilitator reported to the director of secondary curriculum, instruction, and assessment, however, they interacted regularly with the associate superintendent of secondary schools. With the departure of one staff member, the student assessment department was down to two clerical positions. Eventually, another person was added, an office coordinator, bringing the department membership to five. The department had regular meetings involving all staff for update and planning purposes. The primary functions of the department included evaluation and evaluation capacity building, testing, and data support (understanding, analyzing, interpreting and using data). There was more interaction with school leadership than before, but contact with schools directly was still minimal.

In 2003, the department name was changed to the research, evaluation, and testing department, and a director level position was established, filled by a principal on special assignment reporting to the associate superintendent of secondary schools. The department was in a period of expansion, adding six achievement analysts (ToSA positions), a technician, and a district analyst over a short period of time, bringing the department membership to thirteen. The department still had regular meetings, this time only including the achievement analysts and the department leader for check-in, update and planning purposes. The primary functions during this time were testing, data support, and initial phases of supporting school improvement planning. The department expansion included support of achievement analyst positions in the high schools, as well. Regular meetings were held with department staff and the achievement analysts, in addition to department meetings, as a means to collaborate, problem solve and implement building

level support around data use, interpretation, and application. Presentations and trainings were created in these meetings, as well, and shared across schools to support systemic training to all school-level staff. Individual school issues were also discussed and action plans were collaboratively generated to address identified challenges and needs.

In 2006, the first official director of research, evaluation, and testing was named, still reporting to the associate superintendent of secondary schools. The department's priorities were testing, research oversight, school improvement planning, and overall data support. The department continued to expand, adding an educational data coordinator, raising the total to fourteen, which was short-lived, however, when one achievement analyst position was eliminated and two others transitioned into the elementary curriculum, instruction, and assessment department, reverting the departmental membership to eleven. Monthly department meetings for all staff were held and processes, projects and priorities were the main topics. Even though two of the analysts were now supervised in elementary curriculum, school-level support continued through interaction with instructional coaches at the elementary and the analyst positions at the middle and high school levels. The district employed a kind of *train-the-trainer* model of planning and support through collaboration of the full group and then deployment of the analysts and coaches to interact directly with teachers and administrators in the schools. The training and planning loop with schools came full circle through check-ins, collectively at principal meetings and individually through personal conversations with each principal.

In 2009, with structural changes at the central office, the research, evaluation, and testing department began reporting to the associate superintendent of K-12 central

departments. The department added evaluation to its list of priorities and hired two additional ToSA positions to serve as evaluation specialists shortly thereafter. Monthly department meetings with all staff continued but the topics expanded to include such things as a department mission statement – who are we and what do we stand for, development of a strategic plan for the department, determination of criteria and standards by which to measure our effectiveness, goals for the department, and creation of process and communication documents with project point people assigned. Soon after, an intentional learning (professional development) component was added to every meeting. School-level support continued through the tiered train-the-trainer model, with analysts and coaches meeting with the research, evaluation, and testing department and then supporting teachers and administrators with school-level needs. Many common reports and extensive data analysis occurred to inform site decisions pertaining to a wide variety of topics such as interventions, student placement, areas of concern, and school improvement planning.

The next few years brought many additional structural and personnel changes with the department reporting to the associate superintendent of middle schools, then the associate superintendent of high schools, and ebbs and flows of the number of ToSA positions within the department due to budget cuts and realignment of staff to district priorities. Over this time, departmental priorities included evaluation, testing, data support, research oversight, school improvement and goal writing support, and data collection, monitoring, and reporting support for the full organization (not just the curriculum department), along with the initial stages of working with school board members and the superintendent to set goals to monitor and report superintendent

performance. In 2014, the departmental structure settled to include the director, office coordinator, district analyst, technician, educational data coordinator, an achievement analyst ToSA and two clerical positions. Over this time, the analyst positions at the middle and high schools were re-purposed to support common assessments, and supervision shifted to the school principals. With yet another central restructuring, the roles were re-purposed again to incorporate the data support and training aspects performed previously, and connection with the research, evaluation, and testing department was again established.

In addition, research, evaluation, and testing staff routinely were invited to co-present to school leadership teams concerning the school improvement process, understanding and interpreting data, and aligning grade-level and teacher goals to school and district goals, for instance. A train-the-trainer model was still employed with the school leadership team and the school-level achievement/instructional coaching positions training the full staff and serving as liaisons between individual staff questions and research, evaluation, and testing as a means of intentionally building their capacity, as well.

Other avenues for building evaluation capacity in school-level personnel were through sessions offered on district-wide staff development days and through mandatory training for all new teachers and administrators. This mandatory training, which included staffs that were new to the profession and new to the district, was on data accessibility, use, interpretation, and application, as well as strengthening and encouraging evaluative thinking. Staffs were led through a series of activities where different data were

introduced and discussed, providing experience with stages of evaluative thinking and data use.

The different stages of evaluative thinking that were included were first, raising awareness and developing the habit of asking questions about effectiveness and if strategies/programming are producing desired outcomes. Next, staffs were led to data to help answer the posed questions. Then, conflicting data were introduced and discussions occurred about which data to use, how to deal with disparate data, how to decide if data sources are trustworthy and collection methods are sound. Lastly, the group engaged in conversations about using multiple sources of data, increasing validity and reliability in data/collection methods, and where to access resources and support moving forward.

In 2015, the director of the research, evaluation, and testing department began reporting directly to the superintendent, maintaining and strengthening the departmental priorities, including an organization-wide focus. In addition, a second district analyst position joined the department bringing the membership total to nine. Evaluation capacity building efforts at the central and school levels were continuing to grow.

Table 7. *Timeline of Evaluation and Legislative Implications to ECB with Departmental Progression*

Timeline	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010 - 2015
Evaluation Developments	Era of Public Accountability					
	Emergence of Internal Evaluation Defining and Refining of Process Use Surfacing of Organizational Learning				Evaluation Capacity Building	
Department Evolution	Department Name	Assessment Department (est. 4/'86)	Student Assessment Department	Student Assessment Department	Research, Evaluation & Testing	Research, Evaluation & Testing
	Supervision of Department	Director of Finance & Planning	Associate Superintendent of Secondary	Director of Curriculum, Instruction & Assessment ('99-'03)	Associate Superintendent of Secondary (through '08), then Associate Superintendent of K-12	Associate Superintendent of High School ('10-'14) then Superintendent ('15)
	Department members	<ul style="list-style-type: none"> •Student Assessment Coordinator ('86-'92) •3 clerical staff (added in stages) 	<ul style="list-style-type: none"> •Coordinator of Student Assessment ('92-'98) •3 clerical staff (in 1999, there was no direct oversight of the dept.) 	<ul style="list-style-type: none"> •Coordinator of Research and Evaluation ('99-'01) •Student Assessment Facilitator ('99-'03) •3 clerical staff 	<ul style="list-style-type: none"> •Principal on Special Assignment ('03-'06) •Student Assessment Facilitator •Office Coordinator (added '02) •District Analyst (added '06) •6 ToSA Analysts •Technician •Data coordinator (added '08) •2 clerical staff 	<ul style="list-style-type: none"> •Director ('07-present) •Office Coordinator •District Analyst (added another - '15) •1-4 Achievement Analysts •0-2 Evaluation Specialists •Technician •Data Coordinator •2 clerical staff
U.S. Legislation	National Commission on Excellence in Education: A Nation at Risk (1983)	Reauthorization of ESEA to Improving America's Schools Act (1994) Goals 2000: Educate America Act (1994)	Reauthorization of ESEA to No Child Left Behind (2002)	America Recovery and Reinvestment Act (2009) Race to the Top Bill (2013 - Not enacted) Every Student Succeeds Act (12/2015)		

Tables 7 and 8 illustrate the thirty-year progression. The district and departmental findings were added to the evaluation developments and the federal legislation pertaining to education, used as grounding for events in the district.

Table 8. *Timeline of Evaluation and Legislative Implications to ECB – Departmental Processes, Important Events and Internal Influences*

Timeline		1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010 - 2015
Evaluation Developments	Era of Public Accountability						
	Emergence of Internal Evaluation Defining and Refining of Process Use Surfacing of Organizational Learning				Evaluation Capacity Building		
U.S. Legislation		National Commission on Excellence in Education: A Nation at Risk (1983)	Reauthorization of ESEA to Improving America's Schools Act (1994) Goals 2000: Educate America Act (1994)	Reauthorization of ESEA to No Child Left Behind (2002)	America Recovery and Reinvestment Act (2009) Race to the Top Bill (2013 - Not enacted) Every Student Succeeds Act (12/2015)		
District & Department Impacts	Department Leadership	<i>Spawling</i>	<i>Spawling/Hadley/Osborne</i>	Mostenby	Mostenby	Mostenby/Cadence	Cadence/Langley
		Mostenby	Mostenby/ <i>Wheaton</i>	<i>Ritenour/Wheaton</i>	<i>Wheaton</i>	<i>Wheaton/Bennett</i>	<i>Porter/Thomason/Langley</i>
		Mostenby/Eickner	Eickner/Dooren	Dooren/Kincade	Kincade/Anthony/Jordasen	Jordasen/Norhakah	Norhakah
	Internal Evaluation Processes	Content area studies (through curriculum) / Testing	Content area studies (through curriculum) / Testing	Curriculum studies/ Testing/ Evaluation/ Beginnings of others using data (beyond leaders)	Evaluation/ Testing/ Data support/ Curriculum studies	Evaluation/ Testing/ Data support/ Research oversight/ School improvement support	Evaluation/ Testing/ Data support/ Research oversight/ School improvement support/ Data collection support
Important Events	Supt attended a conference on ensuring readiness to graduate Effective Schools Initiative+ Planning, Evaluation and Reporting requirements Partnership with University on study of courses	Significant district enrollment growth Threat of teacher strike - Supt's letter to all staff Elimination of all central administrative content consultants - new organizational model Partnership with University on high school schedule	State Graduation Standards implementation - additional funding available Intentional focus on capacity building under Kincade Much of the evaluation done with external experts	Accountability Focus Compensatory Ed Grant Focus on data structures/technology - Student information system implementation Much of the evaluation done with external experts Threat of teacher strike	ARRA - additional funding available Focus on data accessibility and understandability for all stakeholders Community Involvement expands through more in-depth task force activity	Significant district enrollment decline - school closures Focus on internal evaluation & intentional capacity building Threat of teacher strike Community taskforce involvement - Governmental lawsuit	
Internal Influences	Mostenby's experience with the "Great Cities" schools work	Significant restructuring in Curriculum, Instruction and Assessment (downsizing)	Expansion of technology & network Technical College legislated to state	Significant focus on testing, accountability, internal outcome monitoring & data use	Significant leadership turnover throughout the district Curriculum studies ceased	Significant leadership turnover centrally Significant downsizing	

Adapted from Jorgensen and Hoffman (2003); Fitzpatrick, Sanders, and Worthen (2004); and Russ-Eft and Preskill (2009)

This portion of the timeline, Table 8, includes the evolution of the department.

Included in this portion of the table are the departmental leadership (depicting the superintendent, supervisor of the department head, and the direct department oversight),

departmental process priorities, important events, and internal district influences.

Important events were selected because they stood out as either halting the whole district in some way and/or altering the activities of the department.

Progression of Data Availability

Educational data were always available to the teacher for his or her class, dependent upon what that teacher recorded. These data were typically kept in paper format in a teacher's grade-book and likely included assignment and assessment scores and possible attendance records. The grade book was accessible to the teacher and could be viewed by individual students unless the teacher aggregated it across the whole class and it was kept in hard copy format with the school records for at least the summer months following each school year. Information on the students' past history was kept in cumulative files in the school offices where a teacher, if he or she so chose, could look things up pertaining to the student's past assessment performance, but the accuracy and completeness of the file was purely dependent on the previous year's teacher adding the data.

When the student information system (SIS) was purchased, teachers then had access to information in a systemic electronic format located on the district's network system so it was only accessible on-site. Grades, attendance, student demographic information and personal information was accessible to the teacher for his or her classes and to school level administrators for their schools. In addition, district level administrators could access each schools information, one school at a time. Even with access to multiple students, all users – teachers, school administrators and district administrators – could only access one student at a time. For the first time, parents could

access information on their child's attendance and schedules, along with other minimal information online through a parent portal. A student's history was still kept in paper cumulative files with all the same parameters as described above.

When the data warehouse was implemented, the systems of information were now integrated at the district level, available electronically and was web-based, meaning it was accessible anywhere there was internet access. Users with rights to multiple buildings could access them in the same system without having to log out and log back in to a different version of the application. All staff that had an educational need to know information about a student had access (based on the data security and role process) to that student's complete information with the exception of discipline (behavior) data and free-or-reduced-priced-services qualifications which were suppressed from some users. All the information from SIS was available in the warehouse reporting software with the addition of results from local, state and national assessments, program participation, extra-curricular participation, enrollment history, information on individual plans, information about students' past performance (eliminating the need for paper cumulative files) and even access to the incoming students once enrollment for the upcoming year was complete. Data in the warehouse software could be viewed by individual student, by class, by teacher, by grade level, by school or district-wide. Parents also had the rights to view a complete profile of their child related to the full range of data in the system.

Although student enrollment (the information on students enrolled in the district – entrances and exits) is updated nightly, data in the warehouse was not necessarily meant to be a daily use tool, but filled more of an analytic role, providing information for teachers to analyze the strengths and needs of their incoming students or the performance

of students after an assessment is administered, in other words, anytime you have new students or there is a new district-wide assessment given, new data are loaded into the warehouse.

The warehouse also provides predictive information (how a student is likely to perform on an upcoming assessment based on the results of a current assessment) and placement logic (recommending students for courses based on the data available in the system, allowing for teacher's to input comments if they feel the recommended placement is not indicative of the student's ability). The software has powerful analytical tools that assist in evaluation of programs that are used mainly at the administrator and central levels but are available to all users with the actually students included in the reporting based, again, on their security rights.

The data dashboards were incorporated into the district for school level and select central level administrators only for more just-in-time reporting and were also web-based applications. The dashboards contain all the same data as the warehouse with the addition of teacher absence data, budget data and alerts. The alerts provide information to administrators about which students or classes are not meeting designated benchmarks.

Table 9. *Progression of Available Data in Two-County ISD by Data System*

Data System	Teacher Grade-Book/ Record-Keeping	Student Information System	Data Warehouse	Data Dashboards
Data Available	Scores, Grades, anything captured by teacher in record-keeping	Attendance, Some Test Data, Discipline, Schedule, Course Marks, Demographics, Enrollment History	All SIS data, Achievement Data - national, state & local tests, Health Information, Program Information - interventions or special programs, Extra-Curricular Activities, Predictive tables, Placement Reports	All Data Warehouse data, staff attendance, budget data, alerts for data sets not meeting set benchmarks, progress marks
Format	Paper	Electronic - network based initially, now web based	Electronic - web based	Electronic - web based
Unit of Accessibility	By Student, Classroom	By Student	By Student, Classroom, Grade Level, School, District-wide	Same as Warehouse
Stakeholder Access	Individual teacher - or anyone with access to the location of the paper grade-book	Teacher, Administrators, Some Parental Portal Access - depending on security rights	All Staff and Parents, depending on security rights	Principals, Assistant Principals, Select Central Administrators
Level of Integration	Individual teacher	School-wide	District-wide	District-wide
Location of Student History	Paper Cumulative File that traveled with student from school to school	Same as grade-book plus some in SIS	In Warehouse with other data as an individual student summary page	In Dashboard - linked to report selection and enrollment

Overarching Findings: Document Review

School board agendas/minutes. In analyzing the 753 school board agendas and minutes from January, 1985 – December, 2015, there is evidence of evaluation activity over the full time frame. (See rubric in Appendix C.) Even so, the predominance of activities found to have moderate or strong evaluation connection were throughout the Spawling, Cadence, and Langley eras, with the greatest concentration in the later years of the study period. (See Figure 5.)

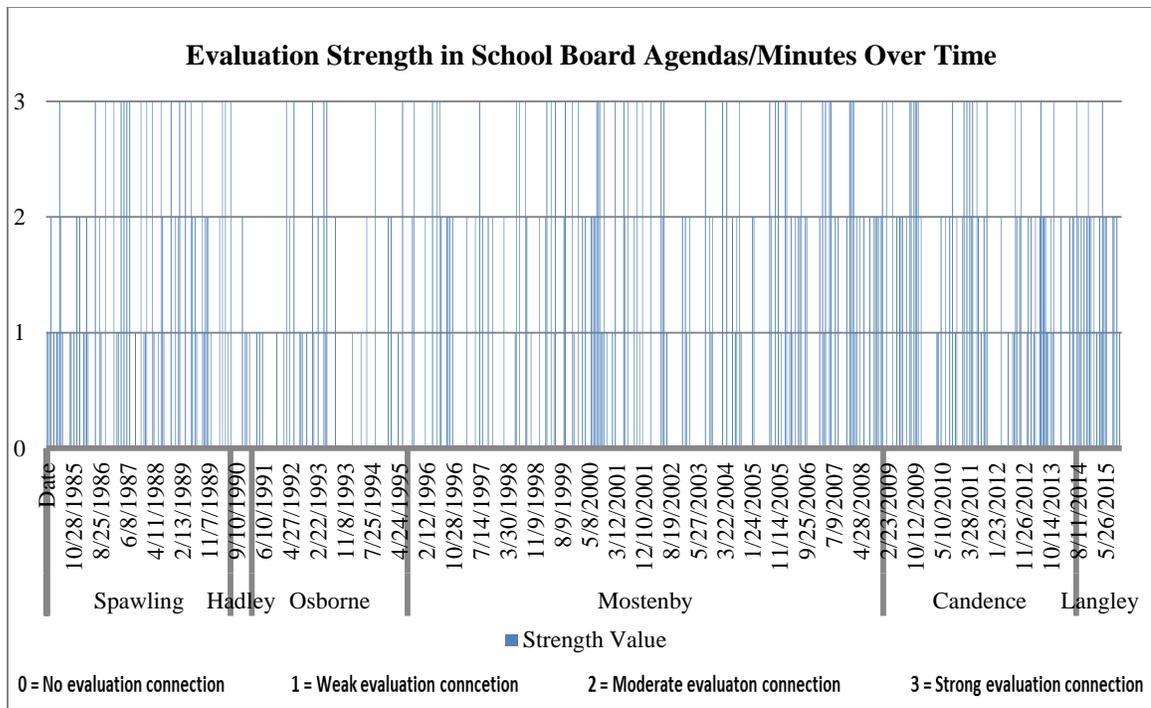


Figure 5. *Evaluation strength in school board agendas/minutes over time.*

There was also a transition of the types of activities across time. The activities reflected in the agendas and minutes in the early years that were deemed to have strong evaluation strength were almost exclusively curriculum-related content area studies, conducted by internal and external evaluation teams, and community committees focused mainly on boundary change decisions. As time passed, the types of studies that were conducted were connected to all segments of the organization, aligned to both instructional and operational departments. There was the lowest percentage of events with evaluation connection during Mr. Osborne’s era and the highest percentage during the Langley era. There is a pattern of decline in the activities found in the agendas/minutes with weak, moderate, or strong evaluation strength (aggregated as showing any level of evaluation connection) between the Spawling and Osborne eras, followed by a pattern of growth between the Osborne and Langley eras. (See Figure 6.)

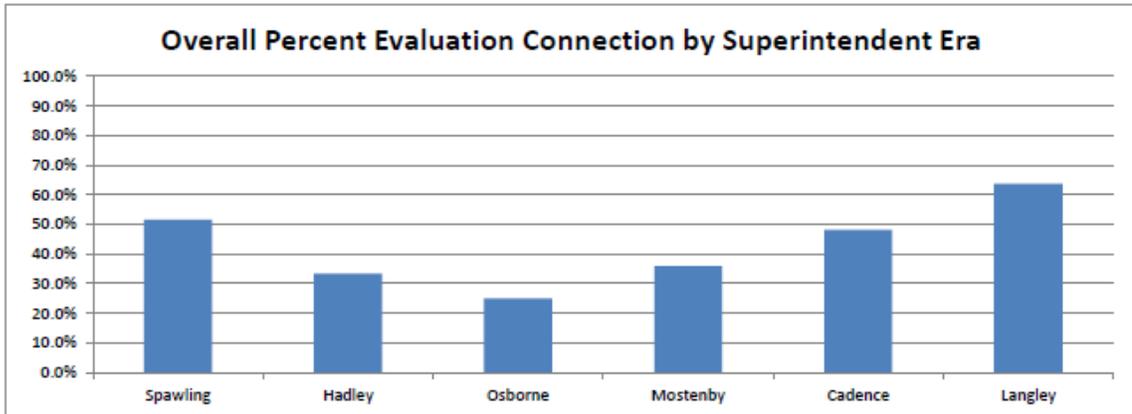


Figure 6. Overall percentage of evaluation connection by superintendent era.

The percentage of activities recorded in the school board agenda/minutes having no connection to evaluation, weak connection, moderate connection, or strong connection grouped by superintendent era (Figure 7) and grouped by connection strength (Figure 8) are illustrated below. These figures show that Osborne’s era contained the prevalence of no connection strength in the agendas/minutes while the Spawling and Hadley eras were nearly tied with the percentage of recorded activities with weak evaluation connection. The Langley era contained the greatest percentage of activities with moderate connection strength, and Spawling’s era had the highest percentage of activities with a strong connection. (Further information in Appendix F.)

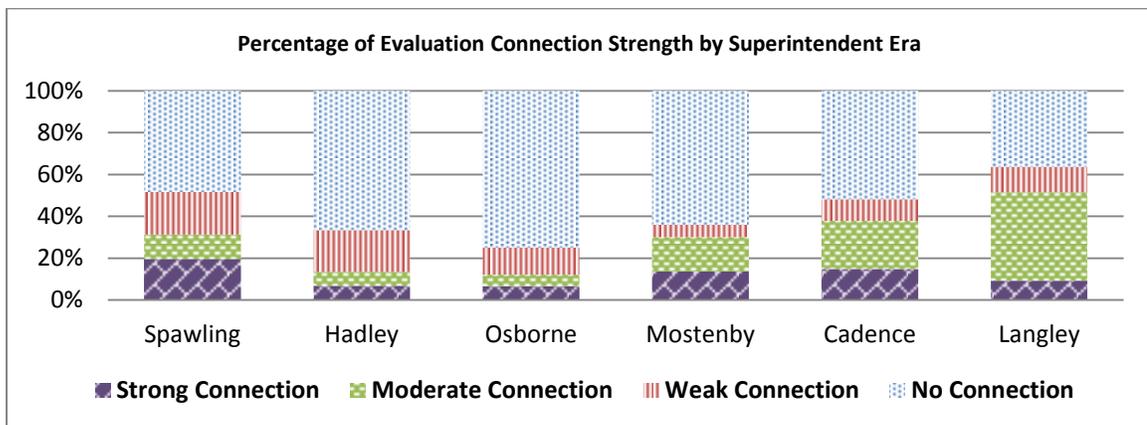


Figure 7. Percentage of evaluation connection strength by superintendent era

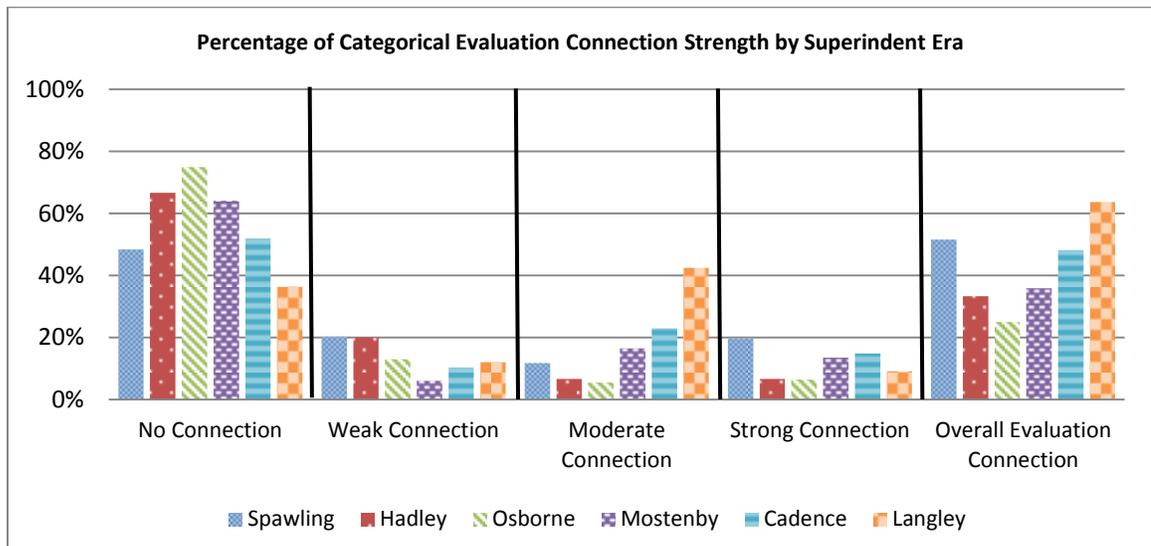


Figure 8. *Percentage of categorical evaluation connection by superintendent era.*

Focus on Two-County community newsletters. Evaluation activity was evident in the 119 district newsletters that were distributed to all community members between autumn, 1984 and fall, 2015, over the study time frame. The publications were typically produced and distributed four times per year, however, that varied some with special publications or other anomalies. Nearly all publications had some level of evaluation connection with the majority rated as weak or moderately connected as measured using the same rubric as for the school board agendas/minutes. (Appendix C.)

The newsletters primarily contained information about the district’s performance on state tests, explanations of community committee findings, or survey results. Nearly all *superintendent columns*, a section in each publication, contained evidence of some level of evaluation connection. Outside of Mr. Hadley’s era, when very few newsletters were published given his six-month interim status, Mr. Osborne’s era contained the lowest percentage of evaluation connection evidence, mirroring the findings from the

school board agendas/minutes. (See Figure 9.) Again, discounting the Hadley and Langley eras due to the low publication counts from their short tenures, the range of publications with either weak, moderate, or strong connection to evaluation ranged from approximately 71% to 94% for the other superintendents' eras. The Spawling and Mostenby eras were the only two with strong evaluation connection, mainly pertaining to content-area study information or community study committee reports. (Additional information in Appendix G.)

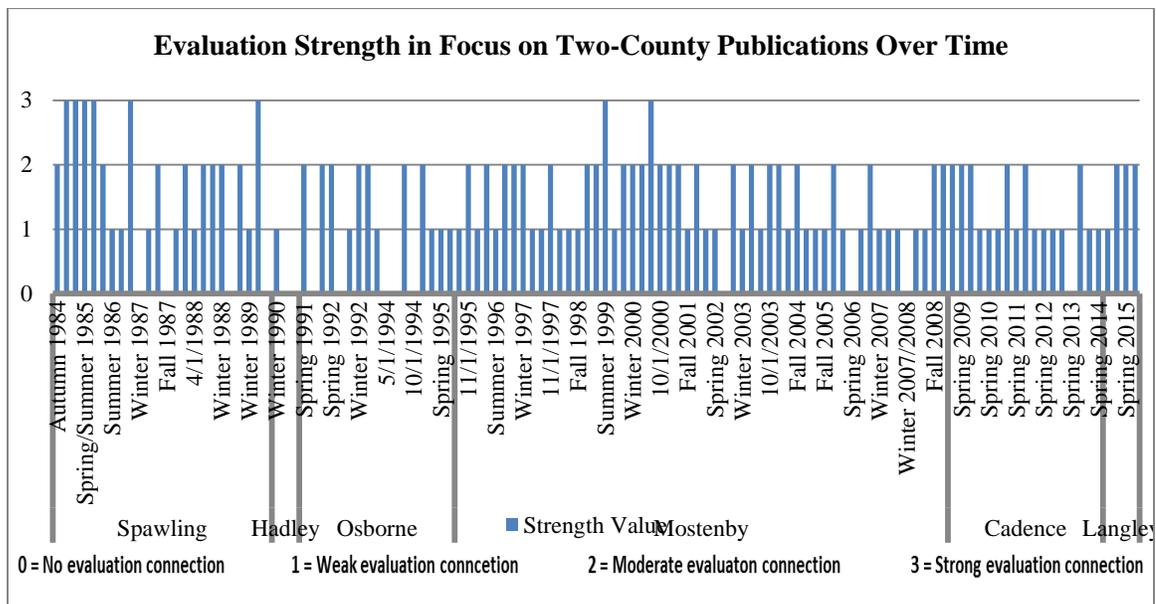


Figure 9. Evaluation strength in Focus on Two-County publications over time.

Ms. Norhakah’s calendar events. Five years of calendar events (from 2011 – 2015) were analyzed, along with a three-month period in 2007 as a basis for comparison from early in Ms. Norhakah’s departmental oversight. Calendar events were coded as having evaluation connection or not, then further delineated by the type of evaluation activity – data use/support, testing, evaluation, or research events. Sixty-two months of events were analyzed, resulting in a total of 3570 meeting events, with a total of 67% of

the overall events related to evaluation-related activities. There was an upward trend of evaluation-related activities over the five-year time frame. (See Figure 10.)

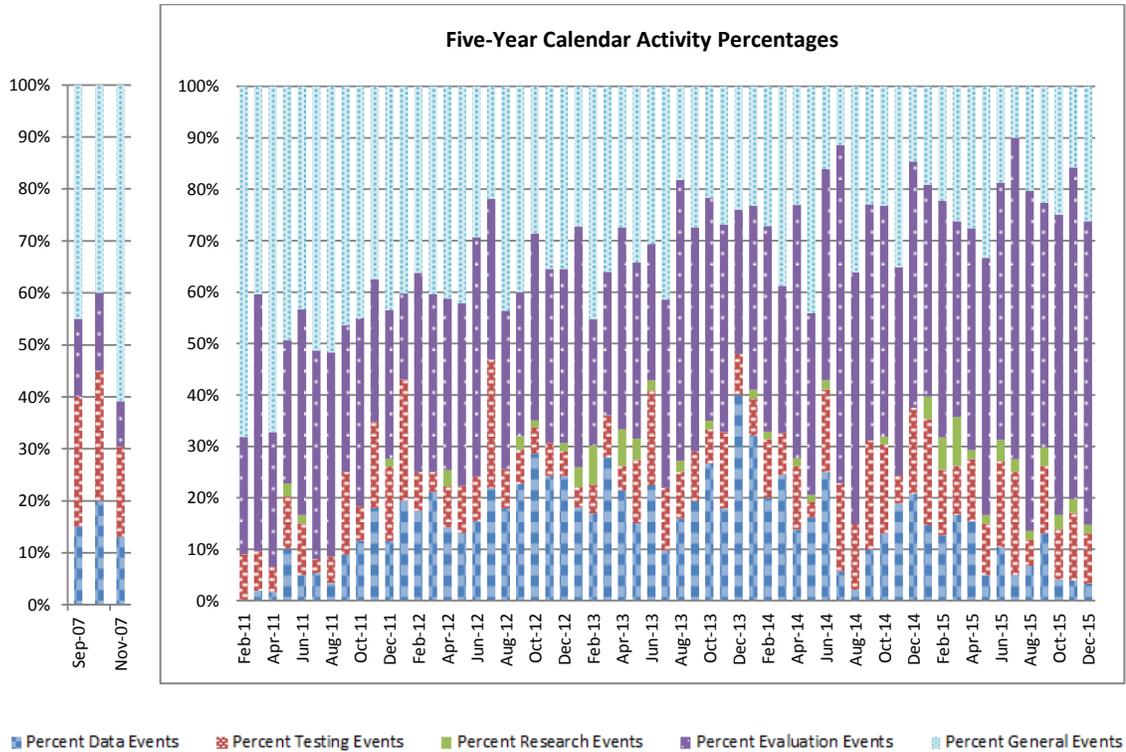


Figure 10. *Five-year calendar activity percentages.*

The average percentage of evaluation-related activities grew from an annual average of 52% of events in 2007 and 2011 to an annual average of 78% of calendar events in 2015, with increases occurring each year. (See Figure 11.) In addition, the percentage of evaluation-specific events increased over time, as well, rising from only 13% in 2007 to an average of 33% in 2011, growing to an annual average of 52% in 2015. Even though the annual average of all evaluative-activities events was the same in the 2007 sample and the 2011 monthly event, there was a substantial shift in the percentage of evaluation-specific meeting that occurred over that time frame.

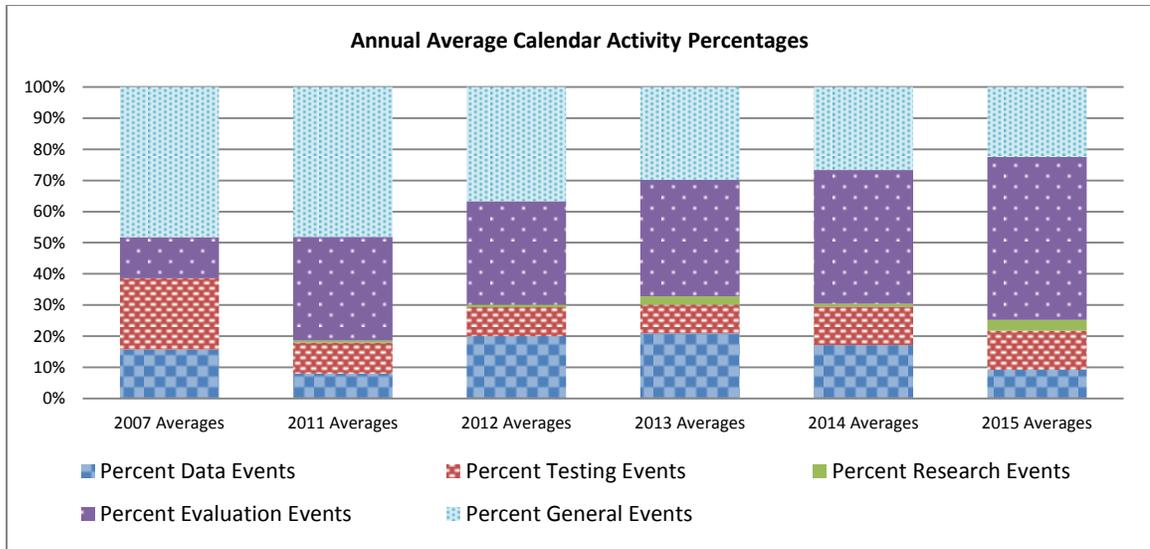


Figure 11. Annual average calendar activity percentages.

When comparing the events over the 2007 sample and the five-year period between 2011 and 2015 by month, there is an overall trend of increased events with subsequent years. (See Figure 12.) This analysis also illustrates that, in general, there were more evaluative activities that occurred in the waning months of the year.

Comparing evaluative calendar events by superintendent era also depicts a general trend of increasing events of interest with each succeeding superintendent.

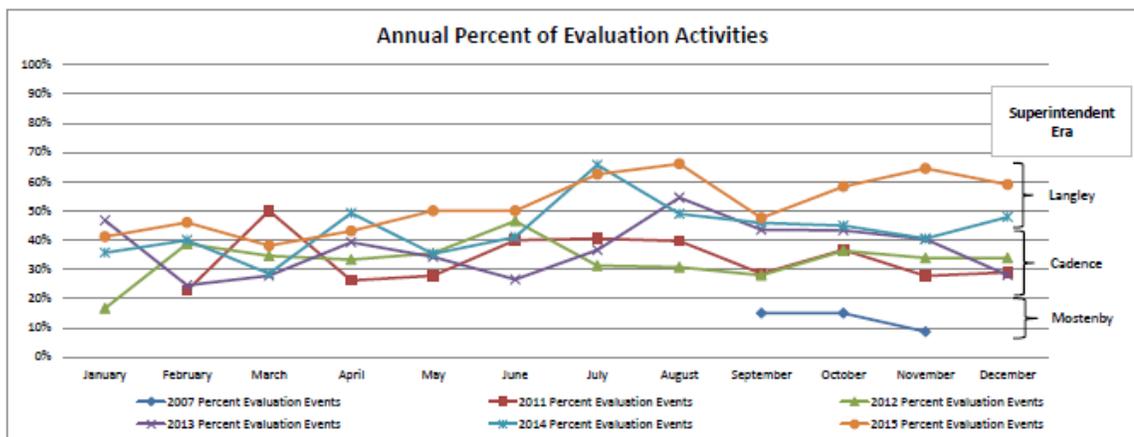


Figure 12. Annual percentage of evaluation activities.

Looking at the change in counts of the types of meetings on Ms. Norhakah’s calendar, whereas the total events grew by just over an average of about 9 events per month, the average number of evaluative activities increased by around 20 events, with 15 of them directly related to evaluation. At the same time, the average number of general events decreased by an average of about 11 meetings per month. (See Table 10.)

Table 10. *Total Annual Average Count of Scheduled Meetings/Events by Activity Type*

Date	Total Number of Events	Data Events	Testing Events	Research Events	Evaluation Events	General Events	Total Evaluative Activities
2007 Avgs	83	13	19	0	11	40	43
2011 Avgs	626	50	62	4	209	301	325
2012 Avgs	748	150	69	6	249	274	474
2013 Avgs	721	151	66	19	270	215	506
2014 Avgs	661	113	82	6	284	176	485
2015 Avgs	731	67	92	25	383	164	567
5-yr Avg	697.4	106.2	74.2	12	279	226	471.4
Approx. Avg change/month (between 2011 & 2015)	9	1	2	2	15	-11	20

The percentage change (or magnitude of change which is dependent upon the starting values of each calendar meeting type) grew by approximately 50% in overall evaluative activities, whereas at the same time, the general events decreased by about the same magnitude. Research events were the most substantial magnitude of change, however, not practically significant, given the very small number of events in occurrence. The evaluation events grew by a magnitude of almost 70% from 2011 to 2015, demonstrating an incredibly heightened focus on evaluation activities over this time frame. (See Table 11.)

Table 11. *Annual Average Percentage of Scheduled Meetings/Events by Activity Type*

Date	Percent Data Events	Percent Testing Events	Percent Research Events	Percent Evaluation Events	Percent General Events	Total Average Annual Evaluative Activities Events
2007 Avgs	16%	23%	0%	13%	48%	52%
2011 Avgs	8%	10%	0.6%	33%	48%	52%
2012 Avgs	20%	9%	0.8%	33%	37%	63%
2013 Avgs	21%	9%	2.6%	37%	30%	70%
2014 Avgs	17%	12%	0.9%	43%	27%	73%
2015 Avgs	9%	13%	3.4%	52%	22%	78%
5-yr % Change	15%	27%	435%	57%	-53%	49%

Interview Findings: Evaluations Importance to Mission and Vision

Related to the importance of evaluation to the mission and vision of a school district, Dr. Kincade unambiguously stated that:

Obviously, [evaluation] is extremely important because if you don't do it, how do you know what's working, what's not working? How do you know what changes to make? It's essential. [adding] What is fascinating to me is the fact that districts don't do it routinely. (Kincade, interview, July 28, 2015)

Dr. Anthony elucidates something similar, citing a well-known source of contemporary quotations:

It's essential, extremely important. Operating without data is like putting a blindfold on and trying to walk through life or walk through an educational system. Without data, you don't know – I like the Yogi Berra approach. If you don't know where you're going, very likely, you're going to end up somewhere else. And without data, you don't know where you're going and you have to justify how you got here or whether what you're doing now is an improvement over what was there before. So basing what a district does on data is absolutely essential. It's the only way you can show you made progress. It's the only way you can be assured that you are actually going in the right direction, and if you have optional directions, it's the only way you know which one to choose. (Anthony, interview, March, 20, 2015)

In a personal reflection on the interview questions, Ms. Norhakah fervently agreed, saying that evaluation is:

Critical – Our mission is to effectively educate each of our students for success. If you aren't evaluating, how are you determining if we are effective or not? In addition, if our vision is to be a public school system of excellence, with high quality staff and programs and successful graduates, we cannot get there without evaluation. (Norhakah, personal reflection, August 12, 2015)

Another example response providing testimony to the importance of evaluation is from Mr. Cadence who said:

It's extremely important – critical! You really [need] to see inside [the organization] – how it's working – what's working, what's not working...I thought all our decisions should be data-driven decisions. You [have to] based them in the context of what the data is – always felt that way. [Evaluation is] critical ...you always have to have the data in context with everything you're doing. (Cadence, interview, July 2, 2015)

Chapter 5:

Discussion and Implications

Stockdill et al. (2002) call for the evaluation of evaluation capacity building (ECB) efforts to inform the practice of program evaluation as well as the practice of ECB. King (2002) adds specificity to this call to track indicators, saying that school districts should monitor such things as data access and use throughout the organization; the reference to data in meetings; increased demand for and support of data; resources aligned to evaluative activities, including dedicated staffing positions; and in-house capabilities to collect, analyze, interpret and report data, involving stakeholders in the processes.

Dr. Kincade expressed that she felt that the department was not well-evaluated, which seemed ironic to her. Despite that, she believes that good work was happening and people were working hard, laying the groundwork for capacity building activities well into the future, benefitting from critical foundational materials like technological advances, leadership development and change, and legislative requirements.

This study examined the emergence and development of evaluative activities and the department charged with their oversight as a means of evaluating the capacity building efforts of Two-County ISD. The following is a discussion of the primary findings, along with implications.

Discussion of Findings

Through this study, eight major themes emerged:

- Leadership matters.
- Evaluation is mission and vision critical.
- There was a shift from evaluation for others (external impetus) to evaluation for the organization (internal impetus).

- A background in teaching is beneficial to ECB.
- Data, and the technology to access, analyze and report them, are imperative to ECB.
- Resources are the foundation of school systems and limitations in resources impact virtually all aspects of ECB.
- Community engagement is a must.
- Legislative awareness and proactivity are paramount.

This chapter will explicate each of these themes.

Leadership matters. Two-County ISD has had eight superintendents, including a six-month interim, since its inception in 1952. These district leaders have ranged in years of service from one and a half years (for the current superintendent) up to 15 years, with an average, discounting the interim position, of nine years of service (See Figure 13.)

This speaks to generally higher rates of continuity compared to the current average United States rate of approximately 3 – 5 years for sitting superintendents (Fullan, 2016) and aligns to one of the top themes in this study – i.e., that leadership matters.

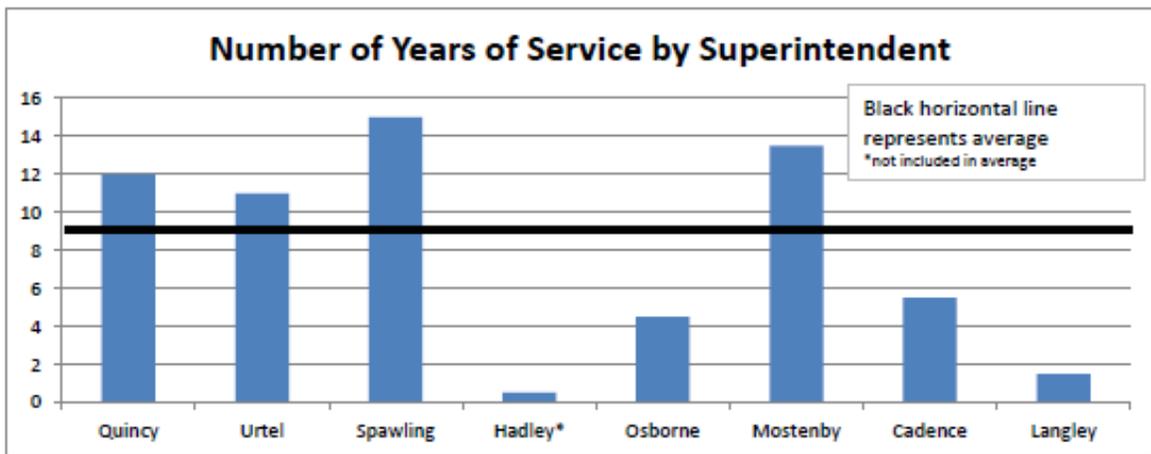


Figure 13. *Number of years of service by superintendent.*

Mr. Cadence recognized the importance of leadership changes in Two-County ISD, as did Dr. Florence, both commenting on how things had changed under each superintendent. Many researchers speak to the importance of leadership in relation to

ECB and internal evaluation, in particular (Baizerman et al., 2002a; King & Volkov, 2005; Love, 1983; Preskill & Boyle, 2008; Sonnichsen, 2000; Taylor-Powell & Boyd, 2008). A leader's visionary, innovative, and creative thinking; prior training and experiences; interests, convictions, and passions – what he or she is willing to stand up and fight for; interpersonal skills and the ability to navigate difficult situation and make difficult decisions all play an important role in moving ECB forward or not. Compton (2009) explicates that “most critical for developing and sustaining an organization's unit is the support of top leadership” (p. 63).

Leaders can also have a negative impact as explained by Dr. Bennett. She described a time with a central office leader put barriers up, not allowing the research, evaluation, and testing department or any other department supervised by Dr. Bennett to interact with any of the principals or departments under his supervision. Such opposition in top leadership positions can be devastating to ECB efforts. Bourgeois and Cousins (2013) indicate that leadership is one of the contributing factors to ECB success, expounding specifically on the leader's potentially positive impact on implementation and sustainability by offering support to the efforts.

Recognizing that, from a system's standpoint, hiring the right leader makes a difference, Ms. Gillian indicated that the school board recognized the need for a change following Mr. Osborne. Significant reductions were made to staff, and a number of people were dissatisfied and disenfranchised with district leadership so they were looking for someone who could bring people together and restore trust. Reflecting on the impact of eliminating central positions and changing structures, Ms. Gillian said she never felt as informed or connected to what was happening in the curriculum department or with

testing after Mr. Osborne dismantled the district structure. Leaders also recognize the importance of evaluation to the mission and vision of the organization.

Evaluation is mission and vision critical. Overwhelmingly, the leaders of Two-County ISD felt that evaluation was extremely important to the mission and vision of the district. Cousins et al. (2006) unequivocally state that one of the “key strategic building blocks found in learning organizations” (p. 158) is having a clear mission and vision as a guide. Kotter (1996) stressed the importance of “[making] structures align with vision” (p.115) to avoid blocking action. Labin et al. (2012) expound that the mission and vision are the essence of program outcomes to which organizations aspire. They found that “program outcomes improved as a result of ECB” (Labin et al., p. 6), citing Wandersman et al. (2005), in that “improved program outcomes have been a long-standing rationale for conducting ECB” (p. 6), providing districts a foundation to routinely conduct evaluation. Dr. Kincade explained that the evaluation supports the mission and vision by providing information about the status of the district’s work – what’s working and what’s not? Dr. Anthony added that without evaluation, a district is blind to what is happening, where improvements should be made, and the progress that is or is not occurring. Ms. Norhakah agreed with these statements, declaring that if a district was not evaluating, there was no way of truly knowing if they were effective in their actions, along with the areas in need of improvement. A mindset of continuous improvement, evaluation capacity building, and conducting evaluations to provide direction toward the espoused mission and vision of the institution is beneficial to organizations.

There was a shift from evaluation for others (external impetus) to evaluation for the organization (internal impetus). Certainly there is an avenue for evaluation to

be used for external accountability and mandates aligned to legislation and grant attainment. Optimistically, there is also growing understanding that evaluation and ECB are extremely beneficial and even critical for internal program improvement, regardless of external requirements. Combining the impetus behind evaluation and accountability, Patton (2011) asserts that:

For vision-and-values-driven social innovators the highest form of accountability is internal. Are we walking the talk? Are we being true to our vision? Are we dealing with reality? Are we connecting the dots between here-and-now reality and our vision? And how do we know? What are we observing that's different, that's emerging? (p. 13)

Huffman et al. (2008) posit that “ideally, an organization has the capacity to engage in evaluation for both external accountability purposes and for the sake of building capacity to improve programs and better serve constituents” (p. 358). Stockdill et al. (2002) differentiate between evaluation as a means of program improvement and accountability and ECB as a shift into making evaluation a routine part of an organization's work. Fullan (2016) outright states that “external accountability does not work unless it is accompanied by development of internal accountability” (p. 51).

The declarations of the scholars above describe one of the findings in this study. In Two-County ISD, between the years 1985 – 2015, there were differences in each superintendent's era related to evaluative activities, but there was evidence of a shift from a primary external impetus to an increasingly internal impetus for evaluation and ECB as a regular part of the district's work.

Mr. Spawling's era was characterized by evaluation-related activities done by and for the curriculum department in the form of content-area studies conducted as part of legislative mandates. The newly-formed assessment department did not interact with

these study processes, but instead focused on developing and implementing classroom assessments that Dr. Eickner explained were not initially well-received; teachers complied begrudgingly because they did not yet see the value of these specific data, or using data in general, along with the fact that this was the first time that systematic data from district-wide common assessments were collected, representing a huge change for many teachers. Overall, in the next era in Two-County district, the Hadley/Osborne period, there were not a lot of evaluative activities that occurred outside of somewhat carrying on the activities that were already occurring.

In Mostenby's era, there was some evidence of evaluation expansion, and external evaluators, including partnerships for evaluation with the state land grant university, became more common. With the passing of the accountability legislation and increased testing aligned to standards, content-area curriculum studies subsided, giving way to focused work on math, reading and, eventually, science curriculum alignment to changing standards. Along with staff trying to keep up with curriculum and materials aligned to standards, there was an increased awareness regarding the necessity for high integrity in data to insure they were accurate and valid; the requisite of the processes and systems embedded with security identifiers to provide appropriate access to data based on user role and protect data privacy; and the importance of technological infrastructure, hardware and software, to support the growing need for data accessibility and analysis.

During the Cadence and Langley eras, internal evaluation emerged to a greater extent than before driven by central and building leaderships' desire to know what was working and what was not in reference to internally identified indicators such as intervention and extension programming, enrollment patterns, predictability, accuracy of

district-developed models, etc.. Processes for research oversight and data collection support were developed and implemented, along with an increased number of annual district-wide surveys, gathering the input and perceptions of students, parents, staff, and community members. Collaboration around evaluative activities expanded across the organization to include both instructional and operational departments instead of support of just the curriculum department. Technological data collection and monitoring and reporting tools continued to be developed and refined across the system, and a focus was placed on alignment of district communication and reporting strategies and products to cultivate a more clarified, shared understanding across stakeholders and advance continuous improvement by everyone moving in the same direction. Strategic partnerships with external evaluators continued, especially related to very political or contentious contexts or when an amplified level of perceived objectivity would be beneficial.

The transition from an almost exclusively outward force for evaluation to one much more internally-driven is evidence of organizational maturity – demonstrating commitment to growth and improvement. Torres and Preskill (2001) explain that:

One way to stimulate organizations' interest and commitment is for their own disciplines... to support and teach about professional practice and leadership which effectively uses evaluation for continuous improvement. It follows that when [staff members], and not evaluators, are defining evaluation purposes, it is for maximally useful feedback about the effectiveness of their work. (p. 391)

Cadence, in speaking about evaluation and the changes in the work of the research, evaluation, and testing department, recalled that the department in the past only worked with the curriculum department and never spanned outside of K-12 programming. In addition, any comparisons that were made concerning performance were to other districts

based on what legislation and the public were directing and identifying. He praised the transition to now identifying the district's areas of need (certainly along with maintaining an external comparison), but in a way that was beneficial and innovative; answering questions that people had and delving into areas the data were directing then, which has made the district a leader amongst state peers, in his experience. He avowed that this shift from an external impetus of what was important to examine and analyze to a more contextual, internal focus was significant and an indicator of growth within the district. Building evaluation capacity can be done through process use – teaching people evaluation by doing evaluation (Patton, 2008).

A background in teaching is beneficial to ECB. Process use, in essence, is learning evaluation by doing evaluation (Patton, 2008), so it stands to reason that training and experience in a classroom setting would be advantageous. Building on the work of Cronbach, King (2007) equates the evaluator to teacher, explaining that the evaluation and the process of conducting evaluation are the curriculum (what the *student*, the participant, is expected to learn and be able to do). Attending to ECB through process use, then, requires skills similar to planning a lesson for the classroom, a task much more natural and surmountable with experience in that realm. Referring to Dr. Eickner's ability to explain data and evaluative findings, Ms. Gillian raved about how well she understood the information, even to the point of being able to explain it to others and to write about it in communications. Dr. Eickner, Dr. Dooren, and Ms. Jordasen all explained that they were hired and experienced success in their roles related to evaluative activities based on their connection to teaching and the classroom. Dr. Kincade recognized the importance of an educational background, as well, explaining that she understood the educators and

they understood her because she was a trained teacher and had been in their shoes. With a teaching background, the implications to and expectations of schools, staff, students, and families related to evaluative activities and data use had the opportunity to be better understood than without a teaching connection because of experience in the school setting and direct interaction with each of these stakeholders.

In reference to the passion and motivation behind the departmental leaders (all of whom possessed classroom teaching experience) for serving each and every student as the primary focus of their work, Dr. Mostenby had the utmost confidence that every leader of the department knew exactly why he or she was there and for whom they were there, never losing focus on students and what was best for them. Having an understanding of how to teach and move the department and district forward involved the accessibility, analysis, interpretation, and use of data.

Data and the technology to access, analyze, and report them are imperative to ECB. Two-County ISD had the incredible opportunity for significant evaluative propulsion forward under the leadership of a professionally-trained expert evaluator in Dr. Kincade, yet her ability to make headway was severely impacted, at least in part, by the inaccessibility of data and the lack of technological tools to effectively manage the data she actually did receive. King (2007) explains that an “accessible and accurate database...is a necessary part of the evaluation infrastructure and cannot be taken for granted” (p. 56). She goes on to explain that “technological advances and a commitment for more immediate access to data have resolved this [challenge]” (King, 2007, p. 57).

Dr. Mostenby agreed, saying that without data, people cannot effectively do their work, and, if they think they can, they are uninformed and extremely mistaken. Cadence

lent agreement to Mostenby's claim, stating you used to be able to get away with not using data and many people did. Even though data were used regularly in the central office to some degree, it was not necessarily commonplace the farther you moved into the heart of the system. Teachers may or may not have used data to make instructional decisions, but now, since the accountability wave had hit, if someone, virtually anyone, in the system was not using data to some degree, he or she would struggle to be successful in their role.

Eddy and Berry (2008) support this assertion saying that "evaluators must guide educators and program personnel through the process of data collection, analysis, and interpretation; additionally, teachers, school officials, and other program administrators must develop the requisite evaluative skills necessary to draw appropriate conclusions" (p. 99). Accompanying the skills and abilities to manage and use the data, educators need appropriate tools to electronically access, organize, and report data findings (Eddy & Berry, 2008). Sutherland (2004) agrees, explaining that regular engagement in evaluative activities, such as those described above, helps schools identify what is going well and what is not, areas of strength and those needing improvement. She also insists that by participating in data-informed decision making, schools and district nurture a culture of continuous improvement (Sutherland, 2004).

Two-County ISD profited from the alignment of values and goals between district leadership and the school board related to data use for decision making, with both groups placing high importance on data use and continuous improvement. In addition, there was shared commitment to gathering and using a wide variety of data, expanding beyond test results to other types of outcome and achievement data, perception data, program data,

and demographic data. Alliance between governance (the school board) and management (district leadership) personnel in the district, along with direct interaction between the research, evaluation, and testing department and the top leadership in governance and management, were advantageous to strengthening evaluative activities throughout the district.

Speaking directly from the superintendent's perspective, Cadence elucidated the need for readily available data, pertinent to the happenings of the district, sometimes at a moment's notice, to explain, dispel or support communication efforts for a wide range of stakeholders. Working closely with the research, evaluation, and testing department and knowing there existed the ability to internally support the superintendent's data requests was invaluable. Adding to the importance of direct connection between the superintendent and the evaluation unit, Langley explained that with the research, evaluation, and testing department so entrenched in the data and the state of the district related to a wide variety of areas, informing change was inevitable if leaders were engaged and collaborating with the department. As evidence to the value of the information, Mr. Langley recalled using data multiple times in addresses to staff and stakeholders in his back-to-school welcome engagements alone.

Mostenby, grasping the significance of the advances in technology, urged that educators and the people actually using the data (central and school administrators, teachers, parents) needed to take an active role in anticipating and shaping the data tools and systems available to ensure they would meet the needs of practitioners instead of letting outside players such as the general public, the business industry, and the vendors themselves direct development. Mostenby cautioned that educators, in the midst of

everything, must not ever lose sight of each individual student in the interest of data aggregation across varying units of analysis. Data are extremely critical, but if resources do not exist to support operations, a system is virtually paralyzed.

Resources are the foundation of school systems, and limitations in resources affect virtually all aspects of ECB. Just as many researchers believe leadership is important in an organization, there is convergence on the significance of resources, as well (Baizerman et al., 2002a; King & Volkov, 2005; Labin et al., 2012, Preskill & Boyle, 2008; Stockdill et al, 2002; Taylor-Powell & Boyd, 2008). Mostenby describes three non-instructional components vital to managing education: demographics, general finances, and personnel costs, explaining that if any one changes the other two and everything else in the district will change because everything else is contained within those. Labin et al. (2012) name resources as one of the foundational components to ECB, as well. Stockdill et al. (2002), whereas not worded as strongly, found that the “promise of ECB is limited by the lack of necessary financial and human resources” (p. 21).

In Two-County ISD, according to virtually every interviewee, budget and fiscal constraints had always been an issue. Cadence explicitly stated that money had always been the driving factor, which was largely based on enrollment data and it was the underlying basis for everything. Money limitations had shaped the core of who Two-County ISD was and how they operated. Because finances were always tight, a greater need existed to seek avenues of effectiveness and efficiency – using every dollar wisely. In addition, continuous improvement and modifying or expanding existing practices, materials, and tools became part of the collective mindset.

Explaining the district's desire to be accountable and committed to being good stewards of limited resources, thus honoring the community support for bond and levy dollars, Mostenby astutely recognized that all district leadership had to be aligned and in agreement with (or at least accepting of) the determined needs of the district prior to going to the community for funding support. He clearly explained that going to the public or even the school board when people were still in disagreement was not advantageous and was, in fact, detrimental to the discussion and request for assistance. In addition, it undermined community confidence in the recommended solutions to budgetary issues proposed by the district.

Mostenby also explained that the district went through at least 10 years of budget cuts between the years of 1985 and 2005, with total community involvement in the process, resulting in at least three levy referendums during that time. When asking monetary support from the community, leaders needed to have a solid rationale as to why the money was needed and a plan for exactly how it would be spend. Two-County had always held the commitment of keeping the cuts away from the classroom (meaning not cutting classroom teachers), and instead targeting central office teachers on special assignment (ToSAs), auxiliary programming, or virtually any other imagined consideration throughout the process.

Related to budget constraints was the difficulty in dedicating funds to evaluation in the midst of limited resources. Bennett illuminated the issue by describing the rotating door of trained evaluation staff in the midst of budget cuts, sending them back to the classroom, if they were teachers, or terminating their employment if they were not--then, if support for evaluation and capacity building existed with decision makers, starting the

process over when funds became available. Florence described money, or lack thereof, as an extremely influential component of the evolution of evaluation use and capacity within the district, citing the example of staffing limitations alone.

Again, Bennett illuminated another connection between limited resources and evaluation, explaining that it was critical to know what leaders were investing in and if it was effective for its intended purposes, giving results that were best for students. Inadequate funding combined with overwhelming accountability for all students lent itself to increased program evaluation. It applied pressure organization-wide, instructionally and operationally, however, Bennett noted that the difference was that the instructional side was under significant public scrutiny whereas the operational side of the organization was not, prompting the felt need for change only on the instructional departments at the time.

Because of consistent fiscal limitations, the search for grants and other alternative sources of funds have been a continual part of the district processes, creating the need for further evaluation capabilities in the form of accountability to funders. The district's sense of fiscal responsibility to be accountable to and engaged with the community transcended the budget into other areas of decision making in Two-County ISD, as well.

Community engagement is a must. Two-County ISD was founded through community input, and community input remained a constant underpinning of district operations and viewed as a valuable aspect of district data and processes. Cousins and Whitmore (1998) provide substantiation for this practice, imparting that involving stakeholders' perspectives provides validation to evaluation data. Gillian revealed that one thing that Two-County ISD did that many other districts do not is to invite students,

parents, and community members to engage with staff in large task forces in an effort to collaborate on a longer-term basis.

In Two-County district, Hadley recalled there had always been a focus on striving toward effective communication with the community, including parents and non-parents in the district boundaries, which took the form of collaborating on study committees of various types, providing multiple means of input about upcoming decisions, and seeking feedback on past district decisions, all relating to a wide array of topics. Sometimes the topics for which community collaboration was sought involved legislative impacts the district faced.

Legislative awareness and proactivity are paramount. Sometimes legislation is seemingly senseless or infeasible, yet schools and districts are bound to the requirements set forth. Eddy and Berry (2008) describe such circumstances related to the No Child Left Behind (NCLB) Act (2002). The mandates and implications left education entities, educators, students, and families running to keep up with the changes imposed by the law. Almost universally, the leaders of Two-County ISD agree that NCLB had created a monumental *sea change* for education.

Describing his grossly mistaken perception of the impact that NCLB would have on local public education systems, Dr. Mostenby recalled telling a reporter that it was extremely unlikely that the legislation would change the way schools served students and families through programming. He was confident that the law could affect resource allocation, but it could not affect policies, practices, and focus at the local level. He recollected thinking how very wrong he was; NCLB changed policy, practice, and focus for every public school system in the nation by the requirements it demanded.

Hadley and Dooren described the beginnings of accountability and a public focus on education with the report *A Nation at Risk*. Hadley remembered that it was not a very flattering report, however, admitted that it was likely needed to instigate systems change. Dooren recalled that data were not collected or used in the same way prior to legislative enactments, explaining that the focus shifted from equitable access to equitable outcomes that, in turn, changed the way schools and districts operated. Bennett suggested that the educational accountability legislation of this time frame (*A Nation at Risk* stemming from legislative commissioning and ESEA reauthorized into NCLB) thrust everybody into a need for evaluation capacity to keep up with the changes, while at the same time, trying to get out in front of it and be proactive.

Jordasen revealed an opportunity that existed by being aware and involved at the legislative level in an effort to be preemptive – the ability to provide influence and insight as to what might be coming. Mostenby strongly advised that, as district officials, leaders had to pay attention to the legislature and legislation (both proposed and enacted). They needed to build relationships with government officials, including the staff at the department of education, the legislators and their staffs, lobbyists, and special interest groups, including parent groups. By doing so, they gained the ability to anticipate and interpret topics of concern from these perspectives, as well as advocate for the issues and solutions deemed most appropriate from the local level. Cadence said that one of the biggest changes he could identify related to the legislature was the expansion in the number of people who were aware and involved throughout the district, which was a good thing.

In addition to being involved with the legislative process, being deliberate and planned with regards to fiscal opportunities through legislative action was astutely crucial. Two-County ISD had routinely made an effort to capitalize on one-time funds in an effort to *build capacity, not cliffs* (funding cliffs) as some districts did with the America Recovery and Reinvestment Act (ARRA) stimulus dollars, for instance, hiring people with no plan for when the money was exhausted. Granted, given the district size, Two-County received proportionately more dollars, however, this principle could be applied regardless the district size.

Gillian shared her view on the importance of the stimulus dollars in that it allowed the district to invest in strategic priorities. For the first time, the district took action to actually intentionally align dollars to initiatives thought to provide forward movement and improvement as a district, even in the middle of declining funds, instead of only desperately trying to maintain what already existed. Some of the strategic investments that were made with ARRA funds included two evaluation specialists to support the work of evaluation capacity building and an investment in technology tool - data dashboards for central and school level administrators.

Implications

There are several overarching implications that emerge from the above discussion that other institutions engaged in or considering evaluation capacity building or simply interested in the experiences of one metropolitan school district in the Midwest may consider. The themes are unavoidably interwoven, with one affecting and connecting to the others, and they could be organized and displayed in a number of different ways,

however, I have chosen to group and order them into three scaffolding categories: resources, people, and processes. (See Figure 14.)

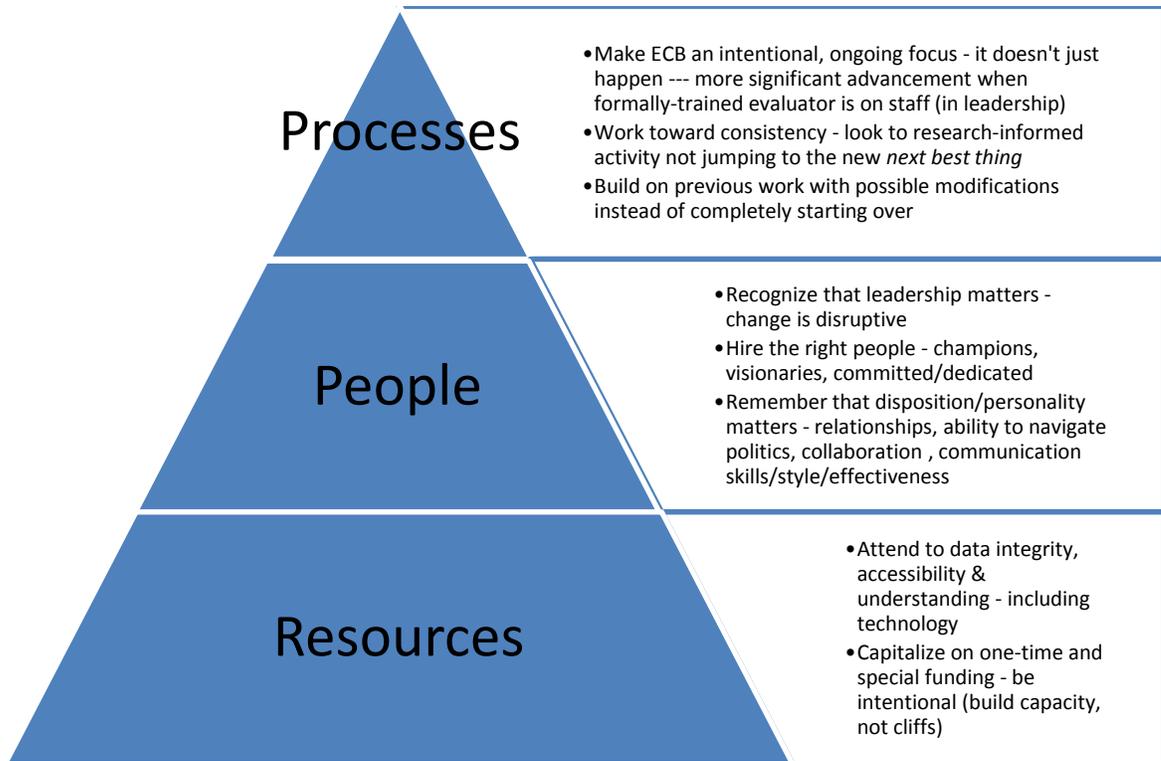


Figure 14. *Implications for evaluation capacity building in schools.*

The implications emerging from this study fell into two categories. Either they were overwhelmingly and universally applicable (e.g., leadership matters, change is disruptive), or they were case-specific (i.e., found to be transformational for Two-County Independent School District) and not impactful to ECB in general. The ideas were layered in the manner in which they were because, as described above, resources are foundational to all other functions within the district, the availability of human resources, people, stems from the footings of district resources and the processes employed are dependent upon the people putting them forth and supporting them. The information within this section is an extension of the material in the discussion section above.

Resources. Resources form the bedrocks from which all activities in an organization arise. Without a firm base of support, stability is difficult, if not impossible to sustain.

Attend to data integrity, accessibility and understanding – including technology.

The analogy that resources are to school districts as data (and the technology to access it) are to evaluations demonstrates how integral data are to conducting evaluations and building evaluation capacity. Cousins et al. (2006) found that a supporting factor for influencing reliance on evaluative inquiry in schools was that when staffs use data, valuing data manifests. They explain that "staff developed deeper appreciation for the power and utility of evaluative inquiry through concrete examples of how data and locally created knowledge can feed into the decision mix" (Cousins et al., 2006, p. 172). The following description is an illustrative example of the necessary progression that must occur to meet the data and technology needs to form the foundation from which evaluation and ECB can emerge.

In Two-County ISD before the turn of the 21st Century, data were only available in paper format and outside of what was directly collected and captured in teachers' record-keeping, staff had to go to cumulative file folders to access whatever data happened to be contained from previous years in a student's educational history. Populating the folders and then accessing them to learn information about the student was completely dependent on individual staff.

This system of paper-based record keeping gave way to electronic data capture when Two-County purchased and implemented a student information system. The system was a vast improvement, however, student data could still only be viewed one student at

a time, and staff could only access a student's personal information, demographics, class schedule, course history, and minimal standardized testing history. In addition, each software program was site-based and thus could only be accessed through individual sites, not from a district-wide view.

Moving to an electronic format prompted other considerations and opportunities for attention such as consistency in data entry and set-up from user to user and system to system, which required training and systems support mechanisms. Additionally, challenges and decisions about access needed to be considered – who should have access to what? What decisions needed to be made regarding data privacy and system security? How might occupational codes be leveraged to provide the appropriate data access, and how do we make provisions for unique user roles, not aligning with one of the set code categories? Do staffs that have access to a student automatically receive access to the student's entire data, or are there data that are limited due to the extreme personal or political nature of the data or situation? All these questions made up the aspects of data security and safety – maintenance of data privacy and protection of sensitive information; data integrity – having accurate and valid data with consistency across all systems and users; and one component of data accessibility – implications materializing with electronic data availability. All of these aspects had to be considered and addressed as part of proper data practices.

At the same time the student information system was implemented and the beginnings of data integrity and access were occurring, Two-County worked on systems integration and technological infrastructure and networking to provide the means to connect as many or all electronic devices and software systems into integrated access

from the user standpoint. Technological infrastructure, just like a transportation system, is paramount to successful technology integration.

Building from the beginnings of electronic data access, the district implemented a data warehouse tool, containing standard and customizable reports, accessible to all staff at a level pertinent to information need, based on role. Again, training was provided along with school and central level support. An intentional decision was to begin with the focus on classroom teacher data needs since teachers were the main point of contact and impact on student achievement within the system.

The warehouse tool allowed staff to have access to a wealth of data by individual student, dynamic or static groupings, roster groupings (classes), and, depending on security rights, by grade, building, or district-wide. Data of all types and sources were available including all the data from the student information system (demographics, personal information, schedules, course history, attendance); student achievement data (grades from the end of marking periods, common assessment data, standardized assessment data – both internally and externally selected, college and career readiness assessment data); program data (interventions and extensions, special course enrollment data, after-school program involvement, participation in specific support programs such as those for first-generation college goers or magnet school programs within schools); extra-curricular activities, enrollment history, attendance history, special plans such as individualized learning plans and health plans, immunization records, etc. Staff could also access all these data for students from the current year, past years, future years (after they had left a particular class or grade level) for actively enrolled students and inactive students.

Predictive tables were embedded in the system to allow staff to determine which students might need intervention or extension, were in jeopardy of not meeting specific performance benchmarks, or had consistently demonstrated low growth between measures. In addition, placement tools, based on data, meant to inform student course enrollment were developed in the tool in place of circulating spreadsheets.

Next dashboards were developed and implemented to provide just-in-time data reporting and alerts for students and classrooms not meeting pre-established benchmarks. Also, structures, communication, and training were established to provide staff access to static internally- and externally- created summary reports on data collected district-wide related mainly to student outcomes and perceptions.

Another aspect of data support was continually monitoring the alignment of staff data needs and the availability of tools and reports to meet those needs in the most efficient and effective manner possible. To do so, evaluation of the systems from a continuous improvement lens had to be part of the process. In fact, Huffman et al. (2006) suggest that ECB may be a beneficial means to manage and apply the vast amount of data currently generated in schools.

Capitalize on one-time and special funding – be intentional. Capitalizing on one-time and special dollars was addressed in the discussion session; however, taking this action when provided the opportunity impelled the district forward in ways that would not have happened without access to additional money, allowing Two-County to attend to infrastructure and capacity of the organization, seemingly not a common approach in other school districts. The message related to unanticipated distinctive opportunities is to be intentional about the decisions that are made, attending to infrastructure and systems

implications, as well as intended and unintended consequences, making sure that the decisions reached align to strategic priorities that support the mission and vision of the district. Efforts should be evaluated along the way for implementation fidelity and effectiveness in reaching intended outcomes. Referring to managing ECB efforts, Compton (2009) elucidates the importance of developing practices and infrastructure to support ECB efforts and sustainability while also strengthening the organization as a result.

People. People are really the essence of any organization. When you ask anyone in Two-County ISD, *What is your greatest asset?* the vast majority of employees would say its staff. Because education is a ‘people business,’ the right people are essential.

Recognize that leadership matters – change is disruptive. Again, the importance of leadership in general, but specifically to evaluation and ECB, is well-established in the literature and was examined in the discussion section of this chapter, nonetheless is essential to the success of ECB. The focus of this section, adding to the prior discussion, is that change is extremely difficult and, in many cases, detrimental to the culture and climate of an organization and to ECB. Furthermore, change is very “resource-hungry” (Fullan & Miles, 1992, p. 750) and for a district that suffers from fiscal famine that is problematic and disruptive. Fullan (2016), drawing on the work of Marris (1975), explains that “real change...whether desired or not, represents a serious personal and collective experience characterized by ambivalence and uncertainty; and if the change works out, it can result in a sense of mastery, accomplishment, and professional growth” (p. 21). Forss et al. (2006) speak to the impact of disruptive events on an organization’s readiness for evaluation.

In reference to Two-County ISD's central office, Jordasen exposed the frustration felt throughout the system, highlighting the perception that there was constant change there – change in structure, change in personnel, even change in focus with different leaders. Just when school staffs thought they knew the system and the contact people, they would change, and eventually, depending on the person's tenacity, he or she might or might not keep seeking support. Florence also recognized that change was disruptive for systems support, especially related to evaluation and ECB efforts, illustrating the inconsistencies in the focus on evaluation due to leadership changes prompting changes in focus affecting the ECB efforts. Speaking about the changes in departmental leadership, Inettow, who was in various roles in secondary curriculum over time, commented that when one person who is well-accepted and successful leaves and is replaced with someone without the same characteristics, perceptions and credence of departmental work diminished, illustrating the difficulty of change, but also the importance of hiring the right people.

Hire the right people. Since education is a people-oriented business, people matter. There is a level of skill, ability, and knowledge necessary to successfully lead evaluation and evaluation capacity building within organizations (Eddy & Berry, 2008; Stockdill et al., 2002; Sutherland, 2004). With rising challenges in the educational world, competing priorities, salaries lower than other industries, and increased scrutiny, a large pool of qualified applicants may be increasingly more difficult to find, but also increasingly more critical. Cadence speaks about the change in the level of expectation for internal educational evaluators since the start of the accountability era, saying that the roles are increasingly more difficult, and, if someone is not able to keep up as well as

think creatively and strategically about what data are needed and the best ways to communicate and display it, he or she will likely not be successful in the role, especially in the long term.

Plowman et al. (2007) provide an interesting insight, describing the impact of the right people in place to recognize, support, and move forward small changes that have the potential to eventually build over the years into large ones, pointing to elements of continuous improvement, vision, and opportunistic awareness. Highlighting the importance of hiring people with visionary skills related to evaluation and capacity building, Florence explicated that people who are visionary can see needs unidentified by others, and thus sometimes face resistance, but also have the skills to push through and beyond it toward the long-term focus.

Remember that disposition/personality matter. Langley described the research, evaluation, and testing staff as being responsive and collaborative, partnering with all departments across the organization on a regular, consistent basis. He acknowledged that they were found to be helpful and supportive in determining best ways to measure performance and collect data and then analyzing the data that resulted, along with general support on evaluative activities. Having a pleasant disposition that engaged and encouraged others was critical to effective collaboration/partnerships that were essential components for evaluation capacity building from a participatory, process use frame.

King (2007) describes the importance of personality, especially in the context of process use, providing an illustration of a scenario dependent upon an even disposition. She cautions that sometimes participants may get frustrated or at odds with each other or the process, even to the point of being noticeably or audibly disparaging. In situations

such as that, it is the evaluator's role to maintain pleasantries and defuse or divert the challenge. If the evaluator reacts differently and inflames the situation further, the whole evaluation and the capacity building efforts may be thwarted completely.

Having the ability to recognize teachable moments, seize opportunities to garner support for evaluative activities and capacity building efforts, and navigate tough political waters, among a host of other skills and abilities, is paramount to effectiveness and success (King, 2007). When seeking an evaluator to hire or with whom to work, attending to the technical skill a person possesses is not uncommon. Paying attention to the other skills, however, is not as commonplace.

King and Stevahn (2013) propose six competency areas in which evaluators and ECB practitioners should demonstrate proficiency, among them, interpersonal or people skills. Collinson et al. (2010) also hone in on interpersonal skills and their impact on the *human relationship* factor of organizational learning. Other researchers also express the importance of personality traits and interpersonal skills in relation to evaluation, particularly internal evaluation, and ECB (Clifford & Sherman, 1983; Preskill et al., 2003; Sonnichsen, 2000). Kincade affirmed that she believed her personality was important in adding to her acceptance and success in the district. She acknowledged that if some of the other professors with whom she worked were to come into the district, they likely would have experienced different outcomes based on their personalities and the fact that they would not be able to relate to teachers or socially and politically navigate the K-12 cultural waters. Personality matters, as does the ability to manage and advance important processes.

Processes. Processes can determine, or at least greatly influence, the effectiveness of an organization. Some processes are highly accepted and yield great results; some are accepted yet do not resolve in favorable outcomes. The converse can be true, as well; there can be generally unacceptable processes with good results or unacceptable processes with poor outcomes. Of the four options described, only one is ideal – accepted processes with positive outcomes. Achieving this condition depends on people making right choices and evaluation of the process and results, which is far more feasible with system’s capacity to think evaluatively.

Make ECB an intentional, ongoing focus – it does not just happen. Chapter 2 explored the idea that ECB is intentional and long-term (Compton, 2009; Cousins et al., 2007; Forss et al., 2006; Huffman et al., 2008). The findings of this study support that notion whole-heartedly; however, in the fast-paced world of education these days, people are interested in quick fixes and easy answers. In light of all that has been expressed within this study, ECB is neither quick nor easy. Contrarily, Fullan and Miles (1992) warn to avoid “impatient and superficial solutions” when approaching change (p. 747). Compton (2009) elucidates the need for ongoing evaluative activity along with the ECB evaluator intentionally making plans, creating structures and processes to bolster ECB, and taking concrete action steps to increase capacity, while consistently interacting with organizational and program leaders to generate demand and support for evaluation as part of his or her standard work.

One point to consider related to intentionality of ECB is that intentionality for the evaluator does not mean intentionality or even awareness for the participant (Patton,

2007). This twist in concept can be avoided if learning targets and ECB intentions are overtly shared with participants as part of the ECB planning, practice, and processes.

Work toward consistency – not the new next best thing. Fullan and Miles (1992) explicitly state that “the temptation is great to latch on to the quick fix, to go along with the trend, to react uncritically to endorse innovations as they come and go” (p. 747); yet they counsel to avoid fads and stay the course. This approach is found to yield much more productive results in the long run (Fullan & Miles, 1992). Over 20 years later, Fullan (2016) appends this advice saying that continuity of good direction is advantageous, but schools and districts too often abandon efforts too quickly, before they have the opportunity to demonstrate measurable gains. The implications to ECB are that evaluators’ plans and efforts must be mindful of the time expectations and they must act accordingly related to evaluation. This attention to time impacts must be given to ECB change expectations, as well.

Build on previous work – seek continuous improvement instead of starting over. Two-County ISD has adopted a continuous improvement mindset as part of the organizational culture. Evaluation activities do not occur in isolation, rather in concert with ongoing organizational changes driven by a commitment to continuous improvement and responsiveness to stakeholder needs. Information gathered through evaluation facilitates discussion and decision making relative to organizational development and change. In the same spirit, Fullan (2016) urges to “stay the course through continuity of good direction by leveraging leadership” (p. 50).

Patton (2011) champions a type of evaluation, *developmental evaluation*, which epitomizes Two-County’s general approach to evaluation. Developmental evaluation

occurs in complex adaptive systems, characterized by unpredictable, emergent situations and innovations, requiring flexibility and fluidity in conducting evaluations (Patton, 2011). The evaluation findings inform and guide adaptations, changes and new innovations in response throughout the long-term, ongoing process. Evaluation in this form does not necessarily have an ending, but is continuous, adaptive, and cyclical, which matches Two-County's embraced ideas about evaluation, as well as its commitment to continuous improvement.

Conceptual Framework of Evaluation Capacity Building Revisited

After considering how the conceptual framework based on the literature (explained in greater depth in Chapter Two) compares to the overarching themes from this study, I have found that there is substantial alignment. The general premise of the framework is that legislated accountability such as that of NCLB was anticipated to increase data use in educational entities, as well as systemic evaluation of programs and processes, in order to increase student achievement. From an evaluation lens, if a greater amount of evaluation studies are conducted, the aim is that they are high quality evaluations. Toward building evaluation capacity, a participatory approach to evaluation is needed. The strategy proposed in this framework, *process use*, is associated with *Utilization-Focused Evaluation* (Patton, 2008). In essence, process use is the act of doing evaluation to learn evaluation, and plans to use this tactic must be intentional to increase the likelihood of success. The realization of learning is also increased through delivery by an internal evaluator given that actually building an organization's evaluation capacity is a long-term endeavor.

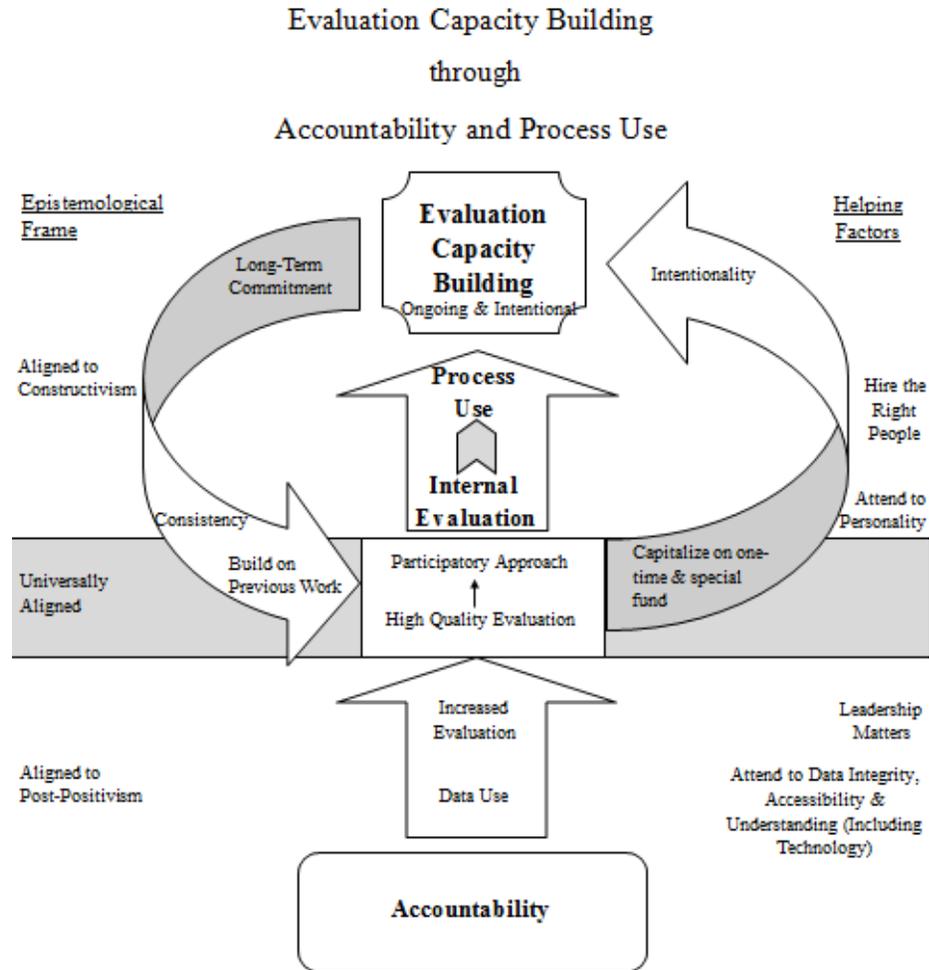


Figure 15. *Conceptual framework of ECB expanded to include study implications.*

Figure 15, above, depicts the conceptual framework of evaluation capacity building through accountability and process use with the added components from the implications section of this chapter, added on the right margin and in the ribbons of *Long-Term Commitment* and *Intentionality*. The connections between the framework and the study implications are described below.

Data use aligns to data integrity, accessibility and understanding. Appropriate data use depends on having integrity – accuracy, validity and reliability – in the data (Cousins et al, 2006). It also requires that data are accessible, which is reliant on technology and

software tools. In addition, technology aids in reporting and understandability by displaying data clearly. Understandability is also enhanced through training and with greater understandability and comfort through a higher sense of competency because of training, data use is likely to rise.

Increased evaluation aligns to leadership matters. As was evident in the literature and the findings of this study, leaders have a significant impact on evaluation (Fullan, 2016, for example). An increase in evaluation has a much greater likelihood when supported and encouraged by leadership, hence, the greater the leadership support, the more sizeable the increase in evaluation.

Attending to personality/disposition of the evaluator is extremely important and linked to the success of internal evaluation. If an internal evaluator is going to be effective, he or she needs to be able to build relationships and collaborate with others in the organization. In addition, he or she must be able to navigate difficult or contentious situations and organizational politics, along with displaying a style of communication that enhances not deters from effectiveness (King, 2007).

Hiring the right people aligns to process use in the sense that individuals with the skills necessary to be able to weave effectual teaching skills intentionally into the activities of conducting evaluations, while engaging participants as active learners, and possessing the technical skill necessary to conduct and lead evaluation is critical (Plowman et al., 2007). The study findings suggest that people with a teaching background have the learned skill and passion for teaching, making the incorporation of preparing lessons more natural.

Capitalizing on one-time and special funds is aligned to the intentionality ribbon because it takes thoughtful and intentional action to strategically use supplemental dollars in a way that creates long-term benefits. The temptation when districts, especially those who are fiscally strapped, is to use additional, non-replicable funding to fill blatant, immediate needs instead of taking the time to intentionally assess and plan where it might make the most significant, sustainable differences. This deliberate planning and action aligns to a continuous improvement mentality which coincides with an evaluative thinking mindset.

Working toward consistency and building on previous work align to the long-term commitment ribbon because both are explicitly associated with ongoing actions over time and neither is necessarily expeditious in nature. Consistency means relying on research-based practices that are known to be effective, not lunging for the next best thing that happens to come into view. It means being practical and thoughtful about when and why changes are made, not following the latest trends (Fullan & Miles, 1992). Building on previous work is similar but speaks more to the principle of assessing process, practices and programs for effectiveness and making modifications as needed as opposed to throwing out the old and completely starting over with the new (which is an approach that is very cost heavy), Building on previous work has an extremely strong connection to evaluation.

Lastly, the implication, make ECB an intentional, ongoing focus, directly aligns with the underpinnings of ECB from the framework. The framework supposition that ECB is long-term commitment and requires intentionality are matched perfectly with the

study implications forming the mechanisms for effective ECB (Forss et al., 2006; Huffman et al., 2008).

Summary

This study outlined eight observations that served as the basis for eight implications. The observations included statements about leadership, the importance of ECB to organizations, the impetus for evaluation, the background and training of the evaluator, the tools and data necessary for ECB, the significance of resources to ECB, and the critical role that legislation, and educational leaders' knowledge of legislation, plays in a district and its relation to ECB.

The implications were grouped into categories relating to resources, people and processes and including aids to increasing ECB effectiveness. These components were pertaining to data and technology, use of one-time and special funding, recognition that leadership is paramount, hiring the right people and paying attention to disposition/personality, working toward consistency and continuity, and committing to ECB as an ongoing, intentional focus. Lastly, a conceptual framework investigating the connection between the existing literature and this study was presented.

Chapter 6:

Conclusions

This study examined the evolution and development of evaluative thinking and evaluation capacity in one metropolitan school district over a thirty-year period through a longitudinal historical case study design. In this chapter, the research questions are answered, limitations of the study are described, and implications for practitioners and future research are suggested. The chapter concludes with final thoughts.

Research Questions and Answers

The study had an overarching research question and three aligning sub-questions that are restated below, followed by study results answering each question.

The overarching research question. What has been the experience of one school district in building evaluation capacity?

Results. The Two-County Independent School District underwent significant transformations related to evaluation capacity and embedding evaluative activities into every day routine over the thirty-year period of this study. Evidence is presented using King's (2007) categories of ECB progress indicators: inputs, processes, and outcomes.

Inputs. From an input standpoint, the department increased sizably, moving from one coordinator and three clerical staff, reporting to the director of finance and planning to its current configuration of a director, six program personnel (four of whom with graduate degrees) and two clerical staff, reporting directly to the superintendent of schools.

Processes. From a process lens, the department increased the amount and types of evaluative activities from only testing support to its current day litany of activities and

support still including testing, but also expanded to in-house research and evaluation capacity, with documented processes for conducting and supporting each; overall and specific data support, including consultation and training on data collection, goal writing, analysis, interpretation, and application; ongoing monitoring and reporting oversight and assistance related to organizational effectiveness, school improvement planning, and central department improvement planning; and direct support and consultation to the school board, superintendent, and cabinet related to all evaluative activities.

Outcomes. In the area of outcomes, at the end of the study period annual reports were created and presented to the school board at regular intervals whereas in years past, interactions with the school board were very infrequent, and data were not collected from a wide variety of measures such as outcome/achievement, demographics, perception of stakeholders, and program. Also, data were shared and discussed at nearly every principal meeting at the elementary, middle school, and high school, and interactions with principals pertaining to evaluative activities were frequent and ongoing; previous department leaders rarely interacted with principals and did not attend their meetings. All district employees had access to the data aligned to their roles and systemic structures were in place to train new and existing staff members.

Along with organizational outputs, each year the research, evaluation, and testing department underwent an internal review, evaluating and reporting progress on self-identified indicators of effectiveness. The department collectively wrote goals and determined the ways in which they would be measured. They also incorporated a staff development component into every monthly department meeting to maintain and increase the knowledge and skills of each member. In addition, they annually evaluated and

discussed reports that were created, ways in which they might improve them, and new reports that could be added to better support and meet the needs of district stakeholders.

Research sub-question 1. What are the perceptions of district leaders as to the importance of evaluation to the mission and vision of the district?

Results. District leaders in Two-County Independent School District overwhelmingly thought evaluation was extremely important to the mission and vision of the district with eleven of the fourteen respondents saying ‘extremely important’ directly along with their rationale as to why they felt that way and the other three providing examples as to why evaluation is important but not overtly stating it was ‘extremely important.’ The findings for this research question stemmed primarily from question one in part two of the individual interviews. The importance of evaluation to the mission and vision of the district was a primary finding of the study and discussed in-depth in Chapters Four and Five.

Research sub-question 2. What is the evidence of evaluation capacity within the organization overtime?

Results. King (2002) appealed to future researchers to assess ECB efforts as part of the process, naming aspects in the areas of inputs, processes, and outcomes. Examples of components in each area were as follows:

Inputs

- increased staffing devoted to evaluation
- decisions to maintain evaluation staff in the midst of declining or reallocated funds

Processes:

- internal capabilities to perform and support evaluative activities
- creation of meaningful and usable reports
- a framework for potential studies and supports for requests

Outcomes:

- increased demand for and use of data
- data referenced and shared in meetings
ongoing data use throughout the organization. (King, 2002, p. 78)

Outcomes. Beginning with the evidential outcomes suggesting increased capacity, currently data were consistently requested, presented, and used in meetings at ever growing rates, which was not observed in early periods of the study time frame. Evidence of this could be found in the analysis of Ms. Norhakah's calendar meetings/events discussed in Chapter Four. In the last five years, for example, whereas the total number of overall events did not change substantially, the average annual percentage of evaluative events changed from 52% to 78%, representing an almost 50% increased change in magnitude. In the same five-year time span, the average annual percentage of general events decreased in magnitude by about 60%. The annual average testing events increased by a magnitude of roughly 30%. Events related to research saw an over 400% change; however, the percentage of calendar events dedicated to research were extremely minimal and, therefore, the values were greatly affected by any change. (See Table 10 in Chapter 4.) In addition, data are part of nearly every principal meeting, shared regularly at district and school level meetings, and shared and reference regularly at cabinet and school board meetings along with at community task force meetings and are a regular part of other community presentations and events.

Processes and inputs. In regard to processes and inputs, Table 12 below summarizes the progression of departmental membership and reporting structure, organizational interactions/collaborations related to evaluative activities, and the expansion of departmental evaluative processes. Given the considerable and notable

growth in each area over the time frame of the study, evidence supported increased evaluation capacity within the district.

Table 12. Evidence of ECB in Two-County ISD Over Time

Evidence of Evaluation Capacity Building in Two-County Independent School District Over Time

Timeline		1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010 - 2015
Departmental Evolution	Department Name	Assessment Department (est. 4/'86)	Student Assessment Department		Student Assessment Department	Research, Evaluation & Testing	Research, Evaluation & Testing
	Supervision of Department	Director of Finance & Planning	Associate Superintendent of Secondary ('92-'98)		Director of Curriculum, Instruction & Assessment ('98-'03)	Associate Superintendent of Secondary (through '08), then Associate Superintendent of K-12	Associate Superintendent of High School ('10-'14) then Superintendent ('15)
	Departmental members	<ul style="list-style-type: none"> •Student Assessment Coordinator ('86-'92) •3 clerical staff (added in stages) 	<ul style="list-style-type: none"> •Coordinator of Student Assessment ('92-'98) •3 clerical staff (in 1999, there was no direct oversight of the dept.) 		<ul style="list-style-type: none"> •Coordinator of Research and Evaluation ('99-'01) •Student Assessment Facilitator ('99-'03) •3 clerical staff 	<ul style="list-style-type: none"> •Principal on Special Assignment ('03-'06) •Student Assessment Facilitator •Office Coordinator (added '02) •District Analyst (added '06) •6 Achievement Analysts (plus supported 6 more housed at HS) •Technician •Data coordinator (added '08) •2 clerical staff 	<ul style="list-style-type: none"> •Director of Research, Evaluation and Testing ('07-present) •Office Coordinator •District Analysts (added a second one - '15) •1-4 Achievement Analysts •0-2 Evaluation Specialists •Technician •Data Coordinator •2 clerical staff
Departmental Leadership		Spawling	Spawling/Hadley/Osborne	Mostenby		Mostenby/Cadence	Cadence/Langley
		Mostenby	Mostenby/Wheaton	Ritenour/Wheaton	Wheaton	Wheaton/Bennett	Porter/Thomason/Langley
		Mostenby/Eickner	Eickner/Dooren	Dooren/Kincade	Kincade/Anthony/Jordasen	Jordasen/Norhakah	Norhakah
Internal Evaluation Interactions / Collaborations		Director of Finance	Curriculum Leaders & Department Members	Secondary Associate Superintendent/ Curriculum Leaders & Department Members	Elementary & Secondary Associate Superintendents/ Curriculum Leaders & Department Members/ Some Principals	All Associate Superintendents/ Curriculum Leaders & Department Members/ All Principals & Some school-level staff	Superintendent & School Board/ Cabinet/ All Central Departments/ All Principals & School-level staff
Internal Evaluation Processes		Content area studies (through curriculum) / Testing	Content area studies (through curriculum) / Testing	Curriculum studies/ Testing/ Evaluation/ Beginnings of others using data (beyond leaders)	Evaluation/ Testing/ Data support/ Curriculum studies	Evaluation/ Testing/ Data support/ Research oversight/ School improvement support	Evaluation/ Testing/ Data support/ Research oversight/ School improvement support/ Data collection support/ full organization & school board support related to data & evaluation

Research sub-question 3. What contextual events expedite or delay evaluation capacity building?

Results. The most expeditious factor toward evaluation capacity building in Two-County ISD was likely the advancement of technology and the infrastructure that sustained it. Both hardware and software provided the means for all members of the

organization to access and use data and, thus, to engage in evaluative activities.

Technology fostered the advancement of analysis and allowed even basic users to perform sophisticated analyses. It also increased time efficiencies, turning tasks that used to take weeks or longer into immediate, on-demand actions.

Another expediting factor was the process of data maintenance and integrity.

Establishing solid and reliable data practices provided accuracy and validity in the data, building confidence in users and trust in the system. It also provided the platform for each generation of leaders to expound on the footings of his or her predecessor, as evidenced in the progression of focus within the superintendents. (See Table 13.)

Table 13. *Technology Advancements and Superintendent Data Focus Over Time*

Timeline	1985-1989	1990-1994	1995-1999	2000-2004	2005-2009	2010 - 2015
Technology Advancement	Computing by high powered calculators First computers are purchased	Computer use expands but still not systemic	All staff have computers but network/infr astructure is minimal	Focus on data structures/ technology Student information system implementation - each site operates as a separate system in the software Developed processes for data security and data integrity - linked systems for increase automaticity Added measures & data collection processes to increase data availability	Focus on data accessibility and understandability for all stakeholders Data warehouse implementation - wide variety of student data in one location - rights and security based on user role - data available in aggregate (at any unit of analysis) and individual student levels Added Common assessment tool	Continued improvement and refinement of data warehouse reporting and functionality Dashboard with reports and alerts for defined thresholds implemented for school and central administrators Upgraded network/ infrastructure for higher functionality and data support
Superintendent Data Focus	Mainly budget	Mainly budget	Budget, Enrollment/ demographics & Student outcomes - mainly testing		Budget, Enrollment/ demographic, Expanded student outcomes & Perception	Budget, Enrollment/ demographic, Expanded student outcomes, Expanded perception, & Alignment through the system

Adapted based on Jorgensen and Hoffman (2003); Fitzpatrick, Sanders, and Worthen (2004); and Russ-Eft and Preskill (2009)

With each superintendent, discounting Osborne’s era (which was much more a time of maintenance, not advancement), the previous data focus was supplemented with a new data set as the use of each area of information became part of the standard work of the organization. For instance, during the Spawling era, the district was extremely focused on the budget as the primary data set used to inform decisions. When Mostenby became superintendent, examining and using the budget data were standard or common

practice, and data review expanded to demographic/enrollment and district-wide testing data.

When Cadence took the helm, analyzing and using enrollment and testing data were commonplace, and as a system, Two-County was able to expand to really digging into other indicators of student achievement, such as a heightened focus on course-taking aligned to college and career readiness, the achievement gap between typically-advantaged and typically-disadvantaged groups of students, and graduation indicators. In Langley's era, the district was able to expand once again, into a greater focus on non-academic indicators such as discipline, attendance, perceptions of stakeholders, and aligning the system of reporting and monitoring tools and practices.

Beyond tangible and process factors, people, specifically the leaders – the superintendent, the associate superintendent (who had most often been the supervisor of the department lead), and the direct departmental oversight – were significant factors in the advancement (or hindrance) of evaluation capacity building. When the leaders were champions or at least supporters of evaluation, efforts improved. In addition, when the direct oversight of the department was an academically-trained evaluator, growth was seemingly accelerated. There was more evidence of evaluative activity, at stronger levels of connection under Dr. Kincade and Ms. Norhakah than any other department heads, indicating that evaluation capacity seemingly increased at greater rates and more effectively when a trained evaluator with commitment and passion for ECB was leading the department.

The main factors that hindered evaluation capacity building included budgetary constraints (and competition for resources), significantly disruptive events that consumed

the whole organization, and lack of leadership support; all of which could be confounded by the others. Certainly, if leadership did not support or appreciate evaluative activities, it would be difficult to gain the necessary momentum and authority to increase evaluations and engage participants to increase evaluation capacity within the organization. In addition, without leadership support, resources would likely not be allocated to support evaluative activities and staff to grow them. Even if funding was initially provided, without the recognition of the value of evaluative activities acknowledged and endorsed, re-distribution or elimination of funding was probable.

Sometimes, especially in school systems under much public scrutiny and attention with many moving (*human*) parts where issues or challenges might arise, there were circumstances or events that halted the organization and required full attention from the masses. In education, these events might come in the form of a death of a renowned employee or community member, other tragedies that significantly impacted students and families, employee strikes (given that people were the commodities of the system and salaries made up the majority of the budget, along with the fact that the accompanying tension and strife eroded trust), and litigations. Events such as these put a halt on the system and ECB efforts. The length of time or magnitude of the cessation was directly dependent on the event.

Lastly, one huge impactful event, albeit a national not local occurrence, was No Child Left Behind (NCLB) Act and the era of accountability. By all accounts, NCLB both helped and hindered evaluative activity and ECB in school districts, and Two-County was no exception. The accountability movement propagated the understanding and acknowledgement that schools needed to change and provided a monumental catalyst

to make that happen. This study had the unique perspective of examining 15 years prior to NCLB and 15 years following. The advancement and the urgency for data use toward effective programming in the two halves of the study were markedly different, painting NCLB as a strong accelerator of ECB. In some ways, however, NCLB was a significant distractor, forcing districts to focus on testing and the management of standards alignment, thus slowing efforts of ECB through time, attention, and resource manipulation. As was the case in Two-County ISD, when staff was able to handle the testing and external accountability implications, they were then able to capitalize on the structures and systems put in place for testing to intentionally re-focus on evaluation and ECB.

Implications for Practitioners

Many researchers called for empirical studies on ECB to add to the existing body of literature (Amo & Cousins, 2007; Baizerman et al., 2002a; Brandon & Higa, 2004; Labin et al., 2012; Preskill & Boyle, 2008; Stockdill et al., 2002). The study of evaluation capacity building in Two-County ISD provides information to add to the discussion. Labin et al. (2012) found that organizations that typically engage in ECB efforts are those from the education and health fields so this research may be particularly of interest to them.

This study examined the context, processes, and activities that affect evaluative activities and ECB in K-12 education. For educational leaders who are in the midst of ECB in their organizations, the experiences of Two-County may provide insight, affirmation, or guidance as they embark on this journey. For practitioners who are trying to decipher why their efforts may or may not be working and are seeking possible

avenues to tread down or avoid, this study may be of service. For historians who are looking for a *good read* about the path toward ECB of one school district, Chapter Four provides a story of growth. This study benefits the knowledge base in education by providing school districts empirical research on the helps and hindrances that can propel an organization toward ECB. It also benefits the knowledge base of evaluation by providing access into the propelling and restraining factors of ECB from the perspective of practitioners. This research also provides a unique opportunity to trace the 30-year history of evaluative activity and ECB work in an educational system.

Implications for Further Research

Harner and Preskill (2007) incite further research on the connection between process use and ECB, which this study examined and found successful, however, process use was a supplemental aspect of the study – the primary means toward ECB and was not the focal concept – suggesting further research may be beneficial on this connection. In addition, comparisons to which strategies of ECB are most effective may benefit the field. Maybe any intentional method or combination thereof would yield positive results (and the dependent variable is the intentionality). Or maybe certain strategies and activities work better in certain contexts. Given that questions exist, further research is warranted.

Amo and Cousins (2007) question the depth of impact of process use and, from the lens of this study, its impact on ECB. They suggest that some participants merely gain conceptual knowledge, where others experience changes in actions or beliefs, and still others undergo attitude or affect transitions. This is an intriguing concept and would be

valuable future research connected specifically to the extent to which ECB is experienced in relation to these levels of change.

Specifically related to this study, one of my underlying premises was that the district's perpetual financial constraints shaped attitudes, actions, and acceptance of evaluation and the urgency to build capacity to engage in evaluative activities, as well as its relationship with the community, which was founded through interviews with stakeholders. This begs the question: Do districts with more fiscal stability or even fiscal abundance experience ECB and the need for evaluation differently? Again, this is an area worth studying.

Another consideration from the study is what level of "capacity" is optimal. Certainly, there is not the need for every staff member to be a trained evaluation expert, but staff should have the underlying thought process to ask questions about their effectiveness and the effectiveness of their program in reaching the anticipated and desired outcomes. So when is "enough capacity"? Stockdill et al. (2002) speak to the idea that ECB is never done, but when does an organization reach a *maintaining* state as opposed to a *building* state?

Lastly, Two-County benefits from the alliance and like-mindedness of the school board and the district leaders related to the importance of evaluative activities and ECB. Some districts likely do not share this support. It would be beneficial to look at this type of study from the lens of the school board as well as from the standpoint of the school board's perception of and confidence in district leadership. What is the effect on ECB when the school board and district leadership have differing views, and how does that change when one is the advocate for ECB over the other or vice versa? Does the evidence

of evaluation and the connective strength of evaluative activities change based on the impetus for the work? Is it a chicken or the egg scenario? Is there more evaluative activity present in school board minutes when there is a greater advocacy for evaluative activity in school leadership, or is there a greater impetus for the development of ECB when school boards are the drivers? What contextual factors make a difference?

Seemingly, the area of evaluation capacity building still has “much room for growth and refinement” (Labin et al., 2012, p. 23). It remains “an area ripe for exploration” (Preskill & Boyle, 2008, p. 457).

Epilogue

As I finish this project, I feel like both a mirror and a light as I sit quietly in the darkness with soft Celtic harp music playing in the background. I am a mirror in the sense that I am reflecting on *from whence I came*, what has brought me to this point, and how it has impacted my work and role as director of research, evaluation, and testing in Two-County School District. As I draw upon my training as an educator, I am able to plan lessons with proven strategies to increase the likelihood of learning through process use. As a trained instructional coach, I am able to employ concepts of reflective practice and effective adult professional development to increase the likelihood of application. As a trained administrator, I am able to navigate the sometimes contentious waters of participation through compliance versus desire to increase the likelihood of ongoing engagement within the process. As a trained evaluator, I am able to provide the expertise to ensure the process and outcomes are in alignment with the program evaluation standards and that attention is paid to intentionality of building evaluation capacity, increasing the likelihood that it actually will.

Torres (1994) speaks to a simpler version of the convergence of experiences and skills that has prepared me for my current role, emphasizing the benefits of possessing both knowledge of the organization and knowledge of evaluation; I completely concur. Having served in all these roles within the district provides me the advantage of years of building relationships with stakeholders in all employee, parent, and community groups, establishing credibility and demonstrating trustworthiness and dedication, affording me the inroads and opportunities to collaborate and partner with the full continuum of stakeholders, thus increasing the likelihood that participatory evaluation and ECB will occur. (Personality is important in laying the groundwork for these things.) Certainly there will be no learning by doing if there is no doing! First steps first...

I am also keenly aware of the capacity of my own department and the need for continual encouragement and team building necessary for their sustainability. The work is hard; the work is stressful; and most often the work is urgent. In addition, my expectations are high, which fortunately is a value shared by my departmental team members, also. Hiring well is critical.

For a department of nine in a district spanning 13 communities, supporting nearly 40,000 students in a K-12 system with over 5000 staff members, it sometimes seems daunting and exciting all at the same. We have had numerous conversations about the demand for data, support, and services, and the strain it causes within the ranks of the research, evaluation, and testing department. Although cognitively we see it as a positive challenge to have, in the face of deadlines, sometimes seeing the positive is blocked from view by the mounting piles of projects. As part of my allegiance to ECB, I see it as part of my role to continually cultivate ways in which the department can interact and support

the organization, thus creating the ongoing learning environment to foster evaluative activity in all areas, adding to the authentically-felt supply and demand conundrum.

I am aware of the possible perils to ECB by building a department with many trained evaluators; whereas I do not see it as a current issue, given that the supply and demand challenge is real and budget issues still plague us. Nonetheless, there is danger in the potential expectation that with a larger department, ECB is not necessary. Why would the department not just do all the evaluative work since they have people who are trained to do it? The concept of building capacity by growing the number of evaluators, not growing the knowledge and skill within all employees is something of which to be aware.

There is danger in the organization thinking ECB is only necessary because of limited resources. There is also danger in the department falling into the ‘I’ll do it myself’ trap; it is many times faster and easier in the short-term, but in the long-run, it is not as organizationally or individually beneficial. Without awareness, commitment and deliberate precautions in place, this could be an unintended outcome (Bourgeois & Cousins, 2013; King, 2007). Contemplating where we are in relation to this issue, propels me into the light.

I am a light in the sense that I am looking forward at the path that lies ahead. As I embody Two-County’s continuous improvement mentality, I cannot help think about the things I need to do tomorrow and in the upcoming days, months, and years to advance and strengthen evaluative thinking and capacity within my district. I believe in the idea that ECB is “continuous sustenance – ongoing intentional work necessary to ensure that evaluation studies and their uses continue to be asked for and carried out; continuous sustenance of supportive organizational structures, cultures and every day practices”

(Stockdill et al., 2002, p. 13). I believe it because I live it and embrace it and, to that end, am continually contemplating how to make it happen, faster, better, more comprehensively across all stakeholder groups. I also realize that one of the next things to do in this journey is to clearly articulate guidance for when we monitor programs, policies, and practices and when we fully evaluate them, especially in light of the supply and demand challenges highlighted above. Duignan (2003) offers insight saying that monitoring should align to accountability and evaluation should be conducted on a selective and strategic basis, underscoring the cost involved in evaluation, however, in evaluation everything is contextual so we will have to wrestle with how this guides and applies to our work in Two-County ISD.

NCLB and, more specifically, the deluge of public accountability, created an enormous watershed moment for public education. Looking forward, we are again at a critical point. Torres (1994) illustrates this idea that when evaluation is externalized (done for others), it is more aligned with accountability, and when it is internalized (done for yourself) is it more aligned with improvement which align to one of the findings of this study. Carden and Earl (2007) talk about evaluative thinking as somewhat of a fulcrum, balancing the opportunity to learn (and I am adding for the betterment of the individual and the organization) with the need for accountability. As I think forward, I am struck by the idea that it is part of my role as the ECB evaluator to illuminate the concept of accountable *to whom*? Anderson (2005) outlines three types of accountability: accountability to a governmental agency – the bureaucracy for adherence to rules (such as teacher licensure); accountability to peers for standards associated with a professional organization (such as the Joint Committee’s Standards of Educational Evaluation); and

accountability to the general public for student learning (such as NCLB). I contend there is a fourth type of accountability – accountability to one’s self – personal and professional integrity. When Patton (2013) wrote, “The highest form of accountability is internal” (p. 13), I am enlightened to believe he was referring to personally internal more than internal to the organization.

On the header wall in the research, evaluation, and testing department is an overarching statement of who we are and for what we are responsible, developed through a collective process involving all department members. The statement says, “Research, Evaluation and Testing: Working collaboratively to support district accountability and integrity in our data, our decisions and ourselves.” ECB is the fulcrum and foundation to making this happening. We will continue forth in Two-County Independent School District because it is the right thing to do, for our students, our families, our staff, and our community – for ourselves.

So what is next for me? Thinking about evaluation connection strength and what constitutes evaluative activities as part of this venture, I am contemplating what a continuum of evaluation would look like and how it might be applicable for school systems. In the world of research, evaluation, and testing, we discuss the idea of districts employing a comprehensive or balanced assessment system – assessments for different purposes to provide information to meet a broad spectrum of needs. I believe this concept applies to evaluation, as well; a comprehensive or balanced evaluation system to inform our decisions and increase effectiveness of all operations within districts along a broad spectrum of need. Maybe conceptualizing and defining that more thoroughly is the “now what?” for me. My advisor says next steps are to “publish!”

I say maybe right now it is time to put the mirror away, shut off the light, and sleep for a couple of hours – just a couple; resting assured that tomorrow is another day, bringing with it all the possibilities and opportunities one might envision...

References

- America Recover and Reinvestment Act – ARRA. (2009). Public Law No. 111-5. 111th Congress. *5 Congressional Record 123*, 182 Stat.
- Amo, C., & Cousins, J. B. (2007). Going through the process: An examination of the operationalization of process use in empirical research on evaluation. In J. B. Cousins (Ed.), *Process use in theory, research and practice. New Directions for Evaluation, 116*, 5-26.
- Anderson, J. A. (2005). *Accountability in education*. Paris, France: International Institute for Educational Planning.
- Baizerman, M., & Compton, D. W. (2009a). A perspective on managing evaluation. In D. W. Compton & M. Baizerman (Eds.), *Managing program evaluation: Towards explicating a professional practice. New Directions for Evaluation, 121*, 7-15.
- Baizerman, M., & Compton, D. W. (2009b). What did we learn from the case studies about managing evaluation? In D. W. Compton & M. Baizerman (Eds.), *Managing program evaluation: Towards explicating a professional practice. New Directions for Evaluation, 121*, 79-86.
- Baizerman, M., Compton, D. W., & Stockdill, S. H. (2002a). New directions for ECB. In D. W. Compton, M. Baizerman, & S. H. Stockdill (Eds.), *The art, craft, and science of evaluation capacity building. New Directions for Evaluation, 93*, 109-119.

- Baizerman, M., Compton, D. W., & Stockdill, S. H. (2002b). Summary and analysis of the case studies: Themes across the cases. In D. W. Compton, M. Baizerman, & S. H. Stockdill (Eds.), *The art, craft, and science of evaluation capacity building. New Directions for Evaluation, 93*, 101-108.
- Balthasar, A. (2008). Institutional design and utilization of evaluation: A contribution to a theory of evaluation influence based on Swiss experience. *Evaluation Review, 33*(3), 226-256.
- Bickel, W. E., Millett, R., & Nelson, C. A. (2002). *Challenges to the role of evaluation in supporting organizational learning in foundations*. Washington, DC: Council on Foundations.
- Blok, H., Slegers, P., & Karsten, S. (2008). Looking for a balance between internal and external evaluation of school quality: Evaluation of the SVI model. *Journal of Educational Policy, 23*(4), 379-395.
- Bourgeois, I., & Cousins, J. B. (2013). Understanding dimensions of organizational evaluation capacity. *American Journal of Evaluation, 34*(3), 299-319.
- Brandon, P. R., & Higa, T. A. F. (2004). An empirical study of building the evaluation capacity of K-12 site-managed project personnel. *Canadian Journal of Program Evaluation, 19*(1), 125-142.
- Carden F., & Earl, S. (2007). Infusing evaluative thinking as process use: The case of the International Development Research Centre (IDRC). In J. B. Cousins (Ed.), *Process use in theory, research and practice. New Directions for Evaluation, 116*, 61-73.

- Chelimsky, E. (1994). Making evaluation units effective. *Handbook of practical program evaluations*. San Francisco: Jossey-Bass, 493-509.
- Chelimsky, E. (1997). Thoughts for a new evaluation society. *Evaluation*, 3(1), 97-118.
- Clifford, D. L., & Sherman, P. (1983). Internal evaluation: Integrating program evaluation and management. In A. J. Love (Ed.), *Developing effective internal evaluation*. *New Directions for Program Evaluation*, 20, 23-45.
- Collinson, V., Cook, T. F., & Conley, S. (2010). Organizational learning in schools and school systems: Improving learning, teaching, and leading. *Theory Into Practice*, 45(2), 107-116.
- Compton, D. W. (2009). Managing studies versus managing for evaluation capacity building. In D. W. Compton & M. Baizerman (Eds.), *Managing program evaluation: Towards explicating a professional practice*. *New Directions for Evaluation*, 121, 55-69.
- Conley-Tyler, M. (2005). A fundamental choice: Internal or external evaluation? *Evaluation Journal of Australia*, 9(1 & 2), 3-11.
- Cousins, J. B., Goh, S. C., & Clark, S. (2006). Data use leads to data valuing: Evaluative inquiry for school decision making. *Leadership and Policy in Schools*, 5(2), 155-176.
- Cousins, J. B., Goh, S. C., Clark, S., & Lee, L. E. (2004). Integrating inquiry into the organizational culture: A review and synthesis of the knowledge base. *The Canadian Journal of Program Evaluation*, 19(2), 99-141.

- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation. New Directions for Evaluation, 80*, 87-105.
- Crossan, M. M., Lane, H. W., & White, R. E. (1999). An organizational learning framework: From intuition to institution. *The Academy of Management Review, 24*(3), 522-537.
- Crotty, M. (1998). *The foundations of social research*. Thousand Oaks, CA: Sage.
- Department of Education – DOE. (2005). Scientifically-based evaluation methods; Notice part II. *Federal Register, 70*(15), 3586-3589.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2009). *Internet, mail, and mixed-mode surveys: The tailored design method (3rd ed.)*. Hoboken, NJ: John Wiley & Sons, Inc.
- Duignan, P. (2003). Mainstreaming evaluation or building evaluation capacity? Three key elements. In J. B. Barnette & J. R. Sanders (Eds.), *The mainstreaming of evaluation. New Directions for Evaluation, 99*, 7-21.
- Eddy, R. M., & Berry, T. (2008). Challenges and opportunities revealed by the impact of No Child Left Behind on evaluation as a discipline. In T. Berry & R. M. Eddy (Eds.), *Consequences of No Child Left Behind for educational evaluation. New Directions for Evaluation, 117*, 95-103.
- Every Student Succeeds Act – ESSA. (2015). Public Law No. 114-95. 114th Congress, *161 Congressional Record S1177*, 129 Stat.

- Fetterman, D. (2003). Fetterman-House: A process use distinction and a theory. In C. A. Christie (Ed.), *The practice-theory relationship in evaluation. New Directions for Evaluation, 97*, 47-52.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd Ed.). Boston, MA: Pearson.
- Forss, K., Rebien, C., & Carlsson, J. (2002). Process use of evaluations: Types of use that precede lessons learned and feedback. *Evaluation, 8*(1), 29-45.
- Forss, K., Kruse, S. E., Taut, S., & Tenden, E. (2006). Chasing a ghost? An essay on participatory evaluation and capacity development. *Evaluation, 12*(1), 128-144.
- Fullan, M. G. (2016). *The new meaning of educational change* (5th Ed.). New York: Teacher's College Press.
- Fullan, M. G. & Miles, M. B. (1992). Getting reform right: What works and what doesn't, *Phi Delta Kappan, 745-752*.
- García-Iriarte, E., Suárez-Balcazar, Y., Taylor-Ritzler, T., & Luna, M. (2011). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation, 32*(2), 168-182.
- Goals 2000: Educate America Act (1994). Public Law No. 103-227. 103rd Congress, 227 *Congressional Record 108*, 125 Stat.
- Gregory, K., & Clarke, M. (2003). High-stakes assessments in England and Singapore. *Theory in Practice, 42*(1), 66-74.

- Guillén-Woods, B. F., Kaiser, M. A., & Harrington, M. J. (2008). Responding to accountability requirements while promoting improvement. In T. Berry & R. M. Eddy (Eds.), *Consequences of No Child Left Behind for educational evaluation. New Directions for Evaluation, 117*, 59-70.
- Hannan, M., & Freeman, J. (1984) Structural inertia and organizational change. *American Sociological Review, 49*, 149-164.
- Hanushek, E. A., Machin, S., & Woessmann, L. (2011). *Economics of education*. The Netherlands, North Holland: Elsevier.
- Harmon, H. L. (2006). Beyond compliance: Building evaluation capacity in the ESA (Educational Service Agency). *Perspectives, 12*, 17-27.
- Harnar, M. A., & Preskill, H. (2007). Evaluators' description of process use: An exploratory study. In J.B. Cousins (Ed.), *Process use in theory, research and practice. New Directions for Evaluation, 116*, 27-44.
- Hoefler, R. (2000). Accountability in action? Program evaluation in nonprofit human service agencies. *Nonprofit Management & Leadership, 11*(2), 167-177.
- Huber, G. P. (1991). Organizational learning: The contributing processes and the literatures. *Organizational Science, 2*(1), 88-115.
- Huffman, D., Lawrenz, F., Thomas, K., & Clarkson, L. (2006). Collaborative evaluation communities in urban schools: A model of evaluation capacity building for STEM education. In D. Huffman & F. Lawrenz (Ed.), *Critical issues in STEM evaluation. New Directions for Evaluation, 109*, 73-85.
- Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation, 29*(3), 358-365.

- Improving America's Schools Act – IASA (1994). Public Law No. 103-382. 103rd Congress, *382 Congressional Record* 108, 3518 Stat.
- Jorgensen, M. A., & Hoffman, J. (2003). History of the No Child Left Behind Act of 2001 (NCLB). *Pearson Assessment Report*, August, Revision 1: December, 1-8.
- Judge, T. A., Thoresen, C. J., Pucik, V., & Welbourne, T. M. (1999). Managerial coping with organizational change: A dispositional perspective. *Journal of Applied Psychology*, *84*(1), 107-122.
- Kim, D. H. (1993). The link between individual and organizational learning. *Sloan Management Review*, *35*(1), 37-50.
- King, J. A. (2002). Building the evaluation capacity of a school district. In M. Baizerman, D. W. Compton, & S. H. Stockdill (Eds.), *The art, craft, and science of evaluation capacity building*. *New Directions for Evaluation*, *93*, 63-80.
- King, J. A. (2007). Developing evaluation capacity through process use. In J. B. Cousins (Ed.), *Process use in theory, research and practice*. *New Directions for Evaluation*, *116*, 45-59.
- King, J. A., & Rohmer-Hirt, J. (2011). Internal evaluation in American public school districts: The importance of externally driven accountability mandates. In B. B. Volkov & M. E. Baron (Ed.), *Internal evaluation in the 21st century*. *New Directions for Evaluation*, *132*, 73-86.
- King, J. A., & Stevahn, L. (2013). *Interactive evaluation practice: Mastering the interpersonal dynamics of program evaluation*. Thousand Oaks, CA: Sage.
- King, J. A., & Volkov, B. (2005). A framework for building evaluation capacity based on the experiences of three organizations. *CURA Reporter*, *35*(3), 10-16.

- Kotter, J. P. (1996). *Leading change*. Boston, MA: Harvard Business Review Press.
- Kyriakides, L., & Campbell, R. J. (2004). School self-evaluation and school improvement: A critique of values and procedures. *Studies in Educational Evaluation, 30*, 23-36.
- Labin, S. N., Duffy, J. L., Meyers, D. C., Wandersman, A., & Lesesne, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation, 33*(3), 307-338.
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Thousand Oaks, CA: Sage.
- Love, A. J. (1991). *Internal evaluation: Building organizations from within*. Newbury Park, CA: Sage.
- Love, A. J. (1998). Internal evaluation: Integrating evaluation and social work practice. *Scandinavian Journal of Social Welfare, 7*, 145-151.
- Love, A. J. (1983). The organizational context and the development of internal evaluation. *New Directions for Program Evaluation, 20*, 5-22.
- Mabry, L. (2008). Consequences of No Child Left Behind on evaluation purpose, design, and impact. In T. Berry & R. M. Eddy (Eds.), *Consequences of No Child Left Behind for educational evaluation. New Directions for Evaluation, 117*, 9-20.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organizational Science, 2*(1), 71-87.
- March, J. G. (1981) Footnotes to organizational change. *Administrative Science Quarterly, 26*(4), 563-577.
- Mathison, S. (1991). What do we know about internal evaluation? *Evaluation and Program Planning, 14*, 159-165.

- McDonald, B., Rogers, P., & Kefford, B. (2003). Teaching people to fish? Building the evaluation capacity of public sector organizations. *Evaluation*, 9(1), 9-29.
- McEwen, N. (1995). Introduction: Accountability in education in Canada. *Canadian Journal of Education*, 20(1), 1-17.
- Merriam, S. B. (2001). *Qualitative research and case study applications in education*. San Francisco, CA: Jossey-Bass Publishers.
- Mills, J. I. (2008). A legislative overview of No Child Left Behind. In T. Berry & R. M. Eddy (Eds.), *Consequences of No Child Left Behind for educational evaluation*. *New Directions for Evaluation*, 117, 9-20.
- Møller, J. (2009). School leadership in an age of accountability: Tensions between managerial and professional accountability. *Journal of Educational Change*, 10(1), 37-46.
- National Commission on Educational Excellence – NCEE. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- Nelson, M., & Eddy, R. M. (2008). Evaluative thinking and action in the classroom. In T. Berry & R. M. Eddy (Eds.), *Consequences of No Child Left Behind for educational evaluation*. *New Directions for Evaluation*, 117, 37-46.
- Nevo, D. (1994). Combining internal and external evaluation: A case for school-based evaluation. *Studies in Educational Evaluation*, 20, 87-98.
- Nevo, D. (2001). School evaluation: Internal or external? *Studies in Educational Evaluation*, 27, 95-106.

- No Child Left Behind Act – NCLB. (2001). Public Law No. 107-110. 107th Congress, *110 Congressional Record 1425*, 115 Stat.
- Patton, M. Q. (2007). Process use as a usefulism. In J. B. Cousins (Ed.), Process use in theory, research and practice. *New Directions for Evaluation*, 116, 99-112.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th Ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: The Guilford Press.
- Peshkin, A. (1988). In search of subjectivity – One’s own. *Educational Researcher*, 17(7), 17-21.
- Plowman, D. A., Baker, L. T., Beck, T. E., Kulkarni, M., Solansky, S. T., & Travis, D. V. (2007) Radical change accidentally: The emergence and amplification of small change. *Academy of Management Journal*, 50(3), 515-543.
- Preskill, H. (2008). Evaluation’s second act: A spotlight on learning. *American Journal of Evaluation*, 29(2), 127-138.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459.
- Preskill, H., & Caracelli, V. (1997). Current and developing conceptions of use: Evaluation use TIG survey results. *American Journal of Evaluation*, 18(3), 209-225.
- Preskill, H., & Torres, R. T. (2000). The learning dimension of evaluation use. In V .J. Caracelli & H. Preskill (Eds.), The expanding scope of evaluation use. *New Directions for Evaluation*, 88, 25-37.

- Preskill, H., Zuckerman, B., & Matthews, B. (2003). An exploratory study of process use: Findings and implications for future research. *American Journal of Evaluation*, 24(4), 423-442.
- Quattrone, P., & Hopper, T. (2001) What does organizational change mean? Speculations on a taken for granted category. *Management Accounting Research*, 12, 403-435.
- Robinson, T. T., & Cousins, J. B. (2004). Internal participatory evaluation as an organizational learning system: A longitudinal case study. *Studies in Educational Evaluation*, 30, 1-22.
- Rodosky, R. J. & Muñoz, M. A. (2009). Slaying myths, eliminating excuses: Managing for accountability by putting kids first. In D. W. Compton & M. Baizerman (Eds.), *Managing program evaluation: Towards explicating a professional practice*. *New Directions for Evaluation*, 121, 43-54.
- Russ-Eft, D., & Preskill, H. (2009). *Evaluations in organizations: A systematic approach to enhancing learning, performance, and change* (2nd Ed.). New York, NY: Basic Books.
- Sanders, J. R. (2002). Presidential address: On mainstreaming evaluation. *American Journal of Evaluation*, 23(3), 253-259.
- Schein, E. H. (1996). Kurt Lewin's change theory in the field and in the classroom: Notes toward a model of managed learning. *Systems Practice*, 9(1), 27-47.
- Schmitt, L. N. T., & Whitsett, M. D. (2008). Using evaluation data to strike a balance between stakeholders and accountability systems. In T. Berry & R. M. Eddy (Eds.), *Consequences of No Child Left Behind for educational evaluation*. *New Directions for Evaluation*, 117, 47-58.

- Schneider, B., Brief, A. P., & Guzzo, R. A. (1996). Creating a climate and culture for sustainable organizational change. *Organizational dynamics*, 24(4), 7-19.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed). Newbury Park, CA: Sage.
- Senge, P. (1990). *The fifth discipline: The art and practice of organizational learning*. New York, NY: Currency Doubleday.
- Shula, L. M., & Cousins, J. B. (1997). Evaluation use: Theory, research, and practice since 1986. *American Journal of Evaluation*, 18(3), 195-208.
- Silberman, M. (2006). *Active training: A handbook of techniques, designs, case examples, and tips*. San Francisco, CA: Pfeiffer.
- Snibbe, A. C. (2006). Drowning in data. *Stanford Social Innovation Review*, Fall, 39-45.
- Sonnichsen, R. C. (2000). *High impact internal evaluation: A practitioner's guide to evaluating and consulting inside organizations*. Thousand Oaks, CA: Sage.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. In D. W. Compton, M. Baizerman, & S. H. Stockdill (Eds.), *The art, craft, and science of evaluation capacity building*. *New Directions for Evaluation*, 93, 7-25.
- Strahan, D., Cooper, J., & Ward, M. (2001). Middle school reform through data and dialogue: Collaborative evaluation with 17 leadership teams. *Evaluation Review*, 25(1), 72-99.
- Stufflebeam, D. L. (1997). Strategies for institutionalizing evaluation: Revisited. *Occasional Papers Series*, 18, 3-38.

- Suárez-Herrera, J. C., Springett, J., & Kagan, C. (2009). Critical connections between participatory evaluation, organizational learning and intentional change in pluralistic organizations. *Evaluation, 15*(3), 321-342.
- Sutherland, S. (2004). Creating a culture of data use for continuous improvement: A case study of an Edison Project school. *American Journal of Evaluation, 25*(3), 277-293.
- Taut, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation, 28*(1), 45-59.
- Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. In M. T. Braverman, M. Engle, M. E. Arnold, & R. A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation, 120*, 55-69.
- Taylor-Ritzler, T., Suárez-Balcazar, Y., García-Iriate, E., Henry, D. B., & Balcazar, F. E. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation, 34*(2), 190-206.
- Torres, R. T. (1991). Improving the quality of internal evaluation: The evaluator as consultant-mediator. *Evaluation and Program Planning, 14*, 189-198.
- Torres, R. T. (1994). Linking individual and organizational learning: The internalization and externalization of evaluation. *Evaluation and Program Planning, 17*(3), 327-337.
- Torres, R. T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation, 22*(3), 387-395.

- Tsoukas, H., & Chia, R. (2002). On organization becoming: Rethinking organizational change. *Organization Science, 13*(5), 567-582.
- Tyler, R. W. (1987). The five most significant curriculum events in the twentieth century. *Educational Leadership, 44*(4), 36-38.
- Tyler, R. W. (2013). *Basic principles of curriculum and instruction*. Chicago, IL: University of Chicago Press.
- Vo, A. D. T. (2013). *Toward a definition of evaluative thinking* (Order No. 3562824). (1377154328). Retrieved from <http://search.proquest.com/docview/1377154328?accountid=891>
- Webb, R., Vulliamy, G., Häkkinen, K., & Hämäläinen, S. (1998). External inspection or school self-evaluation? A comparative analysis of policy and practice in primary schools in England and Finland. *British Educational Research Journal, 24*(5), 539-556.
- Weick, K. E., & Quinn, R. E. (1999) Organizational change and development. *Annual Review of Psychology, 50*, 361-386.
- Winberg, A. (1991). Maximizing the contributions of internal evaluation units. *Evaluation and Program Planning, 14*, 167-172.
- Worthen, B. R., & Sanders, J. R. (1987). The concept of evaluation: An overview. *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman, 21-33.
- Yarbrough, D. B., Shula, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd Ed.). Thousand Oaks, CA: Sage.

Yin, R. K. (2009). *Case study research: Designs and methods* (4th Ed.). Thousand Oaks, CA: Sage.

Zapico-Goñi, E. (2007). Matching public management, accountability and evaluation in uncertain contexts: A practical suggestion. *Evaluation*, 13(4), 421-438.

Appendix A: Evaluation Capacity in School Districts Interview Instrument

Read to participant: *The questions in this interview will help to gather information related to the extent that a school district, namely Two-County School District, uses evaluative strategies in everyday practice. This information will be used to better understand the ongoing needs and challenges districts face in the current era of public accountability and the extent that evaluation capacity has been built over time. The results will be included in a longitudinal case study toward a PhD in Evaluation Studies through the University of Minnesota. Your input is extremely valuable to this study and is greatly appreciated. A copy of the results of the study findings will be provided upon your request.*

Thank you in advance for your time to answer these questions.

The initial questions are about the district context and your role when you were aligned to research, evaluation and assessment activities.

- 1) During what timeframe did you have oversight of the department responsible for research, evaluation and assessment activities in the district? (from _____ to _____)
- 2) What was your role/title in the district?
- 3) What three words or phrases would you use to describe the organizational culture/context during your time in the district?
- 4) What was the organizational structure?
- 5) Who were the dominant leaders and what were the district priorities/initiatives/etc.?

The next set of items is about research, evaluation and assessment activities in the district and the department charged with these activities. Please choose the response that best fits your answer to the following questions. Feel free to elaborate after choosing your answer.

- 1) How important is the ability to evaluate programs, practices and policies to the mission and vision of a school district?
 - a. Extremely Important
 - b. Somewhat Important
 - c. Not at All Important
- 2) Considering all licensed instructional employees (teachers, principals, central office staff, and support staff), to what extent did each of the following evaluation strategies apply to staff during your tenure in the district (during the timeframe in which you had oversight of research, evaluation and assessment activities)?
 - a. To what extent was a variety of data collected and available – demographic, program, achievement, perception?
 - i. Little to no extent
 - ii. Some extent
 - iii. To a great extent

- b. To what extent were the various data types accessible to staff – software and tools existed to easily access data?
 - i. Little to no extent
 - ii. Some extent
 - iii. To a great extent
- c. To what portion of staff were these data accessible – how many staff had access to these data?
 - i. Little to no staff
 - ii. Some staff
 - iii. Most to all staff
- d. What portion of staff were trained to understand and interpret data – staff provided the training and skills to understand data
 - i. Little to no staff
 - ii. Some staff
 - iii. Most to all staff
- e. What portion of staff used data to understand situations/circumstances?
 - i. Little to no staff
 - ii. Some staff
 - iii. Most to all staff
- f. What portion of staff used data to inform decision making – consulting data before making decisions?
 - i. Little to no staff
 - ii. Some staff
 - iii. Most to all staff
- g. To what extent was Action Research used – small scale studies conducted by individuals to determine most effective courses of action in certain circumstances?
 - i. Little to no extent
 - ii. Some extent
 - iii. To a great extent
- h. To what extent was data or input gathered informally pertaining to the effectiveness of full programs/policies/practices – Informal program/policy/practice evaluation?
 - i. Little to no extent
 - ii. Some extent
 - iii. To a great extent
- i. To what extent was formal, systematic evaluation conducted by trained external evaluators – Formal program/policy/practice evaluation (External)?
 - i. Little to no extent
 - ii. Some extent
 - iii. To a great extent
- j. To what extent was formal, systematic evaluation conducted by trained internal evaluators – Formal program/policy/practice evaluation (Internal)?
 - i. Little to no extent

- ii. Some extent
 - iii. To a great extent
- 3) What types of student data were collected and available within the district during your REA oversight timeframe? (choices for each: collected Y/N, available Y/N)
 - a. Demographic data
 - b. Current enrollment data – who is currently enrolled in the district
 - c. Enrollment history data – who was enrolled in the past, the number of years a student has been enrolled, etc.
 - d. Roster data (student schedules)
 - e. Student marks (course grades)
 - f. Course data – courses that are running, students enrolled in particular courses, etc.
 - g. Course history data – courses that students have taken overtime, percentage of students having taken a particular course, courses that have been offered and run overtime, etc.
 - h. Attendance data
 - i. Behavior/discipline data
 - j. Programming data (intervention/special programs)
 - k. State assessment data
 - l. Locally-selected standardized assessment data – i.e.: MAP, ITBS, etc.
 - m. Local common summative assessment data
 - n. Local common formative assessment data
 - o. Student plans (IEPs, behavior, academic intervention, etc.)
 - p. Health/immunization data
 - q. Past student data – data on students from previous years
 - r. Incoming student data – data on students enrolled for the upcoming year
 - 4) Did the district have a separate department dedicated to research, evaluation and assessment during your time in this role? Y/N
 - 5) Did the district have staff dedicated to oversee/support each of the following during your time of REA activity oversight? Was there a formal process for requesting/engaging in each of these? (Y/N for each)
 - a. Research oversight/support? Formal process?
 - b. Evaluation oversight/support? Formal process?
 - c. Assessment oversight/support? Formal process?
 - 6) Did the district have staff that was formally trained in conducting evaluation during your time? (Y/N)
 - 7) What was the focus or major time commitments of the staff/department during your association?
 - 8) What was the vision for the department at that time?
 - 9) What changes, if any, occurred in the department (structure, focus, etc.) during the timeframe you were connected to the department and what prompted them?
 - 10) In what ways did the department collaborate with other departments?

- 11) What legislative issues were prominent and how did they impact the activities of the department and the decisions/direction of the department? Where there any other internal or external events that impacted the activities/direction of the department?
- 12) Is there anything else you would like to share about the functions of the department?

however, is public knowledge. Any person with knowledge of your employment may also identify you as a participant in this study.

There are no direct benefits to you for your participation. The information you share will, however, add to our knowledge about the Evaluation Capacity Building process.

Confidentiality

The records of this study will be kept private. No published report or article will include information that will make it possible to identify you personally. Research records will be stored securely and only the researcher will have access to the records. Audio recordings will be transferred immediately to a password-protected computer file (accessible only to Johnna Rohmer-Hirt) and erased from the recorder.

Voluntary Nature of the Study

Participation in this study is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University of Minnesota or the Anoka-Hennepin School District. If you decide to participate, you are free to not answer any question or withdraw at any time without affecting those relationships.

Contacts and Questions

The researcher conducting this study is Johnna Rohmer-Hirt. You may ask any questions you have now. If you have questions later, you are encouraged to contact her at 612-201-2426 or at rohm0001@umn.edu . Johnna's advisor at the University of Minnesota is Jean A. King, PhD. She may be contacted at 612-626-1614 or kingx004@umn.edu .

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher or her advisor, you are encouraged to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You may keep a copy of this Consent Information Form for your records.

Appendix C: Rubric for Assessing Evaluation Strength

Applied to school board agendas/minutes and Two-County Community Newsletter

Rubric for Non-Departmental Documents

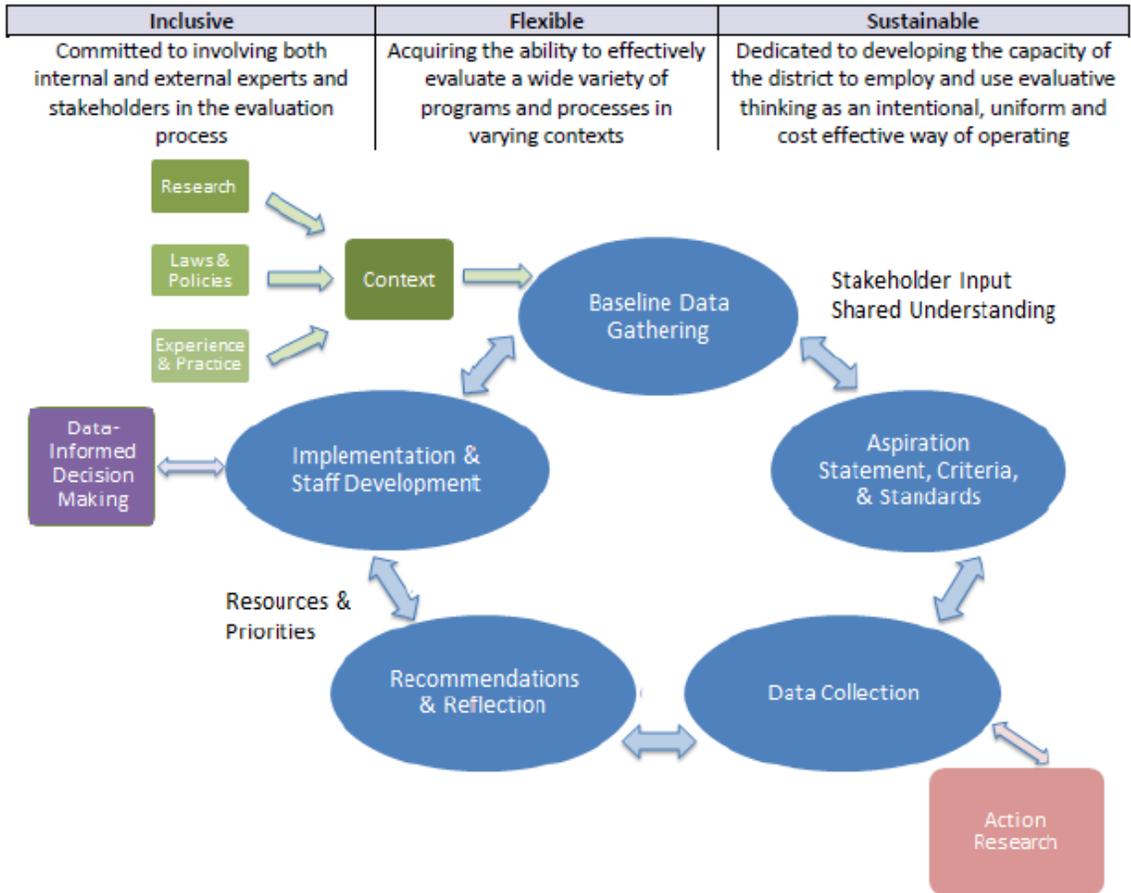
Rating	Description
None	Very little to no evidence of evaluative activity
Weak	Evaluative activity limited to assessment, stating criteria for decision-making, simple data analysis, or other lower levels of evaluation
Moderate	Evidence of formal/systematic data collection, more in-depth data analysis for decision-making, informal evaluation or action research
Strong	Evidence of formal evaluation activities including such things as identifying criteria and standards, systematic data collection, or formally designed research/studies

Appendix D: State Assessment Expansion Over Time

School Year	State Assessment Event
1992-93	State legislation mandating statewide assessments aligned to both basic and high standards
1993-94	Pilot of high standards performance assessments in select districts (Two-County participated)
1995-96	Implementation of 8 th grade basic skills test in math and reading and 10 th grade basic skills test in writing
1997-98	Implementation of state comprehensive assessments in grades 3 and 5 in math and reading
2000-01	Implementation of 9 th grade basic skills test in writing (switching from grade 10)
2002-03	First administration of state comprehensive tests in grade 10 for reading and grade 11 for math (results not reported to public)
2003-04	First operational administration of state comprehensive tests in grade 7 (math and reading), grade 10 (reading) and grade 11 (math)
2005-06	Implementation of state comprehensive tests with increased rigor in grades 3-8 (in math and reading), grade 10 (reading) and grade 11 (math)
2006-07	Implementation of more rigorous graduation-required assessment for diploma (moving beyond basic skills) in grade 9 (writing)
2007-08	Implementation of more rigorous graduation-required assessment for diploma (moving beyond basic skills) in grade 10 (reading) Implementation of a comprehensive state test in science for grades 5, 8 and high school (following biology)
2008-09	Implementation of more rigorous graduation-required assessment for diploma (moving beyond basic skills) in grade 11 (math)
2010-11	Implemented state comprehensive tests aligned to more rigorous standards in math for grades 3-8
2011-12	Implemented state comprehensive tests aligned to more rigorous standards in science for grades 5, 8 and high school (following biology)
2012-13	Implemented state comprehensive tests aligned to more rigorous standards in reading for grades 3-8 and 10
2013-14	Implemented state comprehensive tests aligned to more rigorous standards in math for grade 11

Modified from 2013 state testing technical manual for the graduation-required assessment for diploma. ,

Appendix E: Two-County Program Evaluation Model and Sample Reporting Tool

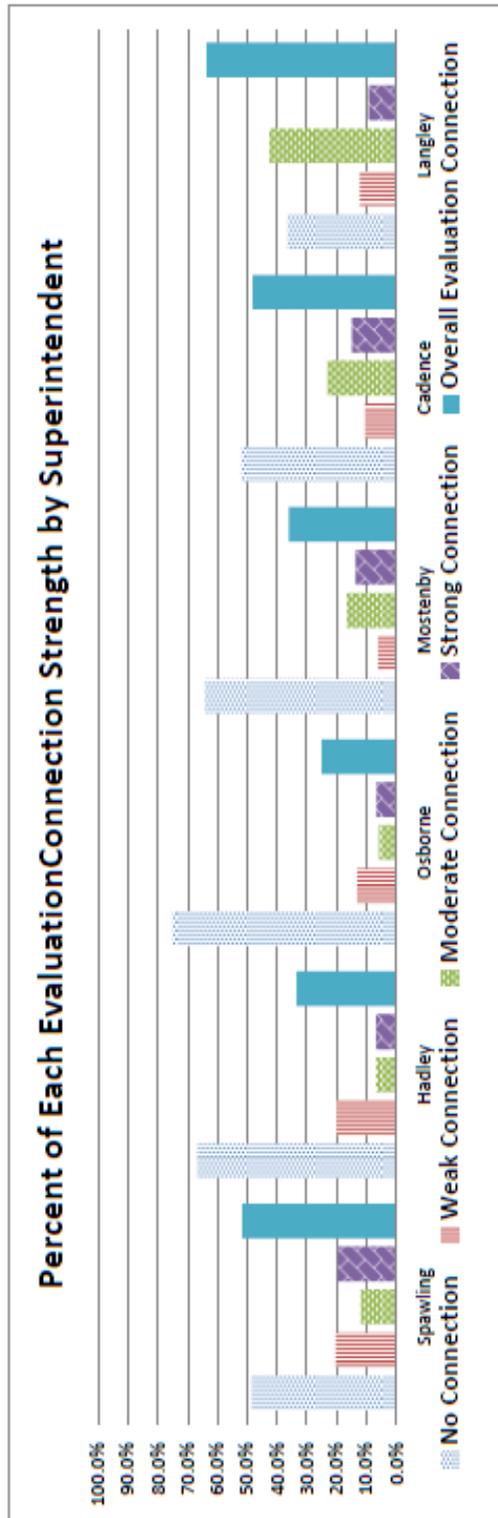


**Example of Reporting Tool Used in conjunction with Two-County ISD’s Program
Evaluation Process – Outcome Measure**

Two-County Anti-Bully/Anti-Harassment Program Evaluation									
<i>Focal Area:</i>	Engage (E.)								
<i>Aspiration Statement:</i>	Educators and staff engage families and community as partners in anti-bullying and anti-harassment efforts and focus on changing behaviors, not beliefs.								
<i>Criteria: (measurement)</i>	<i>Descriptions</i>	<i>Levels:</i>					<i>Trend</i>		
		<i>Intervene</i>	<i>Concern</i>	<i>Baseline</i>	<i>Progress</i>	<i>Vision</i>	<i>2015-16</i>	<i>2016-17</i>	<i>2017-18</i>
E1. Families (Task Force membership list)	Percent of the Community Task Force is made up of family members	≤ 15%	16-20%	21-24%	25-29%	≥ 30%	35%		
E2. Families (Task Force membership list)	Percent of the Community Task Force is made up of students	≤ 15%	16-20%	21-24%	25-29%	≥ 30%	26%		
E3. Families (Task Force meeting notes)	Family members on Community Task Force are represented at percent of the meetings	≤ 75%	76-80%	81-85%	86-89%	≥ 90%	100%		
E4. Families (AB/AH communication plan tools and tactics)	Percent of AB/AH communications designed for parents/guardians are completed as scheduled	≤ 28%	29-50%	51-72%	73-94%	≥ 95%			
E5. Community (Task Force membership list)	Percent of the Community Task Force is made up of community members	≤ 2%	3-4%	5-7%	8-9%	≥ 10%	32%		
E6. Community (Task Force meeting notes)	Community members on Community Task Force are represented at percent of the meetings	≤ 75%	76-80%	81-85%	86-89%	≥ 90%	100%		
E7. Community (AB/AH communication plan tools and tactics)	Percent of Task Force Communication Plans designed for community members are completed as scheduled (<i>needs clarifying</i>)	≤ 28%	29-50%	51-72%	73-94%	≥ 95%			
E8. Community (District events calendar)	District offers __ community event in a __-year period related to anti-bullying/anti-harassment	< 1 in a 2-year	1 in a 2-year period	1 annually	3 in a 2-year	≥ 2 annually	1 annually		

Appendix F: Additional Figures from School Board Agendas/Minutes Analysis

Superintendent	None	Weak	Moderate	Strong	Totals	Total eval	No Connection	Weak Connection	Moderate Connection	Strong Connection	Overall Evaluation Connection
Spawling	62	26	15	25	128	66	48.4%	20.3%	11.7%	19.5%	51.6%
Hadley	10	3	1	1	15	5	66.7%	20.0%	6.7%	6.7%	33.3%
Osborne	81	14	6	7	108	27	75.0%	13.0%	5.6%	6.5%	25.0%
Mostenby	214	20	55	45	334	120	64.1%	6.0%	16.5%	13.5%	35.9%
Cadence	70	14	31	20	135	65	51.9%	10.4%	23.0%	14.8%	48.1%
Langley	12	4	14	3	33	21	36.4%	12.1%	42.4%	9.1%	63.6%
Totals	449	81	122	101	753	304	59.6%	10.8%	16.2%	13.4%	40.4%

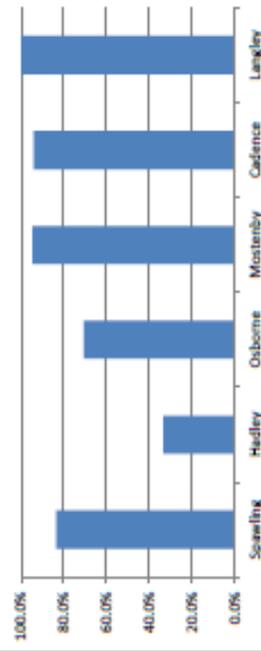


Appendix G: Additional Figures from Two-County Newsletter Analysis

Evaluation Connection Strength of the Focus on Two-County ISD Community Newsletter Publications

Supt	No	Weak	Mod	Strong	Total	Total eval	% No Connection	% Weak Connection	% Mod Connection	% Strong Connection	% Evaluation Connection Totals
Spawling	4	6	8	6	24	20	16.7%	25.0%	33.3%	25.0%	83.3%
Hedley	2	1	0	0	3	1	66.7%	33.3%	0.0%	0.0%	33.3%
Osborne	5	5	7	0	17	12	29.4%	29.4%	41.2%	0.0%	70.6%
Mostenby	3	25	24	2	54	51	5.6%	46.3%	44.4%	3.7%	94.4%
Cadence	1	10	6	0	17	16	5.9%	58.8%	35.3%	0.0%	94.1%
Langley	0	1	3	0	4	4	0.0%	25.0%	75.0%	100.0%	100.0%
Totals	15	48	48	8	119	104	12.6%	40.3%	40.3%	6.7%	87.4%

Overall Percent of Publications with Any Degree of Evaluation Connection



Percent of Evaluation Connection Strength by Superintendent Era

