Fall 12-15-2018

The Impact of a Collaborative Data-Inquiry Culture, as Promoted by the Harvard University Data Wise Improvement Process, on Student Achievement.

Kenneth I. Barrie
Concordia University - Portland, kenneth.barrie@gmail.com

Fall 12-15-2018

The Impact of a Collaborative Data-Inquiry Culture, as Promoted by the Harvard University Data Wise Improvement Process, on Student Achievement.

Kenneth I. Barrie
Concordia University - Portland

Concordia University–Portland

College of Education

Doctorate of Education Program

WE, THE UNDERSIGNED MEMBERS OF THE DISSERTATION COMMITTEE

CERTIFY THAT WE HAVE READ AND APPROVE THE DISSERTATION OF

Kenneth I. Barrie

CANDIDATE FOR THE DEGREE OF DOCTOR OF EDUCATION

Jillian Skelton, Ed.D., Faculty Chair Dissertation Committee

Dion Jones, Ph.D., Content Specialist

Derrick Tennial, Ed.D., Content Reader

The Impact of a Collaborative Data-Inquiry Culture, as Promoted by the Harvard University Data Wise Improvement Process, on Student Achievement.

Kenneth I. Barrie

Concordia University–Portland

College of Education

Dissertation submitted to the Faculty of the College of Education

in partial fulfillment of the requirements for the degree of

Doctor of Education in

Transformational Leadership

Jillian Skelton, Ed.D., Faculty Chair Dissertation Committee

Dion Jones, Ph.D., Content Specialist

Derrick Tennial, Ed.D., Content Reader

Concordia University–Portland

2018

Abstract

The study examined the impact of a collaborative data-inquiry culture, promoted by the Harvard University Data Wise Improvement Process (DWIP), on student achievement. 58 teacher surveys and PARCC scores for 2, 631 students were taken from four middle schools in a minority district. A quantitative method and a quasi-experimental one-sample-pretest-posttest design were used. Analysis involved Chi-Squared tests followed by pairwise comparisons. Statistical significance was determined using standard alpha of $p < 0.05$ and adjusted alphas to reduce type one errors. The DWIP had no significant impact on the proportion of students meeting and not meeting expectations in Reading ($p < 0.05$) and Math ($p < 0.0125$) from pretest to posttest. The DWIP had no significant impact on the proportion of students meeting expectations in Reading at different times of the implementation (2015, 2016, and 2017), $p < 0.05$, and had a negative impact on the number of students meeting expectations in math from 2015 to 2016, $p < 0.00833$. Further, the DWIP had no significant impact on the proportion of students meeting ($p < 0.00556$) and not meeting ($p < 0.05$) expectations in Reading at three implementation levels (initiating, lowly-developing, and developing), but had mixed impact on math outcomes. The highest implementation level had fewer students not meeting expectations in 2016, while the lowest had a higher number of students meeting expectations in 2017, $p < 0.00556$. Future research should focus on a larger school sample, data on the student level, and the impact of the DWIP on school culture.

Keywords: data wise improvement process, collaborative data-inquiry culture, teacher collaboration, data-driven decision-making, reflective learning, reflective practice, supportive leadership, continuous improvement.

Dedication

This dissertation is dedicated to the people who have been an integral part of my life and who have endured this journey with me, especially my children. I hope that the patience, dedication, fortitude, and passion that I demonstrated throughout this journey, and which you witnessed will serve as an example and encouragement for you as you pursue your own academic journey. Thank you for your continued love, support, and encouragement.

Acknowledgements

This dissertation was a collaborative effort of a great team and supporting cast whether they recognized it or not. I owe my gratitude to my kids who sacrificed many family trips, outings, and went without my help with homework to allow me the time and space to complete my coursework and finish this project. To my colleagues who recognized the demands and stress of having a fulltime job and an equally demanding school schedule that they allowed me time from work to rest and to get caught up on school work. Thank you for your care, concern, and generosity. To my dissertation chair, Dr. Skelton, your guidance, encouragement, and accessibility kept the dissertation process on pace, less stressful, and reflective. Your calmness is contagious and kept me reassured and calm during stressful periods. Thank you for making the dissertation process a richly learning and enriching experience. To the other members of my dissertation committee, Dr. Jones and Dr. Tennial, thank you for committing to my personal and professional growth over the last year by providing timely feedback, engaging in discussions, and being readily accessible throughout the process. Finally, I would like to thank Drs. Valentine and Gruenert for allowing me to use the School Culture Survey and to Dr. Boudett for allowing to use the Data Wise Rubric. Both instruments were used to design my survey instrument for the study.

Table of Contents

Abstract………………………………………………………………………………….....ii

Dedication….…………………………………………………………………………....iii

Acknowledgment……………………………………………………………….............iv

List of Tables……………………………………………………………………............ix

List of Figures…………………………………………………………………………..xi

Chapter 1: Introduction ……………………………………………………………….....1

Introduction to the Problem.………………………………………………………………1

Background, Context, History, and Conceptual Framework for the Problem………….3

Statement of the Problem…………………………………………………………………6

Purpose of Study………………………………………………………………………..8

Research Questions and Hypotheses……………………………………………………...8

Rationale, Relevance, and Significance……………………………………………..10

Definition of Terms…………………………………………………………..............12

Assumptions, Delimitations, and Limitations…………………………………………15

Chapter 1 Summary…………………………………………………….................17

Chapter 2: Literature Review…...……...………………………………………………19

Introduction to Literature Review…………………………………………………...19

Conceptual Framework………………………………………………………………20

Review of Research Literature and Methodological Literature………………………...26

The Data Wise Improvement Process……………………………………......26

PARCC Assessments……………………………………………………….......29

Data-Driven Decision-Making…………………………………………………29

Teacher Collaboration……………………………………………………………32
Supportive Learning Environments………………………………………………37
Monitoring and Feedback…………………………………………………………38
Supportive Leadership……………………………………………………………39
Kaizen Theory………………………………………………………………………40
Methodological Literature…………………………………………………......42
Review of Methodological Issues…………………………………………………44
Synthesis of Research Findings…………………………………………………...50
Critique of Previous Research……………………………………………............51
Chapter 2 Summary………………………………………………………………53
Chapter 3: Methodology………..…………………………………………………...56
Introduction to Methodology……………………………………………………...56
Purpose of Study…………………………………………………………………..57
Research Question…………………………………………………………………57
Hypotheses……………………………………………………………………......57
Research Design…………………………………………………………..............58
Target Population, Sampling Method (Power), and Related Procedures………………62
Instrumentation……………………………………………………………………64
Data Collection……………………………………………………………………70
Operationalization of Variables…………………………………………………...72
Data Analysis Procedures…………………………………………………………73
Limitations and Delimitations of Research Design……………………….............75
Internal and External Validity……………………………………………………77

Expected Findings…………………………………………………………………...78

Ethical Issues in the Study……………………………………………….........79

Chapter 3 Summary…………………………………………………………….80

Chapter 4: Data Analysis and Results ……………..…………………………………..82

Introduction ……………………………………………………………………82

Description of Sample…………………………………………….............83

Summary of Results…………………………………………………………….87

Validity and Reliability of Results………………………………………..87

Limitations and Delimitations………………………………………………89

Research Questions and Hypotheses………………………………………...90

Data Analysis Procedures………………………………………….. ………..91

Results…………………………………………………………………………….96

Assumption Tests………………………………………………………………97

Null Hypothesis One …………………………………………………………..97

Null Hypothesis Two …………………………………………………………..102

Null Hypothesis Three………………………………………………………….107

Chapter 4 Summary……………………………………………………………….123

Chapter 5: Discussion and Conclusion……………..…………………………………..125

Introduction …………………………………………………………………….125

Summary of Results…………………………………………………............126

Discussion of Results……………………………………………………………...130

Discussion of Results in Relation to the Literature………………………………….136

Limitations……………………………………………………………………..138

Implications of the Results for Practice, Policy, and Theory…………………………..140

Recommendation for Further Research…………………………………………………...143

Conclusion……………………………………………………………………….........146

References…….…………………………………………………………………………..148

Appendix A: Collaborative Data-Inquiry Survey Instrument…………………………………157

Appendix B: Statement of Original Work…...………………………………………………175

Appendix C: Facility Approval Letter… ……………………………………………………177

Appendix D: Anonymous Click Consent…...……………………………………………........178

Appendix E: Permission to Use School Culture Survey …………………………………........179

Appendix F: Permission to Use Data Wise Rubric……………………………………………180

List of Tables

Table 1: *Breakdown of Gender in Population*……………………………………………….85

Table 2: *Demographic Data*…………………………………………………………….86

Table 3: *School Characteristics*………………………………………………………87

Table 4: *Overall Account of Students Meeting PARCC Assessment Expectations*……………..98

Table 5: *Sample Size Assumption for ELA PARCC*……………………………………….....99

Table 6: *Chi-Square Test for ELA PARCC*……………………………………………..99

Table 7: *Sample Size Assumption for Math PARCC* ……………………………………..100

Table 8: *Chi-Square Test for Math PARCC*……………………………………………101

Table 9: *Pairwise Comparison for Math PARCC*………………………………………102

Table 10: *Sample Size Assumption for ELA PARCC*…………………………………...103

Table 11: *Chi-Square Test for ELA PARCC*…………………………………………....104

Table 12: *Sample Size Assumption for Math PARCC*……………………………………..105

Table 13: *Chi-Square Test for Math PARCC*……………………….......................................105

Table 14: *Pairwise Comparison for Math PARCC*……………………………………….106

Table 15: *Data Wise Implementation Levels*……………………………………..............108

Table 16: *Sample Size Assumption ELA PARCC Met Expectations*……………………….....109

Table 17: *Chi-Square Test for ELA PARCC Met Expectations*………………………………...110

Table 18: *Pairwise Comparison for ELA PARCC Met Expectations*…………………………..112

Table 19: *Sample Size Assumption for ELA PARCC Not Met Expectations*…………………...113

Table 20: *Chi-Square Test for ELA PARCC Not Met Expectations*……………………….....114

Table 21: *Sample Size Assumption Math PARCC Met Expectations*…………………………..115

Table 22: *Chi-Square Test for Math PARCC Met Expectations*……………………………….116

Table 23: *Pairwise Comparison for Math PARCC Met Expectations*…………………………….118

Table 24: *Sample Size Assumption for Math PARCC Not Met Expectations*…………………..119

Table 25: *Chi-Square Test for Math PARCC Not Met Expectations*…………………………...120

Table 26: *Pairwise Comparison for Math PARCC Not Met Expectations*……………………..122

List of Figures

Figure 1: *Conceptual Framework Diagram*……………………………………………………25

Figure 2: *Data Wise Swoosh* …………………………………………………………...27

Chapter 1: Introduction

Introduction to the Problem

Central office staff, administrators, and teachers in many low performing school districts who seek to improve student achievement and close the achievement gap have embraced data-driven decision-making (DDDM) and teacher collaboration in the form of professional learning communities (PLCs) as two of the most important practices (Dougherty 2015; Simms & Penny, 2014). However, while students at some schools have experienced measurable success, most schools have seen little or none (Simms & Penny, 2014). The failure of school leaders and teachers to improve student achievement through these two practices have been attributed to the following reasons: lack of training, lack of structures, inadequate time, limited vision, and poor leadership and organizational support. Cannata, Redding, and Rubin (2016) argued that there is large disparity among schools as to what constitutes effective data-use and teacher collaboration. This was a direct consequence of a lack of training for teachers, administrators, and support staff (Jao & McDougal, 2015; Lashley & Stickl, 2016). Leaders and instructional staff at most schools still rely primarily on accountability data such as test scores rather than perceptual data such as discussions, informal observations, and learning walks that improve teacher practice (Cannata et al., 2016).

The limited vision of schools and teams was another reason why DDDM practices and PLCs have failed to improve student achievement (Simms & Penny, 2014). School leaders who have used data in an effort to improve student outcomes have overlooked the role it plays in guiding professional development. Teachers viewed the use of data and collaboration as ways of improving student achievement and not necessarily to improve their practice through job-embedded professional development (Ezzani, 2015). The lack of structures of teams, schedules, meeting protocols, and monitoring mechanisms also impacted the effective implementation of DDDM and teacher collaboration through PLCs (Brown, 2015; Chow, 2015; Munoz & Branham, 2016; Simms & Penny, 2014). Further, there is a lack of and respect for times set aside for data-utilizations, collaboration, and reflection which also adversely impacted data-use and collaborative efforts (Simms & Penny, 2014). Most teachers do not find time outside of work to regularly reflect on their practice as a way of improving it. Lastly, the lack of supportive, shared, and collaborative leadership (Carpenter, 2015; Edwards, 2015; Ezzani, 2015; Price, 2014; Tschannen-Moran & Garies, 2015) as well as whole organizational support and strategic plans (Ezzani, 2015) negatively impacted the effectiveness of DDDM practices and teacher collaboration through PLCs.

A majority of the literature reviewed for this study were qualitative in nature. The studies showed a strong link between DDDM and student achievement, and between teacher collaboration through PLCs and student achievement (Cannata et al., 2016; Dougherty 2015; Simms & Penny, 2014). The studies also identified factors that influenced or contributed to the findings, namely training, time, structure, vision, and leadership. However, the qualitative nature of the studies made it impossible to make causal inferences about the effect of DDDM and PLCs on student achievement. Additionally, while DDDM and teacher collaboration through PLCs are closely related and dependent on each other, they were studied separately for their impact on student achievement. This is seen a major limitation in the literature findings on the topic.

The Harvard University Data Wise Improvement Process (DWIP) created a blue print for collaborative data-inquiry problem solving in schools to continuously improve teaching and learning. This process combines DDDM and teacher collaboration in the form of PLCs into a single variable, collaborative data-inquiry, and provides training for its use, creates structures and guidance for it implementation, allocates time for reflection and active learning, and guidance for monitoring and assessing actions plans. In essence, it addresses the limitations and confounding factors identified in previous research on DDDM and teacher collaboration in the form of PLCs. The DWIP was first tested in Boston City Public Schools in 2006 and since then it has been adopted by many school districts nationwide (Boudett, City, & Murnane, 2013). In 2015, it was adopted by School District X as an intervention to improve schools by increasing student achievement and closing the achievement gap. Located in a small Mid-Atlantic State, School District X is a large, diverse, moderate to upper income, and minority district, which serves students from urban, suburban, and rural communities. According to the district's five-year strategic plan, the school district is challenged by its positioning in the bottom quartile of student performance within the state. This quantitative quasi-experimental study sought to examine the impact of the DWIP intervention, which is grounded in a collaborative data-inquiry problem solving approach, on student achievement in this unique school district.

Background, Context, History, and Conceptual Framework for the Problem

Student achievement in School District X has suffered over the past 20 years for a number of reasons but primarily because of the lack of stability and consistency in executive leadership, which resulted in instability of instructional initiatives and strategies (Strategic Plan, 2015). Between 2003 and 2012, the school district has seen seven Superintendents, including two Interim Superintendents. Meanwhile, student achievement has remained near the bottom of the state's 24 school systems. The district has enjoyed the longest stability in executive leadership since the new CEO was appointed by the County Council Executive in 2014. In the beginning of the 2014-2015 school year, the new CEO commissioned a transition team, comprised of a diverse group of stakeholders, to conduct a comprehensive internal assessment of the school district to determine its current state. The report served as a source for data-driven and research-based strategy development and prioritization. The report highlighted a number of strengths including a highly qualified pool of teachers, strong alignment of the district curriculum with the state's college and career readiness standards, an expansive portfolio of community partnerships, declining dropout rates, and a four-year record in increasing graduation rates. However, the report highlighted a glaring challenge; students' poor performance on state mandated standardized tests, particularly at the elementary and middle school level, and in ELA and Math.

The CEO and his transition team designed and implemented a five-year strategic plan in March 2015 with the goal of providing outstanding academic achievement for all students. Achievement of the goal will be measured in three areas in the year 2020: Students average SAT and ACT schools will meet or exceed state average, 90% of students will graduate on-time, and 100% of graduates will meet the requirement to enter a two-year or four-year college, a technical school, the military, or will hold a license or certificate that will allow them to enter the workforce within six months. To measure progress towards those goals, the district will use the Kindergarten Readiness Assessments, PARCC, PSAT 8/9, and IB and AP Enrollments. PARCC assessments are the state standardized tests used in elementary, middle, and high schools and are administered in grades three through eleven in Reading (ELA) and Mathematics (Math).

To achieve the goals of providing outstanding academic achievement for all students, the plan focused on the following five areas and strategies: academic excellence, high performing workforce, safe and supportive environments, family and community engagement, and organizational effectiveness. To guide the work around the five focus areas, the plan adopted a coherence framework, which connected the work of teachers, students, and the content with the theory of change; the systemic attributes of culture, resources, stakeholders, systems, and structures; and a process to drive the work. This process is called the Harvard University DWIP.

The DWIP is a collaborative data-inquiry approach to problem solving in schools to improve teaching and learning. It uses student's assessment data in an explorative and collaborative manner to identify problems in instructional practice, design action plans with high impact strategies, and design monitoring tools to assess the action plan and make adjustments. The process is intended to take a ground up approach in which teachers are at the forefront of the improvement process. The process is also cyclical, thus repeating itself in an effort to continuously improve teaching and learning, thereby promoting a collaborative data-inquiry culture. This hybrid approach to problem solving and school improvement combines DDDM and PLCs practices, which were treated as separate factors in a majority of the previous studies on the topic. The strategic plan specifically embraced the DWIP as the systemic improvement approach to academic excellence. It is therefore important, to examine the effectiveness of the DWIP, specifically at the middle school level, to determine if it is helping students make progress towards the goals of 2020.

This study is based on an identified work problem as well as an identified gap in literature. A review of the literature showed that there is an abundance of qualitative research on the topic of DDDM, teacher collaboration in the form of PLCs, and student achievement. While the research showed a link between the two factors and student achievement, it is limited in making possible causal inferences. In addition, DDDM and teacher collaboration in the form of PLCs were treated as two separate variables in the research studies reviewed. The DWIP intervention under study combines these two factors as one variable called collaborative data-inquiry. Also, the DWIP is a refined process that addresses the confounding or limiting variables of training, time, structure, and support, which were identified in previous research. This quantitative, pre-experimental study provides findings that are significant to the evaluation of a work problem, uses a quantitative methodology that is lacking in previous research and which will allow for the determination of possible causal relationships between the hybrid collaborative data-inquiry practice (DDDM and PLCs) and student achievement, and provide more valid and reliable findings on the subject by mitigating for confounding variables.

The quantitative study captures teachers' perception around collaborative data-inquiry practices and examined its impact on student achievement within a continuous improvement framework as supported by *Kaizen's Theory*. Kaizen's theory of continuous improvement is a business model which is grounded in the gradual improvement in practices, efforts, and behaviors to eliminate waste and increase efficiency. It is characterized by three phases of preparation, action, and reflection. The DWIP intervention under study is grounded in the same continuous improvement construct which includes five steps within three phases of prepare, inquire, and assess. Perception of teacher collaborative data-inquiry practices, a hybrid or DDDM, PLCs, and collaborative leadership were captured as a single variable within a five step three phase cyclical process. This approach addressed the limitations in previous research on the topic in which DDDM, PLCs, and collaborative leadership were studied as separate variables and outside of a comprehensive and coherent plan for continuous improvement.

Statement of the Problem

School District X is ranked near the bottom of the 24 school systems in a small Mid-Atlantic state. A comprehensive internal assessment of the district conducted in 2014 by a transition team discovered poor student achievement on the state mandated standardized test as the main challenge (Strategic Plan, 2015). Specifically, there was a decline in ELA and Math scores at the elementary and middle school levels. There was also a widening achievement gap between general education students and students who were English Language Learners (ELL) and Special Education (SPED). Further, the performance gap expanded substantially with each successive year of schooling (Strategic Plan, 2015). The new CEO and his transition team designed and implemented a five-year strategic plan for improvement. Among the five focus areas identified for improvement to reach the goal of outstanding academic achievement for all students by 2020 was academic excellence. The plan embraced the DWIP to drive continuous improvement around academic excellence. Little was known about the process at the time the plan was implemented, except that it was based on DDDM and teacher collaboration in the form of PLCs.

An extensive literature review on the topic of DDDM, teacher collaboration in the form of PLCs, and student achievement showed that there is an abundance of qualitative studies, which show a link between DDDM, PLCs, and student achievement. The qualitative research also showed that there is large variance in the findings because the impact of DDDM and teacher collaboration in the form of PLCs are often viewed separately. Additionally, there is large disparity in how these two practices are implemented at schools giving rise to the variance in outcomes. The literature review also found few studies on the topic using a quantitative methodology. Further, there have only been a handful of studies on the DWIP intervention. In addition, those research studies have been conducted in settings that are geographically different than that of School District X. This study is therefore designed to fill the identified gap of quantitative methodology on the topic, focuses on a unique and underexplored population (underperforming, minority, diverse, and moderate to high socio-economic status), and to measure the effectiveness of the DWIP intervention at the middle school level. This study sought to answer the following question: What is the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement?

Purpose of the Study

The purpose of this study is to determine the significance of a collaborative data-inquiry culture, as promoted by the DWIP, on student outcomes at the middle school level on the state mandated standardized PARCC assessments. The DWIP is a problem-solving approach, which is currently being implemented in School District X as an intervention to increase academic excellence by increasing student achievement in ELA and Math on the PARCC in grades 3-11. Particular focus is placed at the elementary and middle school levels where student performance in ELA and Math is on the decline. This study will measure the effectiveness of the DWIP intervention to guide the strategic planning of the school district as it wrestles with how to continuously improve its schools by improving student achievement. The DWIP intervention is heavily grounded in four popular constructs of school improvement; DDDM, teacher collaboration in the form of PLCs, reflective practice, and collaborative and supportive leadership. As a result, the findings of this study will also contribute significantly to the general body of knowledge around data-driven collaborative culture, school reform, and continuous school improvement.

Research Questions

This quantitative, quasi-experimental, one-sample pretest-posttest study seeks to determine the significance of the DWIP intervention on student achievement. This was done by examining the impact of a collaborative data-inquiry culture, which is promoted by the DWIP intervention, on student outcomes on the PARCC assessments in ELA and Math. The Chi-Squared Test of Homogeneity followed by pairwise comparison of Z tests of multiple proportions were used as the statistical tool to determine if there were any statistically significant differences in the proportion of students who meet or fail to meet performance expectations on the PARCC after the implementation of the DWIP intervention. The study answered the following three research questions with guidance from the null hypotheses, and the alternate hypotheses:

Research Question 1: What is the difference in student outcomes on the PARCC assessments in ELA and Math after the implementation of the DWIP?

Hypothesis 1: There is no significant difference in student outcomes on the PARCC assessments in ELA and Math after the implementation of DWIP.

Alternate Hypothesis 1: There is a significant difference in student outcomes on the PARCC assessments in ELA and Math after the implementation of the DWIP.

Research Question 2: What is the difference in student outcomes on the PARCC assessments in ELA and Math at three different times of the DWIP implementation (Year 0, Year 1, and Year 2)?

Hypothesis 2: There is no significant difference in student outcomes on the PARCC assessments in ELA and Math at three different times during the DWIP implementation (Year 0, Year 1, and Year 2).

Alternate Hypothesis 2: There is a significant difference in student outcomes on the PARCC assessments in ELA and Math at three different times during the DWIP implementation (Year 0, Year 1, and Year 2).

Research Question 3: What is the difference in student outcomes on the PARCC assessments in ELA and Math based on the extent of the DWIP implementation (Not yet started, Initiating, Developing, or Sustaining)?

Hypothesis 3: There is no significant difference in student outcomes on the PARCC assessments in ELA and Math based on the extent of DWIP implementation (Not yet started, Initiating, Developing, and Sustaining).

Alternate Hypothesis 3: Student outcomes on the PARCC assessments in ELA and Math will be significantly different at each DWIP implementation level (Not yet started, Initiating, Developing, and Sustaining).

Rationale, Relevance, and Significance of Study

This study is significant because its findings may serve as a resource for educators, administrators, and policy makers who are trying to use DDDM, teacher collaboration in the form of PLCs, and supportive leadership to drive school improvement efforts. It combined the three variables into a hybrid variable called a collaborative data-inquiry approach. First, participation in the survey will raise a greater awareness among teachers, counselors, and administrators and influence reflection on the collaborative data-inquiry practices in their departments and the school as a whole. While the survey questionnaire will allow teachers to reflect on the implementation of the DWIP intervention, it will also probe reflection of areas of a collaborative culture as suggested by Gruenert (2005). These areas include collaborative leadership, teacher collaboration, professional development, unity of purpose, collegial support, and learning partnership. Reflection will allow teachers to identify areas of strengths, weaknesses, and suggest ways for improvement. Second, the study combines two factors of school improvement that have been studied separately and qualitatively, and looks at them as a single variable and quantitatively. There is an abundance of qualitative studies that showed a link between DDDM and student achievement, and PLCs and student achievement. The qualitative nature of these findings made it difficult to make possible causal comparisons and conclusions about the impact of these variables on student achievement. This limited the generalizability of the studies. This study combines DDDM and teacher collaboration in the form of PLCs as one variable called collaborative data-inquiry approach, promoted by the DWIP. Using a quantitative quasi-experiment approach, the study examined the impact of the DWIP intervention on student achievement. The findings identified possible cause-effect relationships between the variables which may allow educators to assess their instructional practice in terms of data-inquiry and collaboration, administrators in terms of leadership styles that empower teachers and influence collaboration, and policy makers who may use the findings to measure the effectiveness of the intervention and make changes to policy that guide instructional and leadership practices in School District X.

The structure of DWIP also mitigates many of the confounding variables found in previous research, which adversely impacted the validity and reliability of the results, and thereby limited the generalizability of the findings. The DWIP provides uniform training in data-use and collaboration, a structure and system for explorative use of data, staff collaboration, shared decision-making, reflective and action learning, and monitoring mechanisms. By mitigating for these confounding variables, this study will add to the validity and reliability of the findings around the relationship of the hybrid variable of collaborative data-inquiry and student learning. Therefore, the findings will be more applicable to educators, administrators, and policy makers both in and outside the study population. In addition, the study was conducted in a very diverse setting, which serves students from urban, suburban, and rural communities and from all socio-economic backgrounds. This setting is a microcosm of the United States; therefore, the study findings will have applicability in other school districts in their improvement efforts.

Definition of Terms

Definitions of the following terms explain their contextual use in the study:

Collaborative data-inquiry. Collaborative data-inquiry is cyclical process in which teachers collectively and collaboratively explore data on student achievement and instructional practice, identify problems in student learning, reframe the problems in terms of instructional practice, propose and test solutions to address the problem of practice, and collaboratively implement, monitor, and assess the plan (Boudett, City, & Murnane, 2015).

Collaborative inquiry. Collaborative inquiry is a collective approach of focusing on teaching and learning based on student achievement data and instructional practice, and serves as a catalyst for teaching and learning innovations (Carpenter, 2015). It is a four-stage cyclical process in which teachers work together to identify common challenges, analyze relevant data, and test instructional approaches (Carpenter 2015; Ciampa & Gallagher, 2014).

Data-driven decision-making. A set of practices used to improve student outcomes by continuously monitoring and assessing teacher practice and student learning and using the data to make instructional decisions for improvement (Dougherty, 2015).

Data Wise Improvement Process. The DWIP is an approach to school-wide instructional improvement predicated on a "collaborative learning process, data culture, and a culture of inquiry" (Boudett, City, & Murnane, 2014, p. 15). It consists of eight steps designed to enhance the ability of teachers and principals to collaboratively analyze data in order to achieve improved classroom instruction and student learning (Boudett, City, & Murnane, 2015). These steps are organized in three phases: *Prepare*, *Inquire*, and *Act*. Each phase plays an important role in building a school's capacity to use data to improve instruction. Critical features of Data

Wise are (1) creating a School Data Team (SDT) that will meet regularly and guide Data Wise implementation; (2) identifying a gap in skill or understanding common to many students that, if corrected, would have far-reaching implications for students' continued academic growth; (3) examining a wider range of data in order to investigate how teaching practice is contributing to this problem; (4) focusing efforts on solving a specific learner-centered problem related to the identified learning gap; (5) reframing the learner-centered problem as a problem of practice or an instructional change that addresses the learner-centered problem; (6) developing an action plan to document the instructional change, instructional strategies and tasks to support it, and timelines for implementation; (7) and identifying methods of assessing success in implementing the plan and its effect on students.

Improvement science. "The process by which network improvement communities, such as PLCs and data teams, are created to engage in disciplined cycles of inquiry where data is used to understand problems and test solutions" (Cannata, Redding, & Rubin, 2016, p. 1).

Job-embedded professional development. Ongoing practices that promote teacher learning during their normal professional duties which includes coaching, mentoring, collaborative problem solving in PLCs, peer-observations, learning walks, and data-utilization (Carpenter, 2015; Young & Kaffenberger, 2015).

PARCC Assessments. The PARCC or Partnership Assessment for Readiness of College and Career is a standardized test that is aligned to the Common Core State Standards and which is adopted by more than 43 states. The test, which is offered once a year between April and May, measures students' progress towards or readiness for college and career. The test is administered to students in grades three through 11 in ELA and Math. Student performance is measured on a band of 1-5: 1-not meeting expectations, 2-approaching expectations, 3-partially meeting expectations, 4-meeting expectations, and 5-exceeding expectations (Ansel, 2015).

Reflective learning. Reflective learning is the process by which teachers continuously use data about students learning to examine their practice, and critically analyze the data for the purpose of learning to improve their instructional practice. This is done individually and collectively.

School culture. A complex pattern of norms, patterns, attitudes, beliefs, behaviors, ceremonies, traditions, and myths that are deeply ingrained in the core of the organization. Culture is the historically transmitted pattern of meaning that wields astonishing power in shaping what people think and how they act (Barth, 2002). It is also a process that is continually renewed and recreated as new members are taught the old ways and eventually become teachers themselves (Bolman & Deal, 1991).

Student achievement. A measure of student learning based on their performance on the PARCC ELA and Math.

Supportive/collaborative leadership. Leadership approach that gives teachers the autonomy and empowerment to drive collaborative inquiry by taking initiatives, leading teams of teachers, and engaging in the decision-making process. In this approach, leadership is naturally assumed by members of the organization or group, and shared organically between individuals (Brown, 2015; Tschannen-Moran & Garies, 2015).

Teacher collaboration. A set of practices that is aimed at improving student achievement by improving teacher practice. It is achieved by creating PLCs and teams who have shared needs, and vision, and who work collaboratively and interdependently to address those needs and realize the shared vision (Harmon, 2017).

Assumptions, Delimitations, and Limitations

Assumptions. According to Lunenburg and Irby (2008), “assumptions are postulates, premises, and propositions that are accepted as operational for purposes of the research” (p. 135). This research study included the following assumptions:

1. The staff at all six middle schools were adequately trained on the DWIP.
2. Each school followed the suggested implementation protocol for the DWIP.
3. Survey respondents provided answers that were truthful and based only on their work experiences at the specific school.
4. The validity and reliability of the survey instruments were established with fidelity in a research-based, effective, and appropriate procedure and will be maintained in this study.
5. The validity and reliability of the PARCC assessments were established with fidelity in a research-based, effective, and appropriate procedure and that it accurately measures students’ progress towards or readiness for college and career.

Delimitations. Limitations and delimitations are conditions or circumstances that may influence a study. Delimitations are boundaries set for a study by the researcher (American Psychological Association, 2014; Lunenburg & Irby, 2008). The following delimitations were set in this quantitative, quasi-experimental study:

1. This study sought to answer three research questions that examined the significance of a collaborative data-inquiry culture promoted by the implementation of the DWIP on student achievement.
2. The study sample was taken from a population all middle schools in School District X

which consist of approximately 200 schools, 20,000 employees, and 130,000 students.

3. The study sample comprised of four middle schools, 250 teachers, and 2,631.
4. The sample of students were the entire 2015 6th Grade Cohort class of all four middle schools.
5. The interval for the DWIP intervention (IV) was taken for the last three consecutive school years (2014-2015, 2015-2016, and 2016-2017).
6. The measure of student outcomes (DV) was the PARCC assessments for the last three years (2015-2017).

Limitations. Limitations are influences that include conditions or factors that cannot be controlled by a researcher placing restrictions on methodology and conclusions (American Psychological Association, 2009; Lunenburg & Irby, 2008). In this quantitative, ex post facto, quasi-experimental study, the researcher may not be able to directly control the following conditions or factors:

1. The Collaborative Data-Inquiry Survey relied on teacher self -reporting. The accuracy of the reporting was based on the assumption that teachers were honest about the collaborative data-inquiry practices at their respective schools.
2. The ex post facto nature of the study made it impossible to randomly assign groups for the study since all participants had already been exposed to the independent variable.
3. The ex post facto nature of the study made it impossible to truly manipulate the independent variable. As a result, the researcher did not have total control over extraneous or confounding variables that may affect the outcomes of the study.
4. The PARCC assessments used to measure student outcomes is limited to two subject areas; ELA and Math. This placed limits on the analysis of overall student performance.

5. The extent and diligence with which schools implemented the DWIP may be different. As result, a third research question was designed to address this limitation.
6. The Common Core Standard-Based Curriculum and the accompanying PARCC assessments have only been recently implemented in School District X in the last three to four years. This required a shift in instructional pedagogy and student familiarity with a more rigorous assessment, which may also have a significant impact on student outcomes.
7. In a quasi-experimental, one-sample pretest-posttest design, the researcher cannot be certain that differences found are not from a certain variety of unaccounted factors. Therefore, this study could only produce possible cause-and-effect relationships, not statistically verifiable relationships.

Summary

This study sought to determine the significance of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement on the PARCC. To determine if and what relationship exists, the study attempted to answer three research questions; one about the overall impact of the intervention on student outcomes after a three-year implementation period, whether there is significant difference in student outcomes at three different times during the implementation, and whether the extent of the implementation had any significant change on student outcomes. The researcher employed a quantitative methodology to address the identified gap in quantitative studies on the topic. Also, the researcher used a quasi-experimental one-sample pretest-posttest design because of the ex post facto nature of the study. A 46-item survey instrument was used to capture teacher's perception of the collaborative data-inquiry practices, and the extent of the DWIP implementation. PARCC ELA and Math scores from 2015 to 2017 were used to measure student achievement. The Chi-Squared Test of Homogeneity followed by pairwise comparisons were used to analyze the data to answer each research question. This research is important to educators, administrators, and executive leadership in the School District X because it helps determine the effectiveness of the DWIP intervention in improving student achievement in ELA and Math, and its findings can help to guide their strategic planning around school improvement. The study also adds to the general body of knowledge around this new hybrid continuous problem solving approach in the DWIP as part of school reform efforts that can be used by the larger education community.

Chapter 2 follows this chapter and consists of the literature review to provide background information about the Data Wise Problem Solving Approach, and previous research topics that lends to this intervention: DDDM, teacher collaboration, PLCs, collaborative data-inquiry, reflective learning and practice, action learning, supportive and collaborative leadership, and the PARCC assessments.

Chapter 2: Literature Review

Introduction to the Literature Review

The literature review for this study consists of the examination of doctoral dissertation studies, peer reviewed journal articles, and books. The research literature is obtained through several electronic databases including EBSCOhost, ERIC, ProQuest, and Google Scholar. The initial literature search is guided by the terms of the research topic but is later expanded to ensure comprehensive examination of relevant literature. The topic of study is: The impact of a collaborative data-inquiry culture, as promoted by the Harvard University DWIP, on student achievement on standardized test in a low performing urban school district. Therefore, the literature search starts with the following terms, phrases or a combination of terms and phrases: data-inquiry, collaborative culture, and student achievement. The search is expanded using the following synonymous terms, phrases, or a combination of terms and phrases: data-driven instruction, data utilization, data-use, teacher-collaboration, collaborative learning, school climate, teacher development, teacher effectiveness, student outcomes, and school improvement. There is a lack of quantitative research on the topic, which necessitates the expansion of the literature search to include quantitative studies of close analogues of the topic to get perspectives for methodological design.

The choice of topic is influenced by a collaborative data-inquiry problem-solving process currently being implemented, in a large urban school district in a Mid-Atlantic state, in an effort to improve student achievement and drive continuous school improvement. The school district is ranked near the bottom of 24 school districts in the state and is located in one of the most affluent minority counties in the United States. This study is significant because it seeks to explore possible cause-effect relationship between the variables of the initiative, which will provide useful knowledge about its viability and add to the generalized body of knowledge on the topic. The specific problem is that the school district has focused on an initiative to improve student achievement that is based on a collaborative data-inquiry process for which little quantitative research data is available and thus little is known about the possible causes and effects of a collaborative data-inquiry culture on student achievement.

A synthesis of the literature search identifies several common themes or linkages between a collaborative data-inquiry process and student achievement. The themes DDDM, collaboration through PLCs, reflective learning through collaboration, job-embedded professional development, and distributed leadership approach as an important aspect of a supportive environment. The research identifies a lack of quantitative studies on the subject that allow for the identification of correlations and possible cause-effect relationships between variables. A lack of training and time in data-utilization and collaborative practices were also discovered. Lastly, the research identifies the following theory and theoretical constructs that explains data-driven collaboration and which are used in the design of a unique conceptual framework: Kaizen theory (Shang, 2017), collaborative inquiry (Carpenter, 2015; Ciampa & Gallagher, 2015), improvement science (Cannata, Redding, & Rubin, 2016), reflective learning and practice (Gero, 2014), and distributed leadership (Edwards, 2015; Ezzani, 2015).

Conceptual Framework

This study seeks to answer the following question: What is the impact of a collaborative data-inquiry culture, as promoted by the Harvard University DWIP, on student achievement? While there is plenty research that links teacher collaboration in the form of PLCs and DDDM to increased student achievement in the public-school setting, especially in disadvantaged and urban schools, a majority of the studies are qualitative in nature and do not offer the opportunity for possible causal conclusions. This quasi-experimental quantitative study seeks to determine if the collaborative data-inquiry practices of six middle schools in a large affluent urban district in the Mid-Atlantic Region have an impact on student achievement. The six middle schools are in the fourth year of implementation of the DWIP intervention, which is a systemic initiative. The DWIP is a continuous improvement process, which is based on a data-inquiry and collaborative approach to identifying problems with instructional practice through the lens of the learner, designing and implementing action strategies for intervention, and constantly monitoring and assessing those action plans to make needed adjustments (Boudett et al., 2014). Teacher perception of collaborative practices in the form of PLCs and DDDM practices will be captured at the end of the third year of DWIP implementation, and student scores on the state standardized tests will be taken before and after the implementation of the DWIP and will be compared. A comparison of student outcomes before and after the implementation will highlight any differences. However, in order to understand the significance of this difference, the linkages between DDDM, teacher collaboration in the form of PLCs, and student achievement must be examined through the theory of continuous improvement, and the constructs of collaborative inquiry, improvement science, reflective learning, and supportive leadership.

Driven largely by the idea of "improvement science", DDDM is part of a larger set of practices intended to improve student outcomes (Dougherty, 2015) and drive systemic change (Ezzani, 2015). "Improvement science" is the process by which network improvement communities, such as PLCs, are created to engage in "disciplined cycles of inquiry where data is used to understand the problem and test solutions" (Cannata et al., 2016, p. 1). This cycle involves interim monitoring and assessment of data, which is a major challenge, identified by Sims and Penny (2014) in the DDDM process. There are large disparities in how data is used and why. While a significant body of research exists, which has closely looked at schools and districts where the staff utilizes DDDM to maximize student outcomes, few studies have examined the systemic and comprehensive role it plays in professional learning (Ezzani, 2015).

Further, there are large disparities in how staff in different school districts use data to assess teaching, coaching, and supervision of teachers, and guide instructional and management decisions. An important element of effective data-use is the amount and type of data used. Using data from multiple sources allow for triangulation that guides the decision-making process from multiple perspectives. Cannatta et al. (2016) suggests the use of "perceptual data" to guide continuous school improvement because it leads to improvement in teacher practice instead of simple holding them accountable. Perceptual data includes but is not limited to teacher discussions, informal observations, surveys, and learning walks. Although student achievement and teacher evaluation data are used for measuring accountability, they can also be used to guide continuous school improvement efforts (Cannata et al., 2016; Sporte, Jiang, & Luppescu, 2015). Effective data use is also a direct consequence of the infrastructure and organizational support for continuous training and use of data. A distributed leadership framework (Ezzani, 2015), provides the structure, empowerment, and autonomy for teachers to learn how use data and apply it in their daily practice. When there is a uniform and well-tested process for training and implementing the DDDM model, schools are more likely to see improvements in student achievement across the board.

From a "collaborative inquiry" perspective, effective DDDM cannot be done in isolation but rather as an explorative process and in collaborative learning communities or PLCs, which have the potential to fuel ongoing change efforts (Butler, Schnellnert, & MacNeil, 2015). "Collaborative inquiry" is a 4-stage cyclical process in which teachers work together to identify common challenges, analyze relevant data, and test out instructional approaches (Carpenter 2015; Ciampa & Gallagher, 2014). Although there is evidence of the efficacy of PLCs in improving teacher quality and student outcomes, Sims and Penny (2014) warned of many cases in which PLCs had no impact because of the narrow vision, limited data use, lack of training and support, and no interim monitoring mechanisms. Effective PLCs that yield positive student outcomes must have trust, support, time, collaboration, and a shared vision that aligns with the work of the PLC (Sims & Penny, 2014). Moreover, the collaboration must be teacher-driven because it provides greater comfort for teachers to share information (data, instructional practices, and weaknesses) and reflect on practice to improve student learning (Gero, 2014). Lalor and Abawi (2014) agreed that sharing of resources and pedagogical practices have both social and emotional benefits to teachers. Collaboration also improves teacher practice by building relationships that enable job-embedded professional development (Lashely & Stickl, 2016). When teachers collaborate in PLCs that have a clear vision, trusting relationships, and established structures for sharing information it leads to improved teacher practice that will have a positive impact on student achievement.

A distributive leadership approach can also be applied to the impact of a collaborative data-inquiry culture on student achievement (Edwards, 2015; Ezzani, 2015). Given the complex nature of school organizations and the need for close and sustain interactions among teachers, the distributive leadership model is examined to determine the extent to which it facilitates and supports teacher collaboration and DDDM. In this model, leadership is naturally assumed by members of an organization or group and shared organically between individuals. Distributed leadership also gives teachers the autonomy and empowerment to drive the collaborative inquiry process (Brown, 2015; Tschannen-Moran, & Garies, 2015). Chow (2015) found that task, growth, and empowerment oriented leadership ensures participative and interactive learning that is collective, sustainable, engaging, enriching, and which increases the understanding of knowledge that is task-based leading to innovative practices. Further, leadership is identified as “second only to classroom instruction among all school related factors that contribute to what students learn" (Leithwood et al., 2004 as cited in Brown, 2015, p. 11). Therefore, this study explores both interpersonal and task oriented behaviors of school leaders (Tschannen-Moran & Garies, 2015) as a co-variant in an effective data-inquiry collaborative culture.

Finally, the impact a collaborative data-inquiry culture which encompasses DDDM, teacher collaboration in the form of PLCs, reflective learning, and supportive leadership, on student achievement must occur within a cyclical framework that promotes continuous improvement. It borrows from Kiazen’s Theory of continuous improvement, a Japanese business model, which focuses on frequent assessment of a process or product to improve efficiency (Shang, 2017). It is a gradual and daily process that does not only improve a product or process but also improves that manner in which people think, act, and work together to make improvements. The DWIP adopts the continuous framework from Kaizen Theory. It is grounded in the ACE Habits of Mind which promotes the thinking that people should take risks by acting, frequently assessing, and making adjustments to their practice (Boudett et al., 2014). The process seeks to promote continuous improvement by cultivating a collaborative data-inquiry culture through an eight-step process that is cyclical, collaborative, data-driven, and allows for frequent monitoring, assessment, and adjustment (Boudett et al., 2014). The conceptual framework is illustrated in the diagram in figure 1 below.

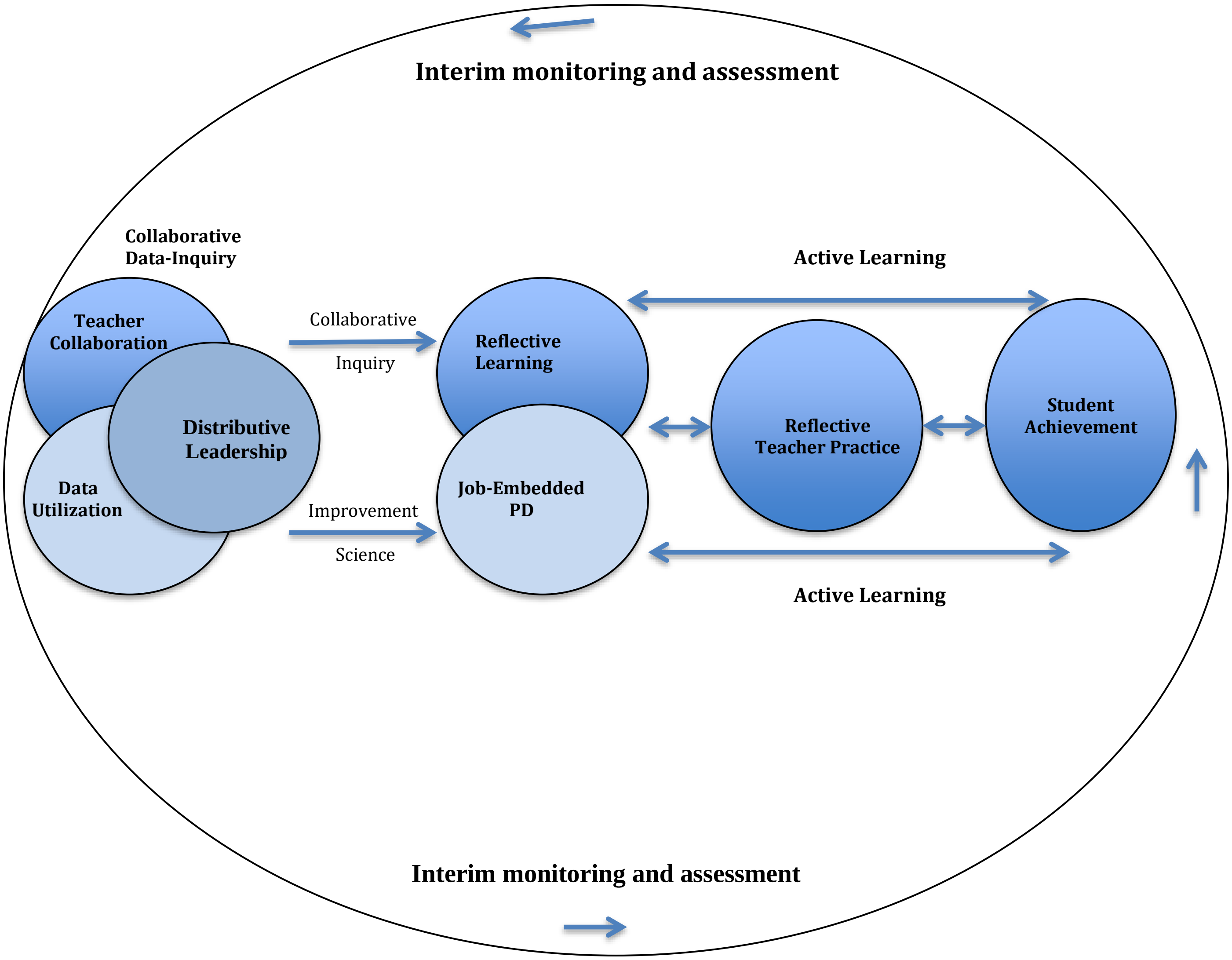

Figure 1. Conceptual Framework for Collaborative Data-Inquiry Practices and Student Achievement

Employing Kaizen's theory of continuous improvement, and the constructs of collaborative inquiry, improvement science, reflective practice, and distributive leadership, this study designs a survey instrument to explore the cause-effect relationship between a collaborative data-inquiry culture promoted by the DWIP and student achievement on a sample population of teachers in six middle schools in the district. The survey examines teachers' perception on DDDM, PLCs, reflective practice, and distributed leadership behaviors (independent variable [IV]) propagated by the implementation of the DWIP. Student achievement data on standards tests pre-and post the DWIP implementation is used to measure the impact (dependent variable [DV]) of the DWIP intervention. By so doing, the researcher will be able to determine the impact of the DWIP on student achievement, which can be used to guide capacity building efforts in data use, collaboration, and leadership that will lead in improved student achievement.

Review of Research Literature and Methodological Literature

The Data Wise Improvement Process. The DWIP is a comprehensive and cyclical model for using a collaborative and data-inquiry approach to solving problems. It was designed to train and coach teachers, principals, and district leaders how to use data, understand and interpret it, and how to use it to improve student learning. The process is predicated on using data to identify common student learning needs, to generate and implement instructional solutions, and to measure those solutions' effectiveness at raising student achievement with a department, grade, or school.

The process is an approach to school-wide instructional improvement developed by a team of educators in the Boston Public Schools and researchers at the Harvard Graduate School of Education (Boudett et al., 2013). It consists of eight steps designed to enhance the ability of teachers and principals to collaboratively analyze data in order to achieve improved classroom instruction and student learning (Boudett et al., 2013). These steps are organized in three phases: *Prepare*, *Inquire,* and *Act*. Each phase plays an important role in building a school's capacity to use data to improve instruction. A graphic of the DWIP is provided in Figure 2.

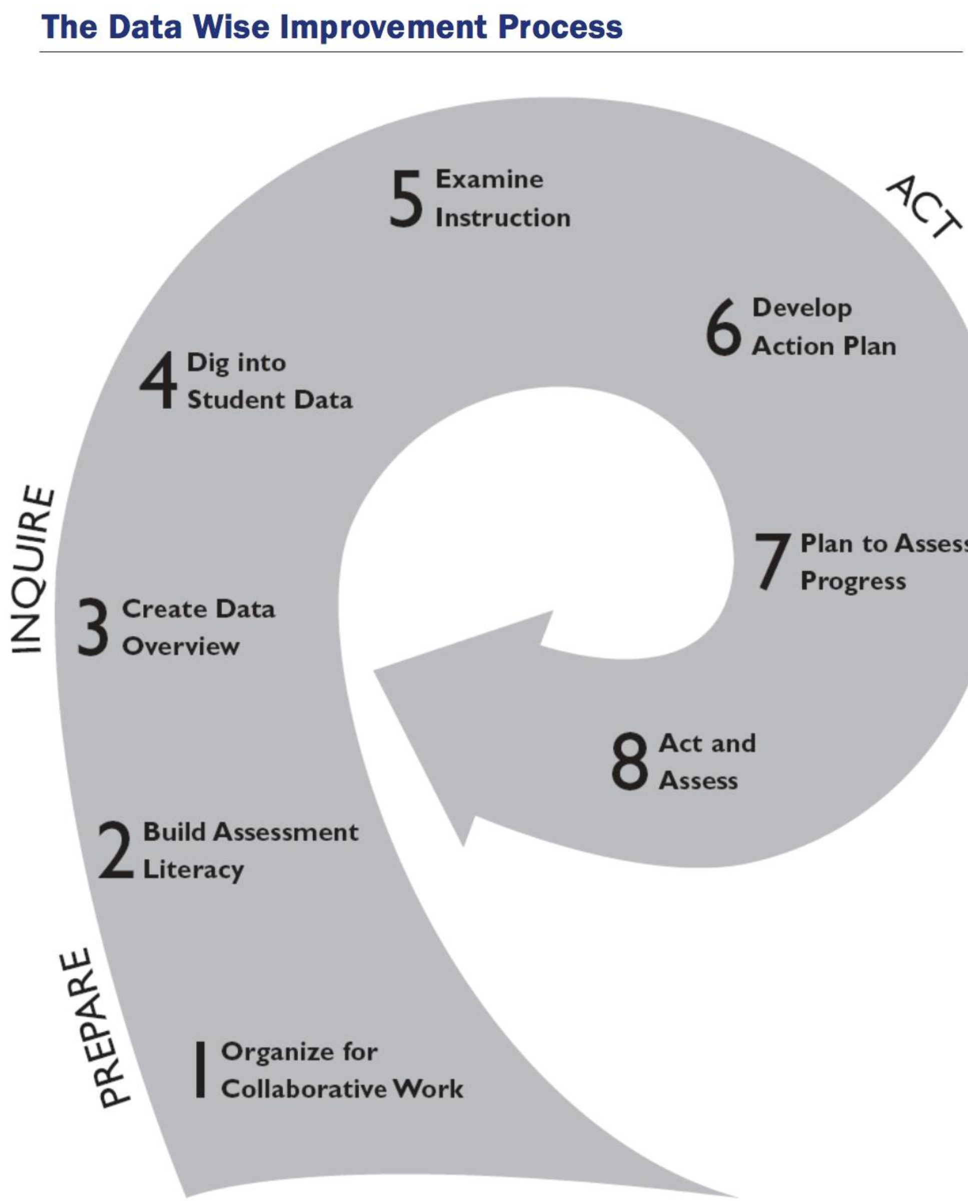

Figure 2: The Data Wise Improvement Process Swoosh. Taken from *A Step-by-Step Guide to Using Assessment Results to Improve Teaching and Learning* (Boudett, City, & Murnane, 2013).

The first phase of the eight-step process is called the prepare phase and consist of two steps. Step 1 of the process involves organizing for collaborative work in which the school organization builds systems of data teams, acknowledges work style preferences, makes time for collaborative work, sets expectations for effective meetings, sets norms for collaborative work, creates a data inventory, and creates an inventory of instructional initiatives. The second step of the process focuses on building literacy around the assessments used to measure student achievement. Teachers, administrators, and central office staff review the skills tested on those assessments, study how the results are reported, and learn principles of responsible data use.

The second phase of the process is called the inquire phase and is composed of the following three steps; create a data overview, dig into student data, and examine instruction. Step three involves creating a data overview by allowing staff to choose a focus area, analyze data and find a story, display the data in staff-friendly ways, and allowing staff members to make sense of the data by creating a priority question. In digging into student data which is step 4 of the process, staff examines a wide range of student data, come to a shared understanding of what the data shows using agreed upon protocols, and identify a learner centered problem. The learner-centered problem is then reframed into a problem of practice in step five after examining a wide range of instructional data and after coming to a shared understanding of what is happening in the classroom.

The Act Phase of the process includes creating an action plan of strategies, a plan for monitoring progress, and a plan for implementing and assessing the action plan. The action plan involves the identification of a repertoire of high impact instructional strategies and decides on what they would look like in the classroom. In planning to monitor progress of the action plan, staff will choose the assessments to be used to measure progress as well as set student-learning goals. The final step of the process is to act and assess. This step involves the following key tasks: implement the action, assess the implementation, assess student learning, adjust the action plan, and celebrate success.

The DWIP is a cyclical process that takes a collaborative and data-inquiry approach to problem solving with the goal of producing continuous school improvement. It contains the critical features of creating school data teams that regularly meet, identifying gap in skill and understanding of students, and examining a wide range of data to investigate how teacher practice contribute to this problem. The process also allows for identifying and solving learner centered problems, reframing the learner centered problem and the problem of practice to create an action plan with high impact instructional strategies, supports, and timelines to address the problem. Finally, the process identifies methods for measuring the progress and success of the implementation of the plan and it effect on student learning.

PARCC Assessment. The Partnership for Assessment of Readiness for College and Careers (PARCC) is the state mandated test used to measure college and career readiness or progress towards college and career readiness. It uses a 5-point performance band with a band of 4 or better indicating college readiness. The test has been used in more than 43 states since 2015 and is administered to students from grades three through 11 in two core subject areas: ELA and Math. The Mid Atlantic State in which School District X is located is a member of the PARCC Consortium and participated in the PARCC field-testing in 2014 in which at least one school in each of the 24 school districts in the state participated in the two-phase testing. The PARCC assessment has been the state mandated assessment in School District X since the 2014-2015 school year and is administered in grades three through 11.

Data-driven decision-making. The DDDM is among a set of practices used to improve student outcomes by continuously monitoring and assessing teacher practice and student learning and using the data to make instructional decisions for improvement (Dougherty, 2015). Cannata et al. (2016) described DDDM as a model used to drive continuous school improvement using “improvement science”. Improvement science is "the creation of network improvement communities, that engage in disciplined cycles of inquiry where data is used to understand the problem and test solutions" (p. 1). Compared to other interventions aimed at improving student achievement at many failing schools, such as state takeover through third party partnerships and charter schools, Klute, Cherasaro, and Arthrop (2015) argued that data-based decisions are more likely to positively impact student achievement. However, Dougherty (2015) acknowledged that while many educators are attempting to use data as part of the improvement efforts, there appears to be confusion in what data to use and how to use it. Much of this confusion is a result of the lack of initial and ongoing training on data collection, analysis, use, and supports. Through a qualitative study, Cannata et al. (2016) discovered that accountability data (used for measuring outcomes) is often misused to drive continuous improvement (evidence of teacher practice). Cannata et al. (2016) found that the use of perceptual data such as surveys, teacher discussions, informal observations, and learning walks lead to improve teacher practice through collective inquiry. While student outcomes and teacher evaluation data are mostly used for accountability purposes, Cannata et al. (2016) argued that student outcomes can be used for improvement purposes. Sporte et al. (2015), in a descriptive quantitative study, found that teacher evaluation data can be used for improvement purposes because it can be used to design embedded and ongoing professional development such as coaching, PLCs, and other systemic training. Another important factor that contributes significantly to the effective use of data to drive instructional decisions is ensuring that data is used collaboratively and consistently throughout the whole organization to set organizational goals, design course of actions, and monitor and assess their implementation (Dougherty, 2015). Providing adequate training and organizational structure and support around data utilization can address the disparity in the application of the DDDM process.

Leaders and staff at many schools are having trouble with implementing DDDM because they have not been well prepared to be effective at data utilization (Lashley & Stickl, 2016). In many cases, teachers, as well as principals, had a problem of interpreting data to turn it into usable knowledge (Cannata et al., 2016). After a yearlong study of the implementation of data teams in one middle school, Baker (2015) discovered that teachers, who were not using data properly, attributed it primarily to inadequate training and forced implementation. However, in a qualitative study on how guidance counselors use data to monitor progress and make data-driven decisions, Young and Kaffenberger (2015) determined that participation in professional development training on data use could influence the perceived data and accountability practices. While there is a significant body of research, mostly qualitative, which has closely looked at DDDM to maximizing student outcomes, few studies have examined the systemic and comprehensive role that professional learning of data utilization plays (Ezzani, 2015).

The literature highlights an important element in the DDDM model that has not been adequately explored and which hinders its effective implementation. There is a paucity of research that has examined the systemic and comprehensive role that professional learning of data utilization plays in the proper implementation of DDDM. A majority of the research on DDDM also points to the disparity and confusion in its implementation as primarily attributed to the lack or inadequate training. This discovery is important since it raises some doubt about the real impact of DDDM on student achievement. The research thus highlights an area in need of further research and provides a sound bases for the current study. This quantitative quasi-experimental study seeks to explore possible cause-effect relationships between a collaborative data-inquiry culture and student achievement by determining if any differences exist between student achievement before and after the systemic implementation of the DWIP intervention.

The DWIP is a collaborative and data-inquiry approach to solving instructional problems (Boudett et al., 2014). The target schools are part of a large district initiative, which received uniform, and ongoing training about data utilization and collaboration as part of the implementation process.

Teacher collaboration. Teacher collaboration is another set of practices that is aimed at improving student achievement by improving teacher practice. It is essential for effective DDDM because it is the foundation for the creation of "network improvement communities" that engage in disciplined cycles of inquiry in which data is used to understand problems and test solutions (Cannata et al., 2016). According to Harmon (2017), "collaboration occurs when individuals within two or more organizations involve in deep and complex interactions in communication to achieve shared goals that are interdependent, long-term, and complex" (p. 1). There are several factors that are important to ensuring effective collaboration in schools-relationships among teachers, formation of teams, knowledge sharing, reflective learning, online strategies, systemic programs and practices, assessment strategies, and leadership approach.

Teacher relationships. Relationships among teachers, administrators, and departments are important to ensuring high quality interactions that allow for collaborative work within and across boundaries to solve shared goals and complex problems (Hislop, 2013). Among the many characteristics for effective collaborative partnership, Hartman (2011) identified trust as very significant. Hallam, Smith, Hite, and Wilcox (2015), in a qualitative case study, found that the principal openness with teams', facilitated interactions that provided opportunities for within team trust to develop. "Greater trust enables greater collaboration" (Hallam et al., 2015, p. 209). Building relationships with external leadership and stakeholders are also important in the continuous improvement efforts because they help make connections between district initiatives and school needs and goals (Ezzani, 2015). “Schools that are characterized by high quality interpersonal relationships, communication, cohesiveness, and belongingness [among] students and teachers are better able to support student psychological needs and promote optimal development in academic domains” (Wang & Degol, 2016, p. 327).

Formations of teams (PLCs). Network improvement communities that facilitate collective data-use, collaborative learning, reflective practice, and joint problem solving are referred to as PLCs. PLCs are recognized as an effective practice in improving teacher collaboration and student achievement (Hallam et al., 2015). There are several characteristics of an effective PLC, but Sims and Penny (2014) highlighted trust, a shared vision, and time for collaboration as critical to its success. Sims and Penny (2014) cautioned against narrowly defining the purpose, mission, and vision of PLCs because that could negatively affect collaboration. For example, referring to PLCs as data teams can be misleading about the larger function of the PLC. Having a clear definition, measuring tool for implementation, and indicators for success are vital for effective PLCs (Munoz & Branham, 2016). Munoz and Branham (2016), in a qualitative study on PLCs, data-use, and student learning in a low achieving urban school district, found that there is a connection among the definition of PLCs, the implementation, and the impact on student learning. The collaborative work in PLCs is a major contributing factor but certainly not the only factor for improving student performance in schools and across school systems. “Integrating learning communities into the work lives of school teachers helps re-culture the teaching profession by changing the ethos of teaching from individualism to collaboration, from conservatism to innovation” (Chow, 2015, p. 303).

Knowledge management. Collaboration among teachers facilitates the sharing of information explicit and tacit knowledge (data and pedagogical practices), and influences reflective practice that improve student learning (Gero, 2014; Hislop, 2013). In a qualitative phenomenological case study, Lalor and Abawi (2014) found that sharing resources and pedagogical resources helped with long-term commitment on action research, focused professional learning, and lifelong learning. The collection, sharing, and storing of data is important for DDDM aimed at improving the instructional program and student achievement. When teachers collaborate within and across small network communities, there is greater ease and comfort with which data is shared. Additionally, Gero (2014) found that teacher-driven collaboration favors the sharing of information in PLCs. Creating a structure for data to be shared and stored so that it is easily accessible is paramount for professional learning and reflective practice.

Reflective learning. DDDM and teacher collaboration in the form of PLCs drive improvement in teacher practice and student achievement through a process of collaborative inquiry. Collaborative inquiry is the process by which teachers work together to identify common challenges, analyze relevant data, and test out instructional approaches (Ciampa & Gallagher, 2014). According to Carpenter (2015), collaborative inquiry focuses on teaching and learning based on student achievement, reflective practice, and serves as a catalyst for teaching and learning innovations. In this process, teachers build shared knowledge through collaborative practice. Teachers are also able to identify gaps in their espoused values and behaviors (cognitive dissonance) which motivate them to make changes to correct those gaps thus improving their instructional practice (Solis, 2015). The process also allows teachers to have courageous dialogues about their beliefs and behaviors as and thus learn from each other's practices. Reflective learning thus allows teachers to identify gaps in their behaviors and values, weaknesses in their practice, and use this information to collectively seek ways to improve.

Embedded professional development. A big advantage of teacher-collaboration is that is allows for job-embedded professional training and development that is easily accessible, ongoing, focused, relevant, and sustainable (Lalor & Abawi, 2015; Young & Kaffenberger, 2015). The joint examination and analysis of student and teacher data allow for reflective learning and exposure to effective pedagogical practices (Solis, 2015). The process of problem solving together allows teachers to share perspectives and build on alternate perspectives thus producing innovative ideas and solutions (Carpenter, 2015). This is perhaps the most powerful learning tool. Additionally, as teachers build relationship and form small learning communities, mentoring and coaching become easier. Hartman (2017) discovered that "academic coaching relies heavily on the formation of collaborative partnerships to bring about instructional improvement" (p. 17). While the study was conducted in a rural setting, where teachers had close familial ties, it can be extended to large urban school settings where small PLCs exist and where there is a positive school culture. Young and Kaffenberger (2015) agreed that professional development is effective when training is reinforced through ongoing coaching. Further, professional learning and development that take place when teachers collaborate are more likely to be effective because they match the goals and needs of the teachers (Young & Kaffenberger, 2015). Since most of this embedded professional development and learning takes place during collaborative meetings, it is important that meetings happen across vertical teams, and that they are interdisciplinary. Al-Saaideh and Al-Zyoud (2015) found that interdisciplinary planning and teaching have a positive impact on student learning and engagement.

Collaborative teaching. Collaborative planning and teaching have a positive impact on student learning and engagement. A mixed study conducted by Al-Saaideh and Al-Zyoud (2015) found that interdisciplinary co-teaching, team teaching, and collaborative teaching have a positive impact on student learning and engagement in both traditional and vocational courses. While the study was conducted in a Jordanian school, the findings have practical applications for the U.S. public schools where an array of courses including vocational courses are offered, and where collaborative teaching is among the set practices embraced to help close the achievement gap (Al-Saaideh & Al-Zyoud, 2015). Co-teaching and team teaching facilitates more purposeful grouping of students that allow for differentiated instruction (Parrott & Keith, 2015). It permits grouping by readiness, interest, and learner profile. In addition, co-teaching and team teaching supports the inclusion model that is highly favored by U. S schools as way of integrating students with learning disabilities into the general education classroom while addressing their specific learning needs. Parrott and Keith (2015) provided an example of effective co-teaching in which reading specialist, librarian, and general educator can co-teach in both a physical and virtual space to promote literacy in schools.

Online Strategies. Online tools can play a major role in supporting and encouraging teacher collaboration because it allows a medium for continuous networking, and a platform for facilitating collaborative inquiry and reflective practice (Ciampa & Gallagher, 2015). Ciampa and Gallagher (2015) discovered that blog use allows for the sharing of information, strategies, and personal and professional learning goals (cognitive learning). Blog use also allows for the sharing of experiences and discoveries related to knowledge gap and ineffective teaching methods (reflective learning). Rolling agendas are also powerful tool that enhances collaboration by influencing whole group participation, and for maintaining meeting information. In rolling agendas, collaborative meeting agendas are made clear, meeting minutes are available to everyone, and action steps for improvement are recorded with specific responsibilities. In addition, participants are able to provide feedback on each meeting and review next steps from the previous meeting ensuring continuity and monitoring of progress (Boudett et al., 2014).

There are other online tools that support collaboration and knowledge sharing which Hislop (2013) referred to as interactive technology devices (ITCs). These devices and tools include webinars, online databases and directories, and video and teleconferencing. Hislop (2013) argued that online databases and directories are powerful in storing and sharing information that can be used to match novice teachers with experts for mentoring and coaching. He warned that databases and directories have some limitations in that it is difficult to code tacit knowledge. Instead, Hislop (2013) suggested face-to-face meetings are the most effective means of sharing tacit knowledge. Other online tools that support collaboration and knowledge sharing by making it accessible beyond the workplace are webinars, telephone and video conferencing, chat rooms, and google hangout. Online tools are however supplementary to face-to-face collaboration (Hislop, 2013).

Supportive learning environments. The quality of the learning environment has a direct impact on how people learn and grow (Branson et al., 2004 as cited in Brown, 2015). Organizational structure plays an important role in ensuring that DDDM and teacher collaboration in the form of PLCs take place in an environment where teachers feel supported, and where the practices are promoted. First, time for collaboration is most important but is recognized as the greatest challenge to effective collaboration in schools. Sims and Penny (2014), in examining the reasons why PLCs fail in schools, identified the lack of time as one of the biggest factors. Collaborative time must therefore be built into the daily schedule of the teachers and must be respected. This will allow for appropriate planning, data analysis, and training. Dougherty (2015) noticed that there is the tendency to focus on content planning but found that vertical or interdisciplinary collaborative planning to be very effective to improving teacher practice and student achievement.

Another structure that helps make collaboration more productive is the adoption of norms that guide collaborative work (Tuttle, 2015). When collaboration protocols are implemented in PLCs, teachers experienced increased positive perception and behaviors (Tuttle, 2015). Protocols included establishing norms, building trust, analyzing student data, and examining teacher and student work to improve instruction. Ezzani (2015) found that data infrastructures that empowered teachers and that gave direct access to the principal existed in two California urban schools that experienced high student achievement after implementing sustainable professional learning in DDDM. Chow (2015) identified "formalized common structures and conditions that facilitated collaborative inquiry, reflective peer review, and dialogue among teachers about their practices, enhance professional autonomy and personal investment in renewing and sustaining those investments" (p. 303). Finally, Ezzani (2015) argued for the dismantling of hierarchical approach and restructuring of communication channels as a way to give teachers access and autonomy to drive the collaborative process.

Monitoring and feedback. Harmon (2017) identified the presence of assessment strategies as important characteristics of effective collaboration. The lack of interim monitoring was identified as one of the main reasons why many PLCs fail (Sims & Penny, 2014). Interim monitoring enables the cycle of inquiry and reflective practice using student and teacher data. Through monitoring and frequent assessment, teachers are able to make adjustments to their instructional practice for improvement (Ezzani, 2015). Munoz and Branham (2016) determined that having a clear measuring tool for PLC implementation and clear indicators for success were vital to effective PLC implementation that lead to improved student outcomes in several K-12 schools in a large urban district with a high poverty rate. In examining teacher evaluation practices, Moss (2015) found that teacher evaluation that is based on evidenced-based feedback, reflection, and shared collaborative experiences with peers was perceived to have greater impact on improvement in instruction and professional practice. Further, evaluation tools that have a structured framework for discussion, and that allow for non-confrontational dialogue are perceived to be more beneficial for improved teacher practice (Moss, 2015). Thus, the monitoring mechanism must be driven by the purposeful collection and use of student and teacher data in collaborative meetings, providing teacher feedback, and in teacher evaluations. Leaders in many school districts, including the one under study, now have student growth measures as part of new teacher evaluations. This study will thus consider ways in which to collect this data for monitoring that is both collaborative and non-threatening.

Supportive leadership. Leadership approach is also an important factor in the effective implementation of PLCs (Harmon, 2017). Shared and distributive leadership approaches are touted as most supportive in facilitating and promoting teacher collaboration (Carpenter, 2015; Ezzani, 2015). A directive leadership style is the less desirable style for promoting a collaborative culture in schools because it does not foster shared decisions making, but fosters mistrust between teachers and administrators (Bennett, YYlimaki, Dugan, & Brunderman, 2014; Carpenter, 2015). A distributed leadership approach is a way to ensure that teachers are leading other teachers in the learning of DDDM and collaborative practices so that they are prevalent school-wide. It empowers teachers to take leadership roles in instructional decision-making, learning, and professional development to drive student achievement. In a qualitative case study, Ezzani (2015) found that several schools in an urban school district in California showed improvement in student achievement over a period of three to five years as a result of the implementation of DDDM where the principals exhibited distributed leadership approaches. Carpenter (2015) identified shared leadership approaches to be most responsible for building a collegial collaborative culture that allowed teachers to train other teachers on how to go about learning communities.

There is also a direct relationship between some leadership behaviors and teacher perception of collaboration. Price (2014) measured a direct relationship between principal interactions and teacher perceptions. “Principals who were more accessible to teachers contributed more positive teacher perceptions of their students’ academic and school engagement” (Price, 2014, p. 129). Further, principals who are more inwardly socially oriented towards their teachers and students positively correlate with teacher’s beliefs regarding support for teachers in their school. Principals’ competence and openness with teams facilitated team interactions that provided opportunities for within-team trust to develop. “Greater trust enables greater collaboration” (Hallam, Smith, Hite, Hite & Wilcox, 2015, p. 209). It is necessary for principals to evidence both interpersonal (friendly, open to input, approachable) and task oriented behaviors (engaged in the instructional program) in order to be trusted by peers (Tschannen-Moran & Garies, 2015, p. 82). When teachers feel that they could turn to the principal for assistance with instructional matters, “teachers perceive their colleagues to be more committed to students, and that they are more cooperative, competent, and supportive” (p. 82).

Kaizen Theory. There is a wide range of meaning of the Japanese term *Kaizen* by scholars and researchers. However, regardless of the perspective, there is consensus that Kaizen means continuous improvement which can be applied to any work or field of work. Some researchers and scholars believe that Kaizen is a principle, method or effort. As a principle, it is a series of basic improvement principles says Lillrank and Kana (1998). Nihon HR Kyokai (1992) believes that Kaizen is not only a method of continuous improvement but also an effort of everyone in the organization trying to improve any work that needs to be improved. Further, William (2001) looked at continuous improvement through the lens of production process and believes that it is one of the most significant and valuable methods to reduce the long-term production cost. Other scholars and researchers such as Imai (1986) considers Kaizen in the workplace to be a process of continuous improvement that includes everyone in an organization so that the concept is grounded in three levels of any organization; management, group, and individual. Kaizen is also viewed as an innovation. Bessant, Caffyn, Gilbert, Harding, and Webb (1994) defined it as focused and continuous incremental innovations that could be applied to the whole company. Whatever, the perspective of Kaizen, it can be extended and can penetrate into methods and applications from theory to practice because it changes organizational routines, involves everyone in the organization, is incremental, and it can be applied to any and every field of work that needs to be improved.

Kaizen takes various forms of application that transfers theory to practice. All the applications are rooted in a systematic and scientific approach to solving problems. Although it originated as a management philosophy, research and practice in many fields have contributed to its methods and application. From a management perspective, the Kaizen process is as follows: select a target process, create teams, set target goals and plan, observe the process, analyze the process, create implementations, implement and make presentations (Dhongade, Singh, & Shrouty, 2013). The most notable practical approach which was derived from Kaizen management perspective, is the planning, doing, checking, and acting (PDCA) cycle. The PDCA cycle is a method of quality control process of a whole cycle which includes planning, doing, checking, and acting (Imai, 1986). This approach constitutes a series of activities pursued for improvement (Imai, 1986). While this approach is used to design tools such as Toyota Quality Control (TQC), QC Statistical Tools, and Total Productive Management (TPM), it can be applied to any field that needs to be improved (Imai, 1986).

Methodological literature. A review of the methodological literature regarding the question of collaborative and data-driven culture and its impact on student achievement revealed that a large percentage of existing research is qualitative in nature. Seventy-five percent of the studies reviewed were qualitative studies, 19% were quantitative, and eight percent were mixed-methods. Further examination of the qualitative studies showed that a majority or 85% were case studies, seven percent were phenomenological studies, and four percent were narratives and ethnographies respectively. A majority of the quantitative studies or 72% was descriptive analysis and nine percent were correlational, experimental, and quasi-experimental. There is thus a paucity of quantitative research in the area of DDDM and teacher collaboration in the form of PLCs. There is even a greater shortage of correlational, experimental, and quasi-experimental studies on the topic.

Case studies provide a deep understanding of a topic through the examination of multiple types of data sources (Creswell, 2014). However, they can be limiting because they do not provide for causal conclusions or for the determination of possible cause-effect relationships (Klune, Cherasaro, & Apthrop, 2015). This makes it difficult to add to the general body of knowledge that is immediately useful to practitioners. The qualitative case studies reviewed used multiple sources of data to conduct the study. While some of the data were collected using self-created tools, many of the studies used existing data captured by the school staff. The data sources used in the case studies were primarily test scores, perception surveys, document reviews, observations, interviews, focus groups, journaling, field notes, and school climate surveys. Many of the studies used cross-case analysis, correlational analysis, statistical analysis, cross-sectional, longitudinal analysis, and abductive analytical reasoning. Sample selections were mixed and some were more effective than others. Sampling methods included random sampling, selective sampling, and whole group sampling. The whole group sample proved particularly problematic. For example, in a descriptive analysis of 24 research studies, Klune, Cherasaro, and Apthrop (2015) found that many studies had serious limitations when only one school was assigned to each condition measured. This was primarily due to the existence of “cofounds”. The main reasons for the use of qualitative case studies was that the topic of DDDM and teacher collaboration in the form of PLCs involves social interaction phenomena, that are complex, and which occurs within a school context undergoing multiply layered initiatives (Butler et al., 2015; Hallman et al., 2015), that qualitative case studies allow for data to be taken from the perspective of the participant, and that they are flexible to the uniqueness of the field sites (Hartman, 2017).

The quantitative studies that sought to explore differences between variables (Jain & Cohen, 2014; Sporte et al., 2015; Tschannen-Morgan, & Garies, 2015) and quantify the impact of one variable on the other (Quinn et al., 2014) used descriptive analysis methodologies. This allowed them to perform correlational and regressional analyses or statistical tests. The data were mostly observational and the data sources included test scores, interviews, school climate survey, and pre/post surveys. The experimental and quasi-experimental studies sought to investigate the benefits of teaching pre-vocational education (PVE) through teacher collaboration, and the effects of the professional development module on DDDM respectively. In the experimental study, an experimental and control group were selected and the experimental group was treated. A variety of data collection instruments were used including attitude tests, student achievement data, and interviews. The researcher used both statistical and qualitative methods to analyze the data. In the quasi-experimental study, 50 teachers where identified for a professional training module on data utilization. The group was given a pre-and post-assessment and statistical analysis was used to measure the significance of the training module on teacher practice in DDDM.

The lack of quantitative studies in the area of DDDM, PLCs, collaborative cultures, and student achievement identifies a need to for more quantitative studies on the subject. Therefore, this study will take a quasi-experimental approach because of the following considerations: (a) qualitative studies have shown that there is a relationship between DDDM and student achievement, however the research methodology does not allow for causal conclusions, or determination of possible cause-effect relationships; (b) DDDM and teacher collaboration via PLCs are not being implemented properly because of poor training; (c) the school district under study, School District X, has already undergone training in data-use and collaborative work and are in year four of implementing using the DWIP intervention; (d) these conditions will make it easy to identify and select a sample of several schools to perform a pre/post assessment to show the significance of data-driven and teacher collaborative culture on student achievement when the lack of training factor is mitigated.

Review of Methodological Issues

A majority of the research literature available for review around the question of DDDM, teacher collaboration in the form of PLCs, and student achievement were qualitative in nature. More than 90% of the literature were qualitative studies compromising primarily of case studies, and a small percentage of phemenological and narrative studies. This overwhelming presence of qualitative studies on the topic can have many strengths but also places many limitations on what can be done with the data, the validity of the studies, the sample sizes to be studied, and time taken to conduct the studies.

One of the observable strengths of the qualitative research reviewed is the ability to identify the factors that affect the areas under study in a rich and descriptive form making use of multiple sources of data and categorizing them into common themes. Another advantage of the qualitative studies is the flexibility to adapt to the natural environment in which the study is taking place thus allowing allow the research to respond to data as it emerges in the study. It allows the study to adjust to the uniqueness of the participants and the environment. However, there were some limitations to the qualitative studies. First, the form in which the findings are presented do not allow for statistical analysis to make observations and trends and draw inferences, and test for validity. Another observable limitation is the limit on the sample size of the study, which makes it difficult to generalize the knowledge gained from the study and apply it to a larger setting. Lastly, the qualitative studies are very time-consuming often making feasibility difficult in situations of time constrains such as this dissertation study. The strengths and weaknesses of qualitative studies reviewed were best captured in the following five studies.

In a qualitative case study conducted by Dougherty (2015) to describe how educators use different types of data to assess teaching and learning, to coach and supervise teachers, and to guide instruction management, the researcher used a sample of nine schools in to two school district in Texas which were heavily populated by disadvantaged students. The researcher used a variety of data such as district benchmarks, school specific common assessments, state accountability tests, college readiness tests, other standardized assessments, classroom observations, grades, attendance, discipline, course completion, graduation rates, transfers, dropout data, and surveys from teachers and students. The vast amount of data were coded into the following six categories to show how data were used: (a) to identify student needs for grouping, interventions, programs, and classrooms, (b) to modify curriculum and instruction, (c) to motivate students and educators, (d) to coach and supervise teachers and other school personnel, (e) to adopt and evaluate programs and management decisions, and (f) to communicate with outside audiences. Although the researchers used the Atlas.ti software for data analysis, the results were still very rich and descriptive. The study had several limitations in that the relatively small sample size made it difficult to generalize from the reports how other school districts across the nation are using data. Another limitation was the self-selection of the participants by district staff and as such did not constitute a randomized sample. Also, the lack of observation did not allow the researcher to cross-reference the data practices that were described in the interviews. Lastly, data were not collected from a large enough sample to make correlations between data practices and school and district performance indicators.

In a second qualitative case study, Sims and Penny (2015) examined a PLC that had a too narrow focus because of the definition as a “Data Team” and failed therefore to affect student achievement. The study gathered data from six interviews with PLC members and three observations of PLC meetings. Interview data were coded thematically, and analyzed based on research questions asked. The observation data were analyzed using predetermined categories. An inductive process of thematic analysis was conducted. The study found that participants perceived the PLC, in its current state, as too focused on a single set of metrics and lacking the time, collaboration, and support needed to be effective. The researchers assessed the credibility of the findings by the process of triangulation, using multiple sources of data collection, both interviews and observations. The researchers also applied the “member check” process (Merriam, 2009 as cited by Sims & Penny, 2015) to check for credibility of the study by taking the analysis back to the participants to check that the interpretations held true and found no discrepancies. The application of the process of triangulation and member check both strengthened the credibility of the results. However, Sims and Penny (2015) recognized that use of a single high school as the setting and limiting the sample size to one PLC affected the generalizability of the study.

A third qualitative study conducted by Ezzani (2015) sought to deepen the understanding of school districts that are implementing sustainable professional learning in DDDM to improve student achievement. The study was done in two urban school districts over a period of five months using one school in each district. "The choice of a small select sample…provided an opportunity to obtain detailed, sensitive and descriptive data for a study of this scope" (Ezzani, 2015, p. 6). The data collection included semi-structured interviews, observations at the school and district levels, and review of artifacts collected from the school, district, and classrooms. In collecting data, Ezzani (2015) recorded and transcribed interviews and conducted "member checks" (Creswell, 2007) with participants to ascertain internal validity. She also took field notes on observations to allow for triangulation with interviews and artifacts. In addition, the researcher made photocopies of documents for the purpose of cataloging, coding, and content analysis. The analysis performed was across districts. To adhere to ethical practice and comply with internal review board procedures, Ezzani (2015) obtained informed consent from all participants, ensured that there was voluntary participation, and ensured that all data collected were kept confidential to maintain anonymity. The findings indicated that the confluence of a focus on DDDM, systemic and comprehensive professional learning, and distributed leadership lead to consistent student achievement gains over a three to five-year period.

In this study, there were several strengths in the methodology, which are note only unique to qualitative studies but are good, ethical research practice. These included the collection of consent forms, ensuring voluntary participation, and safeguarding data to honor confidentiality and anonymity of participants. There were other aspects of the research that were specific to the qualitative nature of study. The selection of a small sample size was purposeful in limiting the scope of the study in order to focus on greater depth. Again, knowledge gained from studies with small sample sizes limits its generalizability and immediate usefulness to the larger stakeholder community. In addition, the processes of triangulation and member check added to the validity of the study. However, the length of time taken to conduct the research will affect the feasibility of these type of studies in many cases. Both the strengths and weakness identified in this research can have huge implications since they will guide improved design in future research studies that are more feasible, valid, credible, ethical, and appropriate.

The abundance of qualitative research on the question of DDDM, teacher collaboration, PLCs, and student achievement, coupled with the scarcity of quantitative studies dictates the need for more quantitative studies on the topic. As a result, this study will use a quantitative quasi-experimental approach that will use the factors identified in the qualitative studies to design instruments that can capture quantitative data for statistical analyses- test for validity, identify trends, differences between groups, and possible cause-effect relationships (Creswell, 2014). Some of the factors identified in the qualitative research that affect the topic under study includes trust, time, training, collaborative inquiry, reflective practice, and supportive and distributed leadership. A quasi-experimental study will be both feasible and appropriate in School District X where a collaborative data-inquiry problem solving intervention, DWIP, is in its fourth year of implementation.

The setting in School District X will mitigate for the concern of the lack of training for data utilization and teacher collaboration in the form of PLCs identified in previous research on the topic. There was both initial and ongoing training throughout the district around data use, building structures of teams, and in engaging in collaborative work. The training will address the issue of lack of training in making DDDM and collaboration in PLCs identified in the literature research and will add greater validity to the study. Further, with the implementation of the DWIP throughout the district, the researcher will not need to create experimental and control groups to be treated. Therefore, a quasi-experimental study that measures teacher perception of the collaborative data-inquiry practices after the implementation of the DWIP intervention and compares student outcomes on the PARCC before and after the intervention would be most appropriate. Also, since the district is very large made up of more than 200 schools, and because the DWIP intervention is a systemic initiative, there is a large enough target population from which an adequate random sample can be obtained. According to Creswell (2014), the large sample size lends to the validity of the research and the generalizability of the findings which are challenges found in qualitative studies. A quantitative quasi-experimental study on the question of a collaborative data-inquiry problem solving approach and its impact on student achievement which utilizes the strengths in designs of previous qualitative studies on the topic will yield an extremely effective research design (Creswell, 2014).

There are two possible limitations that can be anticipated in this quasi-experimental study in this setting. The first limitation is the large turnover rate of teachers in the school district that might reduce the eligibility of the target population. The target population should be teachers and staff who have received training around data-use, PLCs, and collaborative work as part of the implementation of the DWIP intervention for the past three years. However, the large size of the school district of more than 10,000 employees will help mitigate this possible limitation to obtain a large enough sample size. A second limitation might be the feasibility of conducting interviews, focus groups, and possible document reviews to collect data that would allow for triangulation and member check against survey responses to add to the validity of the study (Ezzani, 2015). Mitigating for the possible limitation of sample size and capturing other data regarding the implementation of the various steps of the DWIP intervention in addition to teacher perception on the surveys adds to the credibility and validity of the study.

Synthesis of Research Findings

The literature review reveals that there is extensive research on the topic of data utilization, teacher collaboration, and student achievement, albeit mostly qualitative studies. The research also reveals that most studies examined one variable, factors that impact the variable, problems with implementation of DDDM and teacher collaborative practices, the type of professional learning that promote these practices, the role of the school environment, and the impact of school leadership on those practices. The studies also show that there is a link between these practices and increased student achievement in the K-12 setting. However, there is a lack of research, which looks at the ideas together and as part of a coherent improvement plan, identifies possible cause-effect relationship, or which quantifies the impact of DDDM and collaboration on student achievement.

This study seeks to embrace the identified factors that influence the DDDM, teacher collaboration, and student achievement, and takes a comprehensive look at how it impacts student achievement by connecting all the factors and how they work together, to address concerns identified in the research. First, there is agreement among the research that effective DDDM leads to improved student achievement (Cannata et al., 2016; Dougherty, 2015; Ezzani, 2015; Klute et al., 2015; Lashley & Stickl, 2016; Young & Kaffenberger, 2015). Second, there is plenty research that shows a connection between teacher collaboration through PLCs and improved student outcomes in the K-12 public school setting (Baker, 2015; Butti, 2016; Gero, 2014; Hallam et al., 2015; Harmon, 2017; Hartman, 2017; Jao & McDougal, 2015; Munoz & Branham, 2016; Parrott & Keith, 2015; Sims & Penny, 2014). Third, there is recognition in the research that data utilization and teacher collaboration are practices that must happen together, systemically, and continuously in order have sustained impact on teacher practice and student achievement. These processes are "collaborative inquiry" (Carpenter, 2015) and "improvement science" (Cannata et al., 2016). Fourth, collaborative and collective use of data to drive student achievement leads to continuous improvement in teacher practice. This occurs through reflective learning and job-embedded professional development. Fifth, improvement in teacher practice and student achievement through effective data-use and teacher collaboration is a direct consequence of a supportive environment and one that is predicated on shared and distributed leadership. The themes that evolved in the research show that there are clear linkages and overlaps of the factors. There is also a cycle of professional learning and practice that is evident. Sixth, the evidence points to a collaborative data-driven culture that drives student achievement rather than isolated practices. Finally, distributed leadership facilitates and supports PLCs and DDDM that lead to improved teacher practice and consequently, student achievement.

Critique of Previous Research

The research studies reviewed were dominated by qualitative studies that enabled researchers to gain deep understanding of the topics, and explained and described the topic or phenomenon. However, this qualitative methodological approach does not allow for statistical analyses to be made such as inferential and regression analyses, identify causal relationships between variables and groups, make correlations, or identify cause-effect relationships. For example, Hallam et al. (2015) conducted a qualitative case study using the five facets of trust, to clarify the impact of trust among PLC teachers on their team's collaborative practices. The results of the study showed how trust developed among members of collaborative teams and how principals influenced its development. The findings supported previous research and even added to the depth of existing research by identifying teacher voice as important to the current understanding of trust. While this approach was most appropriate for exploring a complex social issue such as trust, it did not allow for causal inferences to be made thus limiting the generalizability of the knowledge. In this case, a quantitative methodological approach would build on the findings of the study by allowing for possible causal inferences. The identified shortage of quantitative studies on the topic also means that there is a shortage of generalizable knowledge available to the stakeholder community.

An analysis of the methodological designs used in the majority of qualitative studies reviewed showed how the researchers added rigor, credibility, and validity to the research designs by employing "triangulation" and "member check" (Ezzani, 2015; Sims & Penny, 2015). In both qualitative case studies, Ezzani (2015) and Sims and Penny (2015) used multiple sources of data, including surveys, interviews, and document reviews, to analyze their research questions from multiple perspectives and to ensure that there is corroboration among data sources and consistency across data sources. The researchers also used the process of "member check" (Creswell, 2007 as cited in Ezzani, 2015; Merriam, 2009 as cited in Sims & Penny, 2015) by taking the analysis back to the participants to ascertain that the interpretations were true. The processes of "triangulation" and "member check" will be useful in the design of this quasi-experimental study.

A major weakness evident in all of the qualitative studies reviewed was the small sample size, either intentional or limited by methodological approach, which adversely impacted the generalizability of the findings. Ezzani (2015) in the qualitative case study designed to enhance the understanding of districts that are implementing sustainable professional learning in DDDM to improved student achievement, intentionally used a small sample size of nine schools to focus on sensitive, detailed, and descriptive data. A significant limitation for this study was created by the intentional use of a small sample size, which affected the generalizability of the study (Ezzani, 2015). In a separate qualitative case study to examine why a PLC with a narrow data focus failed to improve student achievement, Sims and Penny (2015) used a single PLC at one high school as the sample size. The small sample size affected both the credibility and generalizability of the study. Creswell (2014) highlighted that unlike quantitative studies, qualitative approaches are limited by the small sample size. A quantitative methodology, with a quasi-experimental design will not be limited by sample size thus adding to the validity and generalizability of the study.

Summary

The research identifies known and unknown knowledge about the area of study, which was used to design a unique framework for the study. The research shows that effective data-utilization and DDDM can have a positive impact on student achievement. Second, teacher collaboration through PLCs can have a positive impact of student achievement. Third, effective data utilization and collaboration occurs simultaneously and leads to improvement in teacher practice through reflective learning and job-embedded professional development. Fourth, leaders and staff at most schools are struggling to implement DDDM and teacher collaboration in the form of PLCs because of a lack of comprehensive systemic training, lack of time, and the absence of mechanisms for monitoring their progress and assessing their effectiveness.

The research also identifies one theory and three constructs relevant to the topic. Exploration of the research question using Kaizen's theory of continuous improvement allows the researcher to examine the impact of a collaborative data-inquiry culture of student achievement through a comprehensive, coherent, and cyclical improvement framework. Collaborative inquiry guides teaching and learning through the use of student achievement data and reflective teacher practice thus influencing innovative practice (Carpenter, 2015). Improvement science explains the creation of network improvement communities that engage in disciplined cycles of inquiry in which data is used to understand problems and test solutions (Cannata et al., 2016). Distributed leadership, as a leadership style that empowers teachers to facilitate and support DDDM and collaborative work around teaching and learning through a cycle of inquiry (Ezzani, 2015). Finally, the research reveals that there is an abundance of qualitative studies and a lack of quantitative studies on the question. As a result, there is limited knowledge on the correlation and possible cause-effect relationships between the variables of collaborative data-driven culture and student achievement.

Based on this review of literature, which develops a unique conceptual framework using collaborative inquiry, improvement science, reflective practice, distributed leadership, and the Kaizen theory of continuous improvement to understand student achievement, there is sufficient reason for thinking that an investigation examining the impact of a collaborative data-inquiry culture would yield socially significant findings. Therefore, the claim that the literature review has provided strong support for pursuing a research project to answer the following research question: What is the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement?

Chapter 3: Methodology

Introduction to Methodology

This study examined the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement in a large urban school district located in a Mid-Atlantic state. In an effort to address a systemic problem of poor student achievement, and an increasing achievement gap, the school district adopted and started the implementation of the DWIP as an intervention in the fall of 2015. The DWIP takes a collaborative data-inquiry approach to problem solving by placing teachers at the front of the improvement process using an eight-step approach (Boudett et al., 2015). This research study compared student outcomes on the state mandated standardized test, PARCC, before and after the implementation of the DWIP intervention, at three different times of the implementation, and at three different implementation levels to determine if there were any significant differences.

The literature presented in chapter two showed a lack of quantitative studies on the topic of DDDM, teacher collaboration in the form of PLCs, and their combine effects on student achievement. This scarcity of quantitative studies, specifically pure experimental studies, made it difficult to draw possible causal conclusions about the link between these variables and student achievement. Second, while there is an abundance of qualitative studies that provide depth of knowledge on the topic, the findings are less generalizable and cannot easily be used to support improvement efforts. This study employed a quantitative method and a quasi-experimental one-sample pretest-posttest design. The design was based on three theoretical constructs which are grounded in Kaizen's theory of continuous improvement. These constructs are collaborative inquiry, improvement science, and distributed leadership as part of a larger supportive environment. This chapter outlines how the research was executed and covers the following topics: The problem, Hypothesis, Research Design, Introduction to Chapter 3, Purpose of the Study, Research Questions, Target Population, Sampling Method (power) and Related Procedures, Instrumentation, Data Collection, Operationalization of Variables, Data Analysis Procedures, Limitations and Delimitations of the Research Design, Internal and External Validity, Expected Findings, Ethical Issues in the Study, and Chapter 3 Summary.

Purpose of Study

The purpose of this study was to determine the significance of a collaborative data-inquiry culture as promoted by the DWIP on student outcomes on the state mandated standardized PARCC assessments. The DWIP is a problem-solving approach, which is currently being implemented in School District X as an intervention to address years of poor student achievement, particularly in ELA and Math. This study measured the effectiveness of the DWIP intervention to guide the strategic planning of the school district as it wrestles with how to continuously improve its schools by improving student achievement. The DWIP intervention is heavily grounded in four popular constructs of school improvement; data-use, teacher collaboration in the form of PLCs, reflective practice, and distributed leadership. As a result, the findings of this study contributed significantly to the general body of knowledge around continuous school improvement.

Research Question and Hypotheses

The study sought to answer the following three research questions:

Research Question 1: What is the difference in student outcomes on the PARCC assessment in ELA and Math after the implementation of the DWIP?

Null Hypothesis 1: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math after the implementation of DWIP.

Alternate Hypothesis 1: There is a significant difference in student outcomes on the PARCC assessment in ELA and Math after the implementation of the DWIP.

Research Question 2: What is the difference in student outcomes on the PARCC assessment in ELA and Math at three different times of the DWIP implementation (Year 0, Year 1, and Year 2)?

Null Hypothesis 2: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math at three different times during the DWIP implementation (Year 0, Year 1, and Year 2).

Alternate Hypothesis 2: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math at three different times during the DWIP implementation (Year 0, Year 1, and Year 2).

Research Question 3: What is the difference in student outcomes on the PARCC assessment in ELA and Math based on the extent of the DWIP implementation (Not yet started, Initiating, Developing, or Sustaining)?

Null Hypothesis 3: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math based on the extent of Data Wise implementation (Not yet started, Initiating, Developing, and Sustaining).

Alternate Hypothesis 3: Student outcomes on the PARCC assessment in ELA and Math will be significantly different at each implementation level (Not yet started, Initiating, Developing, and Sustaining).

Research Design

This ex post facto study used a quantitative method with a quasi-experimental design. Specifically, a one-sample pretest-posttest design was used to collect data followed by descriptive and inferential statistical analyses to answer each research question. The intervention called the DWIP, which began in the fall of 2015, is in its fourth year of implementation in School District X. The time of the implementation of the intervention was treated in years with 2015 as year 0, 2016 as year 1, and 2017 as year 3. The pretest and posttest was the PARCC assessment in ELA and Math, which is a state mandated standardized test administered once a year to all students in grades three through 11. This study used archival PARCC data for ELA and Math for a cohort of about 2,500 6th grade students in four middle schools in School District X before and after the implementation of the DWIP intervention. This was the dependent variable (DV). The pretest was the PARCC scores for 6th graders in 2015 and the posttest was the PARCC scores for the same cohort of students who were 7th graders in 2016 and 8th graders in 2017.

The study also involved the administration of a voluntary likert-scaled survey of 58 teachers, counselors, and administrators about their perception of the collaborative data-inquiry practices at their respective schools and the implementation of the DWIP intervention. Archival DWIP data in the form of a *data wise journey* was examined and each of the eight steps of the DWIP was scored using the same survey administered to the teachers. The DWIP and collaborative data-inquiry practices constituted the manipulated or independent variable (IV). Quasi-experiments have been used as far back as the 18th century and continue to be frequently used by researchers today for three primary reasons, one of which is to evaluate the effectiveness of an intervention when the intervention has already been implemented by educators prior to the evaluation procedure having been considered (Shadish, Cook, & Campbell, 2002). According to Adam and Lawrence (2015), when the participants have already been exposed to the independent variable before the study is conducted, the study is classified as an ex post facto study.

The choice of a quasi-experimental one-sample pretest-posttest design for this study was based on following considerations: the purpose of the study, the nature of the study, the identified gap in literature, and literature support for the relevance of the methodology and design. The purpose of the study was to determine the significance of a collaborative data-inquiry approach to problem solving (DWIP) on student outcomes. Hence, a quantitative method would be most suitable for measuring differences in student outcomes and possible cause-effect relationships using inferential statistical measures. Second, a review of the literature on the topic and related topics identified a lack of quantitative studies. More than 75% of the studies reviewed were qualitative, 19% were quantitative, and six percent were mixed methods. A majority, or 72%, of the quantitative studies reviewed were descriptive analyses. The large percentage of qualitative studies made it impossible to draw possible causal conclusions on the topic (Gero, 2014; Hallam, Smith, Hite, Hite, & Wilcox, 2015; Jain, Cohen, Huang, Hanson, & Austin, 2014). Third, the ex post facto nature of the study most heavily influenced the specific type of design selected; the quasi-experimental one-sample pretest-posttest design. This design allowed for the study of an intervention that was implemented before the study was designed.

One reason why the quasi-experimental one-sample pretest-posttest design was selected was because the ex post facto nature of the study did not allow for the random selection of groups. This ruled out the use of a pure or true experimental design. While causal comparative and correlational quantitative designs were considered, both designs required having a control group and an experimental group. However, since the DWIP intervention was implemented in every school in School District X, selecting a control group required the use of a non-equivalent group from a neighboring school district with similar student and teacher characteristics. There were two potential challenges with selecting such a control group. One, the school may have data-use and collaborative practices similar to the DWIP. Second, it would be difficult and time consuming to secure permission from another school district. The quasi-experimental one-sample pretest-posttest design was the most feasible to address the purpose of the study. Mindful of the limitations of this design, steps were taken in the sample selection process to mitigate the effects of the extraneous variables.

The selection of a quasi-experimental one-sample pretest-posttest design for this type of study was supported by previous research studies. Lehman (2015) used a pre-experimental one-group pretest-posttest design to determine the effects of a professional development intervention on teachers' perception in analyzing and using student data. She cited the ex post facto nature of the study as the main reason for the choice in research design. Further, Lehman (2015) acknowledged the use of this design for conducting research studies of most interventions in the education setting. Lehman employed both the ANOVA test and the Paired Sample T-Test to measure differences in sample means between and among groups and differences between pretest and posttest scores (2015). In another quantitative study, which measured the impact of a multi-year (2014-2016) math response to intervention (RtI) on a group of 995 fifth graders as measured by the Smarter Balance Assessments, Park (2017) used a one-sample pretest-posttest design followed by inferential statistics. He (2017) used the ANOVA One-Way Repeated Measures to determine if there was a relationship between the duration of math RtI implementation and math performance as measured by Delaware Comprehensive Assessment System and Smarter Balance. According to Park (2017), the design was the most appropriate because it compared variations across group means with variations within groups. A Repeated Measures ANOVA allowed for the measurement of the same thing at different times, as in the case of a pretest-posttest comparison. The study of 995 fifth grade students in School District A showed that there was a higher level of significant improvement in math scores after RtI was implemented for one year as demonstrated on the Smarter Balance in 2015, but with continued implementation of RtI for the second year, there was very small improvement on the Smarter Balance in 2016 in comparison to the first year of RtI implementation (Park, 2017).

Target Population, Sampling Method (power) and Related Procedures

The setting for this study was School District X, which is a large public school district composed of more than 100,000 students, 15,000 employees, and 200 schools and centers. Among the more than 200 schools and centers, half are elementary schools, and the remaining schools comprise of middle schools, high schools, charter schools, and centers. School District X serves a diverse student population from urban, suburban, and rural communities. The student body includes a growing Hispanic and immigrant population. The composition of the staff is representative of the diversity of the student body.

The researcher employed a clustering approach to sampling (Creswell, 2014) in which four middle schools within School District X were identified for the study. All middle schools within School District X meeting the following criteria were placed in a pool from which six schools were selected using a simple random approach (Creswell, 2014): (a) 85% or better student attendance rate, stability in leadership for the past three years, 85% or more of teachers as highly qualified, evenly mixed student body which includes FARMS, SPED, ESOL, and general education students, and a less than 15% teacher turnover rate. Selection from these criteria allowed for the mitigation of extraneous variables that could impact student outcomes on the PARCC assessments regardless of the DWIP intervention. It also allowed for stratification to ensure that the sample of schools has characteristics that are proportionally representative of the larger school district population (Fowler, 2002). Among the four middle schools, the PARCC scores for the 2015 6th Grade Cohort of students was traced over a period of three years of implementation of the DWIP intervention to determine any significant differences in student scores in ELA and Math. Student performance for the 6th Grade Cohort was traced as follows: 6th Grade, 2015 (Pretest); 7th Grade, 2016; and 8th Grade, 2017 (Posttest). Performance data were taken from the PARCC report card for each school from the state department of education website. The data showed how many students at each grade level scored at each of the five PARCC performance levels. The 6th grade class for all four middle schools combined was approximately 2,500 students.

A simple random sampling approach was used to select a sample of teachers to participate in a survey about their perception of the collaborative data-inquiry practices and the implementation of the DWIP intervention at their respective school. Randomization of the sample allows for the generalizability of the findings (Creswell, 2014). An invitation to participate in an anonymous click consent survey was sent to the work email addresses of all the teachers, counselors, and administrators in the four middle schools. Each teacher had an equal opportunity to voluntarily participate in the study. Roughly 250 invitations were sent out and 58 staff members responded.

The sample size was determined using the equation $N = 1.96^2 \sigma^2 / E^2$ where 1.96 is the confidence level at 95%, σ is the standard deviation, and E is the margin of error. The values for σ and E were determined from previous similar research. Also, a G-Power Analysis 3. 1.9. 2 was used to calculate the sample size (*N*) of teachers and students. The sample size was dependent on the statistical tool used to answer the research questions. In this study, the researcher determined the significance of a collaborative data-inquiry culture, as promoted by the DWIP, on student outcomes on the PARCC ELA and Math. Although, the study had one categorical independent variable, which was the DWIP intervention, and two continuous and related dependent variables, which were the PARCC ELA and Math scores, neither the One-Way MANOVA nor the ANOVA could not be used as the inferential tool because PARCC scores were taken on the school level. Therefore, the Chi-Squared Test of Homogeneity was used. According to G. Power Analysis 3.1.9.2, a sample size of N = 72 teachers and N = 251 students was required for a test at 95% confidence level, with a p of 0.05, two groups, and three measurements.

Instrumentation

Since the study is of an ex post facto nature, archived PARCC data for ELA and Math was retrieved and analyzed. The Mid Atlantic State in which School District X is located is part of the PARCC consortium, has designed state standards that mirror the common core standards, and has been involved in the design and field testing of PARCC assessment items. The PARCC test was field tested in 2014 and was implemented as the state mandated assessment for grades three through 11 since 2015. It was the measure of student performance outcomes in this study. A survey instrument was also used to capture teacher perception of the collaborative data-inquiry practices and the levels of implementation of the DWIP intervention in the four middle schools. The instrument was designed using *Qualtrics* software and was a combination of the Data Wise Rubric used in the Charlotte-Mecklenburg Schools study (Algozzine, Friend, McRae, & Seifert, 2011) and the School Culture Survey (Gruenert & Valentine, 2005). Permission was obtained from Dr. Boudett of the Harvard University Data Wise Project, and from Drs. Gruenert and Valentine to use the instruments. The Data Wise Rubric was used to capture the extent to which schools were implementing the DWIP intervention by surveying teachers and by examining the Data Wise Journey for each school. Each school maintains a Data Wise Journey with artifacts that show how each step of the DWIP intervention is implemented. The School Culture Survey is a six-factor 35-item instrument used to assess teachers' perception of the collaborative nature of the schools' culture in the following six areas: collaborative leadership, teacher collaboration, professional development, unity of purpose, collegial support, and learning partnership.

The PARCC assessment, Data Wise Rubric, and the School Culture Survey have established reliability and validity. Creswell (2014) defined validity as the ability to draw meaningful inferences from the scores of the instrument. He (2014) identified three traditional forms of validity; content validity—do the items measure the content they were intended to measure, predictive validity—do scores predict a criterion measure and correlate to other scores, and construct validity—do items measure hypothetical constructs and do they have a useful purpose and have positive consequences when they are used in practice. Creswell (2014) defined reliability as having internal consistency (are items response consistent across constructs?) and test retest correlation (are scores stable over time when the instrument is administered a second time?).

Data Wise Rubric validity and reliability. To obtain a measure of schools' or teams' ability to demonstrate critical indicators within the eight steps of the DWIP, a Data Wise rubric was developed using content from *Data Wise: A Step-by-Step Guide to Using Assessment Results to Improve Teaching and Learning* (Boudett et al., 2015) and *Data Wise in Action: Stories of Schools Using Data to Improve Teaching and Learning* (Boudett & Steele, 2007). The instrument was used in a 2011 study conducted by the Charlotte-Mecklenburg Schools (Algozzine, Friend, McRae, & Seifert, 2011). The instrument was tested on 93 teams in 48 schools (27 elementary, 14 middle, and seven high), observer reliability checks were conducted, and once 80% or above observer agreement were obtained, the evaluation tool was used in the fall and the spring of school year 2010-2011 by Charlotte-Mecklenburg Schools to measure the extent of the DWIP implementation in their school district. The results of the study in the fall and spring were consistent.

School Culture Survey validity and reliability. Gruenert and Valentine (2005) designed a six factor, 35-item survey to determine if there was correlation between a collaborative school culture and student achievement in Indiana Public Schools. The survey was developed in 1998 and was based on descriptors of collaborative culture found in literature. These six descriptors were collaborative leadership, teacher collaboration, professional development, and unity of purpose, collegial support, and learning partnership. The survey was first piloted on 634 participants as a 79-item survey, and was reduced to 35 items after being placed through a Varimax rotation. Both internal correlation and Chronbach's alpha were established for each of the six-factors. Chronbach's alpha for each of the six factors were as follows: collaborative leadership 0.910, teacher collaboration 0.834, professional development 0.821, unity of purpose 0.867, collegial support 0.796, and learning partnership 0.658. Validity was established using correlated methodology with the National Association of Secondary Principals' CASE-IMS Climate Survey. The survey was the administered to 81 schools using 2750 surveys in the spring of 2002. Using this survey instrument, the research found that "the more collaborative schools tend to have higher student achievement" (Gruenert, 2005, p. 46).

PARCC validity. The Partnership for Assessment of Readiness for College and Careers (PARCC) is the state mandated test used to measure college and career readiness or progress towards college and career readiness. It uses a 5-point performance band with a band of 4 or better indicating college readiness. The test has been used in more than 43 states since 2015 and is administered to students from grades three through 11 in two core subject areas: ELA and Math. The Mid-Atlantic State in which School District X is located is a member of the PARCC

Consortium and participated in the PARCC field-testing in 2014 in which at least one school in each of the 24 school districts in the state participated in the two-phase testing.

The quality and rigor of a quantitative study is enhanced through the measurement of validity and reliability. According to the *Standards for Educational and Psychological Testing*, "validity refers to the degree to which evidence and theory support the interpretations of test scores for uses of tests" (2014, p. 11). Validity must be established for each purpose of an assessment and also depends on technical aspects of the assessment, such as appropriate test administration, scoring and accurate score scaling, equating, and standard setting. Heale and Twycross (2015) defined validity as the extent to which a concept is accurately measured. Phelan and Wren (2006) referred to validity as how well a test measures what is purported to be measured. There are three major types of validity; content, construct, and criterion-related (Creswell, 2015; Heale & Twycross, 2015; Phelan & Wren, 2006).

Content validity. Content validity is the extent to which an instrument accurately measures all aspects of a concept or construct under study (Creswell, 2015; Heale & Twycross, 2015, Phelan & Wren, 2006). The PARCC assessment is most closely aligned to the Common Core State Standards and measures a broad range of knowledge, skills, and behaviors needed by students to be college and career ready. A study on test quality conducted by Thomas B. Fordham Institute and the Human Resources Research Organization (HumRRO), found that PARCC scored higher than three other similar assessments (Massachusetts Comprehensive Assessment System (MCAS), Smarter Balanced Assessment Consortium (SBAC), and ACT Aspire) on alignment to the common core college and readiness standards and assessment of higher order thinking skills (Loschert, 2018). The design and development of PARCC has included the collection and analysis of a variety of data. PARCC has commissioned research that analyzes its 2014 field test, the item development, the test administration, accessibility, quality of items, and comparability of the paper and computer-based assessments. They have used this to make adjustments and improvements to the test. Finally, in the summer of 2015, as the final step in the standard setting process, the consortium used data gathered during the 2015 test administration to compare actual student performance with their earlier estimates and, based on that analysis, established the final cut scores to be used to distribute student performance across the five performance standards.

Construct validity. Construct validity is the extent to which an instrument or tool measures what is actually intended and no other variables. The PARCC assessment is among the best measures of college and career readiness when compared to similar tests such as SBAC, MCAS, and NAEP. A performance score of 4 on the PARCC assessments indicates that the student is ready for college. Similarly, on the Smarter Balance test, a score of a 3 indicates college readiness. When compared to other college readiness test, such as the ACT, SAT, MCAS, and SBAC, the PARCC assessments compares well in determining students' readiness for college.

Criterion or predictive validity. Criterion-related or predictive validity is the extent to which a research instrument is related to other instruments that measures the same variable. PARCC assessments have good predictive validity in preparing students for college and career. In the spring of 2015, the Executive Office of Education in Massachusetts commissioned a study of nearly 850 first-year college students at 11 public colleges and universities throughout the state of Massachusetts to provide objective evidence about the extent to which students' scores on the high school MCAS and PARCC Math and ELA assessments accurately predict success in college. Mathematica Policy Research was then contracted to analyze student scores and correlate them with the students' performance in college (measured by grade point average), college readiness (measured by Accuplacer scores), and placement in remedial courses (measured by course enrollment data). They found that both MCAS and PARCC predicted college readiness as measured by first-year college grades. Both MCAS and PARCC scores provided statistically significant predictions, and both are comparable to SAT scores in predicting first-year college outcomes (Ansel, 2015). Similarly, scores on both MCAS and PARCC provided strong predictions about which students needed remedial coursework in college. PARCC also did a benchmarking study to gather information from other international, national, and state assessments (including NAEP, SAT, ACT, PISA, NY Reagents) to help provide information about the percentage of students who are college and career ready. A performance level of 4 was defined in relationship to the results of the other assessments such as TIMSS, PIRLS, and NAEP. These comparisons helped to inform the initial establishment of performance standards set by PARCC, including the definition of college and career readiness.

PARCC reliability. Reliability is considered the second fundamental element of an assessment and it refers to the "consistency of scores across replications of a testing procedure, regardless of how this consistency is estimated or reported" (Ansel, 2015, p. 33). PARCC offers both a paper-pencil and online version of the test, which produces the same student outcomes (Ansel, 2015). Additionally, when compared to other high stakes assessment such as the SBAC and MCAS, PARCC produces scores that are comparable or better. SBAC, for instance, groups student test scores into four achievement levels and indicates that a score at or above level three suggests a student is ready for college-level course work. PARCC, meanwhile, groups student test scores into five achievement levels and classifies scores at or above level four as a sign of college readiness.

Data Collection

Data collection for this study began after IRB approval was obtained from both University Y and School District X. IRB approval for University Y was sought in the spring of 2018. Immediately after receiving IRB approval from University Y, approval from School District X was sought. Once approval from both institutions was received, a sample of four middle schools with similar characteristics was identified for the study and a letter was sent out to the principal via email to inform him or her of the nature of the study and the selection of the school as part of the study sample. Teachers were informed of the study and were invited to participate with a chance to win prizes.

Three types of data were collected to answer the three research questions about the impact of collaborative data-inquiry culture, as promoted by the DWIP intervention (IV), on student achievement (DV). Data about the collaborative data-inquiry practices, the extent of the DWIP implementation at each school site, and PARCC scores were collected. A two part 46-item survey called the Collaborative Data-Inquiry Culture survey was administered to all teachers, counselors, and administrators in the four middle schools which were named Schools A-D. The first part of the survey, that is questions 1-35, collected data on teachers' perception of the collaborative practices at their respective schools. The second part of the survey, questions 36-46, collected data on teachers' perception of the extent of the implementation of the DWIP intervention. Additionally, the researcher used questions 36-46 of the survey to examine and score the Data Wise Journey of each school which contained artifacts showing how well they were implementing each step of the DWIP intervention (not yet started, initiating, developing, and sustaining).

The survey was an anonymous click consent survey and was sent out to all teachers, counselors, and administrators in Schools A-D on the first day of the study using their work issued email addresses. The consent letter contained information explaining the purpose of the research study and requesting voluntary participation. The researcher used a modified approach of Salant and Dillman (1994) four-phase approach to conducting a mail survey. Approximately seven days after the initial email, a reminder email was sent out to all non-respondents with a second invitation to participate in the survey. This was repeated on day 14, 21, and 28 after the initial invitation email for participation was sent out. The researcher concluded the administration of the survey after four weeks.

In order to examine the Data Wise Journey of Schools A–D, a separate letter was sent out via email to the principal of each of the schools requesting permission for access. The Data Wise Journey is maintained on a school district Google site and stores artifacts of the work on each step of the DWIP. Once access was granted, the Data Wise Rubric, or questions 36–46 of the Collaborative Data-Inquire Culture Survey was used to determine how well each school was implementing the DWIP intervention. The researcher ratings for the DWIP implementation at each school were compared to the average survey score of questions 36–46. This provided an opportunity to perform a "check" on the voluntary survey responses.

PARCC scores in ELA and Math were used as the measure of student outcomes for this study. Each schools' yearly overall performance by grade level is published on the state department of education website which is available to the public. While no permission was required to access the test data, as a courtesy and to be fully transparent, the principal of each of the four schools under study was informed of the intent to use school test data. School performance data on the PARCC was accessed and analyzed for the following school year: 2014-2015 (Pretest), 2015-2016, and 2016-2017 (Posttest). The test data used was for the 6th Grade

Cohort in 2015, 7th Grade in 2016, and 8th Grade in 2017. This allowed the researcher to track the same students throughout the implementation of the DWIP intervention. The number and percentage of students who received scores in the five performance bands were captured for each of the three years (1- expectations not met, 2- expectations partially met, 3-approaching expectations, 4- met expectations, and 5- exceeded expectations).

Operationalization of Variables

In this study, the researcher sought to determine the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement by answering three questions.

Research Question 1: What is the difference in student outcomes on the PARCC assessment in ELA and Math after the implementation of the DWIP? In this question, the independent variable was the overall DWIP intervention, and the dependent variables were student PARCC score in ELA and Math in 2015 (pretest) and 2017 (posttest).

Research Question 2: What is the difference in student outcomes on the PARCC assessment in ELA and Math at three different times of the DWIP implementation (Year 0, Year 1, and Year 2)? In this question, the independent variable was the three different implementation times of the DWIP intervention, and the dependent variables were student PARCC scores in ELA and Math in 2015, 2016, and 2017.

Research Question 3: What is the difference in student outcomes on the PARCC assessment in ELA and Math based on the extent of the DWIP implementation? In this question, the independent variable was the implementation levels of the DWIP intervention (Not yet started, initiating, developing, and sustaining) and the dependent variables were the student scores on the PARCC assessment in ELA and Math in 2015, 2016, and 2017.

Data Analysis Procedures

This section provides the sequence of events for the data analysis for this quantitative, quasi-experimental, one-sample pretest-posttest study. All three research questions and the accompanying null hypotheses were answered using the Chi-Squared Test of Homogeneity followed by pairwise comparisons using a Z-Test of Multiple Proportions with a Bonferroni correction for adjusted alpha values. Although the study contained one categorical independent variable (the DWIP intervention) and two continuous and related dependent variables (PARCC ELA and PARCC Math Scores), the One-Way MANOVA could not be used to measure mean differences in student outcomes because student achievement data were collected on the school level instead of the individual student level (Rockinson-Szapkiw, 2013).

Instead, the Chi-Squared Test of Homogeneity was used to determine if there was a statistically significant difference in the proportion of students who *met* and *did not meet* performance expectations on the PARCC assessment. According to Laerd Statistics (2016), the Chi-Square Test of Homogeneity is the most appropriate test to determine if a difference exists between binomial proportions of two or more independent groups on a dichotomous dependent variable. Further, the Chi-Square Test of Homogeneity allows for the use of total outcomes as in the case of students' PARCC performance data which was obtained on the school level for this study. Statistical significant difference between proportions was determined at the standard alpha of $p < 0.05$. If there was a statistically significant difference in proportions, a pairwise comparison using Z-Test of Two Proportions with a Bonferroni correction were used to determine exactly where the differences lie.

To answer research question 1 about the difference in student outcomes on the PARCC assessment in ELA and Math after the implementation of the DWIP intervention, the following null hypothesis was tested: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math after the implementation of DWIP. Two 2x2 cross tabulations, one for each subject, were performed as follows: Performance expectation (met and not met) x PARCC assessment (pretest and posttest). If it was determined that there was statistically significant difference in the proportion of students who met and did not meet performance expectations on the pretest and posttest at the standard alpha of $p < 0.05$, a pairwise comparison using Z-Test of Multiple Two Proportions with a Bonferroni correction was performed to determine where the differences lie. If it was further determined that there was statistically significant difference between proportions at the adjusted alpha of $p < 0.0125$, the null hypothesis was rejected as it related to that subject.

To answer research question 2 about the difference in student outcomes on the PARCC assessment in ELA and Math at three different times of the DWIP implementation (Year 0, Year 1, and Year 2), the following null hypothesis was tested: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math at three different times during the DWIP implementation. Two 2x3 cross tabulations, one for each subject, were performed as follows: Performance expectation (met and not met) x PARCC assessment (2015, 2016, and 2017). If it was determined that there was statistically significant difference in the proportion of students who met and did not meet performance expectations on the PARCC at the standard alpha of $p < 0.05$, a pairwise comparison using Z-Test of Two Proportions with a Bonferroni correction was performed to determine where the differences lie. If it was further determined that there was statistically significant difference between proportions at the adjusted alpha of $p < 0.00833$, the null hypothesis was rejected as it related to that subject.

To answer research question 3 about the difference in student outcomes on the PARCC assessment in ELA and Math based on the extent of the DWIP implementation (Not yet started, Initiating, Developing, or Sustaining), the following null hypothesis was tested: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math based on the extent of Data Wise implementation. Four 3x3 cross tabulations, two for each subject, were performed as follows: Students meeting PARCC expectations (2015, 2016, and 2017) x DWIP implementation levels (Initiating, lowly developing, and developing). A similar cross tabulation was performed for students not meeting PARCC expectations. If it was determined that there was statistically significant difference in the proportion of students who met and did not meet performance expectations at different DWIP implementation levels at the standard alpha of $p < 0.05$, a pairwise comparison using a Z-Test of Two Proportions with a Bonferroni correction was performed to determine where the differences lie. If it was further determined that there was statistically significant difference between proportions at the adjusted alpha of $p < 0.00556$, the null hypothesis was rejected as it related to that subject.

Limitations and Delimitations of the Research Design

Limitations and delimitations are conditions or circumstances that may influence a study. Limitations are influences that include conditions or factors that cannot be controlled by the researcher placing restrictions on methodology and conclusions (American Psychological Association, 2009; Lunenburg & Irby, 2008). In this quantitative quasi-experimental study, the researcher was not able to directly control the following conditions or factors:

1. The Collaborative Data-Inquiry Culture Survey that was administered to teachers relied on self-reporting. The accuracy of the reporting was based on the assumption that teachers were honest about the collaborative data-inquiry practices at their

respective schools and the implementation of the DWIP intervention.

2. The ex post facto nature of the study made it impossible to randomly assign groups for the study since all participants had already been exposed to the independent variable.
3. The ex post facto nature of the study made it impossible to truly manipulate the independent variable. As a result, the researcher did not have total control over extraneous or confounding variables that might have affected the outcomes of the study.
4. The PARCC assessments used to measure student outcomes was limited to two subject areas; ELA and Math. This set limits on the analysis of overall student performance.
5. The extent and diligence with which schools implemented the DWIP intervention was different. As result, a third research question was designed to address this limitation.
6. The Common Core Standard-Based Curriculum and the accompanying PARCC assessments have only been recently implemented in School District X in the last three to four years. This required a shift in instructional pedagogy and student familiarity with a more rigorous assessment, which also had some impact on student outcomes.

Delimitations are boundaries set for a study by the researcher (American Psychological Association, 2014; Lunenburg & Irby, 2008). The following delimitations will be set in this quantitative quasi-experimental study:

1. This study sought to answer three research questions about the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement.

2. The study sample of four middle schools was taken from a population of more than 30 middle schools within a larger school district of approximately 200 schools, 100, 000 students, and 20, 000 employees.
3. The study sample comprised of 58 teachers and 2,631 students or roughly 15 teachers and 650 students from each of the four schools.
4. The sample of students was the entire 2015 6th Grade Cohort of the four middle schools combined.
5. The interval for the DWIP intervention (IV) was taken for the last three consecutive school years (2014-2015, 2015-2016, and 2016-2017).
6. The measure of student outcomes (DV) was the PARCC assessments for the last three calendar years (2015-2017).

Internal and External Validity

According to Creswell (2009), experimental researchers need to identify threats to the internal validity of experiments and take steps to mitigate them so that questions are not raised about the researchers' ability to conclude that the intervention outcome is due to one factor and not some other extraneous or confounding factor. These considerations include experimental procedures, treatment, and instrumentation. The following steps were taken to minimize the impact of confounding and extraneous factors on the outcome of the study:

1. From the population of middle schools in School District X, a cluster of schools meeting the following criteria was identified: at least three-year stability in core school leadership, 85% of highly qualified teachers, 85% student attendance rate, and a diverse student demographic including socio-economic background, SPED, ELL, and Gen Ed.

2. Using a simple random sampling approach, four schools were selected from the cluster of schools identified above.
3. Teacher survey participants were randomly selected from each of the four schools.
4. The use of the PARCC assessment as a measure of student achievement added validity and reliability to the study because it has been used and tested in a majority of the 50 states including the state in which School District X is located.
5. The examination and scoring of the schools' Data Wise Journey using the Data Wise Rubric allowed for cross checking of voluntary survey responses on the extent of DWIP implementation at each of the four schools. The Data Wise Journey contains artifacts that show the footprint of the implementation of each of the eight steps of the DWIP.
6. The School Culture Survey (Gruenert & Valentine, 2005) has been used in similar and previous studies on the topic of collaborative culture and student achievement. The Data Wise Rubric (Boudett, City, & Murnane, 2016) has been used in a large school district to measure the extent of Data Wise implementation.

Threats to external validity must also be considered and designs must be created to minimize those threats. According to Creswell (2009), external threats arise when researchers draw incorrect inferences from the sample data to other persons, other settings, and past or future situations. The setting in which this study took place has a unique student population which come from mostly middle-income minority households. Further, the student population is diverse with students from urban, suburban, and rural communities. As a result, the claims in this study will be limited to school communities with similar student demographics.

Expected Findings

This quantitative, pre-experimental study sought to determine the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement. A review of the literature showed that there is a shortage of quantitative studies on the topic, which restricts that ability to make possible causal inferences. The study intended to address the gap in methodological approach, and make possible causal inferences between a collaborative data-inquiry culture, DWIP intervention, and student achievement thereby adding to the body of knowledge on the subject. This was achieved by answering three research questions whose answers should reveal the following: (a) collaborative data-inquiry culture as promoted by the DWIP intervention significantly increased the proportion of students meeting performance expectations on the PARCC in ELA and Math, (b) the proportion of students meeting performance expectations on the PARCC in ELA and Math significantly increased in the first two years of implementing the DWIP intervention but levels off in the third year, (c) the proportion of students meeting performance expectations on the PARCC in ELA and Math is significantly higher the greater the extent of the implementation of the DWIP intervention, (d) schools at the same DWIP implementations level should see similar increases in the proportion of students meeting performance expectations on the PARCC assessment in ELA and Math.

Ethical Issues in the Study

This study posed no risk to the participants and there were no ethical concerns. The anonymity and confidentiality of all participants were maintained. PARCC scores, which were obtained from the department of education website, are reported by school and show the total number of students at each grade level who scored at each of the five PARCC performance levels. There were no student names or identification numbers associated to the scores. However, a letter was sent to the principal of each of the four schools informing them of the study, its purpose, and the intent to use the school's data. Second, pseudonyms were used for the four middle schools (School A-D) in the study. Third, a pseudonym was used for the school district in which the study was conducted. The school district was referred to as School District X throughout the study. Fourth, surveys were sent out to teachers using the *Qualtrics* software so as to maintain the confidentiality of the participants. Each participant was assigned a unique identifier.

The procedures for conducting the study as approved by the Concordia University IRB and the School District X IRB were strictly enforced. All participants were informed of the purpose of the study, and participation in the study was completely voluntary. The rights of each participant were clearly communicated in the consent form. Although no one dropped out of the survey, all participants were given the opportunity to opt out freely and without pressure. The purpose of the study and how the findings would be communicated were also shared with the participants before the survey. The findings of the study will benefit all stakeholders in the School District X because it will add to the body of knowledge on how a collaborative data-inquiry culture, as promoted by the DWIP intervention impacts student achievement, thereby informing their school improvement efforts and district wide strategic planning.

Summary

This section focused on the design of the study which used a quantitative method, and a quasi-experimental one-sample pretest- posttest design to measure the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement. It explained data collection procedures, instrument design and use, analytical tools used and provided a rationale for using them. Data were collected using a click consent survey, document review, and collection of archived student test scores. The Chi-Square Test of Homogeneity followed by pairwise comparison tests were used to answer each of the three research questions to determine if there were significant differences in the proportion of students meeting performance expectation on the PARCC in ELA and Math as a result of the DWIP intervention. The section also outlined efforts made to increase validity and reliability of research outcomes, and steps taken to mitigate internal and external validity. It showed how the researcher implemented the research design as approved by the institutional review board to ensure the highest level of ethical behavior that minimized risks to participants and increased the benefits of the study. The following chapter will provide an introduction, followed by a description of the sample, a summary of the results, a detailed analysis of the results, and a chapter summary.

Chapter 4: Data Analysis and Results

Introduction

This study investigated the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement in four middle schools in School District X. The study used a quantitative method with a quasi-experimental one-sample pretest-posttest design to identify any significant difference in student outcomes after the implementation of the intervention. The DWIP is intended to promote continuous improvement in schools by creating a collaborative data-inquiry culture. School District X implemented the DWIP as a systemic intervention at the beginning of the 2015-2016 school year. The study sample included 2,631 students and 58 teachers from four middle school in School District X. The four schools were selected using a cluster sampling method in which all schools with similar characteristics were identified from all middle schools in School District X, and from which four were randomly chosen. This method was used to mitigate for confounding factors that were likely to affect student achievement such as stability in school leadership, teacher quality, student attendance, and students cultural and socio-economic background. The student sample consisted of all students in the four schools who took the PARCC assessments in ELA and Math as sixth graders in 2015 ($N = 698$), seventh graders in 2016 ($N = 785$), and eighth graders in 2017 ($N = 1148$). Twenty three percent or 58 out of 249 of the instructional staff participated in the study by completing an anonymous click consent survey.

Two instruments were used in this research study to measure student achievement and teacher perception of school culture and the implementation of the DWIP intervention. The anonymous click consent survey which was administered to teachers was made up of two existing instruments; the School Culture Survey (Valentine & Gruenert, 2005) and the Data Wise

Implementation Rubric (Boudett et al., 2016). According to Valentine and Gruenert (2005), the School Culture Survey has been used in more than one hundred studies in the United States and the findings of those studies were similar to the findings in their original study which was done in Indiana Public Schools. The Data Wise Rubric, designed by the Harvard University Data Wise Project, is used by many school districts, including School District X, to measure the implementation of the Data Wise Process. Both instruments were used in their current form to maintain the validity and reliability. The PARCC assessment was the instrument used to measure student achievement. The PARCC is used in more than 26 states as the state mandated standardized test. PARCC was field tested in the state in which School District X is located and has been the measure of student achievement since 2015. The test is administered once a year in ELA and Math to students in grades three through 11.

The purpose of the study was achieved by answering three research questions about the differences in student performance on the PARCC assessments after the implementation of the DWIP intervention. Specifically, the research examined the difference in the proportion of students who met and did not meet performance expectations on the PARCC after the implementation of the intervention, at three time periods of the intervention (Year 0, Year 1, and Year 2), and at three different implementation levels of the intervention (initiating, lowly developing, and developing). Descriptive analyses, and Chi-Square Test of Homogeneity followed by pairwise comparisons using a Z-Test of Two Proportions were used to answer each research question. This chapter begins with an introduction, followed by a description of the sample, a summary of the results, a detailed analysis of the results, and a chapter summary.

Description of Sample

The study was conducted in School District X, a large school district in a small Mid-

Atlantic State. School District X serves more than 100,000 students from culturally and socio-economically diverse backgrounds. Students come from urban, suburban, and rural communities. The participants for this study consisted of a total of 2,631students and 58 teachers from four middle schools in School District X. The number of middle schools used in the study was less than the proposed six schools because permission was not granted by the principals of two schools. Students' PARCC performance data were obtained on the school level for students who were in sixth grade in 2015 (N = 698), seventh grade in 2016 (N = 785), and eighth grade in 2017 (N = 1148). The 58 teachers who participated in the survey for the study constituted 23% of the teachers in the four schools. A total of 249 survey invitations were sent out. According to G-Powered Analysis 3.1.9.2, the sample was large enough to measure for a medium effect.

A cluster sampling approach was used to select the schools for the study. All middle schools in School District X meeting the criteria below were placed in a cluster from which four school were selected using simple random sampling. The criteria for the selection of the cluster were 85% student attendance rate, 85% of highly qualified teachers, three-years stability in leadership, and a diverse student population including ethnic and socio-economic diversity. Tables 1, 2, and 3 below give a breakdown of student demographic data and school characteristics.

Table 1

Gender of Students in Sample School Population

School	% Male	% Female
A	51.7	48.3
B	48.1	51.9
C	47.5	52.4
D	52.9	47.1
Average	50.04	49.92

Table 1 indicates the proportion of males and females in each of the four schools used in the study. Overall, there was an equal distribution of males and females among the four schools with 50.04% of students being males and 49.92% of students being females. Although gender was not used as a factor in this study, the proportion of males to females in the schools was most likely represented in the study sample.

Table 2

Ethnicity of Students in Sample School Population

	Percent (%)				
Ethnicity	School A	School B	School C	School D	Average
Am. Indian/AK Native	0	0	0	0	0
Asian	0	2	3.6	2.6	2.1
Black/African Amer.	82.3	30	55.6	69	59.2
Hispanic/Latino	7.7	59	35.6	20	30.5
HI/Pac. Islander	0	0	0	0	0
White	4.5	6	3.3	5	4.9
Two or more races	4.0	2.4	1.7	2.5	2.7

Table 2 indicates that the four schools from which the student sample was taken consisted of two major ethnic groups of which an average of 59.2% were African Americans, and 30.5% were Hispanic/Latino. Table 3 indicates that the four schools from which the student sample was taken shared similar characteristic in student attendance rate with an average of 94.7%, consisted of a distribution of SPED, FARMS, and LEP students, had at least 85% of highly qualified teachers, and an average tenure of the principal at the same school of 4.5 years.

Table 3

Sample School Characteristics

Characteristics	School A	School B	School C	School D	Average
% Student Attendance	>95.0	94.0	94.6	> 95.0	94.7
% SPED	10.6	10.7	9.1	19.2	12.4
% FARMS	35.5	77.8	72.8	63.6	62.4
% LEP	< 5.0	19.8	12.8	7.1	11.2
% Highly Qual. Teachers	85.7	85.7	83.7	85.8	85.2
Years of Stability in Leadership	5	3	5	5	4.5

Note. SPED = students with special education needs; FARMS = students who qualify for free and reduced meals; LEP = students with limited English Proficiency.

Summary of Results

Validity and reliability of results. This study was conducted with fidelity and as proposed to ensure that it measured what it was purported to measure, sufficiently, and consistently. In this case, it was the impact of a collaborative data-inquiry culture, as promoted by the DWIP intervention, on student performance on the PARCC assessments in ELA and Math. Several steps were taken with instrumentation, sampling, data collection, and data analysis to heighten validity and reliability of the results. First, the researcher used previously established survey instruments to measure teachers' perception of school culture and the implementation of the Data Wise intervention. Both the School Culture Survey (Valentine & Gruenert, 2005) and the Data Wise Rubric (Boudett et al., 2016) were used in their current form and without modification. PARCC data were used as the measure of student performance. Since student PARCC data were collected on the school level, it was impossible to use the One-Way MANOVA or the ANOVA of Repeated Measures as the inferential tools to answer the three research questions regarding differences in means of student outcomes. As a result, the Chi-Square Test of Homogeneity was used to look at the difference in proportions of students who met and did not meet performance expectations on the PARCC assessments. This required modification of the research questions to show how student achievement was measured. It also required modification of the hypotheses to make them two tailed to accommodate outcomes that were zero, positive, or negative.

The cluster method used to select the study sample allowed for a degree of randomness while also mitigating for extraneous variables that could affect student achievement. A cluster of middle schools was taken from all middle schools in School District X who met a pre-determined criterion set for student attendance, teacher quality, stability in leadership, and schools' demographic composition (see Table 2 and Table 3). Six schools were then randomly selected from the cluster. However, only four of the six schools granted permission for the study. Although the study was only able to use four schools, it still met the required sample size of teachers and students.

Consistency in data collection procedures also ensured that the data collected was valid and reliable. To ensure this, the researcher pre-screened the email list of potential survey participants and removed any teacher who might have known the researcher. This eliminated potential bias in the survey responses. Additionally, an anonymous and voluntary click consent survey was administered to all participants at the same time and for the same length of time. The survey was conducted over a four-week period. Participants were sent reminders every seven days to complete the survey.

The researcher conducted a "check" on survey responses to increase the validity of the study. This was done by scoring the Data Wise Journey of each school using questions 36-46 of the Collaborative Data-Inquiry Survey instrument. Questions 36-46 came from the Data Wise Rubric (Boudett, et al., 2016). The Data Wise Journey contained artifacts that show the footprint of the implementation of each of the eight steps of the DWIP. The researcher's score for the DWIP steps for each school was factored as one participant score when calculating the DWIP implementation level. This helped to mitigate for any inconsistencies in survey responses.

One potential threat to the validity and reliability of the results was the potential for committing type I errors. The potential for committing a type one error, that is, answering a question as true when it is false, was increased because multiple Chi-Square Tests were required to answer each research question. To address this threat, if there was statistically significant difference in the proportion of students who met and did not meet performance expectations on the PARCC at the standard alpha, $p < 0.05$, pairwise comparisons using a Z-Test of Two Proportions were performed to determine where the differences lie. Statistical significance was therefore determined using adjusted alpha values which were computed using a Bonferroni correction method.

Limitations and delimitations. Limitations and delimitations are conditions or circumstances that may influence a study. Limitations are influences that include conditions or factors that cannot be controlled by a researcher placing restrictions on methodology and conclusions (Adams & Lawrence, 2015; *American Psychological Association*, 2010). There were no limitations encountered in addition to those outlined in the proposal. However, the researcher anticipated a possible difference in the extent and diligence with which schools implemented the DWIP intervention. As result, a third research question was designed to address this limitation, "What is the difference in student outcomes based on the extent of the DWIP implementation?" Delimitations are boundaries set for a study by the researcher (*American Psychological*

Association, 2010; Creswell, 2014). The researcher made a small change to the sample population which impacted the teacher sample size for the study. Four, middle schools were used instead of the proposed six because the researcher did not receive approval from two schools. However, a recalculation of the required sample size using G-Powered Analysis 3.1.9.2 indicated that 58 teachers and 2,631 still exceeded the minimum sample size required to measure for a medium effect. There were no other changes to the original delimitations.

Research questions and hypotheses. Three research questions and the associated hypotheses were addressed in this study. The research questions of this study were:

Research Question1: What is the difference in student outcomes on the PARCC assessments in ELA and Math after the implementation of the DWIP?

Research Question 2: What is the difference in student outcomes on the PARCC assessments in ELA and Math at three different times of the DWIP implementation (Year 0, Year 1, and Year 2)?

Research Question 3: What is the difference in student outcomes on the PARCC assessments in ELA and Math based on the extent of the DWIP implementation (Not yet started, Initiating, Developing, or Sustaining)?

The null hypotheses of this study were:

Null Hypothesis 1: There is no significant difference in student outcomes on the PARCC assessments in ELA and Math after the implementation of DWIP.

Null Hypothesis 2: There is no significant difference in student outcomes on the PARCC assessments in ELA and Math at three different times during the DWIP implementation (Year 0, Year 1, and Year 2).

Null Hypothesis 3: There is no significant difference in student outcomes on the PARCC assessments in ELA and Math based on the extent of Data Wise implementation (Not yet started, Initiating, Developing, and Sustaining).

The alternative hypotheses for this study were:

Alternate Hypothesis 1: There is a significant difference in student outcomes on the PARCC assessments in ELA and Math after the implementation of the DWIP.

Alternate Hypothesis 2: Student outcomes on the PARCC assessments in ELA and Math will be significantly different at three different times during the Data Wise implementation (Year 0, Year 1, and Year 2).

Alternate Hypothesis 3: Student outcomes on the PARCC assessments in ELA and Math will be significantly different at each implementation level (Not yet started, Initiating, Developing, and Sustaining).

Data analysis procedures. The Chi-Square Test of Homogeneity was used to answer all three research questions and their corresponding hypotheses. According to Laerd Statistics (2016), the Chi-Square Test of Homogeneity is used to determine if a difference exists between binomial proportions of two or more independent groups on a dichotomous dependent variable. The test allows the researcher to determine whether the proportions are statistically significantly different in the different groups. Further, the Chi-Square Test allows for the use of group totals as in the case of students' PARCC performance data which were obtained on the school level for this study. Since the unit of analysis for student performance was on the school level, the One-Way MANOVA and the ANOVA of Repeated Measures could not be used as statistical tools to determine mean differences in student outcomes. As a result, the research questions and the associated hypotheses were revised to measure differences in student outcomes in terms of proportions who met or did not met expectations on the PARCC assessments. In addition to the

Chi-Square Test, post hoc tests using the Z-Test of Two Proportions with a Bonferroni correction were used to determine exactly where differences in proportions between groups lie (Laerd Statistics, 2016). This research study also met the following five assumptions needed to run a Chi-Square Test of Homogeneity thus making it the most suitable tool for analysis:

Assumption 1: One dependent variable that is measured at the dichotomous level, that is, the variable has two categorical independent groups. In this study, PARCC assessment scores in ELA and Math were treated as two separate dependent variables, each having two performance levels of met expectations and not met expectations.

Assumption 2: There is one independent variable that has two or more categorical, independent groups. In this study, the DWIP intervention was the independent variable. In question 1, the levels were before and after DWIP implementation; in question 2, the levels were three different implementation times (Year 0-2015, Year 1-2016, and Year 2-2017); and in research question 3, the levels were the extent of DWIP implementation (Initiating, lowly developing, and developing).

Assumption 3: Independence of observation. In this study, there was no relationship between the observations in each group of the independent variable or between groups.

Assumption 4: Groups were randomly selected. The cluster sampling approach used to select the sample of schools for the study allowed for simple random sampling. Further, survey participants were selected using a simple random sampling approach.

Assumption 5: The minimum sample size was met as determined by a Chi-Square test of Homogeneity for expected values.

The procedures used to conduct the data analysis for this quantitative, quasi experimental study required preparation in the scaling of students' performance data on the PARCC assessments, the calculation of DWIP implementation levels for each of the four schools, and the running of an assumption test for expected values to determine if all cells had an adequate sample size to run a Chi-Square Test of Homogeneity. The researcher combined the scoring levels of the PARCC performance from a 5-point scale to a 2-point scale. Students' performance on the PARCC assessments are scored on the following 5-point scale:1-Not met expectations, 2-Partially Met Expectations, 3-Approached Expectations, 4-Met Expectations, and 5-Exceeded Expectations. The researcher combined scores at performance levels 1-3 and labeled them *not met expectations*, and combined performance levels 4-5 and labeled them *met expectations* as indicated in table 4. This created two levels of the dependent variable that allowed the researcher to run two Chi-Square Test of Homogeneity with a 2x2 cross tabulation to answer research one (performance expectations (met and not met) x PARCC assessment (pretest and posttest), and a 2x3 cross tabulations to answer research question two (performance expectation (met and not met) x implementation year (2015, 2016, 2017)).

Second, raw data from the survey instrument questions (36-46) and the observer scoring of each schools Data Wise Journey were used to calculate the level of implementation of the DWIP intervention for each school. Based on the Data Wise Rubric prepared by the Harvard University Data Wise Project (Boudett et al., 2016), schools are scored on the following 4-point scale: 1-Not yet started, 2-Initiating, 3-Developing, and 4-Sustaining. Table 15 shows the Data Wise implementation level by school and students meeting PARCC assessment expectations. Based on the calculation, School A was at the lowly developing level with a score of 2.81, School B was at the initiating level with a score of 2.48, and School C and School D were at the developing level with scores of 2.92 and 2.93 respectively. This allowed the researcher to run two Chi-Square Test with a 3x3 cross tabulation for each subject to answer research question three. One cross tabulation was performed for students meeting expectations and another for students not meeting expectations as follows: PARCC assessment (2015, 2016, and 2017) x DWIP implementation level (Initiating, lowly developing, and developing).

Third, the researcher ensured that the sample size assumption was met by running a Chi-Square Test of Homogeneity for expected values. It was determined that all expected values met the adequate sample size requirement to run the Chi-Square Test of Homogeneity. Thus, the study met all five assumptions needed to run the Chi-Square Test of Homogeneity.

Fourth, using Statistical Package for Social Sciences (SPSS), independent Chi-Square Tests of Homogeneity were performed to answer all three research questions. Statistical significance on each Chi-Square Test of Homogeneity was determined at the standard alpha, $p < 0.05$. If there was statistically significant difference in the proportion of students meeting and not meeting expectations on the PARCC, a pairwise comparison using a Z-Test of Two Proportions was performed to determine where the differences lie. The researcher used an adjusted alpha which was calculated using the Bonferroni correction method to determine statistical significance on the pairwise comparison. If there were statistically significant differences in proportions, the researcher rejected the null hypothesis.

The following procedure was used to answer each research question and the associated hypotheses:

To answer research question one about the difference in student outcomes on the PARCC assessments in ELA and Math after the implementation of DWIP, two separate 2x2 Chi-Square Tests were performed. To determine the differences in the proportion of students who met and did not meet performance expectations on the ELA PARCC, students' pretest data (ELA PARCC 2015) and posttest data (ELA PARCC 2017) from Table 4 were used to run the Chi-Square Test of Homogeneity. A separate test was performed for Math PARCC. Each test contained one dependent variable with two categorical independent groups (met and not met expectations), and one independent variable with two groups (pre and post DWIP intervention). The researcher set the alpha level at the standard of $p = 0.05$. If the p value was less than 0.05, there was statistical significant difference in the proportion of students who met or did not meet expectations on the PARCC. The researcher then performed a pairwise comparison using a Z-Test of Two Proportions to determine exactly where the differences lie. If there was statistically significant difference in the proportions at the adjusted alpha of $p < 0.0125$, the researcher rejected the null hypothesis as it related to the subject.

To answer research question two to determine if there were any significant difference in student outcomes on the PARCC at three different times during the DWIP implementation, the researcher ran two separate 2x3 Chi-Square Test of Homogeneity using data from Table 4. There were two independent levels of the dependent variable (met and not met expectations) and three levels of the independent variable or implementation times (Year 0: 2015, Year 1: 2016, and Year 2: 2017). The researcher set the alpha level at the standard $p = 0.05$. If the p value was less than 0.05, there was statistical significant difference in the proportion of students who met or did not meet expectations on the PARCC assessments. The researcher then performed a pairwise comparison using a Z-Test of Two Proportions to determine exactly where the differences lie. If there was statistically significant difference in the proportions at the adjusted alpha, $p < 0.00833$, the researcher rejected the null hypothesis as it related to the subject.

To answer research question three, four separate 3x3 Chi-Square Tests of Homogeneity were performed using data from Table 15 to determine if there was any significant difference in the proportions of students who met or did not meet performance expectations on the PARCC assessments in ELA and Math in each of the three years of test administration (2015, 2016, and 2017) and at three different DWIP implementation levels (Initiating, Lowly-developing, and Developing). The first test involved the proportion of students who met expectations on ELA PARCC, the second involved students who did not meet expectations on the ELA PARCC, the third involved students who met expectations on the Math PARCC, and the fourth involved students who did not meet expectations on the Math PARCC. The researcher set the alpha at the standard, $p = 0.05$. If the p value was less than 0.05, there was statistical significant difference in the proportions of students who met or did not meet expectations on the PARCC. The researcher then performed a pairwise comparison using a *Z*-Test of Two Proportions to determine exactly where the differences lie. If the difference in proportions were statistically significant at the adjusted alpha of $p < 0.00556$ at both performance levels for the same PARCC test, the researcher rejected the null hypothesis as it related to subject.

Results

In order to investigate the research questions and the corresponding hypotheses, a series of Chi-Square Test of Homogeneity were used. According to Laerd Statistics (2016), the Chi-Square Test of homogeneity is used to determine if a difference exists between the binomial proportions of two or more independent groups on a dichotomous dependent variable. It helps to determine whether proportions are statistically significantly different in the different groups. If there are statistically significant difference in proportions, post hoc tests are used to determine where the differences between these groups lie. In this study, the Z-Test of Two Proportions was used. An adjusted alpha which was calculated using the Bonferroni correction method was used to determine statistical significance on the post hoc tests.

Assumption tests. For each Chi-Square Test of Homogeneity, the sample size assumption was assessed by running a Chi-Square Test of Homogeneity for the expected values. At least 80% of the expected values needed to be greater than five in order for the sample size assumption to be met (Laerd Statistics, 2016). All the sample size assumption tests met the minimum requirement.

Null Hypothesis One. Null hypothesis one states: There is no significant difference in the number of students who met or did not meet expectations on the PARCC assessments in ELA and Math after the implementation of the DWIP. To investigate this hypothesis, data from Table 4 below were used to run two separate 2x2 Chi-Square Test of Homogeneity to determine if there was any significant difference in the proportion of students who met or did not meet performance expectations from the pretest to the posttest on the ELA and Math PARCC. The results are represented in Tables 5 through 9. If the Chi-Square Test of Homogeneity produced statistically significant difference in proportions between independent groups at $p < 0.05$, a pairwise comparison using a *Z*-Test of Two Proportion was performed to determine where the differences lie. Statistical significance was then determined using an adjusted alpha, $p < 0.0125$, which was calculated using a Bonferroni correction.

Descriptive data.

Table 4

Overall Account of Students Meeting PARCC Assessment Expectations

Test type	Year	Subject	Not Met	Partially Met	Approach	Met	Exceed	Total Not Met	Total Met	Total
Pretest	2014- 15	ELA	126	209	179	161	23	514	184	698
DWIP	2015-16	ELA	231	258	296	269	88	785	357	1142
Posttest	2016-17	ELA	312	257	252	269	58	821	327	1148
Pretest	2014-15	Math	154	243	193	89	1	590	90	680
DWIP	2015-16	Math	238	331	178	15	0	747	15	762
Posttest	2016-17	Math	392	265	193	88	1	850	89	939

Note. DWIP = intervention

Table 4 shows the total number of students from four middle schools in School District X who took the PARCC assessments at three different times from 2015 -2017 (before, during, and two years after the implementation of the DWIP intervention. A total of 3,856 students took the ELA PARC of which 868 students (23 %) met expectations while 2,988 students (77 %) did not. A total of 2,575 students took the Math PARCC, of which 194 students (7.5%) met expectations and 2,381 students (92.5%) did not.

Results for English Language Arts (ELA).

Table 5

Sample Size Assumption for Expected Counts

Performance Expectations	ELA PARCC 2015(Pretest)	ELA PARCC 2017(Posttest)	Total
Met	193.2	317.8	511
Not Met	504.8	830.2	1335
Total	698	1148	1846

A Chi-Square Test of Homogeneity was conducted for the number of students who took the ELA PARCC in 2015 and 2017 and those who met and did not meet expectations. All expected cell counts were greater than five. The minimum expected count was 193.2. Therefore, sample size assumption was met.

Table 6

Chi-Square Test of Homogeneity

Performance Expectations		ELA PARCC 2015(Pretest)	ELA PARCC 2017(Posttest)	Total
Met	Count	184	327	511
	Expected Count	193.2	317.8	511
	% within Test Type	26.4	28.5	27.7
Not Met	Count	514	821	1335
	Expected Count	504.8	830.2	1335
	% within Test Type	73.6	71.5	72.3
Total	Count	698	1148	1846
	Expected Count	698	1148	1846
	% within Test Type	100.0	100.0	100.0

* X^2 (1, N = 1846) = 0.978, p = 0.323, two-tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 1846 students from four middle schools in School District X who were exposed to the DWIP intervention took the ELA PARCC assessments in SY 2014-2015 (pretest) as sixth graders and again in SY 2016-2017 (posttest) as eighth graders. The number of students in each group was unequal with *N* = 698 for the pretest and *N* = 1148 for the posttest. After two years of the intervention, more students 327 or (28.5%) met expectations on the posttest compared to 184 students (26.4%) on the pretest. On the posttest, fewer students, 821or (71.5%) did not meet expectations compared to 514 students (73.6%) on the pretest. There was no statistically significant difference in the proportions of students who increased performance from the pretest to the posttest, X^2 (1, *N* = 1846) = 0.978, p = 0.323. Therefore, the researcher failed to reject the null hypothesis as it relates to ELA.

Results for Math.

Table 7

Sample Size Assumption Test for Expected Counts

Performance Expectations	Math PARCC 2015(Pretest)	Math PARCC 2017 (Posttest)	Total
Met	75.2	103.8	179
Not Met	604.8	835.2	1440
Total	680	939	1619

A Chi-Square Test of Homogeneity was conducted for the number of students who took the Math PARCC in 2015 and 2017 and those who met and did not meet expectations. All expected cell counts were greater than five. The minimum expected count was 75.2. Therefore, sample size assumption was met.

Table 8

Chi-Square Test of Homogeneity

Performance Expectations		Math PARCC 2015(Pretest)	Math PARCC 2017(Posttest)	Total
Met	Count	90	89	179
	Expected Count	75.2	103.8	179
	% within Test Type	13.2	9.50	11.1
Not Met	Count	590	850	1440
	Expected Count	604.8	835.2	1440
	% within Test Type	86.8	90.5	88.9
Total	Count	680	939	1619
	Expected Count	680	939	1619
	% within Test Type	100.0	100.0	100.0

*X^2 (1, N = 1619) = 5.661, p = 0.017, two-tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 1619 students from four middle schools in School District X who were exposed to the DWIP intervention took the Math PARCC assessments in SY 2014-2015 (Pretest) as sixth graders and also in SY 2016-2017 (Posttest) as eighth graders. The number of students in each group was unequal with N = 680 for the pretest and N = 939 for the posttest. After two years of the intervention, fewer students, 89 or 9.5% met expectations on the posttest compared to 90 students (13.2%) on the pretest. More students, 850 or 90.5%, did not meet expectations on the posttest compared to 590 students (86.8%) on the pretest. The decrease in the proportion of students who met expectations and the increase in the proportion of students who did not meet performance expectations from the pretest to posttest were statistically significant, X^2 (1, N = 1619) =5.661, p = 0.017. A pairwise comparison followed to determined where the differences in proportions lie.

Table 9

Pairwise Comparison Z-Test of Two Proportions

Performance Expectations		Math PARCC 2015 (Pretest)	Math PARCC 2017 (Posttest)
Met	Count	90a	89b
	Expected Count	75.2	103.8
	% within Test Type	13.2	9.5
	Adjusted Residual *Z*	2.4	-2.4
	Adjusted *p* value	0.01640	0.01640
Not Met	Count	590a	850b
	Expected Count	604.8	835.2
	% within Test Type	86.8	90.5
	Adjusted Residual *Z*	-2.4	2.4
	Adjusted *p* value	0.01640	0.01640

***** X^2 (1, N = 1619) = 5.661, p= 0.0164, two-tailed. Statistical significance was determined at an adjusted alpha using a Bonferroni correction, $p < 0.0125$.

Post hoc analysis involved a pairwise comparison using a *Z*-Test of Two Proportions. Statistical significance was determined at an adjusted alpha of $p < 0.0125$, which was calculated using a Bonferroni correction. The proportion of students who met expectations on the posttest was not statistically significantly lower and the proportion of students who did not meet expectations was not statistically significantly higher, X^2 (1, N = 1619) =5.661, $p > 0.0125$. Therefore, the researcher failed to reject the null hypothesis as it relates to Math.

Null Hypothesis Two. Null hypothesis two states: There is no significant change in student outcomes on the ELA and Math PARCC assessments at three different times during the DWIP implementation. To investigate this hypothesis, data from Table 4 above was used to run two independent 2x3 Chi-Square Test of Homogeneity to determine if there was any difference in the proportion of students who met or did not meet performance expectations on the PARCC at three different times during the DWIP intervention (Year 0: 2015, Year 1: 2016, and Year 2: 2017). If the Chi-Square Test of Homogeneity produced differences in proportions between independent groups that were statistically significantly different at $p < 0.05$, a pairwise comparison using a *Z*-Test of Multiple Proportions was performed to determine where the differences lie. Statistical significance was then determined using an adjusted alpha, $p < 0.00833$, which was calculated using a Bonferroni correction. The results are represented in tables 10, 11, 12, 13, and 14 respectively.

Results for ELA.

Table 10

Sample Size Assumption Test for Expected Counts

Performance Expectations	ELA PARCC 2015(Pretest)	ELA PARCC 2016 (Intervention)	ELA PARCC 2017 (Posttest)	Total
Met Expectations	202.8	331.7	333.5	868
Not Met Expectations	495.2	810.3	814.5	2120
Total	698	1142	1148	2988

A Chi-Square Test of Homogeneity was conducted between the type of ELA test and the number of students who met and did not meet expectations at three different times of the DWIP intervention. All expected cell counts were greater than five. The minimum expected count was 202.8. Therefore, the sample size assumption was met.

Table 11

Chi-Square Test of Homogeneity

Performance Expectations		ELA PARCC 2015 (Pretest)	ELA PARCC 2016 (Intervention)	ELA PARCC 2017 (Posttest)	Total
Met	Count	184	357	327	868
	Expected Count	202.8	331.7	333.5	868
	% within Test Type	26.4	31.3	28.5	29.0
Not Met	Count	514	785	821	2120
	Expected Count	495.2	810.3	814.5	2120
	% within Test Type	73.6	68.7	71.5	71.0
Total	Count	698	1142	1148	2988
	Expected Count	698	1142	1148	2988
	% within Test Type	100.0	100.0	100.0	100.0

Note. *P* is different than in table 4.9 above.
*****X^2 (2, N = 2988) =5.335, p = 0.069, two-tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 2988 students in four middle schools in School District X who were exposed to the DWIP intervention took the ELA PARCC assessments at three different times during its implementation (2015, 2016, and 2017). The number of students who took the PARCC assessments at each time was unequal with N = 698 in 2015, N = 1142 in 2016, and N = 1148 in 2017. The proportion of students who met performance expectations on the ELA PARCC increased slightly from 184 students (26.4%) in 2015 to 357 students (31.3%) in 2016, and then decreased slightly in 2017 to 327 students (28.5%). The proportion of students who failed to meet performance expectations decreased from 514 students (73.6%) in 2015 to 785 students (68.7%) in 2016, but increased slightly to 821 students (71.5%) in 2017. The differences in proportions of students who met and did not meet performance expectations on the ELA PARCC at the three different times were not statistically significant, X^2 (2, N = 2988) = 5.335, p = 0.069. Therefore, the researcher failed to reject the null hypothesis as it relates to ELA.

Results for Math.

Table 12

Sample Size Assumption Test for Expected Counts

Performance Expectation	Math PARCC 2015 (Pretest)	Math PARCC 2016 (Intervention)	Math PARCC 2017 (Posttest)	Total
Met	55.4	62.1	76.5	194
Not Met	624.6	699.9	862.5	2187
Total	680	762	939	2381

A Chi-Square Test of Homogeneity was conducted between the type of Math test and the number of students who met and did not meet expectations at three different times of DWIP implementation. All expected cell counts were greater than five. The minimum expected count was 55.4. Therefore, the sample size assumption was met.

Table 13

Chi-Square Test of Homogeneity

Performance Expectations		Math PARCC 2015 (Pretest)	Math PARCC 2016 (Intervention)	Math PARCC 2017 (Posttest)	Total
Met	Count	90	15	89	194
	Expected Count	55.4	62.1	76.5	194
	% within Test Type	13.2	2.0	9.5	8.1
Not Met	Count	590	747	850	2187
	Expected Count	624.6	699.9	862.5	2187
	% within Test Type	86.8	98.0	90.5	91.9
Total	Count	680	762	939	2381
	Expected Count	680	762	939	2381
	% within Test Type	100.0	100.0	100.0	100.0

***** X^2 (2, N = 2381) = 64.616, $p < 0.001$, two tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 2381 students in four middle schools in School District X who were exposed to the DWIP intervention took the Math PARCC assessments at three different times during the implementation (2015, 2016, and 2017). The number of students who took the PARCC assessments at each time was unequal with $N = 680$ in 2015, $N = 762$ in 2016, and $N = 939$ in 2017. The proportion of students who met performance expectations on the Math PARCC decreased from 90 students (13.2%) in 2015 to 15 students (2.0%) in 2016, and then increased in 2017 to 89 students (9.5%). The proportion of students who did not meet performance expectations increased from 590 students (86.8%) in 2015 to 747 students (98%) in 2016, and then decreased to 850 students (90.5%) in 2017. The differences in proportions of students who met and did not meet performance expectations on the Math PARCC at the three different times were statistically significant, $X^2 (2, N = 2381) = 64.616, p < 0.001$. A pairwise comparison was performed to determine where the difference is proportions in independent groups lie.

Table 14

Pairwise Comparison Z-Test of Two Proportions

Performance Expectations		Math PARCC 2015 (Pretest)	Math PARCC 2016 (Intervention)	Math PARCC 2017 (Posttest)
Met	Count	90a	15b	89c
	Expected Count	55.4	62.1	76.5
	% within Test Type	13.2	2.0	9.5
	Adjusted Residual *Z*	5.7	-7.6	1.9
	Adjusted *p* value	1.20E-08	2.96E-14	0.05743
Not Met	Count	590a	747b	850c
	Expected Count	624.6	699.9	862.5
	% within Test Type	86.8	98.0	90.5
	Adjusted Residual *Z*	-5.7	7.6	-1.9
	Adjusted *p* value	1.20E-08	2.96E-14	0.05743

*$X^2 (2, N = 2381)$, p = adjusted values, two tailed. Statistical significance was determined at an adjusted alpha using a Bonferroni correction, $p < 0.00833$.

Post hoc analysis involved pairwise comparison using a *Z*-Test of Two Proportions. Statistical significance was then determined using the adjusted alpha, $p < 0.00833$, which was calculated using a Bonferroni correction. The decrease in the proportion of students who met expectations from 2015 to 2016 and the increase in the proportion of students who did not meet performance expectations from 2015 to 2016 were statistically significant, X^2 (2, N = 2381) = 64.616, $p < 0.00833$. Therefore, the researcher rejected the null hypothesis.

Null Hypothesis Three. Null hypothesis three states: There is no difference between the extent of the DWIP implementation and student outcomes on the PARCC (Not yet started, Initiating, Developing, and Sustaining). To investigate this hypothesis, data from Table 15 below which shows the levels of DWIP implementation by school and the proportion of students meeting PARCC expectations on three PARCC administrations were used to run four independent 3x3 Chi-Square Test of Homogeneity. A sample size assumption test for expected outcomes was performed for each Chi-Square Test prior to running the Chi-Square Test of Homogeneity. The four independent Chi-Square Tests were performed as follows: students who met performance expectations on ELA PARCC, students who did not meet performance expectations of ELA PARCC, students who met performance expectations on Math PARCC, and students who did not meet performance expectations on Math PARCC. If there were statistically significant differences between proportions in independent groups at the standard alpha of $p <$ 0.05, a pairwise comparison was conducted to determine where the differences lie. Statistical significance was then determined using an adjusted alpha, $p < 0.00556$, which was calculated using a Bonferroni correction. The results are represented in tables 15 to 26.

Descriptive data.

Table 15

Data Wise Implementation Level by School and Meeting PARCC Expectations

			School A		School B		School C		School D	
School Year	Test Type	Subject	Not Met	Met	Not Met	Met	Not Met	Met	Not Met	Met
2014-15	Pretest	ELA	120	36	139	21	190	37	65	90
2015-16	DWIP	ELA	165	56	205	95	243	87	172	119
2016-17	Posttest	ELA	183	50	217	82	253	63	168	132
2014-15	Pretest	Math	140	17	153	11	209	20	88	42
2015-16	DWIP	Math	154	4	228	7	211	1	154	3
2016-17	Posttest	Math	186	16	236	32	250	5	178	36
DWIP Implementation Level			2.81 (L)		2.48(I)		2.92 (D)		2.93 (D)	

Note. DWIP = intervention; I = Initiating level; L = Lowly Developing level; D = Developing level.

Table 15 shows the levels of Data Wise implementation for four schools in School District X and the number of students who met and did not meet expectations on the PARCC assessments in ELA and Math at three test administrations. Schools C and D were at the same developing level of DWIP implementation. School A was at the initiating level and School B was at the lowly developing level. No school was at the “not yet started” level or at the “sustaining” level.

Results for ELA PARCC who MET expectations.

Table 16

Sample Size Assumption Test for Expected Counts

		DWIP Level Initiating	Lowly Developing	Developing	Total
Met Expectations	ELA PARCC 2015(Pretest)	42	30.1	111.9	184
	ELA PARCC 2016(Intervention)	81.4	58.4	217.2	357
	ELA PARCC 2017(Posttest)	74.6	53.5	198.9	327
Total		198	142	528	868

A Chi-Square Test of Homogeneity was conducted between the levels of DWIP implementation and students who met expectations on the ELA PARCC at three different test administrations. All expected cell counts were greater than five. The minimum expected count was 30.10. Therefore, the sample size assumption was met.

Table 17

Chi-Square Test of Homogeneity

Met Expectations		Initiating	DWIP Level Lowly Developing	Developing	Total
ELA PARCC 2015(Pretest)	Count	21	36	127	184
	Expected Count	42	30.1	111.9	184
	% within DWIP Level	10.6	25.4	24.1	21.2
ELA PARCC 2016(Intervention)	Count	95	56	206	357
	Expected Count	81.4	58.4	217.2	357
	% within DWIP Level	48.0	39.4	39.0	41.1
ELA PARCC 2017(Posttest)	Count	82	50	195	327
	Expected Count	74.6	53.5	198.9	327
	% within DWIP Level	41.4	35.2	36.9	37.7
Total	Count	198	142	528	868
	Expected Count	198	142	528	868
	% within DWIP Level	100.0	100.0	100.0	100.0

*X^2 (4, N = 868) = 17.638, p = 0.001, two-tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 868 students from four middle schools in School District X who were exposed to three different levels of DWIP implementation (initiating, lowly developing, and developing) met performance expectations on ELA PARCC at three different test administrations (2015, 2016, 2017). The number of students in each test administration group was unequal with N = 198 in 2015, N = 142 in 2016, and N = 528 in 2017. At the end of the first year of the DWIP intervention, there was an increase in the number of students who met expectations on the ELA PARCC from 2015 to 2016 at all implementation levels. More students, 95 or (48%), met performance expectations on the ELA PARCC 2016 at the initiating level compared to 56 students (39.4%) at the lowly developing level and 206 students (39 %) at the developing level, a statistically significant difference in proportions X^2 (4, N = 868) = 17.638, p = 0.001. Two years after the implementation of the DWIP intervention, there was an increase in the number of students who met performance expectations on ELA PARCC 2017 at the initiating level, and a slight decrease at the lowly developing and developing levels. More students, 82 or (41%), met performance expectations on ELA PARCC 2017 at the initiating level compared to 50 students (35.2%) at the lowly developing level and 195 students (36.9%) at the developing level, a statistically significant difference in proportions X^2 (4, N = 868) = 17.638, p = 0.001. A pairwise comparison was performed to determine where the differences is proportions in independent groups lie.

Table 18

Pairwise Comparison Using Z-Test of Two Proportions

Met Expectations		Initiating	DWIP Level Lowly Developing	Developing
ELA PARCC 2015(Pretest)	Count	21a	36b	127b
	Expected Count	42	30.1	111.9
	% within DWIP Level	10.6	25.4	24.1
	Adjusted Residual *Z*	-4.2	1.3	2.6
	Adjusted *p* value	0.000027	0.193601	0.009322
ELA PARCC 2016(Intervention)	Count	95a	56a, b	206b
	Expected Count	81.4	58.4	217.2
	% within DWIP Level	48.0	39.4	39.0
	Adjusted Residual *Z*	2.2	-0.4	-1.6
	Adjusted *p* value	0.027807	0.689157	0.1095986
ELA PARCC 2017(Posttest)	Count	82a	50a	195a
	Expected Count	74.6	53.5	198.9
	% within DWIP Level	41.4	35.2	36.9
	Adjusted Residual *Z*	1.2	-0.7	-0.6
	Adjusted *p* value	0.230139	0.483927	0.548506

*X^2 (4, N = 868) = 17.638, p = adjusted values, two-tailed. Statistical significance was determined at an adjusted alpha using a Bonferroni correction, $p < 0.00556$.

Post hoc analysis involved pairwise comparison using a *Z*-Test of Two Proportions. Statistical significance was determined at an adjusted alpha, $p < 0.00556$, which was calculated using a Bonferroni correction. Although the proportion of students who met expectations on the 2015 ELA PARCC was statistically significantly higher at the developing level of the DWIP implementation than at the initiating level, X^2 (4, N = 868) = 17.638, $p < 0.\ 00556$, the difference existed at the start of the DWIP intervention. All other differences in proportions were not statistically significant X^2 (4, N = 868) = 17.638, $p > 0.00556$.

Results for ELA who DID NOT meet expectations.

Table 19

Sample Size Assumption Test for Expected Counts

		DWIP Level Initiating	Lowly Developing	Developing	Total
Not Met Expectations	ELA PARCC 2015(Pretest)	136	113.5	264.5	514
	ELA PARCC 2016(Intervention)	207.7	173.3	404	785
	ELA PARCC 2017(Posttest)	217.3	181.2	422.5	821
Total		561	468	1091	2120

A Chi-Square Test of Homogeneity was conducted between the levels of DWIP implementation and students who did not meet expectations on the ELA PARCC at three different test administrations. All expected cell counts were greater than five. The minimum expected count was 113.5. Therefore, the sample size assumption was met.

Table 20

Chi-Square Test of Homogeneity

Not Met Expectations		DWIP Level Initiating	Lowly Developing	Developing	Total
ELA PARCC 2015(Pretest)	Count	139	120	255	514
	Expected Count	136	113.5	264.5	514
	% within DWIP Level	24.8	25.6	23.4	24.2
ELA PARCC 2016(Intervention)	Count	205	165	415	785
	Expected Count	207.7	173.3	404	785
	% within DWIP Level	36.5	35.3	38.0	37.0
ELA PARCC 2017(Posttest)	Count	217	183	421	821
	Expected Count	217.3	181.2	422.5	821
	% within DWIP Level	38.7	39.1	38.6	38.7
Total	Count	561	468	1091	2120
	Expected Count	561	468	1091	2120
	% within DWIP Level	100.0	100.0	100.0	100.0

*****X^2 (4, N = 2120) = 1.540, p = 0.820, two-tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 2120 students from four middle schools in School District X who were exposed to three different levels of DWIP implementation (initiating, lowly developing, and developing) did not meet performance expectations on ELA PARCC at three different test administration (2015, 2016, 2017). The number of students in each test administration group was unequal with N = 561 in 2015, N =468 in 2016, and N = 1091 in 2017. At the end of the first year of the DWIP intervention, there was a similar increase in the number of students who did meet expectations on the ELA PARCC from 2015 to 2016 at all implementation levels. Fewer students 165 (35.3%) did not met performance expectations on the ELA PARCC in 2016 at the lowly developing level compared to 205 students (36.5%) at the initiating level and 415 students (38 %) at the developing level. The differences in proportions were not statistically significant X^2 (4, N =

2120) = 1.540, p = 0.820. Two years after the implementation of the DWIP intervention, there was a slight increase in the number of students who did not meet performance expectations on ELA PARCC 2017 at all DWIP implementation levels. Fewer students 421 (38.6%) at the higher developing level did not meet expectations compared to 217 students (38.7%) at the initiating level and 183 students (39.1%) at the lowly developing level. The differences in proportions were not statistically significant, X^2 (4, N = 2120) = 1.540, p = 0.820. Since the differences in proportions of student who met performance expectations on ELA PARCC were not statistically significant, p > 0.00556, and the differences in the proportion of student who did not meet expectations on the ELA PARCC were not statistically significant p = 0.820, the researcher failed to reject the null hypothesis as it relates to ELA.

Results for Math who MET expectations.

Table 21

Sample Size Assumption Test for Expected Counts

		DWIP Level		
Met Expectations	Initiating	Lowly Developing	Developing	Total
Math PARCC 2015(Pretest)	23.2	17.2	49.6	90
Math PARCC 2016(Intervention)	3.9	2.9	8.3	15
Math PARCC 2017(Posttest)	22.9	17	49.1	89
Total	50	37	107	194

A Chi-Square Test of homogeneity was conducted between the levels of DWIP implementation and students who met expectations on the Math PARCC at three test administrations. Eighty percent of expected cell counts were greater than five. The minimum expected count was 2.9. Therefore, the sample size assumption was met.

Table 22

Chi-Square Test of Homogeneity

Met Expectations		Initiating	DWIP Level Lowly Developing	Developing	Total
Math PARCC 2015(Pretest)	Count	11	17	62	90
	Expected Count	23.2	17.2	49.6	90
	% within DWIP Level	22.0	45.9	57.9	46.4
Math PARCC 2016(Intervention)	Count	7	4	4	15
	Expected Count	3.9	2.9	8.3	15
	% within DWIP Level	14.0	10.8	3.7	7.7
Math PARCC 2017(Posttest)	Count	32	16	41	89
	Expected Count	22.9	17	49.1	89
	% within DWIP Level	64.0	43.2	38.3	45.9
Total	Count	50	37	107	194
	Expected Count	50	37	107	194
	% within DWIP Level	100.0	100.0	100.0	100.0

*X^2 (4, N = 194) = 19.662, p = 0.001, two-tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 194 students from four middle schools in School District X who were exposed to three different levels of DWIP implementation (initiating, lowly developing, and developing) met performance expectations on Math PARCC at three different test administrations (2015, 2016, 2017). The number of students in each test administration group was unequal with N = 50 in 2015, N =37 in 2016, and N = 107 in 2017. After the first year of the DWIP implementation, there was a decrease in the proportion of students who met expectations from the lower to greater implementation level. More students, seven (14%), met performance expectations on the Math PARCC in 2016 at the initiating level compared to four students (10.8%) at the lowly developing level and four students (3.7%) at the developing level, a statistically significant difference in proportions, X^2 (4, N = 194) = 19.662, p = 0.001. Two years after the implementation of the

DWIP intervention, there was an increase in the number of students who met performance expectations on Math PARCC in 2017 compared to the previous year and at all implementation levels. However, more students, 32 (64%), met performance expectations on Math PARCC 2017 at the initiating level compared to 16 students (43.2%) at the lowly developing level, and 41 students (38.3%) at the developing level, a statistically significant difference in proportions, X^2 (4, N = 194) = 19.662, p = 0.001. A pairwise comparison was performed to determine where the differences is proportions in independent groups lie.

Table 23

Pairwise Comparison Z-Test of Two Proportions

			DWIP Level	
Met Expectations		Initiating	Lowly Developing	Developing
Math PARCC 2015(Pretest)	Count	11a	17a, b	62b
	Expected Count	23.2	17.2	49.6
	% within DWIP Level	22.0	45.9	57.9
	Adjusted Residual *Z*	-4	-0.1	3.6
	Adjusted *p* value	0.000063	0.920344	0.000318
Math PARCC 2016(Intervention)	Count	7a	4a	4a
	Expected Count	3.9	2.9	8.3
	% within DWIP Level	14.0	10.8	3.7
	Adjusted Residual *Z*	1.9	0.8	-2.3
	Adjusted *p* value	0.057433	0.423711	0.021448
Math PARCC 2017(Posttest)	Count	32a	16a, b	41b
	Expected Count	22.9	17	49.1
	% within DWIP Level	64.0	43.2	38.3
	Adjusted Residual *Z*	3	-0.4	-2.3
	Adjusted *p* value	0.002700	0.689157	0.021448

* $X^2(4, N = 194) = 19.662$, p = adjusted values, two-tailed. Statistical significance was determined at an adjusted alpha using a Bonferroni correction, $p < 0.00556$.

Post hoc analysis involved pairwise comparison using Z-Test of Two Proportions. Statistical significance was determined at an adjusted alpha, $p < 0.00556$, which was calculated using a Bonferroni correction. The proportion of students who met performance expectations on Math PARCC in 2017 at the initiating DWIP level was statistically significantly higher than the developing level, X^2 (4, N = 194) = 19.662, $p < 0.00556$. The proportion of students who met expectations on Math PARCC in 2017 at the lowly developing and initiating level was not statistically significantly different, $X^2(4, N = 194) = 19.662$, $p > 0.00556$. The proportion of students who met expectations on Math PARCC 2017 at the lowly developing level and at the developing level was also not statistically significantly different, X^2 (4, N = 194) = 19.662, $p >$

0.00556.

Results for Math who DID NOT meet expectations.

Table 24

Sample Size Assumption for Expected Counts

		DWIP Level		
Not Met Expectations	Initiating	Lowly Developing	Developing	Total
Math PARCC 2015(Pretest)	183.9	143.1	351.1	678
Math PARCC 2016(Intervention)	202.6	157.6	386.8	747
Math PARCC 2017(Posttest)	230.5	179.3	440.1	850
Total	617	480	1178	2275

A Chi-Square Test of Homogeneity was conducted between the levels of DWIP implementation and students who did not meet expectations on the Math PARCC at three test administrations. All expected cell counts were greater than five. The minimum expected count was 143.1. Therefore, the sample size assumption was met.

Table 25

Chi-Square Test of Homogeneity

Not Met Expectations		Initiating	DWIP Lowly Developing	Developing	Total
Math PARCC 2015(Pretest)	Count	153	140	385	678
	Expected Count	183.9	143.1	351.1	678
	% within DWIP Level	24.8	29.2	32.7	29.8
Math PARCC 2016(Intervention)	Count	228	154	365	747
	Expected Count	202.6	157.6	386.8	747
	% within DWIP Level	37.0	32.1	31.0	32.8
Math PARCC 2017(Posttest)	Count	236	186	428	850
	Expected Count	230.5	179.3	440.1	850
	% within DWIP Level	38.2	38.8	36.3	37.4
Total	Count	617	480	1178	2275
	Expected Count	617	480	1178	2275
	% within DWIP Level	100.0	100.0	100.0	100.0

* $X^2(4, N = 2275) = 13.739$, $p = 0.008$, two-tailed. Statistical significance was determined at the standard alpha, $p < 0.05$.

A total of 2275 students from four middle schools in School District X who were exposed to three different levels of DWIP implementation (initiating, lowly developing, and developing) did not met performance expectations on Math PARCC at three different test administrations (2015, 2016, 2017). The number of students in each test administration group was unequal with *N* = 617 in 2015, N =480 in 2016, and *N* = 1178 in 2017. After the first year of implementation of the DWIP intervention, fewer students, 365 (31%), did not met performance expectations on the Math PARCC in 2016 at the developing level compared to 154 students (32.1%) at the lowly developing level and 228 students (37%) at the initiating level, a statistically significant difference in proportions, X^2 (4, $N = 2275$) = 13.739, $p = 0.008$. Two years after the implementation of the DWIP intervention, more students did not meet performance expectations on the Math PARCC in 2017 at each implementation level than each of the prior two years. However, fewer students, 428 (36.3%), did not meet expectations at the higher implementation level, the developing level, compared to 186 students (38.8%) at the lowly developing level, and 236 students (38.2%) at the initiating level, a statistically significant difference in proportions, $X^2(4, N = 2275) = 13.739$, $p = 0.008$. A pairwise comparison was performed to determine where the differences is proportions in independent groups lie.

Table 26

Pairwise Comparison Z-Test of Two Proportions

Not Met Expectations		Initiating	DWIP Level Lowly Developing	Developing
Math PARCC 2015(Pretest)	Count	153a	140a, b	385b
	Expected Count	183.9	143.1	351.1
	% within DWIP Level	24.8	29.2	32.7
	Adjusted Residual *Z*	-3.2	-0.3	3.1
	Adjusted *p* value	0.001374	0.764177	0.001935
Math PARCC 2016(Intervention)	Count	228a	154a, b	365b
	Expected Count	202.6	157.6	386.8
	% within DWIP Level	37.0	32.1	31.0
	Adjusted Residual *Z*	2.6	-0.4	-1.9
	Adjusted *p* value	0.009322	0.689157	0.057433
Math PARCC 2017(Posttest)	Count	236a	186a	428a
	Expected Count	230.5	179.3	440.1
	% within DWIP Level	38.2	38.8	36.3
	Adjusted Residual *Z*	0.5	0.7	-1.1
	Adjusted *p* value	0.617075	0.483927	0.271332

*****$X^2(4, N = 2275) = 13.739$, *p*= adjusted values, two-tailed. Statistical significance was determined at an adjusted alpha using a Bonferroni correction, $p < 0.00556$.

Post hoc analysis involved pairwise comparison using a Z-Test of Two Proportions. Statistical significance was determined at an adjusted alpha, $p < 0.00556$, which was calculated using a Bonferroni correction. The proportion of students who did not met performance expectations on the Math PARCC in 2016 at the higher implementation level, the developing level, was statistically significantly lower than at the lowest implementation level, the initiating level, X^2 (4, $N = 2275) = 13.739$, $p < 0.00556$. The proportion of students between the lowly developing and initiating level who did not meet expectations on the Math PARCC 2016 was not statistically significantly different, X^2 (4, $N = 2275) = 13.739$, $p > 0.00556$. The difference in the proportion of students who did not meet performance expectations on the Math PARCC in 2016 at the lowly developing and the developing levels was also not statistically significantly, X^2 (4, N = 2275) = 13.739, $p > 0.00556$. Additionally, the differences in the proportion of students who did not meet performance expectations on the Math PARCC in 2017 at all three implementation levels were not statistically significant. Since the proportion of students who met performance expectations on Math PARCC in 2017 at the initiating implementation level of the DWIP intervention was statistically significantly higher than the developing level, X^2 (4, N = 194) = 19.662, $p < 0.00556$, and the proportion of students who did not met performance expectations on Math PARCC 2016 at developing implementation level statistically significantly lower than at the initiating, X^2(4, N = 2275) = 13.739, $p < 0.00556$, the researcher rejected the null hypothesis as it relates to Math.

Summary

The sample population of this study consisted of 2,631 middle school students from four middle schools in School District X who were in sixth grade in 2015, seventh grade in 2016, and eighth grade in 2017. All students took the PARCC assessments in ELA and Math at three different times during the DWIP intervention (Year 0-2015, Year 1-2016, and Year 2-2017). School District X mandated the start of the implementation of the DWIP intervention in school year 2015-2016. The study attempted to answer three research questions about the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement. The Chi-Square Tests of Homogeneity followed by the Z-Test of Two Proportions were used to answer the three research questions about differences in student outcomes based on varying levels of the DWIP intervention. All tests were performed using Standard Package for Social Sciences, SPSS.

In answering research question one, the researcher failed to reject the null hypothesis that there was no significant difference in the proportion of students who met or did not meet expectations on the PARCC assessments in both ELA and Math from 2015 (pretest) to 2017(posttest). The differences in proportions were not statistically significant. In answering research question two, the researcher failed to reject the null hypothesis as it relates to ELA, that there was no significant difference in student outcomes on ELA PARCC at three different times of the DWIP implementation (2015, 2016, and 2017). However, the researcher rejected the null hypothesis as it relates to student outcomes on the Math PARCC because differences in proportions were statistically significant. In answering research question three, the researcher failed to reject the null hypothesis as it relates to ELA that there was no significant difference in student outcomes on the ELA PARCC based on the extent of DWIP implementation (not yet started, initiating, developing, and sustaining) because the results were no statistically significant. However, the researcher rejected the null hypothesis as it relates to Math because of statistically significant results.

The following chapter, chapter 5, will contain an introduction, a summary of the results, discussion of the results, discussion of the results in relation to the literature, limitations, implications of the results for practice, recommendations for further research, and a conclusion.

Chapter 5: Discussion and Conclusion

Introduction

The purpose of this quantitative, quasi-experimental study was to determine the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement in School District X. The study is intended to add to the body of knowledge around continuous school improvement through a combination of teacher collaboration in the form of PLCs, DDDM, and collaborative leadership within an explorative and continuous framework. The study also serves to inform School District X about the extent of the implementation of DWIP intervention, and its ability to attain district goals of increasing academic excellence by improving student achievement on the state standardized test, the PARCC.

A clustering sampling approach was used to select six middle schools in School District X for the study. However, only four of the six middle schools participated in the study. The sample included 58 teachers who participated in an anonymous click consent survey about their perception of the collaborative data-inquiry practices and the implementation of the DWIP at their schools. The survey was created and distributed using *Qualtrics* Software. The study also used the PARCC assessment data for 2,631 students in ELA and Math as a measure of student achievement.

The first two chapters of this research presented the study from the perspective of teacher collaboration through PLCs, DDDM, reflective practice, job embedded professional training, and supportive leadership within a cyclical structure that is intended to drive continuous improvement in schools. Chapter three and four presented the research plan and results, respectively. Chapter five, presents a review and discussion of the results and provides direction for future use. This is done by summarizing the results of the study by research question followed by a discussion of the results and how they relate to the literature findings. Limitations of the study are presented, followed by a review of the implications of the results. Finally, recommendations for further research are provided, followed by a conclusion of the research study.

Summary of Results

In order to examine the impact of a collaborative data-inquiry culture, as promoted by the DWIP, on student achievement, the researcher attempted to answer three research questions. A quantitative methodology with a quasi-experimental one-sample pretest-posttest design was used for this study. Teacher perception data were collected using the Collaborative Data-Inquiry Survey which was developed by combining two previously used instruments; the School Culture Survey (Gruenert and Valentine, 2005) and the Data Wise Rubric (Boudett et al., 2016). Permission to use both instruments without modification was obtained. PARCC assessment data on the school level were collected and used as the measure of student achievement.

Data analysis included the use of descriptive and inferential statistics. Descriptive statistics was used to describe the sample population and its characteristics and to arrange teacher perception and student achievement data. Inferential statistics was used to answer each research question. The Chi-Squared test of Homogeneity was used to answer all three research questions. Statistical significance was determined at the standard alpha, $p < 0.05$. This was followed by a pairwise comparison using a Z-Test of Two Proportions. Statistical significance was determined at adjusted alpha values which were determined using a Bonferroni correction. To answer research question one to determine whether there was a significant difference in the proportion of students who met and did not meet performance expectations on the PARCC assessment in ELA and Math after the implementation of the DWIP, two separate 2x2 Chi-Squared Test of

Homogeneity followed by a pairwise comparison were performed. To answer research question two to determine whether there was a significant difference in student outcomes on the PARCC assessments in ELA and Math at three different times during the DWIP implementation, two separate 2x3 Chi-Square Test of Homogeneity were performed followed by a pairwise comparison. In order to answer research question three to determine if there was a significant difference in student outcomes on the PARCC assessment in ELA and Math based on the extent of DWIP implementation, four separate 3x3 Chi-Squared Test of Homogeneity followed by a pairwise comparison were performed.

Research Question One.

Research Question 1: What is the difference in the number of students who met and did not meet performance expectations on the PARCC assessment in ELA and Math after the implementation of the DWIP?

Null Hypothesis 1: There is no significant difference in the number of students who met or did not meet expectations on the PARCC assessment in ELA and Math after the implementation of DWIP.

Alternate Hypothesis 1: There is a significant difference in the number of students who met or did not meet expectations on the PARCC assessment in ELA and Math after the implementation of the DWIP.

After analyzing the data, the researcher failed to reject the hypothesis for both PARCC ELA and Math. There was no statistically significant difference in the proportion of students who met and did not meet performance expectation on the ELA PARCC after the implementation of the DWIP, X^2 (1, N = 1846) = 0.978, p = 0.323 or p > 0.05. There was also no statistically significant difference in the proportion of students who met and did not meet performance expectations on the Math PARCC after the implementation of the DWIP, X^2 (1, N = 1619) = 5.661, $p > 0.0125$.

Research Question Two.

Research Question 2: What is the difference in student outcomes on the PARCC assessment in ELA and Math at three different times of the DWIP implementation (Year 0, Year 1, and Year 2)?

Null Hypothesis 2: There is no significant difference in student outcomes on the PARCC assessment in ELA and Math at three different times during the DWIP implementation (Year 0, Year 1, and Year 2).

Alternate Hypothesis 2: Student outcomes on the PARCC assessment in ELA and Math will be significantly different at three different times during the DWIP implementation (Year 0, Year 1, and Year 2).

After analyzing the data, the researcher failed to reject the hypothesis as it relates to ELA but rejected the hypotheses and it relates to Math. The Chi-Square Test of Homogeneity revealed that there was no statistically significant difference in the proportion of students who met and did not meet performance expectations on the ELA PARCC at three different times during the DWIP implementation, X^2 (2, N = 2988) = 5.335, $p = 0.069$ or $p > 0.05$. However, the Chi-Square Test of Homogeneity found that there was a statistically significant difference in the proportion of students who met and did not meet performance expectations on the Math PARCC at the three different times during the DWIP implementation, X^2 (2, N = 2381), $p < 0.001$. Further, a pairwise comparison found that the decrease in the proportion of students who met performance expectations from 2015 to 2016 and the increase in the proportion of students who did not meet performance expectations from 2015 to 2016 to be statistically significant, X^2 (2, N = 2381) = 64.616, $p < 0.00833$.

Research Question three.

Research Question 3: What is the difference in student outcomes on the PARCC Assessments in ELA and Math based on the extent of the DWIP implementation (Not yet started, Initiating, Developing, or Sustaining)?

Null Hypothesis 3: There is no difference in student outcomes on the PARCC Assessments in ELA and Math based on the extent of Data Wise implementation (Not yet started, Initiating, Developing, and Sustaining).

Alternate Hypothesis 3: Student outcomes on the PARCC Assessments in ELA and Math will be significantly different at each implementation level (Not yet started, Initiating, Developing, and Sustaining).

After the analysis of the data, the researcher failed to reject the hypothesis as it relates to ELA PARCC but rejected the hypothesis as it relates to Math PARCC. A Chi-Square Test of Homogeneity followed by a pairwise comparison for ELA PARCC found no statistically significant difference in the proportion of students who met performance expectations, X^2(4, N =868) = $p > 0.00556$, and those who did not meet expectations, X^2 (4, N = 2120) = 1.540, p = 0.820 at different DWIP implementation levels. However, a Chi-Square Test of Homogeneity followed by a pairwise comparison revealed that the proportion of students who met performance expectations on Math PARCC in 2017 at the lower DWIP implementation level, the initiating level, was statistically significantly higher than the higher DWIP implementation level, the developing level, X^2(4, N = 194) = 19.662, $p < 0.00556$. The researcher also found that the proportion of students who did not met performance expectations on the Math PARCC in 2016 at the developing DWIP level was statistically significantly lower than at the initiating level, $X^2(4, N = 2275) = 13.739, p < 0.00556$. This shows that the greater DWIP implementation level had a greater impact on student performance on the Math PARCC in 2016 but not in 2017.

Discussion of Results

Descriptive analyses of the demographic data identified a diversity in the sample population of the students in the study that is not normally reflected in most minority or urban school districts. A majority of the students in the sample of four middle schools in School District X came from mainly two cultural backgrounds; African American (59%) and Hispanic (31%). Students also came from various socio-economic backgrounds. However, the demographic data failed to capture the cultural diversity of the 59% of the student population who are considered African American. There is large diversity in this population represented by migrant students and families from numerous African countries and Caribbean Islands. The demographic data also showed a diversity in the socio-economic backgrounds of students in the schools. While the county in which School District X is located is considered the fifth most affluent minority county in the United States (Brown, 2015), about half of the students from the four schools rely on free and reduced meals. The study was unable to capture the turnover or dropout rate of students at the four middle schools because the school district only maintains dropout data at the high school level. The descriptive data also captured the qualification of the teaching staff of the four schools with 85% of them being highly qualified. However, the study was unable to determine the teacher turnover rate during the three-year period of the study. The biggest reason for the inability to determine the dropout rate of students and the turnover rate of teachers is because data for the study was taken at the school level and not at the individual student and teacher level. This made it impossible to track teachers who were there for the whole three-year period of DWIP implementation, and match students PARCC scores with those teachers who engage in collaborative data-inquiry practice.

Data were analyzed to answer research question one to determine if there was any significant difference in the proportion of students who met and did not meet performance expectations on the PARCC assessments in ELA and Math after the implementation of the DWIP. Two separate Chi-Square Test of Homogeneity was used to answer the question. After analyzing the data, the researcher failed to reject null hypothesis one (Ho1) as it relates to both ELA and Math. There was no statistically significant difference in the proportion of students who met or did not meet performance expectation on the ELA PARCC, $X^2(1, N = 1846) = 0.978$, $p = 0.323$ and Math PARCC, $X^2(1, N = 1619) = 5.661$, $p > 0.0125$, after the implementation of the DWIP. The results indicate that the DWIP intervention had no overall significant impact on student achievement on the PARCC assessments in both the ELA and Math for the three-year period of the DWIP intervention.

There are three possible reasons for retaining the null hypothesis. They are because of the sample size of schools, the unit of analysis, and the inferential tool. The research plan proposed the use of a sample of six middle schools in School District X, 72 teachers, and 1,500 students. The proposed sample size was higher than the calculated size using G-Powered Analysis for a medium sample effect and using the MANOVA test. Although only four of the six middle schools and 58 teachers participated in the study, and the achievement data for 2, 631students were used, the minimum sample size requirement for the proposed inferential tool was met. However, since student achievement data were captured on the school level, the MANOVA test could not be used. The MANOVA test would have allowed the researcher to compare the effect of the independent variable (DWIP) against a linear composite of the two dependent variables

(PARCC ELA and Math) first, before comparing them independently. This would reduce the possibility of committing type one errors (Rockinson-Szapkiw, 2013). Instead, the Chi-Square Test of Homogeneity was used since it was the most appropriate tool to measure differences in student outcomes using student achievement data on the school level. Two separate tests were performed, one for each subject test, thus the independent variable (DWIP) was examined separately for its effect on each dependent variable. A pairwise comparison using a *Z*-Test of Two Proportions with a Bonferroni correction was used. The adjusted alpha value of $p = 0.0125$ was used instead of the standard alpha of $p = 0.05$ to determine statistical significance and thereby limit the potential of a type one error. Finally, an assumption test for an adequate sample size was performed before performing the Chi-Square Test of Homogeneity. While the sample size assumption for teachers and students was met, the sample size for the number of schools of four consistently fell below the minimum requirement of five. The small sample of schools, the unit of analysis of the school level, and the use of multiple Chi-Square Test instead of the One-Way MANOVA might have contributed to the statistically insignificant difference in proportions.

To answer research question two to determine if there was any significant difference in student outcomes on the PARCC assessments in ELA and Math at three different times during the DWIP implementation (Year 0, Year 1, and Year 2), two separate 2x3 Chi-Square Test of Homogeneity were performed instead of the One-Way MANOVA or the Repeated Measures ANOVA. The researcher failed to reject the null hypothesis as it relates to ELA PARCC because there was no statistically significant difference in the proportion of students who met and did not meet performance expectations on the test in year 0, year 1, and year 2, X^2 $(2, N = 2988) = 5.335$, $p = 0.069$. This means that the DWIP intervention had no significant impact on student performance on the ELA PARCC from year zero to year one and again from year one to year two. Again, the possible explanation for retaining null hypothesis two, could be attributed to the same reasons for retaining null hypothesis one. Namely, a small school sample of four, using a unit of analysis on the school level rather than on the individual student level, and using a Chi-square Test of Homogeneity for analysis instead of the MANOVA. While the Chi-square Test of Homogeneity was the most suitable test given the unit of analysis of student achievement data on the school level, it did not allow the researcher to compare the effect of the independent variable with a linear composite of the two dependent variables. Instead multiple Chi-square Tests of Homogeneity were performed thus increasing the chances of committing a type one error.

However, the researcher rejected the null hypothesis as it relates to Math PARCC because there was a statistically significant decrease in the proportion of students who met expectations from year zero to year one and a statistically significant increase in the proportion of students who did not meet performance expectations for the same period, $X^2\ (2, N = 2381) = 64.616$, $p < 0.00833$. There was no significant change in the number of students meeting and not meeting performance expectations in Math from year one to year two. This means that the DWIP intervention had a negative impact on student achievement in Math in the first year of its implementation and no impact in the second year.

There are three plausible explanations for the negative impact of the DWIP intervention on students' performance on the Math PARCC. One is that the PARCC assessment is a more rigorous test which replaced the state designed standardized assessment in 2015. Students are still adjusting to the rigor and format of the new assessment which requires more fluency, written explanations, strategy, and quantitative arguments. Second, the state in which School District X is located adopted the Common Core State Standards in 2012-2013 which required the school district to revamp the reading and math curriculum. The new curriculum requires a shift in instruction to which many teachers are still in the transition phase. Third, the standardized math score in the school district has been historically lower than that of reading and science. These three reasons justify the need to use a linear composite of the dependent subject variables in the form of a MANOVA test for future research on an intervention containing two or more variables and which is not content specific.

Finally, to answer research question three to determine if there was a significant difference in student outcomes on the PARCC assessment in ELA and Math based on the extent of the DWIP implementation (Not yet started, Initiating, Developing, or Sustaining), the researcher performed four separate Chi-square Test of Homogeneity using a 3x3 cross tabulation followed by a Z-Test of Two Proportions with a Bonferroni correction. The researcher failed to reject hypothesis three (Ho3) as it relates to ELA but rejected the hypothesis as it relates to Math. The difference in the proportion of students who met performance expectations on ELA PARCC at three different implementation levels of the DWIP intervention was not statistically significant based on a pairwise comparison, X^2 (4, N = 868) = 17.638, $p > 0.00556$. Further, the difference in the proportion of student who did not meet expectations on the ELA PARCC were not statistically significant, X^2 (4, N = 2120) = 1.540, $p = 0.820$.

The results indicate that the extent to which the DWIP intervention was implemented had no significant impact on student achievement. School A was at the lowly developing implementation level, School B was at the initiating implementation level, and School C and D were at the developing implementation level. There were no schools at the not yet started and sustaining implementation levels. Although the use of the Chi-Square Test instead of the One-Way MANOVA or the Repeated Measures ANOVA, the unit of analysis of student achievement data on the school level, and the small sample of four schools were plausible explanations why the null hypothesis was retained as in question one and two. The small sample of schools was probably the most plausible reason for this result. The small sample of four schools only allowed the researcher to examine one or two schools at each DWIP implementation level and therefore did not lend to the possibility that the implementation data for each school might have been an outlier.

The researcher rejected hypothesis three (Ho3) as it relates to Math that there was no difference in student outcomes on the PARCC assessment based on the extent of DWIP implementation. The proportion of students who met performance expectations on Math PARCC in 2017 at the initiating implementation level was statistically significantly higher than the developing implementation level, $X^2(4, N = 194) = 19.662$, $p < 0.00556$, and the proportion of students who did not met performance expectations on Math PARCC in 2016 at the initiating level was statistically significantly higher than the developing level, $X^2 (4, N = 2275) = 13.739$, $p < 0.00556$. Although there was statistically significant difference in the proportion of students meeting performance expectations at the initiating and developing levels on the 2017 PARCC, more students met performance expectations at the lower implementation level. On the 2016 PARCC, fewer students failed to meet performance expectations at the higher implementation level. The desired outcome would be to increase the number of students meeting performance expectation while decreasing the number of students not meeting performance at the higher implementation level in each of the implementations years.

The most plausible explanation for this result could be the lack of uniformity in the implementation of the DWIP intervention in addition to the explanations provided in questions one and two. Three years after the implementation of the DWIP, there remained large disparity in the levels of DWIP implementation on the school level. None of the four schools were at the sustaining level, which is the level at which the DWIP is fully implemented. One school was still at the initiating level almost four years into the systemic adoption of the DWIP while three schools were mostly at the lowly developing level. Second, as in the case of research question two, the small sample size of the schools probably most adversely impacted the outcome because there were only 1-2 schools at each implementation level. This did not allow the researcher to make adjustments for possible outlier data.

Discussion of Results in Relation to the Literature

The study found that the DWIP intervention, which is purported to promote a collaborative data-inquiry culture (Boudett et al., 2014), had no significant impact on the overall student outcomes on the PARCC assessment in both ELA and Math. This result is not supported by research on the topic. Although there were no previously conducted studies that measured the actual impact of this intervention on student achievement, there were studies that linked the factors of the intervention, namely DDDM, teacher collaboration in the form of PLCs, reflective practice, professional development, and collaborative leadership, to student achievement. Gruenert and Valentine (2005) found a positive correlation between a collaborative school culture and student achievement. Collaborative culture was traced through six factors which included collaborative leadership, teacher collaboration, collaborative partnership, professional development, unity of purpose, and collegial support. In addition, Dougherty (2015) linked DDDM to improved teaching and student learning which is in contrast to the findings of the study. Further, Ezzani (2015) found that in two urban school districts where teachers and teacher leaders used interventions that focused on DDDM, systemic and comprehensive professional learning, and distributed leadership, saw consistent gains in student achievement over a 3-5 year period. However, Simms and Penny (2014) found that interventions where DDDM and PLCs are narrowly focused, and lack training, time, support, and interim monitoring mechanisms, failed to have any impact on student achievement. This suggests that the manner in which training, support, and monitoring occurred during the implementation of the DWIP intervention is an area for further exploration.

In response to research question two, the finding that there was no statistically significant difference in student outcomes on the ELA PARCC at three different times of the DWIP implementation is not support by a majority of the research. Research on the factors that make up the hybrid collaborative data-inquiry culture promoted by the DWIP intervention show that there is a positive link to student achievement. Dougherty (2015) and Simms and Penny (2014) found a strong link between DDDM, PLCs, and student achievement. Ezzani (2015) found that schools that implement DDDM, PLCs, with systemic professional development and which is directed by collaborative leadership see improvements in student achievement within 3-5 years. However, Parker (2017) in a research study to measure the impact of a response to intervention on math performance found that interventions that work have the greatest impact in the first two years and the impact levels off in the third year. To the contrary, the DWIP intervention had a negative impact on student outcomes on the Math PARCC from year zero to year one and no significant impact in the third year. The negative impact of the DWIP intervention on student achievement could be further linked to a research conducted by Turtle (2015). Turtle (2015) found that there was a decline in student performance in middle schools who once had a vibrant PLC model. Among the reasons provided why PLCs were no longer as effective were teacher attrition, poor training on the use of PLCs, poorly skilled new staff, and other economic factors.

The level of DWIP implementation had no significant impact on the overall student outcomes on the ELA PARCC over a three-year period while there was significant impact on the Math PARCC but within the same year. There was a negative impact on student outcomes on the Math PARCC at each implementation level from year zero to year two. Although there was significant difference in student outcomes on the Math PARCC, the results were mixed with a more positive impact on student outcomes at the higher DWIP implementation level in 2016 but a more positive impact on student outcomes at the lower implementation level in 2017. A research study conducted by Algozzine et al. (2011) on the evaluation of the implementation of the DWIP in Charlotte-Mecklenburg Schools found challenges in the implementation of the DWIP steps that involved data use and teacher collaboration that contribute most to improving instruction and student learning. Algozzine et al. (2011) found that after three years of implementation, a majority of the schools were at high implementation levels for steps one and two but struggled with steps three through eight that involved creating data overviews, using data to identify problems with student learning and teacher practice, creating action plans, and monitoring and assessing action plans to make adjustments. A similar pattern was observed in the four schools in this study sample. In addition, none of the schools were at an overall sustaining level, which is the highest implementation level. The Charlotte-Mecklenburg Schools study did not look at the impact of the DWIP implementation on student achievement.

Limitations

There were at least three areas that could have enhanced the outcomes of the study. The unit of analysis of student achievement data were the greatest limitation of the study. Student achievement data were collected on the school level by going to the state department of education website and collecting PARCC scores for students at four middle schools in School

District X. The scores were captured for all students who took the PARCC assessment as six graders in 2015, seven graders in 2016, and eighth graders in 2017. The data provided the total number of students at each school who scored at each of the five performance levels on the PARCC assessment. The data were limiting because it did not match students with a particular teacher but with a school. Consequently, the researcher was not able to use the preferred inferential tool, the MANOVA for analyzing the data.

The MANOVA test was the preferred instrument for data analysis because the study contained two categorical and continuous dependent variables and one independent variable. The MANOVA would have allowed the researcher to study the effect of the independent variable on a linear composite of the dependent variables, thereby limiting the potential for committing type one errors. However, because student achievement data were collected on the school level, multiple Chi-Squared Test of Homogeneity were performed to examine the effect on the independent variable on each of the two dependent variables, and separately. While the Chi-Square Tests were followed by pairwise comparisons with a Bonferroni correction to account for the numerous tests performed, the multitude of tests and the examination of the dependent variables separately increased the potential for committing a type one error.

Second, the research questions were modified because of the change in the inferential tool. The change involved measuring student outcomes on the PARCC assessments based on the proportion of students who met and did not meet performance expectations rather than looking for differences in mean scores after the DWIP implementation. Measuring outcomes using the differences in proportion of students who met and did not meet performance expectations on the PARCC did not allow the researcher to capture differences in student outcomes between performance levels. There are five performance levels on the PARCC: 1-not met expectation, 2-

partially met expectations, 3-approached expectations, 4-met expectations, and 5-exceeded expectations. The researcher grouped performance levels 1-3 and called them not met expectations, and grouped performance levels 4-5 and labeled them met expectations. As a result, this measurement scale did not allow the researcher to track differences between performance levels and thus possible incremental improvements in student outcomes on the PARCC.

A third limitation was the small sample size of schools used. The researcher proposed using six middle schools for the study. However, permission was only received to conduct the study in four schools. While the sample of 58 teachers and 2,631 students met the minimum requirement for a medium sample size effect, the small number of four schools was limiting when it came to answering research question three. Research question three sought to measure the differences in student outcomes on the PARCC assessments in ELA and Math based on the extent of the implementation of the Data Wise Intervention. One school was determined to be at the initiating level, one at the lowly developing level, and two schools were at the developing level. Having only one or two schools at each implementation level did not allow for the possibility of having any outliers. Further, the sample of four schools fell slightly below the minimum level of five required to meet the adequate sample size for a Chi-Square Test of Homogeneity.

Implications of the Results for Practice, Policy, and Theory

The results of this study suggest that the DWIP intervention as implemented in four middle schools in School District X is not effective in increasing student achievement. This has implications for practice, policy, and theory which can be extended to the whole school district as well as other school districts in which teachers and teacher leaders are embracing the same approach to improving student achievement. The approach includes DDDM, teacher collaboration in the form of PLCs, continuous and comprehensive professional development, and collaborative leadership which are implemented within a coherent and cyclical framework.

Implications for Practice. The data suggest that the DWIP intervention as implemented in four middle schools in School District X is not effective in increasing student achievement in ELA and Math. Three years into the implementation of this problem- solving approach, there was no significant change in the proportion of students who met performance expectations on the ELA PARCC, while there was a significant decrease in the proportion of students who met expectations on the Math PARCC. All four schools were at an overall lower implementation level, that is, the initiating or developing level. Further, a review of the proficiency with which each school implemented each of the eight steps of the DWIP intervention showed that all schools scored at a high implementation level for steps one, two, and three but scored on a low implementation level for steps four through eight. The latter steps involve the practices of using data to identify learner-centered problems and problems in teacher practice, designing action plans to address the identified problems, and designing and implementing monitoring and assessment mechanisms.

Therefore, the areas of the intervention that required DDDM, PLCs, ongoing professional learning, and collaborative leadership that are most attributed to increasing student achievement (Dougherty, 2015; Ezzani: 2015; Gero, 2015; Hallam et al., 2015) were not well implemented. In order to realize the true potential of the intervention, all eight steps must be properly implemented. This study underscores the need for teachers and teacher leaders to engage in more training around data-use and teacher collaboration in order to properly implement all the steps of the DWIP intervention. Further, school and executive leadership need to reassess their approach for monitoring and assessing the implementation of the intervention. The researcher strongly encourages that more time and resources be allocated for ongoing professional development that is grounded in data-use, teacher collaboration in the form of PLCs, support and structure for ongoing monitoring to ensure that interventions that are grounded in this practice succeed (Simms & Penny, 2014).

Implications for Policy. In response to *Every Student Succeed Act*, the state in which School District X is located recently adopted a new accountability measuring system for school performance. The measure consists of four critical areas which include academic achievement, academic progress, English language proficiency, and school quality and student success. Academic achievement accounts for more than twenty-five percent of the current measure. The results of this study provide useful insights for policy makers in both the school district and the state level for addressing the areas that count towards student academic achievement. Like this study, student achievement on the state accountability measure is based solely on PARCC scores in ELA and Math. Further, the measure of student achievement is based on the same met and not met expectations measures used in the study. The results of the study show that students in School District X are not meeting performance expectations in ELA and Math at the levels needed to score at an above average level on the state accountability measure. Therefore, policy makers can use the results of this study to guide budgetary decisions around this intervention as it relates to systemic professional development around data-use, teacher collaboration, and leadership training that promote the type of distributed or collaborative leadership needed to drive continuous school improvement (Ezzani, 2015).

Implications for Theory. Kaizen's theory of continuous improvement (Shang, 2017) and the related characteristics are represented in the DWIP intervention that was central to this study.

The DWIP is a cyclical problem solving approach which is intended to promote a collaborative data-inquiry culture (Boudett et al., 2014). Like Kaizen's theory, the DWIP focuses on constantly assessing practice and making adjustments to improve it. It focuses on how people think, act, and adjust. The DWIP embraces four popular factors used in school improvement efforts, namely DDDM, teacher collaboration in the form of PLCs, continuous and comprehensive professional learning, and collaborative leadership. While the researcher intended to examine the hybrid collaborative data-inquiry factor within a cyclical structure, the lack of implementation of the latter steps of the intervention did not allow the researcher to capture the cyclical nature of the intervention. The researcher thus impresses on the need to implement all the eight steps of the DWIP adequately and continuously in order to measure the true impact of the intervention on student achievement. This will allow all staff members in the school to continuously assess the way they think and act and allow them to make adjustments to refine their practice through the lens of student achievement.

Recommendations for Further Research

One area for improvement in future research is in the collection of student data. This quantitative, quasi-experimental study addressed a gap in methodology on this topic. The majority of the literature reviewed on the topic (82%) were qualitative studies. Further, a majority of the small amount of quantitative study were descriptive in nature. Only nine percent of the quantitative studies reviewed were correlational, comparative, or experimental. Hence, there was a need to explore the topic using a quantitative approach. However, the researcher collected student achievement data on the school level rather than on the individual student level. This data collection approach placed limits on the inferential tool that could be used to most accurately analyze the data. The preferred One-Way MANOVA could not be used to analyze the effect of the independent variable on a linear composite of the two dependent variables thereby reducing the possibility of committing type one error. Instead, multiple Chi-Square Test of Homogeneity were used to analyze student achievement data but the large number of tests increased the probability of committing type one errors.

A second area for future research on the topic to consider is the sample size of schools. In this study, the researcher used four middle schools for the study. While the number of teachers and students from the four middle schools who participated in the study met the minimum requirements for a medium sample size effect, the small number of schools posed a problem. In answering research question three about the differences in the impact of the extent of DWIP implementation on student achievement, there was one school at the initiating level, one school at the lowly developing, and two schools at the developing level. The small number of schools at each of the three levels of DWIP implementation was not sufficient in that they could have been outliers. It is recommended that future studies have a large enough sample of schools that would allow for more than one school to be at each implementation level.

A third area for research on this topic is to look at the differences in student achievement between performance levels on the PARCC assessment. Looking at the differences between the five performance levels after the intervention will allow the researcher to identify incremental differences in student performance. It is equally important for the teacher, administrator, and policy maker to know if students increased their performance from level 1-not met expectations to level-3 partially met expectations even if they did not meet the required level 4-met expectations. This research was limited in that it only looked at the differences in the proportion of students who either met or did not meet performance expectations on the PARCC assessment.

Finally, future research on the topic should consider to what extent does the DWIP impact a collaborative school culture. Although this study collected data on school culture using the School Culture Survey (Valentine & Gruenert, 2005) in survey in questions 1-35 of the survey instrument, it did not explore the effect of various levels of the DWIP implementation on school culture. This would be an area in need of study. The DWIP is an eight- step cyclical approach to continuous school improvement that promotes a collaborative data-inquiry culture (Boudett et al., 2014). The process is grounded in DDDM, teacher collaboration in the form of PLCs, reflective practice, ongoing professional development, and collaborative leadership. According to qualitative studies conducted by Butler et al. (2015) and Simms and Penny (2014), there is a link between teacher collaboration and student achievement. Dougherty (2015), Cannata et al. (2016), and Ezzani (2015) found a strong link between DDDM and student achievement. Additionally, organizations in which employees engage in reflective practice through collaboration show stronger gains in student achievement (Gero, 2014). Further, organizations with leaders who engage in distributed or shared leadership show greater support for collaboration by empowering teachers (Ezzani, 2015; Leithwood et al., 2002). It is therefore important to examine if the DWIP which is encompasses all these variables a in refined framework continues to cultivate a collaborative school culture which was the main framework that guided this study.

Conclusion

The purpose of this quantitative, quasi-experimental study was to determine the impact of a collaborative data-inquiry culture, as promoted by the Harvard University DWIP, on student achievement in School District X. Specifically, the study used a one sample pretest-posttest design with an intervention to answer three research questions. By answering the three research questions, the study was able to add to the body of knowledge around school improvement through teacher collaboration in the form of PLCs, DDDM, reflective practice, continuous and comprehensive professional development, and collaborative leadership within a continuous framework called the DWIP.

Analysis of the data revealed that the DWIP either had no impact or had a negative impact on student achievement. In response to research question one, the researcher found no statistically significant difference in the proportion of students who met or did not meet performance expectation on both the ELA and Math PARCC after the implementation of the DWIP. As a result, the researcher failed to reject the null hypothesis. In response to research question two, the researcher found that there was no statistically significant difference in student outcomes on the ELA PARCC at three different times during the implementation but found statistical significant differences in student outcomes as it relates to Math PARCC. As a result, the researcher failed to reject the null hypothesis as it relates to ELA but rejected the null hypothesis as it relates to Math. The results showed that fewer students met performance expectations on the Math PARCC each year from 2015 to 2017. Finally, in response to research question three, the researcher found no statistically significant difference in student outcomes on the ELA PARCC based on the level of implementation of the DWIP (initiating, lowly developing, and developing). However, the researcher found statistically significant difference in student outcomes on the Math PARCC. As a result, the researcher failed to reject null hypothesis three as it relates to ELA but rejected the null hypothesis as it relates to Math. More students met performance expectations on the Math PARCC at the lower implementation level in 2016 and fewer students met performance expectations from year zero to year two at all three implementation levels.

The study provided empirical evidence that showed that PLCs, DDDM, reflective practice, continuous and comprehensive professional learning, and collaborative leadership when implemented as a coherent and cyclical approach through the DWIP intervention had no significant impact on student achievement in ELA and a mostly negative impact on student achievement in Math. However, the intervention was not sufficiently implemented to have a measurable impact on student achievement. Further research on the topic must focus on a larger sample size of schools, and must use achievement data on the individual student level in order to make the best inferential analyses and to limit the potential of committing type one errors. Additionally, future research must explore the effects of various levels of the DWIP on a collaborative school culture.

References

Adams, K. A., & Lawrence, E. K. (2015). *Research methods, statistics, & applications*. Thousand Oaks, CA: Sage Publications, Inc.

Aldred, G. (2014). Quasi-experimental designs in education. Retrieved from https://cupo.blackboard.com/bbcswebdav/pid-786029-dt-content-rid-9027195_1/courses/20182021373-cup/rooms/resources/Quant-Experimental/Quasi-Experimental%20Design%20in%20Education.pdf.

Algozzine, A. K., Friend, C., McRae, D., & Seifert, A. (2011, October 01). Evaluation of the Data Wise Improvement Process. Retrieved from http://www.cms.k12.nc.us/cmsdepartments/accountability/cfre/Pages/default.aspx.

Al-Saaideh, M. A., & Al-Zyoud, M. S. (2015). Benefits of teaching interdisciplinary subjects collaboratively in Jordanian pre-vocational education. *Educational Research and Reviews, 10*(20), 2702–2712. doi:10.5897/err2015.2461.

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education (AERA/APA/NCME). (2014). *Standards for Educational and Psychological Testing*. Washington, D.C.: American Educational Research Association.

Anwar, N. P. (2016). Action research a tool to build capacity of teacher educators. *Journal of Educational Research; Bahawalpur, 19*(2), 105–116. Retrieved from search.proquest.com/openview/24542c766b78e494f3ee1f1b966391ce/1?pq...cbl.

Baker, W. J. (2015). *Middle school teachers' perception of data use within professional learning communities* (Doctoral dissertation). Retrieved from ProQuest databases (UMI No. 3701247).

Barth, R. (2002). The culture builder. *Educational Leadership, 59*(8), 6–11.

Beckett, L. (2014). Raising teachers' voice on achievement in urban schools in England: An introduction. *The Urban Review, 46*(5), 783–799. doi:10.1007/s11256-014-0301-x.

Bennett, J. V., Ylimaki, R. M., Dugan, T. M., & Brunderman, L. A. (2013). Developing the potential for sustainable improvement in underperforming schools: Capacity building in the socio-cultural dimension. *Journal of Educational Change*, *15*(4), 377–409. doi:10.1007/s10833-013-9217-6.

Bessant, J., Caffyn, S., Gilbert, J., Harding, R. and Webb, S. (1994). Rediscovering continuous improvement, *Technovation*, *14*(1), 17–29.

Bolman, L., & Deal, T. (1991). *Reframing organization: Artistry, choice, and leadership*. San Francisco, CA: Jossey-Bass.

Boudett, K. P., City, E. A., & Murnane, R. J. (2015). *Data wise: A step-by-step guide to using assessment results to improve teaching and learning*. Cambridge, MA: Harvard Education Press.

Boudett, K. P., City, E.A., & Murnane, R.J. (2016). Data wise rubric [Pdf]. Cambridge, MA: Data Wise Project.

Boudett, K.P., & Steele, J.L. (Eds.). (2007). *Data Wise in action: Stories of schools using data to improve teaching and learning*. Cambridge, MA: Harvard Education Press.

Cannata, M., Redding, C., & Rubin, M. (2016). Continuous improvement in action: Evidence use for school improvement. Retrieved from https://files.eric.ed.gov/fulltext/ED567214.pdf.

Carpenter, D. (2015). School culture and leadership of professional learning communities. *International Journal of Educational Management, 29*(5), 682–694. doi:10.1108/ijem-04-

2014-0046.

Chatmon, B. N. (2016). *The impact of mentoring on job satisfaction and occupational commitment among African American nurse educators at historically black colleges and universities* (Doctoral dissertation). Retrieved from ProQuest databases (ProQuest No. 10103218).

Chi-square test of homogeneity. (n.d.). Retrieved from https://statistics.laerd.com/premium/spss/cstoh2xc/chi-square-test-of-homogeneity-2xc-in-spss.php.

Chow, A. W. (2016). Teacher learning communities: The landscape of subject leadership. *International Journal of Educational Management, 30*(2), 287–307. doi:10.1108/ijem-07-2014-0105.

Ciampa, K., & Gallagher, T. L. (2015). Blogging to enhance in-service teachers' professional learning and development during collaborative inquiry. *Educational Technology Research and Development, 63*(6), 883–913. doi:10.1007/s11423-015-9404-7.

Cornman, L. A. (2017). *The effectiveness of Missouri title I schoolwide programs in relation to the reading achievement of Missouri elementary students* (Doctoral dissertation). Retrieved from ProQuest databases (ProQuest No. 10255666).

Coughlan, M. (2015). *Teacher leadership: Connecting learning to practice* (Unpublished doctoral dissertation). Saint Mary's College of California, CA.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles, CA: Sage.

Dhongade, P. M., Shrouty, V. A., & Singh, M. (2013). A review: Literature survey for the implementation of kaizen. *International Journal of Engineering and Innovative*

Technology, 3(1), 57–60.

Dougherty, C. (2015). Use of data to support teaching and learning: A case study of two school districts. *ACT, 1*(1), 6–34. Retrieved from www.act.org.

Edwards, S. T. (2015). *Narratives of teacher leadership development: The voices of high Schoolteacher leaders* (Doctoral dissertation). Retrieved from ProQuest databases (ProQuest No. 3730313).

Ezzani, M. (2015). Coherent district reform: A case study of two California school districts. *Cogent Education, 2*(1), 1–20. doi:10.1080/2331186x.2015.1018698.

Fowler. F. (2002). Survey research methods (3rd ed.). Thousand Oaks, CA: Sage.

Gero, G. (2015). The prospects of lesson study in the US. *International Journal for Lesson and Learning Studies, 4*(1), 7–25. doi:10.1108/ijlls-02-2014-0007.

Gruenert, S. (2005). Correlations of collaborative school cultures with student achievement. *NASSP Bulletin, 89*(645), 43–55. doi:10.1177/019263650508964504.

Hallam, P. R., Smith, H. R., Hite, J. M., Hite, S. J., & Wilcox, B. R. (2015). Trust and collaboration in PLC teams. *NASSP Bulletin, 99*(3), 193–216. doi:10.1177/0192636515602330.

Hallowell, E. M. (2011). *Shine: Using brain science to get the best from your people*. Boston, MA: Harvard Business Review Press.

Harmon, H. L. (2017). Collaboration: A partnership solution in rural education. *The Rural Educator, Winter,* 1–5. Retrieved from epubs.library.msstate.edu/index.php/ruraleducator/article/view/409/371.

Hartman, S. (2017, January 01). Academic coach and classroom teacher: A look inside a rural school collaborative partnership. Retrieved from epubs.library.msstate.edu/index.php/ruraleducator/article/view/411/378

Heale, R., & Twycros, A. (2015). Validity and reliability in quantitative studies. Retrieved from https://www.bing.com/cr?IG= A01573E820DC459C877046E70F16128C&CID=1263AFAD75246034263DA405748B61C0&rd=1&h=SA-.

Hislop, D. (2013). *Knowledge Management in Organizations: A Critical Introduction*. Oxford, UK: Oxford University Press.

Imai, M. (1986). *Kaizen:The key to Japan's competitive success*. New York, NY: McGraw-Hill.

Jain, S., Cohen, A. K., Huang, K., Hanson, T. L., & Austin, G. (2015). Inequalities in school climate in California. *Journal of Educational Administration, 53*(2), 237–261. doi:10.1108/jea-07-2013-0075.

Jao, L., & McDougall, D. (2015). The collaborative teacher inquiry project: A purposeful professional development initiative. *Canadian Journal of Education*, *38*(1), 1–21. Retrieved from https://eric.ed.gov/?id=EJ1057950.

Jones, A., & Shindler, J. (2016). Exploring the school climate-student achievement connection: Making sense of why the first precedes the second. *Educational Leadership and Administration*, *27*, 35–51. Retrieved from https://eric.ed.gov/?id=EJ1094419.

Karwacinski, B. (2017). *The influence of cognitive and psychological well-being factors on freshmen community college student GPA: A prediction model* (Unpublished doctoral dissertation). Andrews University, Berrien Springs, MI.

Klute, M., Cherasaro, T., & Apthorp, H. (2015). Summary of research on the association between state interventions in chronically low-performing schools and student achievement. *Institute of Education Sciences, U.S Department of Education*, 2–25.

Kyōkai, N. H (1992). *Guiding continuous improvement through employee suggestions*. New York, NY: Productivity Press.

Lalor, B., & Abawi, L. (2014). Professional learning communities enhancing teacher experiences in international schools. *International Journal of Pedagogies and Learning, 9*(1), 76–86. doi:10.1080/18334105.2014.11082021.

Lashley, C., & Stickl, J. (2016). Counselors and principals: Collaborating to improve instructional equity. *Journal of Organizational & Educational Leadership, 2*(1), 1–24. Retrieved from http://digitalcommons.gardner-webb.edu/joel/vol2/iss1/6.

Lehman, T. (2015). *Professional development effects on teachers' perceptions in analyzing and using student data* (Doctoral dissertation). Retrieved from ProQuest databases (UMI 3680310).

Lillrank, P. M., & Kano, N. (1989). *Continuous improvement: Quality control circles in Japanese industry*. Ann Arbor, MI: Center for Japanese Studies.

Lund Research Ltd. (2018). Retrieved from https://statistics.laerd.com/premium/spss/cstoh2xc/chi-square-test-of-homogeneity-2xc-in-spss.php.

Mills, D. L. (2015). *A qualitative, causal-comparative study of teacher autonomy and the effect on student achievement.* (Doctoral dissertation). Retrieved from ProQuest databases (ProQuest No. 10004874).

Mills, H., O'Keefe, T., Hass, C., & Johnson, S. (2014). Changing hearts, minds, and actions through collaborative inquiry. *Language Arts*, *91*(1), 39–51. Retrieved from epubs.library.msstate.edu/index.php/ruraleducator/article/view/409/371.

Moss, J. T. (2015). *The Danielson model of teacher evaluation: Exploring teacher perceptions* *concerning its value in shaping and improving instructional practice* (Unpublished doctoral dissertation). Seton Hall University, South Orange, NJ.

Munoz, M. A., & Branham, K. E. (2016). Professional learning communities focusing on results and data-use to improve student learning: The right implementation matters. *Planning and Changing, 47*(1/2), 37–46.

Park, M. M. (2017). *The impact of multi-year math response to intervention as measured by smarter balanced assessments* (Doctoral Dissertation). Retrieved from ProQuest Databases (ProQuest No. 10591267).

Parrott, D. J., & Keith, K. J. (2015). Three heads are better than one: Librarians, reading specialists, and classroom teachers in the learning common. *Teacher Librarian, 42*(5), 12–18. Retrieved from Teacherlibrarian.com.

Phelan, C., & Wren, J. (2006). Exploring reliability in academic assessments. Retrieved from https://chfasoa.uni.edu/reliabilityandvalidity.htm.

Price, H. E. (2015). Principals' social interactions with teachers. *Journal of Educational Administration, 53*(1), 116–139. doi:10.1108/jea-02-2014-0023.

Publication manual of the American Psychological Association. (2010). Washington, DC: American Psychological Association.

Qualtrics Survey Platform (Version 2018) [Web-based software]. Provo, UT: Qualtrics, LLC.

Quinn, M. A., Southerland, J. L., Richards, K., Slawson, D. L., Behringer, B., Johns-Womack, R., & Smith, S. (2016). Quantifying collaboration using Himmelman's strategies for working together. *Health Education, 116*(1), 34–49. doi:10.1108/he-03-2014-0034.

Rockinson-Szapkiw, A. (2013). The doctoral journey: ANOVA vs. MANOVA [Video]. Retrieved from https://www.youtube.com/watch?v=mp5IRMiqpOU.

Salant, P., & Dillman, D. A. (1994). How to conduct your own survey. New York, NY: John Wiley.

Salkind, N. (1990). Exploring research. New York, NY: MacMillan.

Shadish, W., Cook, T., & Campbell, D. (2002). Experimental and quasi-experimental designs for generalized causal inference. Boston, MA: Houghton Mifflin Company.

Shang, H. (2017). A literature review on kaizen: What have we learned over the past 30 years? *Yokohama Journal of Social Sciences, 22*(4), 4–6. Retrieved from https://ynu.repo.nii.ac.jp/?action=repository_uri&item_id=8931&file_id.

Sims, R. L., & Penny, R. G. (2015). Examination of a failed professional learning community. *Journal of Education and Training Studies, 3*(1), 39–45. doi:10.11114/jets.v3il.558.

Solis, R. (2015). *Teachers' experiences of learning through a reflective inquiry process focused on the relationship between teaching beliefs and behaviors* (Doctoral dissertation). Retrieved from ProQuest databases (ProQuest No. 3725349).

Sporte, S. E., Jiang, J. Y., & L, S. (2015). Teacher evaluation in practice: Exploring relationships between school characteristics & evaluation scores. *SREE Conference, Spring* (2015), A1–B4.

SPSS statistics 17.0 (Version V25) [Computer software]. Armonk, NY: IBM

Thomas, T., Herring, M., Redmond, P., & Smaldino, S. (2013). Leading change and innovation in teacher preparation: A blueprint for developing TPACK ready teacher candidates. *TechTrends, 57*(5), 55–63. doi:10.1007/s11528-013-0692-7.

Tschannen-Moran, M., & Gareis, C. R. (2015). Faculty trust in the principal: An essential ingredient in high-performing schools. *Journal of Educational Administration, 53*(1), 66–92. doi:10.1108/jea-02-2014-0024.

Tuttle, C. (2015). *The impact of collaborative learning protocols on professional learning communities: An action research study* (Doctoral dissertation). Retrieved from ProQuest databases (ProQuest No. 3737251).

Wang, M., & Degol, J. L. (2015). School climate: A review of the construct, measurement, and impact on student outcomes. *Educational Psychology Review*, *28*(2), 315–352. doi:10.1007/s10648-015-9319-1.

Williams, M. (2001). Maximum cost reduction minimum effort. *Manufacturing Engineer*, *80*(4), 179–182.

York, K. (2017). *A causal comparative study on the effect of proficiency-based education on school climate* (Doctoral dissertation). Retrieved from ProQuest databases (ProQuest No. 10271282).

Young, A., & Kaffenberger, C. (2015). School counseling professional development: Assessing the use of data to inform school counseling services. *Professional School Counseling*, *19*(1), 46–56. doi:10.5330/1096-2409-19.1.46.

Appendix A: Collaborative Data-Inquiry Culture Survey

Collaborative Data-Inquiry Culture Survey

Start of Block: Valentine-Gruenert School Culture Attributes

Q1 Teachers utilize professional networks to obtain information and resources for classroom instruction.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q2 Leaders value teachers' ideas.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q3 Teachers have opportunities for dialogue and planning across grades and subjects.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q4 Teachers trust each other.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q5 Teachers support the mission of the school.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q6 Teachers and parents have common expectations for student performance.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q7 Leaders in this school trust the professional judgment of teachers.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q8 Teachers spend considerable time planning together.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q9 Teachers regularly seek ideas from seminars, colleagues, and conferences.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q10 Teachers are willing to help out whenever there is a problem.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q11 Leaders take time to praise teachers that perform well.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q12 The school mission provides a clear sense of direction for teachers.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q13 Parents trust teachers' professional judgment.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q14 Teachers are involved in the decision-making process.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q15 Teachers take time to observe each other teacher.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q17 Teachers' ideas are valued by other teachers.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q16 Professional development is valued by the faculty.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q18 Leaders in our school facilitate teachers working together.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q19 Teachers understand the mission of the school.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q20 Teachers are kept informed on current issues of the school.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q21 Teachers and parents communicate frequently about student performance.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q22 My involvement in policy or decision-making is taken seriously.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q23 Teachers are generally aware of what other teachers are teaching.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q24 Teachers maintain a current knowledge base about the learning process.

- ◯ Strongly Disagree (1)
- ◯ Disagree (2)
- ◯ Undecided (3)
- ◯ Agree (4)
- ◯ Strongly Agree (5)

Q25 Teachers work cooperatively in groups.

- ◯ Strongly Disagree (1)
- ◯ Disagree (2)
- ◯ Undecided (3)
- ◯ Agree (4)
- ◯ Strongly Agree (5)

Q26 Teachers are rewarded for experimenting with new ideas and techniques.

- ◯ Strongly Disagree (1)
- ◯ Disagree (2)
- ◯ Undecided (3)
- ◯ Agree (4)
- ◯ Strongly Agree (5)

Q27 The school mission statement reflects the values of community.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q28 Leaders support risk-taking and innovation in teaching.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q29 Teachers work together to develop and evaluate programs and projects.

- ○ Strongly Disagree (1)
- ○ Disagree (2)
- ○ Undecided (3)
- ○ Agree (4)
- ○ Strongly Agree (5)

Q30 The faculty values school improvement.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q31 Teaching performance reflects the mission of the school.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q32 Administrators protect instruction and planning time.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q33 Teaching practice disagreements are voiced openly and discussed.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q34 Teachers are encouraged to share ideas.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

Q35 Students generally accept responsibility for their schooling, for example they engage mentally in class and complete homework assignments.

- Strongly Disagree (1)
- Disagree (2)
- Undecided (3)
- Agree (4)
- Strongly Agree (5)

End of Block: Valentine-Gruenert School Culture Attributes

Start of Block: Data Wise Implementation Rubric

Q36 **Data Wise Step 1**: Organizing for Collaborative Work **Click for Rubric 1.1-1.4**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
1.1 The faculty and staff adopt an improvement process.	○	○	○	○
1.2 The faculty and staff build a strong system of teams.	○	○	○	○
1.3 The faculty and staff make time for collaborative work.	○	○	○	○
1.4 The faculty and staff set expectations for effective meetings.	○	○	○	○

Q37 **Data Wise Step 1**: Organizing for Collaborative Work **Click for Rubric 1.5-1.8**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
1.5 The faculty and staff set norms for collaborative work.	○	○	○	○
1.6 The faculty and staff acknowledge work style preferences.	○	○	○	○
1.7 The faculty and staff create data inventory.	○	○	○	○
1.8 The faculty and staff create data inventory of instructional initiatives.	○	○	○	○

Q38 **Data Wise Step 2**: Building Data Literacy **Click for Rubric 2**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
2.1 The faculty and staff review skills tested.	○	○	○	○
2.2 The faculty and staff study how results are reported.	○	○	○	○
2.3 The faculty and staff learn principles of responsible data use.	○	○	○	○

Q39 **Data Wise Step 3**: Creating Data Overview **Click for Rubric 3.1-3.3**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
3.1 The faculty and staff choose a focus area.	○	○	○	○
3.2 The faculty and staff analyze data, find the story.	○	○	○	○
3.3 The faculty and staff display the data.	○	○	○	○

Q40 **Data Wise Step 3**: Creating Data Overview **Click for Rubric 3.4**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
3.4 The faculty and staff make sense of the data and identify a priority question.	○	○	○	○

Q41 **Data Wise Step 4**: Digging into Data **Click for Rubric 4**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
4.1 The faculty and staff examine a wide range of student data.	○	○	○	○
4.2 The faculty and staff come to a shared understanding of what student data show.	○	○	○	○
4.3 The faculty and staff come identify a learner-centered problem.	○	○	○	○

Q42 **Data Wise Step 5**: Examine Instruction **Click for Rubric 5.1-5.3**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
5.1 The faculty and staff examine a wide range of instructional data.	○	○	○	○
5.2 The faculty and staff get clear about the purpose of observation.	○	○	○	○
5.3 The faculty and staff come to a shared understanding of what is happening in classrooms.	○	○	○	○

Q43 **Data Wise Step 5**: Examine Instruction **Click for Rubric 5.4**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
5.4 The faculty and staff identify a problem of practice.	○	○	○	○

Q44 **Data Wise Step 6**: Developing an Action Plan **Click for Rubric 6**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
6.1 The faculty and staff decide on instructional strategies.	○	○	○	○
6.2 The faculty and staff agree on what the plan will look like.	○	○	○	○
6.3 The faculty and staff put the action plan in writing.	○	○	○	○

Q45 **Data Wise Step 7**: Planning to Assess Progress **Click for Rubric 7**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
7.1 The faculty and staff choose assessments to measure progress.	○	○	○	○
7.2 The faculty and staff set student-learning goals.	○	○	○	○

Q46 **Data Wise Step 8**: Acting, Assessing, and Adjusting **Click for Rubric 8**

	Not started (1)	Initiating (2)	Developing (3)	Sustaining (4)
8.1 The faculty and staff assess implementation of the action plan.	○	○	○	○
8.2 The faculty and staff assess student learning.	○	○	○	○
8.3 The faculty and staff adjust the action plan.	○	○	○	○
8.4 The faculty and staff celebrate success.	○	○	○	○

End of Block: Data Wise Implementation Rubric

Appendix B: Statement of Original Work

The Concordia University Doctorate of Education Program is a collaborative community of scholar-practitioners, who seek to transform society by pursuing ethically-informed, rigorously- researched, inquiry-based projects that benefit professional, institutional, and local educational contexts. Each member of the community affirms throughout their program of study, adherence to the principles and standards outlined in the Concordia University Academic Integrity Policy.

This policy states the following:

Statement of academic integrity.

As a member of the Concordia University community, I will neither engage in fraudulent or unauthorized behaviors in the presentation and completion of my work, nor will I provide unauthorized assistance to others.

Explanations:

What does "fraudulent" mean?

"Fraudulent" work is any material submitted for evaluation that is falsely or improperly presented as one's own. This includes, but is not limited to texts, graphics and other multi-media files appropriated from any source, including another individual, that are intentionally presented as all or part of a candidate's final work without full and complete documentation.

What is "unauthorized" assistance?

"Unauthorized assistance" refers to any support candidates solicit in the completion of their work, that has not been either explicitly specified as appropriate by the instructor, or any assistance that is understood in the class context as inappropriate. This can include, but is not limited to:

- Use of unauthorized notes or another's work during an online test
- Use of unauthorized notes or personal assistance in an online exam setting
- Inappropriate collaboration in preparation and/or completion of a project
- Unauthorized solicitation of professional resources for the completion of the work

Statement of Original Work (Continued)

I attest that:

1. I have read, understood, and complied with all aspects of the Concordia University- Portland Academic Integrity Policy during the development and writing of this dissertation.

2. Where information and/or materials from outside sources has been used in the production of this dissertation, all information and/or materials from outside sources has been properly referenced and all permissions required for use of the information and/or materials have been obtained, in accordance with research standards outlined in the *Publication Manual of The American Psychological Association*

Kenneth Barrie

Digital Signature

Kenneth Barrie

Name (Typed)

October 26, 2018

Date

Appendix C: Facility Approval Letter

June 1, 2018

Mr. Kenneth Barrie

XXXXXXXXXXXXX
XXXXXXXXXXXXX

Dear Mr. Barrie:

The review of your request to conduct the research titled, *"The Relationship between a Collaborative Data-Inquiry Culture, as Promoted by the Data Wise Improvement Process, and Student Achievement"* has been completed. Based on the examination, I am pleased to inform you that the Department of Testing, Research and Evaluation has granted you authorization to proceed with your study.

Authorization for this research extends through the 2017-2018 school year only. If you are not able to complete your data collection during this period, you must submit a written request for an extension. We reserve the right to withdraw approval at any time or decline to extend the approval if the implementation of your study adversely impacts any of the school district's activities.

All documents requiring this office's approval are enclosed. Please be aware that the content of these documents must be exactly as that of the version approved by our office. Further, the participant consent forms distributed to your target research subjects must have the **XXXX** – IRB office "APPROVED" stamp. Should you revise any of these documents or change the procedure, the revisions and the revised procedure must be approved by this office before being used in this study.

An abstract and one copy of the final report should be forwarded to the Department of Testing, Research and Evaluation within one month of its completion. Do not hesitate to contact me if you have any questions. I can be reached at **XXX-XXX-XXXX** ext. **XXXX** or by e-mail, **XXXXXXXXX**. I wish you success with your study.

Sincerely,

XXXXXXXXXXXXXXX

Acting Supervisor, Office of Testing, Research, and Evaluation

XXX: Enclosure

Appendix D: Anonymous Click Consent

CONSENT FOR ANONYMOUS SURVEY

Dear Participant,

The purpose of this study is to examine the relationship between a collaborative data-inquiry culture, as promoted by the Harvard University Data Wise Improvement Process, and student achievement. I expect approximately 150 teacher volunteers to participate in the study. No one will be paid to be in the study but participants will be entered into a raffle to win up to ten $25 gift cards. Enrollment will begin on May 15, 2018 and end on June 15, 2018. To be in the study, you will complete this online survey. The survey will ask you questions about your perception of the data-use and collaborative practices at your school. While completing the survey should take less than 20 minutes, you will be able to stop and resume at a later time by clicking the green Bookmark at the top of the survey. You will also be able to complete the survey using any electronic device including a smartphone.

There are no risks to participating in this study other than the everyday risk of you being on your computer as you take this survey. The benefit is that your answers will help us understand the relationship between a collaborative data-inquiry culture and student achievement, which could guide future school improvement efforts. You could also benefit by reflecting on your own beliefs and practices around data-inquiry and teacher collaboration.

All data is collected anonymously. The online survey is anonymous. I will not ask you any personal identifying information and your survey responses will be not be linked to your email address. If you were to write something that made it to where I predict that someone could possibly deduce your identity, I would not include this information in any publication or report. Any data you provide will be held privately and will be destroyed three years after the study ends.

You Have the Right to Withdraw. You can stop answering the questions in this online survey at any time and for any reason.

Please print a copy of this for your records. If you have questions, you can talk to or write the principal investigator, Kenneth Barrie at kenneth.barrie@**XXX.XXX**. If you want to talk with a participant advocate other than the investigator, you can write or call the director of our institutional review board, Dr. **XXXXXXXXXX** (email **XXXX@XXX.XXX** or call **XXX-XXX-XXXX).**

Click the link below to consent and take this survey.

Appendix E: Permission to Use School Culture Survey

Fr: Kenneth.Barrie@**XXX.XXX**
To: **XXXXX@XXX.XXX**
Cc:**XXXX @XXX.XXX, XXXX@XXX.XXX**

Good morning Dr. **XXXX** and Dr. **XXXX**,

My name is Kenneth Barrie, a doctoral student at Concordia University-Portland. I am requesting permission to use your 35-item School Culture Survey to design an instrument for my dissertation study. My proposed quantitative study will examine the relationship between a collaborative data-inquiry culture, as promoted by the Harvard University Data Wise Improvement Process, and student achievement. I am copying my Dissertation Chair, Dr. **XXXX** on this request. She can be reached at **XXXXX@XXX.XXX.**

I affirm that all responses garnered via the SCS will be anonymous and/or confidential. I also affirm that all respondents' privacy rights will be protected and that no data will be used in any manner for the purposes of personnel evaluation, supervision or employment review. Further, I affirm that upon completion of the study, the findings will be shared, either via a copy of the full study or a summary of the findings and conclusions of the study with Professor **XXXX** and **XXXX**.

Please find attached a one page explanation of the study design, including the population, sample, and variables to be studied and statistical treatments for this quantitative study.

Kindest Regards,
Kenneth Barrie

Fr: **XXXX@XXX.XXX**
To: me, **XXXX@XXX.XXX, XXXX@XXX.XXX**

Kenneth,

I write to provide you with permission to use the School Culture Survey in your dissertation research. This permission is granted based upon the explanation you provided to us by email on April 2, 2018.

We wish you the very best of luck with your study. We look forward to reading your findings.

XXXXX,

XXXXX, Ph.D.
Professor Emeritus
XXXXXXXXXXXXXX
XXXXXXXXXXXXXX
XXXXXXXXXXXXXX
(XXX) XXX-XXXX
XXXX@XXX.XXX
www.XXXX.XXX
www.XXXX.XXXX.XXX/XXX/XXX

Appendix F: Permission to Use Data Wise Rubric

From:Kenneth.barrie@**XXX.XXX**
To: **XXXX@XXX.XXX**

Good morning Dr. **XXXX**,

Thank you for your work on improving teaching and learning through purposeful collaboration and data use. I am a doctoral candidate at Concordia University Portland and a **XXXX XXXX XXXX XXXX**. As part of my dissertation research, I am investigating the significance of a collaborative data-inquiry culture, as promoted by the Harvard University Data Wise Improvement Process, on student achievement in my school district. As you may be aware, **School District X** has used the process as a district-wide initiative since 2015.

During the course of my literature review, I discovered a *Draft Data Wise Guidance Rubric* and a *Data Wise Implementation and Protocol Rubric* which was used by the Charlotte-Mecklenburg Schools in a 2011 study on the Data Wise Improvement Process. I have attached both rubrics for your review. I believe that the Harvard University Data Wise Project owns the right to both of these instruments. I am therefore seeking the permission of the Data Wise Project to use one or both of the rubrics to design a survey instrument for my study. The rubric will be used in its current form and in conjunction with Steve Gruenert's 1998 35-item collaborative culture survey which was used in a 2005 study on Correlations of Collaborative School Cultures with Student Achievement. I intend to use the rubrics in their current form in order to preserve their validity and reliability. I have cited your work in both my literature review and methodology and will ensure that I give proper credit for the use of the rubrics.

The plan is to complete the study by October 1, 2018. After the successful completion and defense of my study, I will be happy to share the finding of my research with you.

Thank you for your consideration and I look forward to hearing from you.

Kindest Regards,
Kenneth Barrie

From: **XXXX@XXX.XXX**
To: me, **XXXX@XXX.XXX**

Thank you for your message, Mr. Barrie. Yes, it is fine for you to use the rubric. I recommend using the attached version, which includes a small correction. Please do share your research with me when it is complete… I am very curious to know what you learn!

XXXX XXXX XXXX
XXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
www.XXXX.XXXX.XXX/XXXX