The Relationship between Student Math Achievement and Teachers Utilizing a Process Involving Interim Instructional Assessments

Douglas B. Greer
Western Michigan University, greerdouglasb@gmail.com

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Leadership Commons, and the Science and Mathematics Education Commons

Recommended Citation
Greer, Douglas B., "The Relationship between Student Math Achievement and Teachers Utilizing a Process Involving Interim Instructional Assessments" (2019). Dissertations. 3493.
https://scholarworks.wmich.edu/dissertations/3493

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
THE RELATIONSHIP BETWEEN STUDENT MATH ACHIEVEMENT AND TEACHERS UTILIZING A PROCESS INVOLVING INTERIM INSTRUCTIONAL ASSESSMENTS

by

Douglas B. Greer

A dissertation submitted to the Graduate College in partial fulfillment of the requirements for the degree of Doctor of Philosophy
Educational Leadership, Research, and Technology
Western Michigan University
August 2019

Doctoral Committee:

Jianping Shen, Ph.D., Chair
Louann Bierlein Palmer, Ed.D.
Michael Burde, Ph.D.
ACKNOWLEDGMENTS

I could not have accomplished this journey on my own accord as I have many whom I need to acknowledge. First and foremost, I need to praise the God whom I serve who “said to me, ‘My grace is sufficient for you, for My power is made perfect in (your) weakness’” (2 Corinthians 12:9 ESV). There were numerous times along this journey that I felt too weak to continue and I would experience encouragement from Him. Second is my beautiful wife, Kim, whose patience and support was beyond anything I ever expected. She has been by my side encouraging and supporting me since we met at Western Michigan University over 20 years ago. Kim and I are blessed with four healthy boys and I appreciated their understanding when I went away to work many evenings and about one weekend per month. On many of those weekends, I was able to stay in a cottage tucked back in the wilderness on a lake. The cottage is owned by my brother, Lance, and I appreciate his generosity, prayers, and support. I am also thankful for my parents, other family, and friends who have prayed.

On a more professional level, I appreciate the guidance and support of Dr. Shen, my chair, and Dr. Palmer as they were both instrumental in the creation of this research study. In addition, Dr. Burde was an encouragement along the way and it was helpful having him on my committee. I would be remiss not to mention Dr. Patricia Reeves, Dr. Jessaca Spybrook, Mary Ebejer, and Rebecca Bush who were also helpful with certain aspects of this study and they were always willing to help without asking anything in return. Finally, I would like to thank Pete Haines and Dr. Kyle Mayer at the Ottawa Area ISD who supported this work as we built
interim assessments and later used this study to launch a secondary math network across the region.

Douglas B. Greer
According to recent national and state level assessments, only about one-third of Michigan students are proficient in secondary math. Previous studies have been inconsistent in demonstrating the impact of an interim assessment process on student achievement, especially with high school mathematics. Moreover, previous studies were not found to utilize Hierarchical Linear Models (HLM) to test such a relationship, especially a three level HLM that links secondary students to a primary math teacher. Therefore, the purpose of my study was to ascertain the extent to which math teachers utilized an interim instructional assessment (IIA) process within middle and high schools, and how such utilization levels connected to student achievement. Another purpose was to ascertain the role that the intensity of teacher training had on the levels of utilization within these schools.

This quantitative study focuses on how the levels of implementation of the IIA process relate both to levels of teacher training and math achievement of secondary students. The sample included 13,494 students nested within 165 teachers and 35 middle or high schools. The first two research questions examine the variation of the data and the relationship between levels of teacher training and seven research-based components of the IIA process. The final research question examines the relationship between seven components of the IIA process and the student achievement in secondary math.
The data analysis reveals differences at the teacher level across the region regarding their utilization of the IIA Process, as well as differences among the student-level math achievement data provided by the state. Such differences were found to have a positive relationship (0.25*) between regional training and higher, self-reported practices of utilizing interim assessments in the classroom. The analysis also shows a positive relationship (0.11**) between growth in student achievement on the state assessment and interim assessment utilization such that a teacher with a lower IIA Process score of 2.8 would expect the students to have an average z-score of 0.09 on the state math test. Whereas a similar teacher with a higher IIA Process score of 5.4 would expect the students to have an average z-score of 0.37 on the state math test.

Likely the greatest implication to educational leaders is the impact teachers have on student learning even when controlling for at-risk factors such as poverty. This is evident in the positive relationship between the growth in student math achievement and some of the individual components, in particular, those related to professional learning communities (assessment design, data analysis, supportive structures and relationships) and those related to high impact instruction (student discourse and distributed practice). Educational leaders would want to assess how their school utilizes interim instructional assessments that follow the scope and sequence of instruction and the structures in place to allow for data analysis within a collaborative environment. In other words, those who collaborate around assessment results seem to have a greater impact on student learning. In addition to teachers talking with each other, this research supports the notion that students learn better when talking with each other as well.
TABLE OF CONTENTS

ACKNOWLEDGMENTS .................................................................................................................................................. ii

LIST OF TABLES .......................................................................................................................................................... viii

LIST OF FIGURES ......................................................................................................................................................... x

CHAPTER

1. INTRODUCTION .............................................................................................................................................................. 1

   Background ........................................................................................................................................................................ 1

   Problem Statement ........................................................................................................................................................... 9

   Purpose Statement and Research Questions .................................................................................................................. 14

   Conceptual Framework .................................................................................................................................................... 15

   Methods Overview .......................................................................................................................................................... 20

   Significance ....................................................................................................................................................................... 21

   Chapter 1 Closure .......................................................................................................................................................... 22

2. LITERATURE REVIEW .......................................................................................................................................................... 23

   Operational Definition of Assessment Types and Purposes ......................................................................................... 24

   Conflicting Studies around Formative and Interim Assessments ................................................................................. 28

   Theories around the Process to Utilize Assessments to Improve Achievement ..................................................... 34

   Additional Factors to Consider When Utilizing Interim Assessments ................................................................... 39

   Chapter 2 Closure .......................................................................................................................................................... 42

3. METHODS ......................................................................................................................................................................... 44

   Research Design, Approach, and Rationale .................................................................................................................. 45

   Chapter 3 Closure .......................................................................................................................................................... 50
Table of Contents — continued

CHAPTER

Population, Sample, and Setting ........................................................................................................ 48
Instrumentation ................................................................................................................................. 50
  M-STEP ........................................................................................................................................ 50
  PSAT/SAT ..................................................................................................................................... 53
  Teacher Survey Regarding Interim Assessments ........................................................................... 56
Operationalization of Variables ......................................................................................................... 58
  Teacher Survey Regarding Interim Assessments ........................................................................... 58
  M-STEP, PSAT, and SAT .................................................................................................................. 59
Demographic Variables ...................................................................................................................... 60
  School Site, Teachers, and Students ............................................................................................... 61
Data Collection Procedures .............................................................................................................. 64
Data Analysis .................................................................................................................................... 65
  Research Question 1 ......................................................................................................................... 65
  Research Question 2 ......................................................................................................................... 65
  Research Question 3 ......................................................................................................................... 69
Limitations and Delimitations ........................................................................................................... 72
  Chapter 3 Closure ............................................................................................................................ 72

4. RESULTS ..................................................................................................................................... 74
  Research Question 1: Descriptive Statistics ............................................................................... 74
  Teacher Level Statistics .................................................................................................................... 74
Table of Contents — continued

CHAPTER

| Building Level Statistics | .......................................................... | 85 |
| Student Level Statistics | ......................................................................... | 86 |
| Research Question 2: Training across the Region Related to Survey Results | ........................................ | 88 |
| Unconditional Model HLM Level 2 | ......................................................................... | 89 |
| Parsimonious Covariant Model | ......................................................................... | 89 |
| Professional Learning Levels Related to Levels of Implementation | ...................................... | 91 |
| Research Question 3: Relationship between Survey Components and Math Achievement | ................................ | 94 |
| Unconditional Model HLM Level 3 | ......................................................................... | 94 |
| Parsimonious Covariant Model | ......................................................................... | 96 |
| Survey Components in Relation to Math Achievement | ....................................... | 99 |
| Chapter 4 Summary | ......................................................................... | 102 |

5. DISCUSSION

| Discussion of Major Results | ......................................................................... | 106 |
| Levels of Training had a Positive Relationship with Levels of Utilization | .... | 107 |
| Levels of Utilization had a Positive Relationship with Student Achievement | ..... | 110 |
| Relationship of Results to Existing Studies | .......................................................... | 114 |
| Implications for Educational Leaders | ......................................................................... | 117 |
| Recommendations for Future Research | ......................................................................... | 120 |
| Concluding Thoughts | ......................................................................... | 123 |

REFERENCES | ......................................................................... | 125 |
### APPENDICES

A. Research Project Invitation to Administrators ................................................................. 133

B. Research Project Invitation to Teachers ............................................................................. 136

C. HSIRB Approval Letter ...................................................................................................... 139

D. Teacher Survey Regarding Interim Assessments ................................................................. 141

E. Additional Models for Research Questions 2 and 3 .......................................................... 152

F. Administrator Overview of Training .................................................................................. 167

G. Permission from J. Hattie .................................................................................................... 170

H. Permission to Use De-identified Data for Research ............................................................ 173
LIST OF TABLES

1. M-STEP Proficiency Levels by Scale Score ................................................................. 51
2. M-STEP Spring 2017 Mathematics Results for Baseline Data ........................................... 52
3. Redesigned SAT Suite Proficiency Levels and Scale Score Ranges ..................................... 54
4. SAT Spring 2017 11th Grade Results Publically Available .................................................. 55
5. Data Analysis Cross-Walk Table ....................................................................................... 63
6. Component (i): Prioritization of Standards ......................................................................... 76
7. Component (ii): Supportive Structures and Relationships .................................................. 77
8. Component (iii): Autonomy and Assessment Literacy ......................................................... 78
10. Component (v): Analysis that Leads to Action .................................................................... 80
11. Component (vi): Instructional Goals and Feedback ............................................................ 81
12. Component (vii): Student Discourse and Distributed Practice .......................................... 82
13. Survey Reliability Summary by Component and IIA Process Composite ............................ 83
14. Professional Learning Levels and Distributions .................................................................... 85
15. Student Level Descriptive Statistics ................................................................................... 87
16. Two Level HLM Unconditional Model Results .................................................................. 89
17. Results of the Parsimonious Level 2 Survey Model ................................................................ 91
18. Results of the Professional Learning Model ......................................................................... 92
19. Results of the Professional Learning Model on the Components ........................................ 93
20. Three Level HLM Unconditional Model Results ................................................................. 95
<table>
<thead>
<tr>
<th>Table Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Three Level HLM Block 1 Covariants Model Results</td>
<td>96</td>
</tr>
<tr>
<td>22</td>
<td>Three Level HLM Parsimonious Covariants Model Variations</td>
<td>98</td>
</tr>
<tr>
<td>23</td>
<td>Results of the Parsimonious Covariant Model</td>
<td>98</td>
</tr>
<tr>
<td>24</td>
<td>Results of the Composite Survey Model</td>
<td>100</td>
</tr>
<tr>
<td>25</td>
<td>Results of the Student Achievement Model and Components</td>
<td>101</td>
</tr>
<tr>
<td>26</td>
<td>Summary of Greer (2019) Research Results on Research Question and IIA Process</td>
<td>104</td>
</tr>
<tr>
<td>27</td>
<td>Summary of Greer (2019) Research Results on the Components of the IIA Process</td>
<td>105</td>
</tr>
<tr>
<td>28</td>
<td>Comparison of Greer Research with Previous Research Findings</td>
<td>116</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

1. NAEP Results for Michigan and the National Average .................................................. 3
2. 2017 Proficiency Results for Michigan ........................................................................ 4
3. Illustration of Four Types of Assessments and the Primary Purpose .............................. 7
4. Conceptual Framework around Interim Instructional Assessments ............................... 16
5. Illustration of Four Types of Assessments and Common Samples ............................... 27
6. Barometer of Influence .................................................................................................. 29
7. Updated Conceptual Framework with Results around the IIA Process ......................... 107
8. Highest and Lowest Rated Components in the Region ................................................. 118
CHAPTER 1

INTRODUCTION

Dr. Seuss once wrote about the concerns of a school around testing:

We also have a principal; his name is Mr. Lowe.

He is the very saddest man that any of us know.

He mumbles, ‘Are they learning this and that and such and such?’

His face is wrinkled as a prune from worrying so much.

He wrung his hands, he cleared his throat, he shed a single tear.

Then sobbed, ‘I’ve something to announce, and that is why I’m here.

All schools for miles and miles around must take a special test.

To see who’s learning such and such, to see which school’s the best’ (Seuss, Prelutsky, & Smith, 1998, pp. 10, 22).

High-stakes testing, with their real consequences for students, teachers, principals, and schools have increased for the past 15 years. Their increased frequency, however, has brought little increase value. This is likely because most tests merely occur in a moment of time, usually for the purpose to evaluate someone or something, and they do very little to improve instruction.

Occasionally, a series of assessments with a different design are embedded into a collaborative process that will improve teaching and learning. When teachers utilize a process around quality assessment data, will their students do “the best” on the “special test?” These “special tests” usually refer to high-stakes national assessments given annually versus the series of classroom assessments that usually follow the scope and sequence of instruction.

Background

High-stakes national assessments such as the National Assessment of Educational
Progress (NAEP) and a variety of state named assessments (e.g., M-STEP in Michigan) based on a national assessment consortium are typically used to evaluate. During the past two decades, significant pressure had come from federal and state legislation to utilize these assessments to measure student achievement. Since the re-authorization of the Elementary and Secondary Education Act (ESEA), called the No Child Left Behind Act of 2001 (NCLB), school buildings have been measured and labeled based on standardized testing, and student subgroups have been tracked for Adequate Yearly Progress (AYP). NCLB was the first major federal effort that eventually led to the creation of a new set of national standards for English Language Arts and Mathematics, known as Common Core State Standards (CCSS). The Smarter Balance Assessment Consortium was one of two national consortia that built test item banks to measure the CCSS. Michigan adopted the CCSS in 2010 when the legislators pushed through additional bills tying student growth to educator evaluation in a failed attempt to secure Race to the Top federal funds from the Obama administration (“Governor Granholm Signs,” 2010). Although Michigan was not successful in securing a portion of these federal funds, the state laws on educator evaluation and school building accountability remained in place.

With the emergence of clear standards and the demand to link educator evaluation and school accountability to student achievement and growth, various stakeholders, including legislators, have examined test results to evaluate not only teachers and schools but also the performance of education within the state of Michigan. The NAEP had been a stable measurement to compare states for nearly three decades as states compare to the national average and often ranked by other organizations. Twenty years ago, Michigan 8th grade students were significantly above the national average in mathematics with a scale score of 277 compared to the nation’s 272, a difference of five scale points. In 2005, Michigan remained at 277 points;
however, the nation climbed to 279 points, leaving Michigan two points behind the national average. In the recent NAEP results of 2015, Michigan 8th grade students climbed one point in mathematics to 278 while the national average increased to 282, dropping Michigan even further below the national average (NAEP, 2015) as seen in Figure 1.

Figure 1. NAEP Results for Michigan and the National Average (Greer, 2018).

Another issue that traditionally affects schools is low academic performance among different groups of students at the national level. With the amount of research and publications to support best practice around assessment data and instructional practice, it seems math achievement gaps between disadvantaged students should have decreased over recent years. Yet, NAEP 8th grade mathematics data shows a gap of 28 points between students who qualify for free/reduced lunch and those who do not qualify for both the 1996 and 2015 assessments; over 20 years without any impact in decreasing the gap at the national level. A disappointing trend had also been observed when looking at the Michigan students who qualify for free/reduced lunch. In 1996, Michigan 8th grade students who qualified for free/reduced lunch scored five points above the national average on the NAEP, while those who did not qualify also scored five
points above the national average. By 2005, however, free/reduced lunch students in Michigan were four points behind the national average, whereas those Michigan students who did not qualify were three points behind the nation. This is an astonishing drop in achievement, from five points above average, to three and four points below average. By 2015, the gap in Michigan grew to five and six points below the national average for the two respective groups (NAEP Results, 2015).

Today at the state level in Michigan, there are two assessments depending on the grade level. The Michigan Student Test of Educational Progress (M-STEP), which aligns with the national CCSS, covers elementary and middle school mathematics. The high school level test utilizes a college entrance exam, the SAT, by the College Board. Three different assessments in 2017 estimate that only about one-third of Michigan secondary students are proficient in mathematics as seen in Figure 2.

Similar proficiency across NAEP, M-STEP, and SAT show only about one-third of Michigan students are proficient in math.

![Figure 2. 2017 Proficiency Results for Michigan (Greer, 2018).](image)

According to the College Board (2017), Michigan ranked 48th in the nation on the SAT Math sub-score, with a 495 scale score. According to MI School Data (2018), only 36.8% of 11th-grade students in Michigan met the SAT benchmark for college readiness. Similarly, only 33.5% of Michigan 8th grade students, 36.2% of the 7th grade students, and 34.2% of the 6th grade
students were proficient on the M-STEP.

Students who qualified for free/reduced lunch show even lower performance with on the M-STEP and SAT, with only 18% of students proficient. The concerns with poverty and achievement gaps do not seem to escape any region in Michigan. Across one of Michigan’s intermediate school districts, virtually every one of the nearly 100 school buildings had a significant achievement gap between students who qualify for free/reduced lunch and those who do not. Based on 2017 M-STEP scores, 8th-grade students who qualified for free/reduced lunch had 29.8% fewer students proficient. Likewise, on the 2017 SAT scores, 11th grade showed a 27.5% gap between those who qualify or not for free/reduced lunch (MI School Data, 2018). As Michigan continues to lag behind the nation, proficiency scores on state tests show only about a one-third of students are proficient, and achievement gaps persist, a sense of urgency should exist in the field of education.

Compounding the low achievement concerns, there exists a misconception that high-stakes tests have the ability to improve teaching rather than simply evaluate a snapshot of student learning (Perie, Marion, Gong, & Wurtzel, 2007). Researchers and practitioners often refer to these high-stake, “special” tests as summative assessments. Such summative assessments occur at the end of the instructional interval and tend to evaluate a large amount of student learning over a pre-determined set of criteria (DuFour, DuFour, Eakert, & Many, 2010; Reeves, 2007; Stiggins & DuFour, 2009). Perie et al. (2007) defined a summative assessment based on three criteria: it covers a large amount of content, it has annual administration, and it has a primary purpose of evaluating achievement. On the other end of the testing spectrum is a formative assessment, which Perie et al. define based on similar criteria: it covers a small amount of content, it has a frequent administration (daily), and it has a primary purpose of driving
instruction. Formative assessments occur during the learning cycle and cover a small amount of content, usually delivered by teachers within a lesson to check for understanding (DuFour et al., 2010; Perie et al., 2007; Reeves, 2007; Stiggins & DuFour, 2009). Later in Dr. Seuss’s story of Mr. Lowe’s concerns on the “special test,” a courageous teacher, Miss Bonkers, steps forward and assures him and the students, “you’ve learned the things you need to pass the test and many more, I’m certain you will succeed” (Seuss, 1998, p. 26). There are likely teachers across the region like Miss Bonkers who are making a difference for students by having very little concern for the summative tests as they utilize not only formative assessments but also assessments that fall between formative and summative to improve teaching and learning. Unfortunately, there is no universal agreement on the terminology of these assessments that fall between formative and summative.

Indeed, DuFour et al. (2010) define four types of assessments on a continuum from formative to summative rather than just two types with nothing between formative and summative. They refer to small-scale, on-going assessments as the “most” formative usually delivered throughout a lesson on a daily basis. Assessments that are “more” formative, given on a weekly or monthly basis are called Common Formative Assessments (CFAs). CFAs are usually developed collaboratively among teachers and will often be given at the end of a unit or multiple units of instruction. DuFour et al. move along the continuum to “more” summative assessments administered two to three times per year, called Interim Benchmark assessments, also referred to as Comprehensive Benchmark or District Benchmark Assessments. Finally, summative assessments, as previously defined, are considered “most” summative. Assessment types that fall between formative and summative are also lumped under the term of interim assessments (Bambrick-Santoyo, 2010; Marshal, 2008; Perie et al., 2007), and Perie et al. (2007)
note that the primary purpose of this broad category of interim assessments varies greatly depending on the assessment design.

As a summary, Figure 3 combines the right triangle illustration of three types of assessments used by Perie et al. (2007), with the four categories of DuFour et al. (2010). The vertical axis of the right triangle represents the scope of content covered by each type of assessment; whereas the horizontal axis represents the frequency of administration.

![Four Types of Assessments](image)

**Four Types of Assessments**

Defined by the amount of material covered (scope), how often the assessments are given (frequency), and the primary purpose.

- **Summative** (primarily evaluative)
- **Interim Benchmark** (evaluative or predictive)
- **Interim Instructional or Common Formative Assessment** (primarily instructional)
- **Formative Assessment** (daily, checking for understanding, integrated into each lesson)

*Increasing frequency of administration*

Annually → Quarterly → Monthly → Weekly → Daily → Hourly → more...

*Figure 3.* Illustration of Four Types of Assessments and the Primary Purpose (Greer, 2018).

Note: Previous iterations of Figure 3 have been shared by Greer with practitioners since 2013.

Summative assessments cover a large amount of content and are infrequently administered. Whereas, formative assessments cover a small amount of content and are used within the lesson on a daily basis. Figure 3 breaks the broad category of interim assessments into
two categories: interim benchmark and interim instructional (also known as Common Formative Assessments (CFAs)). Interim benchmark assessments are usually given two to three times per year to measure growth by giving the same test throughout the year (Perie et al., 2007). Interim instructional or CFAs are usually given every five to nine weeks, follow the scope and sequence of instruction, and have a primary purpose of driving instruction (Bambrick-Santoyo, 2010; DuFour et al., 2010). The name interim instructional assessments imply both the type and the purpose of the assessment.

Beyond knowing the type of assessments, it is important to note that merely administrating an interim instructional assessment on a frequent basis will not intuitively increase student achievement. Using a medical analogy, imagine a patient whose doctor discovers he has high blood pressure at an annual physical. To assure the high blood pressure measurement is not an anomaly, the doctor requests the patient to measure his blood pressure on a daily basis and at different times throughout the day to keep a log. The simple act of measuring something more frequent only verifies the patient has high blood pressure on a regular basis and does nothing to decrease it to a healthy level. One month later, the doctor reviews the entries in the log to determine the patient needs intervention. The patient decides the option of diet and exercise is more appealing than medication. The doctor provides the patient with three researched based strategies that tend to reduce blood pressure: 1) reduce his sodium intake, 2) eliminate caffeine, and 3) begin running at least 20 minutes a day for 3 to 5 days per week. The patient continues to measure his blood pressure daily and keeps a log, hoping after six months, the log will show a steady decrease in blood pressure by the implementation of three strategies. If a corrective course of action is not taken, the patient’s blood pressure will not decrease only from the act of measuring it. Likewise, if corrective action is not taken when students struggle
on frequent assessments, student achievement is not likely to improve. Therefore, absent corrective action after taking interim instructional assessments, math scores will not likely improve on the “special” summative tests used at the state and national level.

**Problem Statement**

The data is evident on the achievement gaps and struggling performance of lower-income students and “after decades of effective schools research, the empirical support for the idea that schools can diminish the effects of social background on schooling outcomes is weak” (Vasquez, 2005, p. 2). Yet, there are schools who have found ways to achieve higher results despite poverty levels. “Unlike reading achievement, which is strongly influenced by family practices and the educational resources in the home, math achievement – especially at the high school level – is more likely the result of the opportunities to learn math that schools provide” (Reeves, 2012, p. 889). Given such findings, assisting teachers to embed research-based practices should allow schools to significantly outperform similar schools when controlling for certain demographics, such as poverty.

One of the more varied and complex research-based practices shown to lead to impact student achievement in some cases was the *utilization of interim instructional assessments (IIA)* within a collaborative environment or *IIA Process* (Alkhas, 2012; Bambrick-Santoyo, 2010; Konsantopoulos & Miller, 2013; Stiggins & DuFour, 2009). Indeed, researchers and practitioners suggested that in order to improve student achievement there are different components that should be implemented in order to effectively utilize interim instructional assessments within a collaborative environment. The first component rests on research Marzano (2003) published from a meta-analysis that ranked the top school-level factor for increasing student achievement as the establishment and prioritization of a *guaranteed and viable*
curriculum. Stiggins and DuFour (2009) explained prior to effective interim instructional assessment creation, teachers need to prioritize the learning targets and the level of student attainment. The prioritization of standards are a crucial first step to establish a guaranteed and viable curriculum.

Bailey and Jakicic (2012) implied one of the first components of the CFA or IIA Process are a clear understanding of assessment literacy; as they emphasized the knowledge and utilization of Common Formative Assessments. A certain level of autonomy concerning the creation of these assessments is suggested (Pink, 2009), and practitioners suggest the creation of interim assessments by collaborating with teachers (Bailey & Jakick, 2012; DuFour et al., 2010). Therefore, the second component for the IIA Process should be a clear understanding of the different types and purposes of assessments along with the autonomy at the district level to determine how to maximize effectiveness and efficiency. Furthermore, as a third component of the IIA Process, a quality interim assessment design must have sufficient evidence for the prioritized standards covered and alignment to the rigor of the high-stakes, summative test (Bailey & Jakick, 2012; Bambrick-Santoyo, 2010; Chappulus, Stiggins, Chappulus, & Arter, 2012).

Regardless of the high quality of the assessment design, learning will not likely improve without a diagnostic analysis that leads to action planning as a fourth component (Bambrick-Santoyo, 2010). If the quality of assessment holds importance, then indeed the quality of instruction, as a fifth component, would be at least as important. There must be high impact instruction and intervention strategies to improve student achievement, such as those found by Marzano’s (2003) meta-analysis and Hattie, Fisher, and Frey (2017). While these strategies provide guidance on how to teach, assessments help guide what needs to be taught or re-taught.
The balance of assessment and instruction are found in a collaborative environment that are supported by both the relationships within the school and the structures that support collaboration. As a final component, *supportive relationships and structures* are necessary for an effective team of teachers to diagnose assessment data that will lead to high-quality instruction (DuFour et al., 2010; Olivier & Hipp, 2010).

In order to assure the positive results seen in the school buildings that Bambrick-Santoyo (2010) has studied, he described four similar key ideas: assessment, analysis, action, and culture. According to Bambrick-Santoyo, the first step was to create interim instructional assessments that are rigorous and provide meaningful data. The next step was to analyze the results of the assessments to identify the causes of perceived strengths and weaknesses. Teacher action was the third step, which needs to be a response to what students need to learn the most based on the analysis. The last piece required a component of school culture that required the data-driven creation of an environment conducive to thriving instruction, the foundation of which was the combination of relationships and structures supporting the utilization of interim instructional assessments. Bambrick-Santoyo believes schools that implement these four principals with fidelity will avoid mistakes made by others when interim assessments fail to make an impact on student achievement. Bambrick-Santoyo also observed several mistakes made by schools, the use of time being among the primary errors; schools do not make time for collaborative data-dialogues nor do some schools make time for effective re-teaching. Therefore, his book elaborated on these four key principals and provided examples of professional development for training teachers.

However, veteran teachers often need more than directives from a book or an administrator to implement research-based strategies, such as the utilization of interim
instructional assessments within a collaborative environment, as they may feel they have found success in their current practice. Heath and Heath (2010) explain that people often have a “knee-jerk skeptical response to ‘imported’ solutions, (saying) ‘our situation is more complicated than that, those ideas will not work here’” (p. 31). Sparks (2014) made a similar connection that solutions are often discovered rather than imported. When addressing school reform, Sparks recommends four basic steps: define, determine, discover and design. The first step is to define the problem, such as math achievement on high-stakes tests, especially for students in Michigan and even more so for students in poverty. The next steps are to determine if there exist any positive deviants, or bright spots, then discover what strategies seem to have the greatest impact on teaching and learning, and design new programming based on knowledge learned from the bright spots.

Indeed, by investigating the stories of the other teachers with promising bright spot data, Heath and Heath (2010) found that leaders need both stories to motivate the “emotional elephant” while simultaneously directing the “rational rider” with the bright spot data. Ross (2016) emphasized the need to stop comparing ourselves to the state average and instead set higher expectations by studying the bright spots in order to raise academic expectations. Bambrick-Santoyo (2010) used stories of transformation with high poverty, bright spot schools who made a drastic improvement in mathematical achievement by focusing on interim assessments in a collaborative environment.

By examining the math data linked directly to teachers and the accompanying stories of success around interim assessments, schools across a region may come together around a common purpose of improving student outcomes in mathematics. Senge (2012) found educators should “learn to nourish a sense of commitment in a group or organization by developing shared
images of the future they seek to create and the strategies, principles, and guiding practices by which they hope to get there” (p. 7). However, a given group will need first to be brutally honest about the culture of the school system, and then need to establish a clear vision of where they want to be in the near future. Autonomy should be offered within the implementation of a balanced assessment system that focuses on the utilization of interim instructional assessments.

Of course, the simple creation of interim instructional assessments will likely fall short if the culture does not support the utilization of the assessments. Despite research on the components needed for teacher utilization of interim assessments in a collaborative environment, findings are inconsistent. Bambrick-Santoyo (2010) sites specific schools where the process has worked and summarized the mistakes likely made by schools who failed to show the impact on student achievement. The process of implementing interim assessments was complex and disappointing results might attribute to several causes (Bambrick-Santoyo, 2010; Marshall, 2008). The vast majority of studies seem to have focused on a specific school building; there does not seem to be sufficient research in the field that an entire region might be impacted by the proper implementation of interim instructional assessments.

Also, research still may be needed that directly links secondary teachers to math achievement and the role that the utilization of interim assessments within a collaborative environment may have played in producing higher achievement results. In addition, research around interim assessments was not found that utilized hierarchical linear models to detect the nested effects on achievement at the teacher level nor to identify the components that may add the most significant value. Some studies have used HLM to demonstrate results on the state tests can be predicted with interim benchmark assessments at the school level (Konstantopoulos & Miller, 2013) and schools perform significantly better when using interim benchmark assessment
data to drive instruction compared to schools who do not have benchmarks (Carlson, Borman, & Robinson, 2011). However, most studies neglected to use HLM, and the various results from other approaches showed inconsistent results where most grade levels did not show a significant impact from interim assessments (Alkhas, 2012; Goertz, Olah, & Riggin, 2010; Oliver, 2013; Pon, 2013).

**Purpose Statement and Research Questions**

The purpose of my study was to ascertain the extent to which math teachers utilized an interim instructional assessment (IIA) process within middle and high schools, and how such utilization levels connected to student achievement. Another purpose was to ascertain the role that the intensity of teacher training had on the levels of utilization within these schools. In order to accomplish these purposes, my study had three research questions. The first research question examined the different levels of research-based components related to teacher utilization of interim instructional assessments in a collaborative environment across all schools within a specific region using descriptive statistics. The next question grouped the schools by those who participated to some degree in regional teacher training and those who did not, to see if the levels of utilization differ between the two groups. The final question accounts for the student, teacher, and school level demographics to determine if at least one of the components will predict improved student scores in mathematics on high-stake, state-level tests linked directly to teachers. The specific research questions for this study included:

1. Within secondary math classrooms in a specific region of a Midwestern state, what were the levels of teacher implementation of seven research-based components regarding the utilization of interim instructional assessments within a collaborative environment, specifically:
i. prioritization of standards for a guaranteed and viable curriculum;

ii. supportive relationships and structures;

iii. assessment literacy and autonomy;

iv. quality interim assessment design;

v. an analysis that leads to action planning;

vi. high impact instruction: learning goals; and

vii. high impact instruction: engagement involving practice and discourse?

2. To what degree are there differences in these components between teachers who participated to some degree in training around interim instructional assessments and those teachers who did not? To what extent do buildings differ based on their decision to send teachers to participate in some degree of training?

3. To what extent are there relationships between the level of teacher implementation of these seven components involved in the utilization of interim instructional assessments and math achievement score improvements over time, controlling for certain student, teacher, and school building factors (i.e., poverty, ethnicity, special education, and gender)?

**Conceptual Framework**

The seven components around the utilization of interim assessments and the two relationships seen in research questions two and three are illustrated in Figure 4. The first relationship sought how regional teacher training influenced the effective utilization of the components around interim assessments, as seen in research question 2. The second relationship examined how the levels of teacher utilization of these seven components impacted math achievement on the high-stakes, state tests, as seen in research question 3, by tying student
The hope was that knowledge from this study regarding such relationships could be used with future training that would increase the likelihood of the conditions being met for utilizing interim instructional assessments within a collaborative environment for secondary teachers across a region, which was implied by the arrow between the left and center boxes of Figure 4.

**Professional Learning: Build & utilization of interim assessments**
- Collaborated to prioritize the standards
- Built interim instructional assessments
- Data analysis training
- Provided instructional resources and guidance on high impact strategies
- Encouraged districts to set-aside time analysis and intervention

**Components around the utilization of interim assessments**
1. Prioritization of standards for a guaranteed and viable curriculum
2. Supportive relationships and structures
3. Assessment literacy and autonomy
4. Quality interim assessment design
5. An analysis that leads to action planning
6. High impact instruction: Learning goals
7. High impact instruction: Engagement around student discourse and distributed practice

**Student math achievement on high-stakes, state tests**

*Increase in mean standardized math scores when controlling for both student, teacher, and school level factors and in relation to the utilization of instructional interim assessments*

*Figure 4. Conceptual Framework around Interim Instructional Assessments (Greer, 2018).*

Furthermore, a similar study at the regional center might reveal bright spot data for teachers with high fidelity of implementation in some, if not all, of the seven components around the
utilization of interim assessments. In other words, components around the teacher utilization of interim instructional assessment might have shown a positive impact over time for student scores on the math portion of the state assessment. The arrow between the center and the right boxes of Figure 4 suggested a possible positive impact.

The center box of Figure 4 shows the seven components around the teacher utilization of interim instructional assessments discussed earlier. These conditions for effective utilization of interim instructional assessments included: (i) prioritization of standards for a guaranteed and viable curriculum, (ii) supportive relationships and structures, (iii) assessment literacy and autonomy, (iv) quality interim assessment design, (v) an analysis that leads to action, (vi) high impact instruction: learning goals, and (vii) high impact instruction: engagement around student discourse and disbursed practice. My study relied on the work of Marzano (2003) and Hattie et al. (2017) to operationalize the seven components around the utilization of interim instructional assessments as seen in the center box of Figure 4. In addition, supportive relationships and structures surrounding the teacher that are likely indicators of success to improve teaching and learning (Olivier, Hipp, & Huffman, 2010). Some of the critical aspects of these supportive components include trust and respect among both staff and students, within a culture that embraces risk. Other school structure supports include proximity, time, and resources to support collaboration. The hope was that training was provided that increased the likelihood of the components being met for utilizing interim instructional assessments within a collaborative environment for secondary teachers across a region.

The left box of Figure 4 shows the basic steps for teacher training related to building and utilizing interim instructional assessments as well as a clear understanding of the professional learning outcomes. Such assessment literacy around the types and purposes of assessments was
addressed in the background and in more detail found in Chapter 2. The first step of any training seen in the left box would be to understand the state-adopted standards that would be taught in different classes (i.e., 8th grade math, algebra 1, geometry). DuFour et al. (2010) and Buffum, Mattos, and Weber (2009) began their collaborative process with the question: What do we expect students to learn? They suggest a process for prioritizing the student learning, specifically the state-adopted standards, based on five criteria: teacher intuition, high-stakes tests, readiness, leverage, and endurance based on the original work of Reeves (2007). Therefore, one of the first stages of any training would be around the collaborative prioritization of standards to establish a guaranteed and viable curriculum, as shown in the left box of Figure 4. For the purposes of building interim instructional assessments in a training series, teachers needed to agree on what knowledge and skills to measure.

Once the standards were prioritized, the training guided teachers through a process to build enough assessment items to have sufficient evidence for each essential knowledge or skill (Bailey & Jakicic, 2012). In order to build interim instructional assessments, the second bullet in the left box of Figure 4 illustrated the levels of mastery. Teachers then determined the levels of mastery to establish the success criteria through scoring guides and rubrics (Hattie, 2009). In addition, teachers who studied each standard in relation to the expectations of the high-stakes assessments would better understand the rigor of each standard, usually placed on a continuum based on Dr. Webb’s Depth of Knowledge (DOK) (2005). Sufficient evidence, success criteria, DOK, and alignment to high-stakes testing were components of quality assessment design (Bambrick-Santoyo, 2010; Chappulus, Stiggins, Chappulus, & Arter, 2012).

Hattie et al. (2017) cited the importance of examining the success criteria when analyzing student assessment data, often at the standard level. Bambrick (2010) concurred with a
standards-level analysis. However, he promoted item-level analysis as the next level, “it is absolutely essential that assessment analysis be done test-in-hand. Teachers should constantly ask why students bombed given questions” (Bambrick-Santoyo, 2010, p. 45). By examining the distractors on a multiple choice test or the mistakes a student made on an open-ended response, teachers might establish a causal theory for student errors. If teachers understood why students struggled while analyzing the data, then they would be more likely to understand how to re-teach the essential knowledge and skills (Bambrick, 2010; Chappulus et al., 2012; DuFour et al., 2010). In other words, a quality analysis must lead to action planning, as illustrated by the fourth bullet in the left box of Figure 4.

Such action planning should lead to high impact instruction and intervention. Re-teaching usually requires using alternative strategies rather than continuing with the original attempts to teach the content. Likewise, student learning will not improve if teachers merely measured it and failed to make the time to respond with high impact instruction and intervention. Hatti et al. (2017) and Marzano (2003) highlighted several high impact instructional strategies, such as clear learning targets, student discourse, specific feedback, distributive practice and students assessing their learning progression. Therefore, any training would provide instructional resources and guidance on high impact strategies, as illustrated in the sixth and seventh bullets in the left box of Figure 4.

Practitioners Bambrick-Santoyo (2010) and Marshall (2006) observed in their work that a shortfall of implementation had been the lack of time set-aside to analyze data collaboratively and re-teach to struggling students. Therefore, any data analysis training, as illustrated in the third bullet of the left box of Figure 4, would have emphasized the importance of setting aside time as well as what to do with the time set aside. Therefore, districts and school leaders were
encouraged to set aside time for teachers to analyze the data and time for interventions for struggling students, as illustrated in the fifth bullet of the left box in Figure 4. The vision that was cast for teachers in training would involve the seven components of interim instructional assessments found in the center box of Figure 4. Furthermore, these specific components might predict bright spot data for schools or teachers with high fidelity of implementation. In other words, components around the utilization of interim instructional assessment might show a positive impact over time for student scores on the math portion of the state assessment.

**Methods Overview**

My study examined 35 middle and high schools across a specific region in a Midwestern state. Teachers within these school buildings were surveyed on how well they perceived the seven components around interim instructional assessments were utilized in their school or the IIA Process. Some of these teachers received regional training on building interim assessments and recommendations for utilization. Since the training was regional, there was no established level of fidelity that was mandated for teacher implementation, so a wide range of variation among the research-based components was expected among teachers, regardless of any regional training. The autonomy built within the training likely resulted in various levels of implementation of these key components of utilizing interim assessments to the extent that some schools may incorporate the strategies to analyze data but not implement the interim assessments built. To that end, other schools in the region that were not part of the training process had the same conditions in place that research suggested may improve student outcomes. Further information on this survey tool that measures the collaborative utilization of quality assessments was offered in the conceptional framework of this chapter and within Chapter 3.

In addition to the survey results, student math achievement for all schools was also
analyzed using the state assessment given every spring. For grades 6, 7, and 8, the M-STEP mathematics test was standardized, with students tracked over time. For grades 9, 10, and 11, the College Board suite of assessments known as the PSAT 8/9, PSAT 10, and SAT was used to measure student progress over time. A hierarchical linear model was utilized to set the most recent math achievement score as the dependent variable while measuring the seven components around interim instructional assessments in a collaborative environment as the primary independent variables of interest. The hierarchical linear model also controlled for previous student achievement data, and certain demographic factors at both the building level and student level such as ethnicity, economically disadvantaged, students with disabilities, English Learners, and gender. The hope was to isolate the impact these seven components have on improving math achievement for students in secondary schools.

**Significance**

Based on the struggles of Michigan students in mathematics according to the NAEP and promising practices of utilizing interim instructional assessments for students in other parts of the nation, there existed a need to determine what worked in one particular region of Michigan. With nearly 22,000 students in the region prepared to take a “special test to see which school is the best,” the time was now to learn from each other. The literature around the utilization of interim assessments and the process that involved analyzing the data varied from study to study. Some show an impact on student achievement where others did not. There seems to be a lack of studies that link teachers to student achievement data and measure the extent that teachers utilize interim assessments within a collaborative environment. Maybe where other studies have failed, this study would succeed. If nothing else, this study would add a unique seven-component analysis to determine the strengths and weaknesses of an area that was not performing to the
expectations of policymakers, parents, or many educators. Did the process of building interim assessments and the utilization of assessment data in a collaborative culture help improve student achievement and what would surface as the most significant components? If bright spots are determined, teachers in West Michigan would want to replicate the process or learn to adhere to the components that may prove to have the highest impact on student achievement.

Chapter 1 Closure

Based on the disappointing trends of the NAEP mathematics, state data in Michigan, and the alarmingly low proficiency rates in Michigan on the M-STEP, PSAT, and SAT, schools have looked for ways to improve student achievement, and educators were concerned about student growth scores. Some research has pointed to the utilization of interim instructional assessments. With a preliminary background on different types and purposes of assessments found in this chapter, Chapter 2 takes a deeper look into assessment literacy, especially the difference between interim benchmark and interim instructional assessments. This deeper look helps to discern the varied results attributed to interim assessments as found in the literature review. Chapter 3 expands on the survey that measured the seven components around the utilization of interim instructional assessments and provides details on the hierarchical linear models (HLM). Previous studies were not found that used HLM to link teachers with student achievement when researching the components around utilizing interim instructional assessments. Chapter 4 reports the results of the HLM models and Chapter 5 discusses the major results and the implication to educational leaders.
CHAPTER 2
LITERATURE REVIEW

While summative assessments typically do not make claims about improving teaching and learning, a growing amount of research over the past decades has focused on the impact of formative assessments. This is likely why interim instructional assessments were sometimes referred to as Common Formative Assessments (CFAs); however, the research on these type of assessments have shown mixed results. Unfortunately, there seemed to be a lack of understanding in the educational community between what distinguished an interim assessment and how these interim assessments might be categorized based on the primary purpose and assessment design. Therefore, in this chapter, a clear definition of the different types of assessments will be established, followed by a brief review of formative and interim assessment literature. Although the amount of research was growing around interim assessment, there did not appear to be consistency in terms of the positive impact it may have on student achievement. Often researchers point to different factors for why interim assessments did not demonstrate the anticipated impact on student learning.

However, there did seem to exist universal agreement that interim and formative assessments cannot stand alone and must be accompanied by analysis that leads to teacher action (Black & William, 1998; Goertz, Olah, & Riggan, 2009; Nichols, Meyers, & Burling, 2009; Stiggins & DuFour, 2009). Therefore, the third section of this chapter addresses how researchers have offered different frameworks for how assessment fits into an instructional cycle of improvement to make that connection between assessment, analysis, and action. Finally, a few researchers had identified additional factors to support the successful utilization of interim assessments to drive instruction and truly impact teaching and learning, though not everyone
agrees on an exclusive list of essential components. Therefore, my literature has been categorized into the following themes: define the types of assessments, established research on formative and interim assessments, the cycle of instructional improvement, and additional factors to consider when implementing interim assessments.

**Operational Definition of Assessment Types and Purposes**

Perie et al. (2007) have established the most widely accepted definition in the educational field to categorize three primary types of assessment: summative, interim, and formative. These assessments, according to Perie et al., vary based on the primary purpose, the scope of the content, and the frequency of the assessment. On the continuum from summative to formative, Figure 3 in Chapter 1 shows the relationship between the scope of content and frequency of assessment(s). Simply stated, formative assessments cover a small amount of content, often a learning target, and frequently occur within a lesson because their primary purpose is always to drive instruction. Summative assessments, on the other hand, cover a large amount of content and are administered less frequently, usually at the end of the term or annually. Summative assessments are often referred to as “high-stakes tests,” and their primary purpose is to evaluate student achievement, which may, in turn, evaluate teachers, schools, districts, and states, such as the NAEP. The third type, interim assessments, falls somewhere in between formative and summative assessments, since interim assessments usually cover a subset of the standards, not the entire year, or simply a single learning target. Also, interim assessments are given less frequently than formative and more frequently than summative. A growing number of studies claim interim assessments occur in an interval that ranges from five to nine weeks (Bambrick-Santoyo, 2010; Frey & Fisher, 2009; Goertz et al., 2009; Marshall, 2006; Stiggins & DuFour, 2009). In sum, high-stake, summative assessments are primarily intended to evaluate and are not
intended to drive instruction. Formative assessment tends to be instructional and interim assessments may vary considerably depending on three possible purposes: predictive, evaluative, or instructional (Perie et al., 2007).

Educators often fail to distinguish the purpose of the assessment and legislators often believe that one type of assessment can be used for an array of purposes. Perie et al. (2007) explained to the contrary: “Unfortunately, one of the truisms in educational measurement is that when an assessment system is designed to fulfill too many purposes—especially disparate purposes—it rarely fulfills any purpose well” (p. 11). Interim assessments can be categorized based on each of these three primary purposes and often take on unique names. One specific purpose based on assessment design is to predict future achievement. A common example is the ACT assessment suite, since the primary purpose is to “determine each student’s likelihood of meeting some criterion score on the end-of-year tests” (Perie et al., 2009, p. 8). Prior to 2015, ACT provided three assessments to schools in Michigan that were marketed to predict college readiness. The EXPLORE test (7th-8th grade) was able to predict success on the PLAN test (9th-10th grade), which was able to predict success on the ACT (11th grade), which claims the ability to predict college readiness. Also, benchmark scores were provided for each test to assess if the student was on the pathway to achieve college readiness based on a nationally normed trajectory. Michigan’s current high school state assessment, the SAT by College Board, has a similar design and purpose as the ACT. Both the ACT and SAT are usually administered annually; however, some schools gave the same assessment both in the fall and the spring to measure student growth.

When an interim assessment was given two to three times per year to measure growth, it was often referred to as a comprehensive benchmark assessment. Also known as interim
benchmark or simply benchmark. It was usually given three times per year: fall (week 1), winter (week 18), and spring (week 36) and was usually evaluative in nature. Note this violated other research that suggested the frequency should be at least quarterly, or every nine weeks (Bambrick-Santoyo, 2010). These benchmark assessments give a similar set of questions three times throughout the year to measure student growth typically on the set of standards the student was expected to learn during the course of the year or instructional interval. Perie et al. (2009) state the only instructional gains would be to view the benchmark as a programmatic assessment designed to change instruction not necessarily in mid-term but over the years. The students benefiting from the information gleaned from these assessments would not necessarily be the students assessed, but the students receiving the instruction in the future. (p. 8)

Therefore, the fall benchmark acted as a pre-test, while the spring benchmark served as a post-test and the primary purpose of these benchmark assessments was clearly evaluative, or evaluating student growth over a defined period of time.

Of course, all assessments evaluate student performance at a moment in time. However, another type of interim assessment sometimes referred to as Common Formative Assessments, use the data primarily to drive instruction; therefore, I have referred to these as interim instructional assessments (IIA) to attach the purpose to the term. In fact, such assessments should rarely appear as an unchangeable grade in a teacher’s grade book. One major difference between interim instructional assessments and the interim benchmark assessment is the content follows the scope and sequence of instruction, approximately every five to nine weeks (Bambrick-Santoyo, 2010; Frey & Fisher, 2009; Marshall, 2008; Stiggins & DuFour, 2014). Interim instructional assessments are designed to adjust instruction to better meet the needs of
students. In other words, the items should inform the teacher of where the student was struggling so that instructional strategies can be taken to help students reach the learning goal. This may sound similar to formative assessment; however, there was a significant difference in frequency, such that formative occurs much more frequently than interim assessments (Perie et al., 2009). Interim assessments intended for instructional purposes need to follow the scope and sequence of instruction and allow for in-depth item analysis to inform re-teaching of the whole group or small intervention groups. In sum, there are interim assessments that maintain the same content, like a pre-test and post-test, often referred to as a benchmark. An example of a benchmark would be the ACT and SAT college readiness exams. The other type of interim assessments adjusts the content of the assessment with the scope and sequence of instruction. An example might be the unit tests found in a math textbook. This is why assessment literacy was important since not all interim assessments have a primary purpose of driving instruction. Figure 5 below provides additional samples to accompany the continuum illustration.

![Four Types of Assessments with Samples](figure5.png)

**Figure 5.** Illustration of Four Types of Assessments and Common Samples (Greer, 2018).¹

¹ Examples do not explicitly address performance based assessments.
A common issue among secondary teachers is that fact that chapter or unit tests are often treated more as a summative assessment since there is no time set aside to re-teach the essential concepts from the previous chapter. Teachers often move on to the next chapter if the majority of the students performed satisfactorily on the previous chapter test (Bailey & Jakicic, 2012). Therefore, if chapter tests are considered as interim instructional assessments, they should be used to assess students in a moment of time and re-teach the essential content that a group of students may have struggled with. Once a student has shown mastery of the concept, the grade should be replaced for that portion of the chapter test (O’Conner, 2011). Much of the research on the impact of formative and interim assessments focus on small-scale formative assessments and interim benchmark assessments. The few studies that clearly used interim instructional assessments or CFAs did not show consistent gains in academic achievement.

**Conflicting Studies around Formative and Interim Assessments**

Long before the enactment of Race to the Top placed a tremendous amount of emphasis on standardized testing, Black and Wiliam (1998) wrote a ground-breaking literature review on *Assessment and Classroom Learning*. The study set out to answer one fundamental question: Does there exist evidence that formative assessment will increase student achievement? After reviewing 250 articles by researchers from several countries, the authors published a follow-up article that confidently concluded: “few initiatives in education have had such a strong body of evidence to support a claim to raise standards” (Black & Wiliam, 2004, p. 9). Though the term formative assessment may be relatively new, the concept dates back as early as, and likely earlier than, Dr. Madeline Hunter (1916 - 1994). Hunter is well-known for her lesson plan template that articulates a process that should occur in every lesson and requires teachers to check for student understanding (formative assessment) prior to any guided or independent practice.
Marshall (2006) found a common thread in his literature review of assessments, noting that exceptional teachers have “taken responsibility for ensuring that all their students learn (by) checking for understanding on a minute-by-minute, day-by-day basis and tenaciously working with students who are confused or unsuccessful” (p. 4). Numerous studies have cited Black and Wiliam (1998) while adding to the volume of research that supports the practice of formative assessment. Hattie (2009) included a scale with his published work, “Visible Learning: A Synthesis of Over 800 Meta-analyses Relating to Achievement,” which ranked over 130 factors that show a positive impact on student achievement. Formative assessments ranked in the top three of the 130 positive factors and had an effect size (0.90) that was more than double the expected growth of students that Hattie found in his meta-analysis. As seen in Figure 6, Hattie provided a scale for the effect size of each component, where his research has found the typical effect size over the course of one year was 0.40.

*Figure 6: Barometer of Influence. Hattie (2009) used with permission. (See Appendix D).*

While there exists a plethora of research around formative assessment, Perie et al. (2009) found that research was lacking on interim assessments, especially depending on the design of the interim assessment. Likewise, a research report commissioned by the Consortium for Policy Research in Education (CPRE) noted that “much of the belief in the potential of interim assessments to improve student learning comes from the growing body of research on formative
assessment” (Goertz, Olah, & Riggan, 2010, p. 1). To understand the research, a clear knowledge of assessment literacy was needed. The studies that do exist on interim assessment tend to use a variety of terms for these assessments, such as interim assessments, benchmark assessments and common formative assessments (CFAs). Regardless of the label used for the assessments, studies and publications show mixed results for improving student learning by using interim assessments.

For example, a quantitative study on *Common Formative Assessments for Mathematics Reform* by Alkhas (2012) looked specifically at the implementation of CFAs across high school courses for English language arts and mathematics in a large urban district with about 2,000 students, 88% minority. The researcher used the term common formative assessments as defined by Stiggins and DuFour (2009) as assessments “created collaboratively by teams of teachers who teach the same course or grade level and represent a powerful tool in effective assessment in professional learning communities” (p. 640). Hord (1997) goes further to define the term professional learning community (PLC) as an “ongoing process through which teachers and administrators work collaboratively to seek and share learning and to act on their learning, their goal being to enhance their effectiveness as professionals for students benefit” (p. 50). Alkhas found at the conclusion of his research that further studies would be needed to determine why his results varied drastically between subject areas, particularly why improvement was so much higher in English Language Arts classes compared to those of Algebra 1. Unfortunately, Algebra 1 proved to be the only math class with empirical evidence to demonstrate a significant improvement in the California State Test after implementing CFAs; whereas in Geometry, Algebra 2 and Summative Mathematics, “the results showed no statistically significant increase in the means between the two cohorts in three out of four mathematics courses” (Alkhas, 2012,
p. 144). The study opens the door to further studies in secondary mathematics involving CFAs or interim instructional assessments.

In another study, Pon (2013) compared two different types of interim assessment in two large districts located in California. Her study focused on mathematics achievement for students in grades 4-10. The first district used a comprehensive benchmark test that was “more summative in nature” (p. xi). The second district used an interim instructional assessment that followed the sequence of instruction in six-week intervals. The mixed methods research design showed that both types of assessments were predictive of the state test. Though her study showed that the final assessments of both the interim instructional assessment and the benchmark assessments both held strong positive correlations to the state assessments that was attributed to predictability, there was also a comparison using qualitative methods. From interviews and participation in data dialogues, the researcher could clearly identify that the assessments that follow the sequence of instruction were perceived as having a greater impact on adjusting instruction versus the benchmark assessments. However, the study did not link the teachers’ process of analysis and instructional adjustment to math achievement on the state test, simply noted the correlation and predictability.

Oliver (2013) researched the relationship between interim assessments for 8th grade mathematics and the Georgia state test for 13 middle schools. The Georgia school system administered the interim assessment every nine weeks, referring to them as Quarterly Based Assessments. Likewise, Carlson, Borman, and Robinson (2011) conducted a multistate district-level cluster randomized trial based on the implementation of Quarterly Based Assessments. Neither study specified the scope of content covered on each quarterly assessment, simply that Oliver collected teacher perception on the usefulness of these “long” assessments and Carlson et
al. used a commercial interim assessment. Therefore, it was not known if the quarterly
assessments followed the scope and sequence of instruction, but it seemed more plausible that
these assessments were more like a comprehensive benchmark test. However, survey questions
regarding if the assessments were beneficial to identify prior knowledge implied they were
benchmark type tests. Similar to Pon (2013), Oliver found that the quarterly assessment had a
strong correlation to the state assessment and could be used for predictability. Oliver also
surveyed the teachers on their perceived effectiveness of the quarterly assessment and sought a
relationship with state achievement scores. Carlson et al. did find the implementation of a
benchmark assessment along with “extensive training on interpreting and using the data to guide
reform” (p. 378) resulted in a significant increase on the state test compared to those schools who
had no training or use of benchmark assessments. Surprisingly, Oliver found different results as
“the school expressing the highest confidence in the QBAs also had the lowest percentage of
students who passed” the state test and “the school with the lowest confidence rating in the
QBAs had the second highest percentage of students who passed” the state test (p. 132). The
relationship was not significant and actually seemed to be counter-intuitive.

Math and ELA outcomes seem to be common areas of focus for researchers studying
interim assessments. Another report commissioned by the American Educational Research
Association (AERA) also found the impact varied among subject areas and grade levels (K-8) in
59 schools across a large region in Indiana. This study used a large-scale school level cluster
randomized experiment to examine “the impact of two well-known commercial interim
assessment programs on mathematics and reading achievement in Indiana” (Konstantopoulos &
Miller, 2013, p. 481). These ‘commercial interim assessments’ are not described in the study,
but were likely benchmark assessments that did not follow scope and sequence of instruction and
clearly were not teacher created. Although the use of interim assessments in this study resulted in improvement in student achievement, again not all results were significant. In fact, the impact was insignificant in lower grades (K-2) and varied substantially in the upper grades (3-8), where “significant treatment effects (were) detected in grades 3 to 8, especially in third- and fourth-grade reading and fifth- and sixth-grade mathematics” (Konstantopoulos & Miller, 2013, p. 481). The fact that 8th grade math has not been mentioned as showing a positive impact in any of these studies parallels the disturbing trends seen in 8th grade mathematics on the NAEP in Michigan.

Of the studies examined thus far, there are numerous studies that show interim assessments of different types or purposes that may have a positive result in one school but not another (Alkhus, 2012; Goren, 2010; Konstantopoulos & Miller, 2013; Marshall, 2006). They tend to show predictability but fail to show a relationship between teacher fidelity of a process around assessment data and student impact on a high-stakes test. The process of implementing interim assessments is complex and disappointing results may attribute to several causes such as failure to set aside time for collaborative analysis (Bambrick, 2010). Marshall (2006) also noted that interim assessments do not always lead to significant achievement gains. “As I have watched well-intentioned, hard-working educators make these mistakes, I have realized that interim assessments are a lot harder to implement well than a lot of us thought” (Marshall, 2008, p. 65). In another noteworthy study commissioned by the Consortium for Policy Research in Education (CPRE), Goertz et al. (2010) conducted an extensive study in Pennsylvania focusing on 3rd and 5th-grade mathematics. The researchers specifically looked at both an urban and suburban district and concluded that “in the end, more than anything, we are convinced that nearly countless stars must align in order for teachers to use interim and formative assessments
to teach mathematics for understanding” (p. 219). A few of the key mistakes that Marshall (2006, 2008) observed included a lack of alignment to the standards, delay in scoring, and time not built into the calendar for both data meetings and re-teaching. These are three key components seen in the conceptual framework in Chapter 1.

A report commissioned by the Council of Chief State School Officers (CCSSO) warned that what many educators call formative assessment does not lead to the impact found with Black and Wiliam’s research (Heritage, 2010). One issue that may have arisen from the work of Black and Wiliam was that they used a very broad definition for formative assessment, “encompassing all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (Black & Wiliam, 1998, p. 7). This definition had opened the door to vendors labeling a single activity or a single test falsely as a formative assessment. Heritage (2010) specifically concludes that the “core problem lies in the false, but nonetheless widespread, assumption that formative assessment is a particular kind of measurement instrument, rather than a process that is fundamental and indigenous to the practice of teaching and learning” (Heritage, 2010, p. 1). Clark (2010) confirmed this emphasis on process calling formative assessment a pedagogical process “that links instruction and curriculum with assessment to support individual learning in the social setting of the classroom” (p. 341). Therefore, it seemed vital to include how various researchers have described the process utilizing various assessment within what many would call a cycle of improvement.

**Theories around the Process to Utilize Assessments to Improve Achievement**

Several researchers have articulated the components around utilizing interim assessments as a process involving teachers and students. One study conducted by Goertz et al. (2010)
articulated such a process as a cycle of improvement, as the researchers focused on the utilization of two distinct interim assessments, likely benchmark, for elementary mathematics in both an urban and suburban district in Pennsylvania. The four-stage cycle of improvement parallels the well-known Deming’s Plan-Do-Study-Act (1992). First, educators gather evidence about student learning on the district-wide interim assessment. Next, grade level groups study the evidence by interpreting the evidence on the assessment. Third, individual teachers in collaboration with other educators design a course of action to improve instruction, which results in the action plan for re-teaching. Finally, the plan was carried out in either the whole group or small group intervention. In addition to the four-stage cycle, Goertz et al. consider several other factors that “influence how teachers access, manage, interpret, and act on data, as well as the types of data available to them” (p. 3). The researchers argue that the impact on student achievement may vary based on the influence of factors such as professional development, data systems, time allocation, and instructional support. However, they were not able to state conclusively that one or more of these factors made a significant impact on student achievement outcomes.

With so many outside factors that may impact the cycle of improvement or the process around interim assessments, it is important to consider how others have viewed the essential components of this process. Fisher and Frey (2009) conducted an ethnography on teacher utilization of common formative assessments (CFAs) over four years with a large urban elementary school that showed significant increases in student achievement. They also demonstrated how these CFAs or interim instructional assessments need to be a part of an instructional cycle that spanned every six weeks in this study. Teachers were given two hours every other week based on an early release of students. Therefore, there were three grade level
meetings for each six-week cycle. “As part of the cycle, teachers meet with their grade-level peers to agree on pacing, or curriculum guides (that) focus on specific content standards” (Frey & Fisher, 2009, p. 675). Prioritizing standards seems to be one of the essential components that are not articulated in many of the studies that expressed disappointment in the lack of statistical significance. Frey and Fisher explain how teachers develop their common assessments based on the “standards they have been focusing according to the pacing guide” (p. 675). The researchers emphasize the need to align items to the standards and items should resemble the types of items on high stakes test, as well as performance tasks and extended writings.

The final meeting in the six-week cycle analyzes the results of the assessment items and the individual performances of students. An important step was to establish a causal theory as the teachers “discuss their hypotheses for the incorrect responses and plan future instruction and focused intervention plans” (Frey & Fisher, 2009, p. 675). Bambrick-Santoyo (2010) had a similar recommendation that teachers need to focus on why students may have struggled on certain items in order to develop an action plan on how to support struggling students. A report commissioned by CPRE, Goertz, et al. (2010), after examining the impact of interim assessments on 3rd and 5th grade mathematics, had a similar conclusion about teachers attempting to determine what the student might have been thinking when their answer was incorrect. The study went even further to say, “we see this moment of analysis as a critical juncture between the reporting of assessment data and modification of instruction” (p. 124). This transition from analysis to action seemed to be one that solicited much discussion in the field and will likely be an area of interest when searching for bright spot data within any research study around interim instructional assessments, including this study, as I studied the analysis component (v) from research question number one.
The process around assessment data of collecting student responses on an assessment, scoring the assessment, then interpreting the data may seem simple. However, Nichols, Meyers and Burling (2009) articulated the complexity of moving from assessment to instruction based on data interpretation, which “involves reasoning from a handful of particular things students say, do, or make in particular circumstances, to their status on more broadly construed knowledge, skills, and abilities that constitute the student model” (p. 17). Therefore, it was important to remember that any assessment represents only a sliver of the knowledge and skills of any student at the moment in time. The complexities of what they referred to as a student model that illustrates an “individualized representation of a student’s current understanding of the subject matter (that are) inferred from observable behavior” (p. 17). The true challenge then comes in transitioning this inferred knowledge of the student into the instructional practices used to “alter the gap between a student’s current understanding and the targeted understanding” (p. 17).

Making this transition can be difficult for teachers. As Goertz et al. (2010) explained, “many teachers’ use of formative assessment information (including interim assessments) identified what content to re-teach and to whom, but did not necessarily delve into the reasons for individual students’ mistakes” (p. 124). While studying two large districts in Philadelphia, Goertz et al. discovered similar factors that influenced the analysis and interpretation of interim assessment data. These factors included the perception of student performance based on past data, the performance of the whole class on that particular assessment and teacher’s perception when it dealt with individual items and how the item was constructed. In order to translate this analysis into an actionable response in the classroom, the researchers conjectured that the level of support available to each classroom might have also impacted the analysis and action plan. “In sum, teachers’ analyses of interim assessment data varied both within and across districts,
served multiple purposes, and unfolded at different levels. The level at which the data were analyzed was strongly related to the instructional strategies employed in response” (Goertz et al., 2010, p. 150).

Another study conducted by Heritage, Vendlinski, and Herman (2009) sampled 118 6th-grade teachers across Los Angeles County and found similar results when conducting research around the measures of teacher content knowledge for teaching mathematics developed at the National Center for Research on Evaluation, Standards, and Student Testing (NCRESST). The researchers found that although teachers were capable of drawing inferences from interim assessments regarding student levels of understanding, teachers struggled with determining the next instructional steps. Heritage et al. found that teachers need a clear knowledge of the learning progression in a particular unit of study. Specifically, the teachers must possess a deep understanding not only of the content taught but also the foundational concepts that students would need to learn the new content. This includes the ability to explicitly communicate both the learning target and the criteria for students demonstrating mastery of the content. Heritage et al. concluded that “until teachers have better conceptions of learning to work with and deeper knowledge of how the elements of student learning are manifested, then the movement from evidence to action as a seamless process will remain a somewhat distant goal” (p. 31).

While Heritage et al. (2009) were able to make such a profound conclusion about teacher knowledge about the student learning progression, Goertz et al. (2010) went one step further to eliminate the belief the results would be more about content knowledge. Specifically, they sought to determine both the variation and the relationship between mathematical knowledge for teaching and knowledge of student thinking practices. In fact, their study falsely assumed that there would be a quantitative relationship between math knowledge and knowledge of student
thinking. Regardless of how high teachers revealed their mathematical knowledge, their sample reported similar occurrences of practices that demonstrated their knowledge of student thinking practices. Goertz et al. then took a more qualitative approach through observing classroom instruction and found that association was not based on a survey of reported frequency of classroom activities that are affected by math knowledge, “but rather the quality of those activities” (p. 217). Therefore, the impact of interim assessments seemed to depend on the quality of instruction that followed the analysis of the assessment. The next logical question was if there existed additional factors that would promote quality instruction based on data-driven analysis.

**Additional Factors to Consider When Utilizing Interim Assessments**

Marshall (2006) summarized from his literature review 23 conditions to assure success when implementing interim assessments, which he organized under four categories: antecedents, assessments, analysis, and action. Under the antecedent category, Marshall addressed teacher trust around clear learning expectations as measured by frequent, common assessments. Similar to the prioritization and autonomy found in the conceptual framework of my study. Marshall’s assessment category emphasized alignment to the rigor of state tests, high-quality items that would be diagnostic, yet aligned to the school’s curriculum sequence. His analysis included teacher scoring and analysis of user-friendly data reports and team data meetings supported by the administration. Marshall’s final category of action called for follow-up with students and enlisting students to monitor their own progression. Unfortunately, there did not seem to be a research study that supports the implementation of all 23 conditions.

Likewise, Stiggins and DuFour (2009) have published numerous best-selling, educational works that referenced successful cases. Based on one example at a small, rural elementary
school in Virginia, Stiggins and DuFour concluded, “to build a balanced and effective assessment system to work productively at all levels, four essential conditions must be satisfied: clear learning targets, commitment to standards-based instruction, high-quality assessment, and effective communication” (p. 642). Another practitioner expanded on these conditions surrounding interim assessments with the support of multiple case studies to claim, “The first step on the path to high student achievement (must be to) establish: transparent, common, rigorous assessments” (Bambrick-Santoyo, 2007, p. 44). However, in order to assure the positive results seen in the schools Bambrick-Santoyo (2010) had studied, he defined specific criteria for interim instructional assessments:

- Common across all teachers of the same grade level or course
- Content shall follow the scope and sequence of instruction
- Teachers interact with the assessments before instruction
- Frequent assessments, following every 6 – 8 weeks of instruction
- Aligned to the format and content of the state level assessment
- Aligned to the rigor of the state level assessment and college readiness
- Standards re-appear on subsequent interim assessments

In addition to this comprehensive list, two more essential elements to consider based on other case studies and literature reviews would be prioritizing standards and involve students in the process (Marshall, 2006; Frey & Fisher, 2009; Stiggins & DeFour, 2009).

These key components are supported by other researchers who have synthesized hundreds of meta-analysis studies to quantify the impact of the components observed by Marzano (2006), Bambrick-Santoyo (2010), Frey and Fisher (2009), and Stiggins and DuFour (2009). In addition, Marzano (2003) ranked establishing a guaranteed and viable curriculum as the top
school-level factor for increasing student achievement. Stiggins and DuFour support this additional factor by explaining, “effective assessment requires a framework of clear learning targets that are centered on the best thinking about the most important learnings and manageable given the resources and time to teach and learn them” (p. 642). This process of prioritizing standards was not seen in much of the research reviewed and cited, including Goertz et al. (2009), Bambrick-Santoyo (2010), Alkhas (2012), and Konstantopoulos et al. (2013), and might be a cause of the frustration of some teachers not knowing what to re-teach. This was a primary area of deficit in the research and was clearly an essential component (i) of my study.

Marzano (2003) established, as the second highest factor for improving student achievement, the process of teachers setting challenging goals (based on the prioritized standards) and providing effective feedback to students, which could be based on formative and interim assessments results. Hattie (2014) has stated publically that virtually everything he had studied works to improve student achievement; however, certain strategies simply work better than others. Hattie (2009) found that out of the 138 influences that he studied, only five had a negative effect, and only two had an effect size higher than 1.00.

Therefore, the goal of schools would be to incorporate those strategies that exceed the typical effect size as seen earlier in Figure 6. In fact, 72 influences were below the hinge point of 0.40 effect size, and 66 influences exceeded this threshold with the top 25% of influences having an effect size of 0.55 or higher. Factors with the highest effect sizes that could be incorporated into the process of implementing interim assessments included: mastery learning (.58), effective feedback (.73), teacher clarity (.75), providing formative evaluation (.90), and student self-reported grades (1.44) (Hattie, 2009). In a later study, Hattie et al. (2017) emphasized additional factors specifically for mathematics that included student discourse and
distributed practice with an effect size of 0.71. The concept of promoting increased student involvement in tracking their progress on the learning targets was supported by Marshall (2008), Stiggins and DuFour (2009), and others.

**Chapter 2 Closure**

Marshall (2006), who served as a teacher, principal, policy writer and curriculum director for 33 years, offers a simple analogy regarding the utilization of interim instructional assessments. He wrote, “at its best, this process generates the ‘Nintendo effect,’ paralleling what kids experience when they play computer games: they master a level and then move up, constantly honing their skills and improving their scores” (Marshall, 2006, p. 15). Likewise, Many (2013), who served as a teacher, principal, curriculum director and 20 years as a superintendent, addressed the process surrounding interim assessments (or CFAs) by boldly stating,

We know what works; in fact, there has never been a clearer consensus or greater agreement on what schools must do to positively impact student learning. The importance of a guaranteed and viable curriculum, common formative assessments (or interim assessments) and systemic pyramids of intervention is not up for debate. Neither is the idea that teachers should work together interdependently on collaborative teams. (p. 8)

Therefore, if a process was taught to build interim assessments (Bambrick-Santoyo, 2010; Marshall, 2006; Stiggins & Dufour, 2009) that prioritized a guaranteed and viable curriculum (Marzano, 2003; Frey & Fisher, 2009), teachers might be better equipped to provide effective feedback (Marzano, 2003; Hattie, 2009) that will allow students to track the essential standards (Hattie, 2009). If the right type of assessments are built, and an implementation process was
taught, those teachers who experience the process may be able to incorporate other research
based elements into their instruction such as mastery learning, student self-reported grades
(Hattie, 2009), student discourse, and distributive practice (Hattie et al., 2017).
CHAPTER 3

METHODS

Previous literature around interim instructional assessments within a collaborative environment had suggested a plethora of components that might support the use of these assessments (Bambrick-Santoyo, 2010; DuFour et al., 2010; Marshall, 2008). Previous studies have been inconsistent in demonstrating an impact on student achievement whether using interim benchmark or interim instructional assessments (Alkhas, 2012; Carlson, Borman, & Robinson, 2011; Goertz, Olah, & Riggan, 2010; Konstantopoulos & Miller, 2013; Oliver, 2013; Pon, 2013). Beyond the inconsistent results found in research for secondary math, studies were not found that deconstructed the components around the utilization of assessment data using hierarchical linear models nor a direct link to teachers and student achievement. Therefore, the purpose of my study was to ascertain the extent to which math teachers utilizing research-based components related to the utilization of interim instructional assessments were in place within middle and high schools, and how such utilization levels connect to student achievement score improvements. Another purpose was to ascertain the role of training with teachers to utilize interim instructional assessments had on the levels of utilization within these schools. The research questions for this study included:

1. Within secondary math classrooms in a specific region of a Midwestern state, what were the levels of teacher implementation of seven research-based components regarding the utilization of interim instructional assessments within a collaborative environment, specifically:
   i. prioritization of standards for a guaranteed and viable curriculum;
   ii. supportive relationships and structures;
iii. assessment literacy and autonomy;

iv. quality interim assessment design; an

v. an analysis that leads to action planning;

vi. high impact instruction: learning goals; and

vii. high impact instruction: engagement involving practice and discourse?

2. To what degree were there differences in these components between teachers who participated to some degree in training around interim instructional assessments and those teachers who did not? To what extent did buildings differ based on their decision to send teachers to participate in some degree of training?

3. To what extent was there a relationship between the level of teacher implementation of these seven components involved in the utilization of interim instructional assessments and math achievement score improvement over time, controlling for certain student, teacher, and school building factors (i.e., poverty, ethnicity, and special education)?

Chapter 3 describes the research design and provided a detailed description of the methodology used in this study. Information on the schools across one region in a Midwestern state are described, including the demographic makeup of the nearly 22,000 students in the region. To further address these three research questions, data collection procedures, analytical methods, and limitations are also discussed in this chapter.

**Research Design, Approach, and Rationale**

Since this quantitative study focused primarily on levels of implementation of the process around interim instructional assessments and how such levels related both to levels of training and math achievement of secondary students, I chose a nonexperimental ex-post facto design. Such a design looks back on existing data to explain relationships such as those surrounding
utilization of interim assessments and the researcher manipulated no variables. My study postulated there will be differences among teachers across the region regarding the seven components and that regional training would have increased the utilization of the seven components among teachers. Furthermore, that teachers who excelled in one or more of the seven components would correspond to higher student achievement results on the state assessment. The first research question looked at descriptive data of these components across a specific region. Research question two looked back in time to determine if past training of teachers had a significant impact on the components around the utilization of interim assessments within a collaborative environment. Kerlinger and Lee (1999) explain the nonexperimental ex-post facto design begins with the outcome data, such as math achievement on the state assessment, then examines any relationships between the primary, independent variables of interest, e.g., the seven components around utilization of interim instructional assessments for research question three in this present study.

Within my study, I believed there would be differences that exist among teachers and the buildings for at least one of the seven components and the composite score. Research question two focused on the components surrounding the utilization of interim instructional assessments in a collaborative environment and how these components related to levels of professional learning. Hierarchical Linear Models (HLM) were the most appropriate for determining if there are differences across the seven components of utilizing interim assessments related to levels of training (Lomax & Hahs-Vaughn, 2012). Provided there would be differences in both the components around assessment utilization and math achievement on the state test, research question three would use HLM to determine if the IIA Process IIA Process composite score and which of seven the components, if any, were statistically significant for predicting student
Many of the other studies utilized a simple linear, multiple regression, or ANOVA to determine differences and to predict results and were unable to include the teacher level (Alkhas, 2012; Burde, 2016; Oliver; 2013; Pon, 2013). However, a simple linear analysis has difficulties with heterogeneity of regression and misestimated standard errors. In addition, analysis of multilevel data in a single level regression often results in aggregation bias. Raudenbush and Bryk (2002) explain how the average SES of a school building may have an effect on math achievement above and beyond the effect of the level of the individual student. At the student level, the poverty status provides a measure of the “intellectual and tangible resources in a child home environment” (p. 99). However, when SES was aggregated to the building level, it was a “proxy measure of schools’ resources and normative environment. Hierarchical Linear Models help resolve this confounding by facilitating a decomposition of any observed relationship between variables, such as achievement and social class, in separate level 1 and level 2 components” (p. 100). Therefore, I used a three-level HLM to create an enhanced model to control for other covariants at the student level, teacher level, and the building level for the third research question.

Of course, there were certain shortcomings to strictly using a quantitative approach and not seeking meaning through a qualitative approach. Creswell (2014) explains the benefits of qualitative research as exploring the meaning of something like the use of interim assessments for teachers. A qualitative approach would have allowed for observing how teachers utilize assessment data in a collaborative environment. My study followed a deductive approach by examining the relationship between teachers utilizing a research-based set of seven components and the student math achievement on the state assessment. Other factors may have attributed to
an increase or decrease in achievement scores. Therefore my study was not attempting to explain causation of any of the seven components.

**Population, Sample, and Setting**

My study examined the perceptions of approximately 170 teachers in 35 middle and high schools regarding their utilization of interim assessments within a collaborative process, along with the math achievement on the state test, assessing approximately 22,000 students. The entire sample resided in the western region of a Midwestern state. Across the 35 middle and high schools, buildings had different grade configurations, including K-8, 5-6, 6-8, 7-8, 8-12, and 9-12. State assessment data was available for grades 6-8 based on the Michigan Educational Assessment Program (MEAP) prior to 2015, and the Michigan Student Test of Education Progress (M-STEP) since 2015. Prior to the 2015/16 school year, only 11th grade was measured using the Michigan Merit Exam (MME) based on ACT College Readiness Exam and a Michigan specific mathematics test as a wraparound to complete the MME. In 2015/16, the College Board replaced the MME with the SAT in 11th grade, and students in 9th and 10th grade were required to take a similar state assessment using the PSAT for the first time. Based on the state assessment using M-STEP in grades 6th – 8th and PSAT/SAT in grades 9th-11th, this study considered a grade span of 6th through 8th grade to represent middle schools and 9th through 11th to represent high schools, though not all buildings had this exact configuration. Of the 35 secondary schools, 16 are considered a high school, 19 are considered a middle school, of which, five are charter schools or Public School Academies (PSA). In 2016/17, there were approximately 11,600 students in grades 6th – 8th and 11,800 students in grades 9th – 11th who were assigned a state mathematics assessment in the region. An invitation was sent to 38 schools in the region and this study included the 35 schools that chose to participate in the survey, as three of the eight PSAs
chose not to participate.

The regional center that supported these schools had created a process that helped 64 teachers from 18 of these schools via training to build and implement interim instructional assessments over a 12 to 15 months timeframe, with the utilization of the assessments to drive instruction as part of an ongoing collaborative process. The professional learning offered across the region built interim assessments for grades 6 – 8, Algebra 1, Geometry, and Algebra 2. A large portion of schools, 18 of 35 schools, sent some teachers to the regional training. However, the majority of teachers, 115 of 179 teachers, across the region did not receive direct support by attending the training. The process to build the interim assessments for Algebra 1 occurred from November 2013 to January 2015; implementing and responding to the analysis of assessment data during the 2014/15 school year. Algebra 2 and Geometry were built from January 2015 to December 2015 with implementation, analysis and responding action plans occurring during the 2015/16 school year. Participating schools were allowed to expand implementation and participate in ISD wide collaborative dialogues in the 2016/17 school year. Middle school 6th, 7th, and 8th grades built interim assessments from January 2016 to February 2017, with full implementation during the 2017/18 school year. Therefore, all teachers and schools reached full implementation by the 2017/18 school year or earlier. My study looked at the teacher implementation during the 2017/18 school year and the achievement of students on the state assessment in the spring of 2018.

Research question three controlled for certain demographic factors such as ethnicity (NonCA), economically disadvantaged (ED), students with disabilities (SWD), English learners (EL), and gender (Female) at the student level and building level. All demographic data at the building level was publically available on a cloud-based data warehouse maintained by the
Michigan Department of Education known as MI School Data. All student-level data was de-identified and provided by the regional center for research purposes. Ethnicity was grouped into two categories: Non-Caucasian/Non-Asian (NonCA) and Asian/Caucasian. Caucasian students made up the majority of the region (75%), and Asian students were one of the smallest portions (3% of the region). These two ethnicities were grouped to create a dichotomous variable since both of these ethnic groups tend to be the highest performing across this nation, state, and region. NonCA students were made up of Hispanic students (16% of the entire region), African American (2% of the region), and any other ethnic group (4% of the region). Compared to the state population, this specific region of the state had fewer students who qualify for economically disadvantaged (36% versus 51%), more Hispanic students (16% versus 8%), and fewer African American students (2% versus 18%). All other demographic variables were about the same as the state that standardized the assessment scores.

Instrumentation

M-STEP

The Michigan Student Test of Educational Progress (M-STEP) was first administered state-wide in Michigan in the spring of 2015. M-STEP was based on the item bank and assessment blueprint of the Smarter Balanced Assessment Consortium (SBAC). The M-STEP based on SBAC “provides valid, reliable, and fair assessments of the deep disciplinary understanding and higher-order thinking skills increasingly demanded by a knowledge-based global economy” (Smarter Balanced 2014-15 Technical Report, 2016, p. vi). This next generation assessment used an adaptive approach that adjusts the sequence and difficulty of questions for each student based on the student response to previous questions. Adaptive tests are “customized to each student, thereby yielding lower error and greater reliability than fixed-
form tests of the same length” (Smarter Balanced 2014-15 Technical Report, 2016, p. 37). The M-STEP assessed students in English Language Arts and mathematics for third through eighth grade. Scores were reported as scale scores that were adjusted from the raw score based on the complexity and difficulty of the items students were given and scored correctly. Each grade level was given a unique range of scale scores above or below the proficiency benchmark of 1600 of sixth grade, 1700 for seventh grade, and 1800 for eighth grade. For the purpose of this study, these scale scores were standardized into z-scores for comparability across grade levels.

The M-STEP scores were reported as both a scale score and one of four proficiency levels: not proficient, partially proficient, proficient, and advanced. Table 1 showed the ranges for the proficiency levels based on scale scores at the middle school grades for both English language arts and mathematics. As Table 1 illustrates, the range for the proficient level was much smaller in mathematics compared to ELA.

Table 1

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>Not proficient</th>
<th>Partially proficient</th>
<th>Proficient</th>
<th>Advanced</th>
</tr>
</thead>
<tbody>
<tr>
<td>6th grade ELA</td>
<td>1508 – 1578</td>
<td>1579 – 1599</td>
<td>1600 – 1623</td>
<td>1624 – 1655</td>
</tr>
<tr>
<td>Math 6th grade</td>
<td>1518 – 1578</td>
<td>1579 – 1599</td>
<td>1600 – 1614</td>
<td>1615 – 1650</td>
</tr>
<tr>
<td>7th grade ELA</td>
<td>1618 – 1679</td>
<td>1680 – 1699</td>
<td>1700 – 1726</td>
<td>1727 – 1753</td>
</tr>
<tr>
<td>Math 7th grade</td>
<td>1621 – 1679</td>
<td>1680 – 1699</td>
<td>1700 – 1716</td>
<td>1717 - 1752</td>
</tr>
<tr>
<td>Math 8th grade</td>
<td>1725 – 1780</td>
<td>1781 – 1799</td>
<td>1800 – 1814</td>
<td>1815 – 1850</td>
</tr>
</tbody>
</table>

Source: MI School Data, 2018

M-STEP data was used both as the dependent outcome data for math achievement, but will also be used to control for previous math achievement. By utilizing previous state-level data, this study concentrated more on how the students improved rather than just how well they
achieve on a single state test. Traditionally, the region studied performed higher than the state average. Table 2 shows the academic achievement results for the region in this study compared to the state as a proficiency level, average scale score, and standardized z-score. As Table 2 illustrates, the region had nearly 19% more students proficient in 8th-grade mathematics when combining percent proficient and advanced. The same 8th-grade group of students scored 12.7 scale points higher than the state, which equated to a z-score of 0.4885. Note 5th-grade scores were included as they represent baseline data for the 6th-grade students in my study. There was a consistent pattern within the data at each grade level with about 12% more advanced and 5% more proficient students in the region than in the state.

Table 2

*M-STEP Spring 2017 Mathematics Results for Baseline Data*

<table>
<thead>
<tr>
<th>Mathematics</th>
<th>Not Proficient</th>
<th>Partially Proficient</th>
<th>Proficient</th>
<th>Advanced</th>
<th>Scale Score</th>
<th>Std. Dev.</th>
<th>z-Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional 5th Grade</td>
<td>18.3%</td>
<td>29.0%</td>
<td>24.1%</td>
<td>28.6%</td>
<td>1500.0</td>
<td>23.7</td>
<td>0.456</td>
</tr>
<tr>
<td>State 5th Grade</td>
<td>35.2%</td>
<td>29.8%</td>
<td>18.4%</td>
<td>16.6%</td>
<td>1488.5</td>
<td>25.2</td>
<td>0.000</td>
</tr>
<tr>
<td>Regional 6th Grade</td>
<td>18.4%</td>
<td>30.3%</td>
<td>23.6%</td>
<td>27.7%</td>
<td>1598.7</td>
<td>22.7</td>
<td>0.424</td>
</tr>
<tr>
<td>State 6th Grade</td>
<td>33.4%</td>
<td>32.4%</td>
<td>18.8%</td>
<td>15.4%</td>
<td>1588.1</td>
<td>25.0</td>
<td>0.000</td>
</tr>
<tr>
<td>Regional 7th Grade</td>
<td>22.2%</td>
<td>25.8%</td>
<td>24.7%</td>
<td>27.3%</td>
<td>1698.2</td>
<td>24.7</td>
<td>0.360</td>
</tr>
<tr>
<td>State 7th Grade</td>
<td>35.6%</td>
<td>28.2%</td>
<td>19.5%</td>
<td>16.7%</td>
<td>1688.8</td>
<td>26.1</td>
<td>0.000</td>
</tr>
<tr>
<td>Regional 8th Grade</td>
<td>22.6%</td>
<td>25.1%</td>
<td>22.6%</td>
<td>29.7%</td>
<td>1799.2</td>
<td>24.5</td>
<td>0.489</td>
</tr>
<tr>
<td>State 8th Grade</td>
<td>39.9%</td>
<td>26.6%</td>
<td>16.4%</td>
<td>17.1%</td>
<td>1787.9</td>
<td>26.0</td>
<td>0.000</td>
</tr>
<tr>
<td>Regional 5-8 Avg.</td>
<td>20.38%</td>
<td>27.55%</td>
<td>23.75%</td>
<td>28.33%</td>
<td>1649.0</td>
<td>23.9</td>
<td>0.432</td>
</tr>
<tr>
<td>State 5-8 Average</td>
<td>36.03%</td>
<td>29.25%</td>
<td>18.28%</td>
<td>16.45%</td>
<td>1638.3</td>
<td>25.6</td>
<td>0.000</td>
</tr>
</tbody>
</table>

*Source: MI School Data, 2018*
For M-STEP and other SBAC assessments with an adaptive component, simulations are conducted using the item pool to estimate test reliability. The reliability of SBAC was verified by both the National Center for Research on Evaluation, Standards, and Student Testing (CRESST) and the American Institutes for Research (AIR) in 2015 (Smarter Balanced 2014-15 Technical Report, 2016). Furthermore, the Smarter Balanced Technical Report (2016) laid out a validity framework that aligned with the AERA, APA, and NCME’s Standards for Educational and Psychological Testing (2014). The Technical Report concludes with confidence that the SBAC assessments “essentially cover the breadth and depth of assessable standards” (p. 21) from Common Core State Standards, which “are widely recognized for being on track for college and career readiness in lower grades” (p. 21). Likewise, the high school state assessments are considered among the few assessments used for college entrance.

**PSAT/SAT**

The SAT (formerly known as the Scholastic Aptitude Test) has been in use for over 90 years across the nation (Lawrence, Rigol, Van Essen, & Jackson, 2002). Michigan adopted the SAT as the high school state test to be administered to 11th grade students for the first time statewide in the spring of 2016. At the same time, the Preliminary SAT (PSAT) was required for both 9th and 10th grade students for the first time in recent history. The SAT had two primary sections since 1930 assessing both verbal aptitude and mathematical aptitude. In 1974, the math portion was introduced to quantitative comparisons in order for items to be less dependent on verbal skills for the student. Further changes were made in 1994 when the SAT added a calculator portion and introduced open-ended questions on the math portion (Lawrence et al., 2002). In 2015, College Board (2015) released the redesigned SAT where the math portion “focuses strongly on algebra, which research shows was disproportionately important for college
and career readiness and success” (p. 6). Therefore, it was important to track the course taught by teachers as other content areas such as geometry were not covered well on the SAT.

Table 3 shows the proficiency levels related to the scale scores on the redesigned SAT, PSAT 10, and PSAT 8/9. Green represents a section score that met or exceeded the benchmark, where yellow represents a student score likely within one academic year of growth to meet the benchmark, and red represents more than one-year’s growth needed. Table 3 illustrates how much higher the math portion requires to meet benchmark compared to the evidenced-based reading and writing (EBRW). The benchmarks are based on correlation studies to align PSAT/SAT scores to the likelihood of success in a freshman college course in English or mathematics. It was important to note that scores have a 10 point increment for each assessment and a range of 600 points. The SAT has the highest range of 200 – 800 scale points and the PSAT 8/9 has the lowest range from 120 – 720 scale points.

Table 3

*Redesigned SAT Suite Proficiency Levels and Scale Score Ranges*

<table>
<thead>
<tr>
<th>Grade Level</th>
<th>SAT</th>
<th>PSAT 10</th>
<th>PSAT 8/9</th>
<th>PSAT 8/9</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBRW Green</td>
<td>480 – 800</td>
<td>430 – 760</td>
<td>410 – 720</td>
<td>390 – 720</td>
</tr>
<tr>
<td>Math Green</td>
<td>530 – 800</td>
<td>480 – 760</td>
<td>450 – 720</td>
<td>430 – 720</td>
</tr>
<tr>
<td>EBRW Yellow</td>
<td>460 – 470</td>
<td>410 – 420</td>
<td>390 – 400</td>
<td>370 – 380</td>
</tr>
<tr>
<td>EBRW Red</td>
<td>200 – 450</td>
<td>160 – 400</td>
<td>120 – 380</td>
<td>120 – 360</td>
</tr>
<tr>
<td>Math Red</td>
<td>200 – 500</td>
<td>160 – 440</td>
<td>120 – 420</td>
<td>120 – 400</td>
</tr>
</tbody>
</table>

*Source: College Board website (2018)*
The region studied in my research again seemed to score higher than the state collectively on the SAT in 11th grade. Unfortunately, PSAT results for 9th and 10th-grade students were not publicly available because they are not utilized for school accountability. I needed to request these scores from the Michigan Department of Education specifically for one region. Table 4 illustrates the publicly reported scores for 11th-grade students on the SAT.

Table 4

*SAT Spring 2017 11th Grade Results Publicly Available*

<table>
<thead>
<tr>
<th>11th Grade SAT</th>
<th>SAT Score</th>
<th>Benchmark</th>
<th>Met or exceed</th>
<th>Did not meet</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regional EBRW</td>
<td>530.9</td>
<td>480</td>
<td>69.3%</td>
<td>30.7%</td>
<td>3,458</td>
</tr>
<tr>
<td>State EBRW</td>
<td>509.9</td>
<td>480</td>
<td>60.3%</td>
<td>39.7%</td>
<td>105,905</td>
</tr>
<tr>
<td>Regional Math</td>
<td>522.2</td>
<td>530</td>
<td>47.8%</td>
<td>52.2%</td>
<td>3,458</td>
</tr>
<tr>
<td>State Math</td>
<td>497.6</td>
<td>530</td>
<td>36.8%</td>
<td>63.2%</td>
<td>105,905</td>
</tr>
<tr>
<td>Regional Total Score</td>
<td>1053.1</td>
<td>n/a</td>
<td>45.1%</td>
<td>54.9%</td>
<td>3,458</td>
</tr>
<tr>
<td>State Total Score</td>
<td>1007.6</td>
<td>n/a</td>
<td>34.9%</td>
<td>65.1%</td>
<td>105,905</td>
</tr>
</tbody>
</table>

*Source: MI School Data, 2018*

The redesigned SAT and PSAT were recent, yet the College Board had been a stable predictor of college readiness. The reliability of the SAT in the past had been rather high. The internal consistency, which measures the reliability of the questions on the assessment against a standard, was 0.93 for reading and 0.92 for the math section (Ewing, Huff, Andrews, & King, 2005). The College Board (2015) assured not only the continued strength of reliability but also implies a greater validity, “the predictive validity of the exam — its ability to estimate the likelihood of success in postsecondary education — has been redesigned to maintain if not strengthen this predictive validity” (p. 205). It was important to note that the College Board
emphasizes the purpose of the assessment is to predict college readiness rather than assess all high school standards. In this regard, the College Board has narrowed the focus of the assessments to have a greater emphasis on algebra and less of an emphasis on geometry and trigonometry. This might offer some insight into the teacher population and the content taught by those participating in the study.

**Teacher Survey Regarding Interim Assessments**

Math teachers at each of the participating schools were administered a researcher-developed survey, as seen in Appendix D. This survey captured the seven components around the utilization of interim assessments in a collaborative environment as seen in the research questions. The paper version of the survey was transferred into Qualtrics to collect response and sort into a cloud-based spreadsheet. Data collected by Qualtrics was secure in the sense that the creator may set the security so that only the creator’s login had access to the results and the paper copies will be handled by the researcher. The descriptive data from the survey and achievement data from the Michigan Department of Education was examined in IBM SPSS version 25 for missing data, outliers, and descriptive statistics to check for assumptions. The data was then imported into HLM version 7.01 for further analysis into the models for research questions two and three.

The survey questions were developed based on the literature review found in Chapter 2. Specific publications that influenced the creation of the items include Bambrick-Santoyo (2010), DuFour et al. (2010), Hattie et al. (2017), Marzano (2003), Olivier and Hipp (2010), and Perie et al. (2009). The survey instrument was grounded in research as well as experienced in practice while working with teachers to develop assessments and utilize assessment data. Building the survey in this manner provided the reliability necessary for others to administer a similar
collection of data and analysis. To test the reliability of these components, I used Cronbach’s alpha to check the internal consistency among the items for each subscale and the IIA Process composite score (Field, 2013). In order to help increase the validity, the survey was reviewed by my doctoral committee and piloted with practicing math teachers and consultants. The feedback received helped to improve the readability of the questions and clarify the intended purpose of each section and each question.

The teacher survey gathered information on the seven components regarding interim assessments such that each component contains four to six items for a total of 37 items. Math teachers provided a response on one of two six-point scales. For components regarding the collective math department such as (i) prioritization of standards, (ii) supportive relationships and structures, and (iii) assessment literacy and autonomy, a six-point, forced-choice Likert scale was utilized:

- 1 – strongly disagree
- 2 – moderately disagree
- 3 – slightly disagree
- 4 – slightly agree
- 5 – moderately agree
- 6 – strongly agree

These components of knowledge, processes, and structures either exist or do not exist in the culture of the school. Whereas, the remaining components may vary in the extent to which they are used week by week or unit by unit in an individual classroom. Therefore, regarding component (iv) the quality of assessments, (v) the depth and frequency of analysis, and (vi and vii) high impact instruction, the survey used a six-point scale measuring the extent of expertise
on a continuum:

- 1 – no expertise at all
- 2 – novice
- 3 – advanced beginner
- 4 – competent
- 5 – highly proficient
- 6 – regional expert

The survey instrument was divided into seven components around teacher utilization of interim instructional assessment in a collaborative environment, as well as a demographic section. The survey results kept respondents identities confidential, as certain demographics were collected but not shared with anyone except in aggregate form to protect confidentiality. Demographics included the type of math classes the teacher has taught the past three years, the number of years teaching, highest degree obtained, school name, and participation in specific professional learning related to content in this study. Finally, the regional center collected the teacher name to link student data with teacher roster information and the results of their survey. The regional center then provided me with a de-identified data set for the purpose of this research study. Permission to use the de-identified data can be found in Appendix H.

**Operationalization of Variables**

The two primary variables of interest that were operationalized in my study were teacher survey data and student achievement data. In addition, covariates such as demographics were used for control within the models.

**Teacher Survey Regarding Interim Assessments**

The survey tool questions were each placed on a six-point scale from one to six. Each of
the seven components aggregated four to six questions that teachers rated on the six-point scale yielding a mean between 1.0 and 6.0 as well as an overall composite score on the same scale. Therefore, there were a total of eight scores: (a) prioritization of standards for a guaranteed and viable curriculum (Prioritize), (b) supportive structures and relationships to promote collaboration (Support), (c) assessment literacy and autonomy (AssessLit), (d) quality assessment design to inform instruction (Design), (e) analysis that leads to action (Analysis), (e) high impact instructional strategies: learning goals (LearnGoal), (f) high impact instruction: engagement around student discourse and distributed practice (Engage), and (h) the IIA Process composite score of the previous seven components (Survey). Cronbach’s alpha was conducted for each of the eight scores. The adjusted mean scores for each of the components was used as a primary variable of interest for research questions two and three. Scores were centered on the grand mean so that a zero exists and holds meaning for the analysis.

**M-STEP, PSAT, and SAT**

Since all teachers who had training started at different periods of time, full implementation of the training and accessibility to the interim assessments was first accomplished in 2017/18 for grades six through 11. Therefore, the most recent math achievement scores on the state tests from the spring of 2018 was used as the dependent variable for research question three, and the previous year math achievement scores for each student in 2017 was used as a covariate. The primary focus of the study was on the relationship the seven components may have had on student achievement over the course of one year to minimize the number of teachers the student experienced. The Michigan Department of Education reported the state level scores as scale scores using the M-STEP for grades five through eight, PSAT for grades nine through 10, and the SAT for grade 11. The department also reported the state mean
and standard deviation, therefore, all scores were converted to z-scores. Z-scores are a continuous variable with the majority of scores ranging from approximately -2.0 to 2.0. In addition to the most recent student scores, the previous year test scores were standardized (PreTz) to control for the previous achievement to have an indication of growth over time.

**Demographic Variables**

All student and building level demographic variables were dichotomous in nature. For gender, the variable named Female, assigning a value of one for females and coding males as zero. Social-economic status was reported by the Michigan Department of Education as Economically Disadvantaged (ED). Therefore, students qualifying as Economically Disadvantaged coded as one, others coded as zero. Likewise, Students with Disabilities (SWD) coded with a one for those who qualified as disabled and zero for those who did not. Finally, Asian/Caucasian students coded as zero, and NonCA students (Hispanic, African American, and other) were coded as one, as explained earlier in this chapter.

The grade or content level the teacher taught the most recent year was dichotomous such that one indicated the content of the course aligned well to the content alignment to the state test. For example, courses such as 6th through 8th grade, pre-algebra, and algebra I were covered well by the corresponding state test (MSTEP and PSAT). However, geometry, algebra II and pre-calculus/trigonometry appear on the PSAT/SAT but only as a small portion. Since the content from AP Statistics and AP Calculus was not measured by the PSAT/SAT, these courses will not be used in this study. Since the study revealed some teachers taught multiple courses that may have aligned or not aligned to the state test, the alignment of the course was moved to the student level as a covariant.

Level-2 teacher variables included “Years” as the number of years the teacher has taught
math as a continuous variable. The educational attainment was also dichotomous as teachers with merely a minor in mathematics were assigned zero and teachers with a major in mathematics or additional coursework were coded with one. Finally, teacher participation in part or all of the professional learning offered by local ISD were indexed based on the amount of time invested. Teachers who spent two years creating and piloting the interim assessments were assigned a one. Those who participated in the professional learning communities throughout the first year of implementation but did not invest time creating the assessments were indexed at 0.7. If teachers did not interact with the interim assessments in either of these ways but they had other training such as through the Delta Math program then the index was set at 0.3, whereas an index of zero given to those lacking any professional learning around assessment utilization.

**School Site, Teachers, and Students**

Each student at level one was linked to a teacher as a level 2 variable in the HLM model, and each teacher was linked to a school as a level 3 variable; however, all data was de-identified by the regional center. The focus of my study was on the instructional strategies at the teacher level. Research question 1 looked at how these instructional strategies vary across the region and the second research question investigated if professional learning had an impact on instruction. Research question three analyzed the relationship between the instructional strategies of the teacher and student achievement. In order to gauge the relationship between what a teacher does and the impact on student achievement, it was thought best to assure the teacher spent adequate time with the student. Unlike the elementary students who tend to spend the majority of the day with one teacher for the entire school year, secondary students rotate to different content teachers throughout the day. In addition, students often rotate classes at the end of the semester or trimester. This meant that some students had multiple teachers who instructed them on the
content of Algebra 1 or another math course. The issue of multiple memberships could be dealt with in a non-hierarchical methodology. However, many students completed the course with the teacher in a traditional hierarchical model by having only one teacher for the course. In other words, each student took a mathematical course with one teacher, and each teacher was associated with one school building. Among the 22,000 potential students, this study focused on the portion of those students who had the same teacher for the completion of a mathematical course, approximately 13,500 students. Each teacher linked to at least 10 students who had both a valid state test in both 2017 and 2018 to be included in the study, which resulted in 165 teachers. The achievement and demographics of the students who had primarily only one teacher will be compared with those that had multiple teachers to assure a random selection had occurred from each building. Therefore, Hierarchical Linear Models (HLM) was utilized for most of the research questions.

Table 5 connects the variables to my three research questions and the statistical analysis.
Table 5

Data Analysis Cross-Walk Table

<table>
<thead>
<tr>
<th>Research Questions</th>
<th>Measurement of Variables</th>
<th>Data Analysis Approach</th>
</tr>
</thead>
</table>
| 1. Within secondary math classrooms in a specific region of a Midwestern state, what were the levels of teacher implementation of seven research-based components regarding the utilization of interim instructional assessments within a collaborative environment, specifically:  
  i. prioritization of standards for a guaranteed and viable curriculum;  
  ii. supportive relationships and structure;  
  iii. assessment literacy and autonomy;  
  iv. quality interim assessment design;  
  v. analysis that leads to action planning;  
  vi. high impact instruction: learning goals; and  
  vii. high impact instruction: engagement around student discourse and distributed practice? | DV: Seven Components on a six-point rating scale (1 to 6) converted to contain zero, with four to six items per component:  
  i. Prioritize  
  ii. Support  
  iii. AssessLit  
  iv. Design  
  v. Analysis  
  vi. LearnGoal  
  vii. Engage and the total Survey score. | Descriptive statistics to examine the number of math teachers, aggregated mean teacher level scores for each of the seven components and total score, the standard deviation, and normality of the distributions |
| 2. To what degree were there differences in these components between teachers who participated to some degree in training around interim instructional assessments and those teachers who did not? To what extent did buildings differ based on their decision to send teachers to participate in some degree of training? | DV: Teacher average of the seven components from RQ#1 and the total Survey score  
IV: School participated in training (0 or 1). Teacher participated in training (index 0, 3, 7, or 1).  
CV: Teacher level of education (Degree), Years of teaching | HLM used with the teacher scores at level one along with teacher level covariates. Schools at level-2 with building level covariates. |
| 3. To what extent was there a relationship between the level of teacher implementation of these seven components involved in the utilization of interim instructional assessments and math achievement score improvement over time, controlling for certain student, teacher, and building factors (i.e., economically disadvantaged, ethnicity, English learners, special education, and pre-test)? | DV: Math achievement z-scores for the most recent school year.  
IV: Seven component scores and IIA Process composite score at the teacher level 2.  
CV: Math achievement z-score on a previous math test, student and school demographics (ED, EL, SWD, NonCA, Female), and teacher demographics (Years, Degree) at level 2. | HLM used with the student scores at level one, teacher factors at level two, along with building average component scores at level-3. Control for building level demographics and student level demographics. |

Key: DV = Dependent Variable, IV = Independent Variable, and CV = Covariant for control
Data Collection Procedures

In order to address the research questions for my study, I needed to collect data from two sources. The first data source came from the Michigan Department of Education in terms of math scores for individual students on the state assessment (M-STEP, PSAT, and SAT). Both building and student data were coded by the regional center in compliance with the Family Educational Rights and Privacy Act (1974) and the Human Subjects Institutional Review Board (HSIRB). Since students included in this study took different state assessments, the math achievement scores were standardized based on the state average. The difference was calculated between the state average scale score and each student scale score. Then the difference was divided by the state standard deviation to create individual $z$-scores. These individual math $z$-scores were used in the HLM models and also aggregated at the building level.

The second data source came from math teacher responses on a survey instrument regarding the seven components around the utilization of interim instructional assessments in a collaborative environment. Every teacher was given a randomly generated four-digit code by the regional center.

The regional center sought permission from local superintendents and curriculum directors to survey the secondary math teachers, obtain teacher-student data links, and certain demographic data from students, teachers, and buildings. Based on the potential benefits of aggregate information this study may produce, I anticipated the majority of schools would participate in the study, as secondary mathematics had been a popular topic of conversation recently. The regional center followed up with each school in the fall of 2018 to explain the benefits of the study to teachers and administrators. In addition, I reviewed the procedural protocols associated with the distribution and administration of the teacher survey.
**Data Analysis**

**Research Question 1**

The data collected for my study was analyzed by a variety of methods corresponding to each research question. In order to address the overall data collected for this study and to specifically address the first research question, descriptive statistics were utilized. The benefit of descriptive research is to describe, summarize, and interpret a mass amount of data (Glass & Hopkins, 1996). Therefore, I used histograms, scatterplots, and tables to better understand the survey, demographic, and achievement data. Central tendency, range, standard deviation, skewness, and kurtosis were examined for each of the primary variables of interest and the control variables.

The first research question sought to understand the levels of implementation of seven research-based components regarding the utilization of interim instructional assessments (IIA) within a collaborative environment. Once the survey results had been gathered, the data was cleaned by searching for missing data and outliers. Students who did not have a pre-test or other essential missing data were removed from the dataset. Outliers were investigated to determine if the data point was appropriate to retain or remove from the data set. Data was cleaned and analyzed using Microsoft Excel and IBM SPSS software. For this question, I aggregated the scores up to the building level as well as looked at the individual results for distribution. The aggregate at the building level compared central tendencies, skewness, and kurtosis.

**Research Question 2**

For the remaining two research questions, I used hierarchical linear models (HLM) as the most appropriate methodology for education when hierarchical structures exist from the student to teacher to building level and the inadequacy of traditional statistical techniques for modeling
Specifically, a 3-level HLM model was applied in which students represented the first level, teachers at the second level, and buildings at the third level, if appropriate. The use of traditional models to study hierarchical structures, such as multiple regression and ANCOVA, often have misleading results. Using HLM allowed for the separation of between-building and within-building error terms, which allows for interpretations of both the teacher effect and the building effect (Raudenbush & Bryk, 2002).

For research question number two, HLM was used to determine the degree of differences of the research-based components based on an index of participation for teachers who participated in training at some level around assessment utilization and those teachers who did not, presuming some difference exists. The null hypothesis was that the survey scores will not be significantly different between teachers nor between schools who participated in professional learning or not. The alternative hypothesis was there exists a significant difference between the survey scores between the index of those who participated in the professional learning and those who did not, when controlling for teacher factors in a nested environment. All calculations and analysis of the survey data utilized SPSS version 25 and HLM version 7. If differences existed, the model provided data both within a school building and between school buildings. Such an analysis will involve an unconditional model 2.0 without any predictors or covariates:

\[
IIA_{Process_{jk}} = \beta_{0k} + r_{jk}
\]

\[
\beta_{0k} = \gamma_{00} + u_{0k}
\]

where \(IIA_{Process_{jk}}\) was the composite survey result for all seven components for teacher \(j\) in building \(k\), \(\beta_{0k}\) was the mean survey score in building \(k\). The model has two random effects, the first component, \(r_{jk}\), which was the random effect associated with each teacher \(j\) in building \(k\) and the second random effect, \(u_{0k}\), was the random effect associated with the average survey
score for each school building $k$. The purpose of the unconditional model was to determine how much variance was at each level so that the full model can examine the variation at level 2. The unconditional model allowed for the calculation of the intra-class correlation (ICC) that measures the extent the variance of the outcome partitioned between the teacher level and the building level (Raudenbush & Bryk, 2002).

A 2-level hierarchical linear model was used to investigate the effect of that different levels of professional learning had on the composite score of the IIA Process from the teacher survey. The investigation into research question two used the following level-1 model:

$$IIA\_Process_{jk} = \beta_{0k} + \beta_{1k}(ProfLearn_{jk}) + \beta_{2k}(Years_{jk}) + \beta_{3k}(Degree_{jk}) + r_{jk}$$  \hspace{1cm} (2.1)

where $IIA\_Process_{jk}$ was the composite survey result for all seven components for teacher $j$ in building $k$, $\beta_{0k}$ was the adjusted mean survey score in building $k$ for teacher with a minor in mathematics who had taught an average number of years and did not experience any professional learning around assessment utilization. $\beta_{1k}$ was the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building $k$ and $\beta_{2k}$ to $\beta_{3k}$ are the covariate effects for building $k$. By grand centering $Years$ the teacher reported, the intercept $\beta_{0k}$ was appropriately interpreted for a teacher with the average number of years in the region. $Degree$ measured if teachers had a minor in mathematics or more extensive college course work in the content area. The level-1 model had one random effects component, $r_{jk}$, which was the random effect associated with each teacher $j$ in building $k$.

The building level or level-2 model was:

$$\beta_{0k} = \gamma_{00} + \gamma_{01}(BuildPD_{0k}) + \gamma_{02}(HS_{0k}) + \gamma_{03}(cED_{0k}) + \gamma_{04}(cNonCA_{0k}) + \gamma_{05}(cEL_{0k}) + \gamma_{06}(cSWD_{0k}) + u_{0k}$$  \hspace{1cm} (2.2)

$$\beta_{1k} = \gamma_{10}$$
\[ \beta_{2k} = \gamma_{20} \]
\[ \beta_{3k} = \gamma_{30} \]

where \( \gamma_{00} \) was the overall adjusted mean survey score, \( \gamma_{01} \) was the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and \( \gamma_{10} \) to \( \gamma_{30} \) are controlled covariate effects. The level-2 model had one random effect components; \( u_{0k} \) was the random effect associated with the average survey score for each school after controlling for the participation of some teachers in a professional learning series within a school. This model addressed research question 2 about whether the school participation had an effect on school-level teacher participation index after controlling for teachers’ professional characteristics and schools’ demographics and grade span. The school level demographics included the general grade span of the building to be a high school (HS), coded as one, or a middle school, coded as zero. Student level demographics for economically disadvantaged (ED), non-Caucasian and non-Asian (NonCA), English Learners (EL), students with disabilities (SWD), and gender (Female) were grand centered.

For research question 2, the IIA Process composite score was the primary variable of interest, however, each of the seven components reported as a subscale were analyzed as well, the models follow model 2.1 and 2.2 with the exception that the aggregate survey score was replaced with each component score. The models 2.3 through 2.9 are found in Appendix E. Where the second question specifically looked at the relationship with training and the IIA Process, research question three moves to the heart of my study by looking at the relationship with student achievement.
Research Question 3

For research question three, I not only examined the IIA Process composite score, but also examined each of the seven components to determine to what extent the level of implementation of each component might predict math achievement score improvement over time. Therefore, the dependent variable was the most recent math achievement on the state test as a z-score and the unconditional model for a 3-level hierarchical linear model follows:

\[
\begin{align*}
\text{MathZ}_{ijk} &= \pi_{0jk} + e_{ijk} \\
\pi_{0jk} &= \beta_{00k} + r_{0jk} \\
\beta_{00k} &= \gamma_{000} + u_{00k}
\end{align*}
\]

where \(\text{MathZ}_{ijk}\) was the standardized math z-score for a student \(i\) taught by teacher \(j\) in building \(k\), \(\pi_{0jk}\) was the mean math achievement of teacher \(j\) in building \(k\), \(\beta_{00k}\) was the mean math achievement in building \(k\), and \(\gamma_{000}\) was the grand mean for the state which was zero based on z-scores. The model contains three random effect components such that \(e_{ijk}\) was the random effect associated with each student \(i\) connected to teacher \(j\) in building \(k\), \(r_{0jk}\) was the random effect associated with the average math achievement score for each teacher \(j\), and \(u_{00k}\) was the random effect associated with the average math achievement score for building \(k\).

Since each of the seven components around interim assessments in a collaborative environment was evaluated in separate models in addition to the IIA Process composite score, there were eight full models to address the third research question. Model 3.1 shows a 3-level HLM analysis where the independent variable of interest was the composite score of the seven components at level 2; the other independent variables act as covariates:

Level 1:

\[
\begin{align*}
\text{MathZ}_{ijk} &= \pi_{0jk} + \pi_{1jk}(\text{PreT}_{z_{ijk}}) + \pi_{2jk}(\text{ED}_{ijk}) + \pi_{3jk}(\text{SWD}_{ijk}) + \\
&+ \pi_{4jk}(\text{PUD}_{ijk}) + \pi_{5jk}(\text{SA}_{ijk}) + \pi_{6jk}(\text{CLC}_{ijk}) + \pi_{7jk}(\text{MCA}_{ijk})
\end{align*}
\]
\[ \pi_{4jk}(EL_{ijk}) + \pi_{5jk}(NonCA_{ijk}) + \pi_{6jk}(Fem_{ijk}) + \pi_{7jk}(Course_{0jk}) + e_{ijk} \]

Level 2:
\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(IIA\_Process_{0jk}) + \beta_{02k}(Years_{0jk} - 1) + \beta_{03k}(Degree_{0jk}) + r_{0jk} \]
\[ \pi_{pjk} = \beta_{p0k} \text{ for all covariates where } p = 1 \text{ to } 7 \]

Level 3:
\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(cPreTz_{00k}) + \gamma_{002}(cED_{00k}) + \gamma_{003}(cSWD_{00k}) + \gamma_{004}(cEL_{00k}) + \gamma_{005}(cNonCA_{00k}) + \gamma_{006}(HS_{00k}) + u_{00k} \]
\[ \beta_{01k} = \gamma_{010} + u_{01k} \]
\[ \beta_{0qk} = \gamma_{0q0} \text{ for all covariates where } q = 2 \text{ to } 3 \]
\[ \beta_{p0k} = \gamma_{p00} \text{ for all covariates where } p = 1 \text{ to } 7 \]

where MathZ_{ijk} was the standardized math z-score for a student i taught by teacher j in building k, \( \pi_{0jk} \) was the adjusted mean math achievement of teacher j in building k for an Asian or Caucasian, male student with an average pretest, who does not qualify for disabilities, nor English Learner, nor economically disadvantaged. \( \beta_{00k} \) was the adjusted mean math achievement in building k for a first year teacher, with a minor in mathematics, teaching a course not covered well on the state test. \( \beta_{01k} \) was the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \( IIA\_Process_{0jk} \) was the survey composite score for all seven components for teacher j in building k, and \( \gamma_{000} \) was the adjusted mean for the average middle school building with average pre-test scores and an average percentage of NonCA students, English Learners, students with disabilities, and economically disadvantaged. The model contains covariates at the student level and grand centered at the building level that control for the student’s previous math score on a state test (PreTz), for students who qualify as economically disadvantaged (ED), students with disabilities (SWD),
English Learners (EL), female students (Female), and Non-Caucasian/Non-Asian students (NonCA). At the teacher level-2, the model controls for the number of years teaching (Years), whether the teacher has a minor or major in mathematics (Degree), and whether the majority of content of their course of instruction was covered on the state test (Course). The model also contains four random effect components such that $e_{ijk}$ was the random effect associated with each student $i$ connected to teacher $j$ in building $k$, $r_{0jk}$ was the random effect associated with the average math achievement score for each teacher $j$, $u_{00k}$ was the random effect associated with the average math achievement score for building $k$, and $u_{01k}$ was the random effect associated with the slope of the IIA Process composite score for building $k$.

In order to conserve space, Models 3.2 through 3.8 were moved to Appendix E. Each of the models used the same covariates and dependent variables as the 3-level hierarchical linear model of 3.1 with the exception of the primary variable of interest at teacher level-2, the IIA Process composite survey score. In Model 3.2, the primary variable of interest at teacher level-2 was the supportive relationships and structures ($Supportive_{ojk}$) for each teacher $j$ in building $k$. For Model 3.3, the variable of interest at level-2 was the knowledge of assessment literacy ($AssessLit_{ojk}$) for each teacher $j$ in building $k$. For Model 3.4, the variable of interest at level-2 was the quality of interim assessment design ($Design_{ojk}$) for each teacher $j$ in building $k$. For Model 3.5, the variable of interest at level-2 was the teacher analysis that leads to action planning ($Analysis_{ojk}$) for each teacher $j$ in building $k$. For Model 3.6, the variable of interest at level-2 was the high impact of instructional strategies around learning goals ($LearnGoal_{ojk}$) for each teacher $j$ in building $k$. For Model 3.7, the variable of interest at level-2 was another set of high impact instructional strategies engagement around discourse and practice, ($Engage_{ojk}$) for each teacher $j$ in building $k$. Finally, Model 3.8, used the component score for prioritization of
standards for a guaranteed and viable curriculum as the variable of interest at level-2

\( P_{\text{Prioritize}_{jk}} \) for each teacher \( j \) in building \( k \).

**Limitations and Delimitations**

A nonexperimental ex-post facto design has notable limitations, such as the lack of random assignment of a treatment variable. Therefore, the components tested predictability and cannot make a leap to causality. The study was limited to one geographical location since the aim was to link student data to teachers and teachers to buildings. Since the survey was given to a relatively small number of schools based on convenience sampling, generalizations are limited to the region with similar demographics. In 2017/18, the state of Michigan tested approximately 699,000 secondary students whereas my study used a specific region that tested approximately 22,500 secondary students or 3.2% of the state. The specific region in this study had fewer students who qualified for economically disadvantaged, more Hispanic students, and fewer African American students as sited in the population section. As the data hubs in Michigan develop, additional studies may be able to research a larger population that might include the entire state.

**Chapter 3 Closure**

The applied methodology and research design was a robust analytical design to address my study’s preeminent question: is there a relationship between student math achievement scores and teachers utilizing the seven components around interim assessments in a collaborative environment? I was able to apply an HLM model with students nested in teacher classrooms who were nested in school buildings. Using a 3-level model, I was able to separate the random effects between teachers and between buildings. It seemed rare to find a study with access to 22,000 students that might link up to 170 teachers across 38 buildings. Though the study has the
potential for a larger sample size, the results produced in chapter 4 might be intriguing depending on the components that demonstrate significance.

Many of these schools allowed individual teachers to participate in a series of training around the creation and utilization of interim instructional assessments. Those who participated heard the inspiring stories of individual schools that utilized interim assessments within a collaborative culture to improve math achievement on high-stakes tests. Robert Meehan once said “the most valuable resource that all teachers have is each other. Without collaboration, our growth is limited to our own perspectives.” This group of teachers were supported by their administration to attend the training and were committed to learning the process of building and utilizing interim instructional assessment for secondary math. The incentive for teachers to participate in this study was the assurance that what might be learned about improving teaching and learning for secondary mathematics would be shared across the region.
CHAPTER 4

RESULTS

Chapter 4 provides the results of the analysis described in the previous chapter to address the three research questions of this study. Research question one called for descriptive statistics to examine the teacher survey results. Research questions two and three required the utilization of descriptive statistics to check for assumptions associated with the hierarchical linear models (HLM) and then a thorough analysis of the HLM models. This chapter is divided into three primary sections: (a) descriptive statistics related to school-level data, teacher survey data, and student-level data, (b) an HLM analysis of teachers who participated in training, and (c) an HLM analysis of the relationship between the teacher survey data and math achievement from one year to the next.

Research Question 1: Descriptive Statistics

This first section covers descriptive statistics to respond to research question one and the assumptions of research questions two and three.

Teacher Level Statistics

The purpose of my study was to ascertain the extent to which math teachers utilized an interim instructional assessment (IIA) process within middle and high schools, and how such utilization levels connected to student achievement. The first research question examined the different levels of research-based components related to teacher utilization of interim instructional assessments in a collaborative environment across all schools within a specific region using descriptive statistics. Specifically, research question one asked:

1. Within secondary math classrooms in a specific region of a Midwestern state, what were the levels of teacher implementation of seven research-based components regarding the
utilization of interim instructional assessments within a collaborative environment, specifically:

i. prioritization of standards for a guaranteed and viable curriculum;

ii. supportive relationships and structures;

iii. assessment literacy and autonomy;

iv. quality interim assessment design;

v. an analysis that leads to action planning;

vi. high impact instruction: learning goals; and

vii. high impact instruction: engagement around discourse and practice?

The survey tool measured the seven components above and was offered to 179 secondary math teachers, of which 172 teachers completed the survey and consented to the study. However, seven teachers who completed the survey did not have a sufficient number of students, at least 10, who qualified for the study. In order for a student to qualify, they must have spent a majority of the school days with the teacher of record and taken both the 2017 and 2018 state assessment with a valid math score. Therefore, 165 teachers had sufficient data to be used in this study.

Teachers responded to each of the seven components by completing a self-rating on a six-point Likert scale. For the first three components, the scale ranged from 1-Strongly disagree to 6-Strongly agree, without the option of neutral. Table 6 shows the five items for component (i) Prioritization of Standards for a Guaranteed and Viable Curriculum, along with the descriptive statistics. The average score for all 37 items was 4.18 and Table 6 shows that each of the five items for component (i) were relatively high compared to the overall average. Most of the items had an average near 5.0 with a standard deviation close to 1.0; therefore, about 84% of the responses fell between 4-slightly agree and 6-strongly agree, which is also seen by the
distribution of percentages in Table 6. The highest item focused on the identification of essential knowledge and skills (5.36, s.d. 0.92) and the lowest item was based on the distinction between teaching those essential skills to mastery whereas only exposing students to other standards.

Table 6

Component (i): Prioritization of Standards

<table>
<thead>
<tr>
<th>Component (i)</th>
<th>1- Strongly Disagree</th>
<th>2- Moderately Disagree</th>
<th>3- Slightly Disagree</th>
<th>4- Slightly Agree</th>
<th>5- Moderately Agree</th>
<th>6- Strongly Agree</th>
<th>% (n)</th>
<th>% (n)</th>
<th>% (n)</th>
<th>% (n)</th>
<th>% (n)</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clearly identified essential knowledge and skills</td>
<td>1.2 (2)</td>
<td>0.6 (1)</td>
<td>3.0 (5)</td>
<td>5.4 (9)</td>
<td>35.2 (58)</td>
<td>54.5 (90)</td>
<td>5.36</td>
<td>0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently discuss content that hinders learning</td>
<td>0.6 (1)</td>
<td>2.4 (4)</td>
<td>5.4 (9)</td>
<td>20.0 (33)</td>
<td>34.5 (57)</td>
<td>37.0 (61)</td>
<td>4.96</td>
<td>1.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ample opportunity to learn all essentials</td>
<td>0.6 (1)</td>
<td>4.2 (7)</td>
<td>5.4 (9)</td>
<td>15.2 (25)</td>
<td>43.6 (72)</td>
<td>30.9 (51)</td>
<td>4.90</td>
<td>1.07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depth and breadth of standards addressed</td>
<td>2.4 (4)</td>
<td>4.2 (7)</td>
<td>7.3 (12)</td>
<td>15.8 (26)</td>
<td>48.5 (80)</td>
<td>21.8 (36)</td>
<td>4.69</td>
<td>1.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clear distinction between mastery and exposure</td>
<td>1.8 (3)</td>
<td>7.3 (12)</td>
<td>11.5 (19)</td>
<td>26.1 (43)</td>
<td>38.2 (63)</td>
<td>15.2 (25)</td>
<td>4.37</td>
<td>1.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: survey items re-ordered to be from highest to lowest mean.

Table 7 shows the six items for component (ii) Supportive Structures and Relationships, along with the descriptive statistics. Most of the items had a relatively high score between 4.47 and 5.21. The highest items focused on the perception that the math department had a unified and sustained commitment to improving student learning, felt comfortable taking risks, and felt supported in data dialogues with colleagues. Therefore, supportive relationships rated relatively high. Whereas only the proximity of classrooms rated relatively high for supportive structures. On the contrary, other structures such as the school schedule and the school setting aside
adequate time for collaborative work did not rate above the overall average of 4.18.

Table 7

**Component (ii): Supportive Structures and Relationships**

<table>
<thead>
<tr>
<th>Component (ii)</th>
<th>Supportive Structures and Relationships within the Math Department</th>
<th>1- Strongly Disagree</th>
<th>2- Moderately Disagree</th>
<th>3- Slightly Disagree</th>
<th>4- Slightly Agree</th>
<th>5- Moderately Agree</th>
<th>6- Strongly Agree</th>
<th>% (n)</th>
<th>% (n)</th>
<th>% (n)</th>
<th>% (n)</th>
<th>% (n)</th>
<th>% (n)</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unified and sustained commitment to learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Safe culture to take risks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proximity allowed for ease in collaborating</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vulnerable and supportive data dialogues</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adequate time for collaborative work</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School schedule promoted collective learning</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: survey items re-ordered to be from highest to lowest mean.

Table 8 shows the four items for component (iii) Autonomy and Assessment Literacy, along with the descriptive statistics. Only one item had a relatively high score of 4.87, which shows that 90.9% of teachers agreed at some level that students are allowed to make up points on an interim assessment. This may suggest a shift in practice where it was not uncommon for teachers in the past to move on after a chapter test without the opportunity for students to re-take a portion of the test. The other two items around assessment literacy were not rated as high, where 34.5% of teachers disagreed at some level that more time was spent on interim assessments rather than summative assessments. The lowest item was the feeling that teachers
had the autonomy to personalize any training around assessment utilization and data dialogues to their course. This suggests that over one-third of our teachers feel they only receive a “one-size-fits-all” type professional learning or maybe none at all.

Table 8

*Component (iii): Autonomy and Assessment Literacy*

<table>
<thead>
<tr>
<th>Component (iii)</th>
<th>1-Strongly Disagree</th>
<th>2-Moderately Disagree</th>
<th>3-Slightly Disagree</th>
<th>4-Slightly Agree</th>
<th>5-Moderately Agree</th>
<th>6-Strongly Agree</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students allowed to make up points on interim assessments</td>
<td>2.4 (4)</td>
<td>1.2 (2)</td>
<td>5.4 (9)</td>
<td>21.2 (35)</td>
<td>37.6 (62)</td>
<td>32.1 (53)</td>
<td>4.87</td>
<td>1.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>More time spent on interim rather than summative assessments</td>
<td>4.8 (8)</td>
<td>7.3 (12)</td>
<td>22.4 (37)</td>
<td>21.8 (36)</td>
<td>26.1 (43)</td>
<td>17.6 (29)</td>
<td>4.10</td>
<td>1.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shared practices around formative assessment use</td>
<td>3.0 (5)</td>
<td>8.5 (14)</td>
<td>18.2 (30)</td>
<td>27.9 (46)</td>
<td>33.3 (55)</td>
<td>9.1 (15)</td>
<td>4.07</td>
<td>1.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autonomy to personalize training around assessments</td>
<td>5.4 (9)</td>
<td>12.1 (20)</td>
<td>20.1 (34)</td>
<td>32.1 (53)</td>
<td>21.2 (35)</td>
<td>8.5 (14)</td>
<td>3.77</td>
<td>1.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: survey items re-ordered to be from highest to lowest mean.

For the final four components, the scale was changed, along with the appearance of the survey, to measure the level of expertise occurring in each teacher’s classroom. The scale ranged from 1-No Expertise at All to 6-Regional Expert. The four components based on this scale seemed to have very few teachers who felt comfortable rating themselves at one or six. Table 9 shows the five items for component (iv) Quality Assessment Design for Instruction, along with the descriptive statistics. Most of the items had a very similar distribution and mean score, with the means ranging from 4.18 to 4.41. This indicates that all five items were at the overall average of 4.18 or higher. Teachers level of expertise for varying both the complexity and
difficulty of assessment items had the tightest distribution (s.d. = 0.77) and shows that 87.3% of teachers feel competent or highly proficient. Note the items below do not distinguish between selected response items and performance based assessment items.

Table 9

*Component (iv): Quality Assessment Design*

<table>
<thead>
<tr>
<th>Component (iv)</th>
<th>Quality Assessment Design for Instruction at the Classroom Level</th>
<th>1- No expertise at all</th>
<th>2- Novice</th>
<th>3- Advanced beginner</th>
<th>4- Competent</th>
<th>5- Highly proficient</th>
<th>6- Regional expert</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timely feedback based on interim assessments</td>
<td>% (n)</td>
<td>% (n)</td>
<td>% (n)</td>
<td>% (n)</td>
<td>% (n)</td>
<td>% (n)</td>
<td></td>
<td>4.41</td>
<td>0.96</td>
</tr>
<tr>
<td>Success criteria for every motivated student established</td>
<td>0.6 (1)</td>
<td>3.6 (6)</td>
<td>12.1 (20)</td>
<td>27.9 (46)</td>
<td>49.1 (81)</td>
<td>6.7 (11)</td>
<td></td>
<td>4.35</td>
<td>0.84</td>
</tr>
<tr>
<td>Sufficient evidence on interim assessments</td>
<td>0 (0)</td>
<td>4.2 (7)</td>
<td>6.1 (10)</td>
<td>39.4 (65)</td>
<td>48.5 (80)</td>
<td>1.2 (2)</td>
<td></td>
<td>4.35</td>
<td>0.81</td>
</tr>
<tr>
<td>Ensuring interim items vary in complexity and difficulty</td>
<td>0 (0)</td>
<td>3.0 (5)</td>
<td>6.7 (11)</td>
<td>49.7 (82)</td>
<td>37.6 (62)</td>
<td>3.0 (5)</td>
<td></td>
<td>4.31</td>
<td>0.77</td>
</tr>
<tr>
<td>Ensuring interim items align with state content and rigor</td>
<td>1.2 (2)</td>
<td>6.1 (10)</td>
<td>14.5 (24)</td>
<td>35.2 (58)</td>
<td>38.2 (63)</td>
<td>4.8 (8)</td>
<td></td>
<td>4.18</td>
<td>1.02</td>
</tr>
</tbody>
</table>

Note: survey items re-ordered to be from highest to lowest mean.

Table 10 shows the six items for component (v) Analysis that Leads to Action, along with the descriptive statistics. Only two items had a mean score higher than the overall mean of 4.18. Most of the items had a relatively lower score, with some concern that 24.3% of teachers do not feel competent at providing additional time for struggling students nor 36.3% did not feel competent providing other accommodations for struggling learners. The lowest item focused on the expertise of teachers frequently leading data analysis dialogues around interim assessments, where only 51.6% felt at least competent and only one of 165 teachers felt as though they were a
regional expert. Furthermore, 20 teachers felt as though they had no expertise at all to lead a data dialogue, often referred to as a Professional Learning Community.

Table 10

*Component (v): Analysis that Leads to Action*

<table>
<thead>
<tr>
<th>Component (v) Analysis that Leads to Action at the Classroom Level</th>
<th>1- No expertise at all</th>
<th>2- Novice</th>
<th>3- Advanced beginner</th>
<th>4- Competent</th>
<th>5- Highly proficient</th>
<th>6- Regional expert</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td>%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Determine why students are struggling on assessments</td>
<td>1.2</td>
<td>4.2</td>
<td>9.1</td>
<td>41.8</td>
<td>41.8</td>
<td>1.8</td>
<td>4.24</td>
<td>0.90</td>
</tr>
<tr>
<td>(2)</td>
<td>(7)</td>
<td>(15)</td>
<td>(69)</td>
<td>(69)</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interim assessments used to identify struggling students</td>
<td>0.6</td>
<td>3.6</td>
<td>14.5</td>
<td>38.2</td>
<td>41.8</td>
<td>1.2</td>
<td>4.21</td>
<td>0.88</td>
</tr>
<tr>
<td>(1)</td>
<td>(6)</td>
<td>(24)</td>
<td>(63)</td>
<td>(69)</td>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment data builds on strengths and weaknesses</td>
<td>0</td>
<td>6.1</td>
<td>13.9</td>
<td>41.8</td>
<td>36.4</td>
<td>1.8</td>
<td>4.14</td>
<td>0.90</td>
</tr>
<tr>
<td>(0)</td>
<td>(10)</td>
<td>(24)</td>
<td>(69)</td>
<td>(60)</td>
<td>(3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide additional time for struggling students</td>
<td>1.2</td>
<td>7.3</td>
<td>15.8</td>
<td>43.0</td>
<td>29.7</td>
<td>3.0</td>
<td>4.02</td>
<td>0.99</td>
</tr>
<tr>
<td>(2)</td>
<td>(12)</td>
<td>(26)</td>
<td>(71)</td>
<td>(49)</td>
<td>(5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Provide struggling students with other accommodations</td>
<td>0.6</td>
<td>12.7</td>
<td>23.0</td>
<td>38.8</td>
<td>23.6</td>
<td>1.2</td>
<td>3.76</td>
<td>1.01</td>
</tr>
<tr>
<td>(1)</td>
<td>(21)</td>
<td>(38)</td>
<td>(64)</td>
<td>(39)</td>
<td>(2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequently leading analysis of interim assessments</td>
<td>12.1</td>
<td>20.1</td>
<td>15.8</td>
<td>35.2</td>
<td>15.8</td>
<td>0.6</td>
<td>3.24</td>
<td>1.30</td>
</tr>
<tr>
<td>(20)</td>
<td>(34)</td>
<td>(26)</td>
<td>(58)</td>
<td>(26)</td>
<td>(1)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: survey items re-ordered to be from highest to lowest mean.

Table 11 shows the six items for component (vi) Instructional Goals and Feedback, along with the descriptive statistics. There were three items the scored higher than the overall mean, with the highest two indicators beginning each unit with clear learning targets and linking the previous unit with the next unit. Teachers also felt competent in providing timely feedback on student progress toward learning goals. However, teachers did not feel nearly as competent in leading their students to track their own progress towards learning goals that is one of the highest indicators leading to student achievement according to Hattie (2014) as discussed in Chapter 2.
The lowest scoring item on the entire survey was found in Table 11 based on the expertise of teachers to involve students in complex projects that address the learning goals. This may suggest fewer performance based assessments that were not addressed specifically in the survey.

Table 11

Component (vi): Instructional Goals and Feedback

<table>
<thead>
<tr>
<th>Component (vi) Instructional Goals and Feedback at the Classroom Level</th>
<th>1- No expertise at all</th>
<th>2- Novice</th>
<th>3- Advanced beginner</th>
<th>4- Competent</th>
<th>5- Highly proficient</th>
<th>6- Regional expert</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begin each unit with clear learning targets</td>
<td>1.8 (3)</td>
<td>2.4 (4)</td>
<td>6.7 (11)</td>
<td>38.8 (64)</td>
<td>45.5 (75)</td>
<td>4.8 (8)</td>
<td>4.38</td>
<td>0.91</td>
</tr>
<tr>
<td>Next unit based on foundations of the previous unit</td>
<td>0 (0)</td>
<td>3.6 (6)</td>
<td>9.7 (16)</td>
<td>45.5 (75)</td>
<td>39.4 (65)</td>
<td>1.8 (3)</td>
<td>4.26</td>
<td>0.80</td>
</tr>
<tr>
<td>Timely feedback on student progress toward learning goals</td>
<td>0.6 (1)</td>
<td>5.4 (9)</td>
<td>9.7 (16)</td>
<td>42.4 (70)</td>
<td>38.8 (64)</td>
<td>3.0 (5)</td>
<td>4.22</td>
<td>0.91</td>
</tr>
<tr>
<td>Ending each unit with feedback on learning goals</td>
<td>2.4 (4)</td>
<td>12.7 (21)</td>
<td>23.6 (39)</td>
<td>37.6 (62)</td>
<td>23.0 (38)</td>
<td>0.6 (1)</td>
<td>3.68</td>
<td>1.06</td>
</tr>
<tr>
<td>Students track their own progress of learning goals</td>
<td>4.8 (8)</td>
<td>17.6 (29)</td>
<td>22.4 (37)</td>
<td>40.0 (66)</td>
<td>13.9 (23)</td>
<td>1.2 (2)</td>
<td>3.44</td>
<td>1.12</td>
</tr>
<tr>
<td>Students involved in complex projects addressing the learning goals</td>
<td>13.9 (23)</td>
<td>0 (0)</td>
<td>33.9 (56)</td>
<td>20.6 (34)</td>
<td>7.9 (13)</td>
<td>0 (0)</td>
<td>2.78</td>
<td>1.19</td>
</tr>
</tbody>
</table>

Note: survey items re-ordered to be from highest to lowest mean.

Table 12 shows the six items for component (vii) Engagement around Student Discourse and Distributed Practice, along with the descriptive statistics. All of the items were relatively low with means ranging from 3.48 to 3.92. Ironically, the lowest item was around teacher expertise in using at least 25% of class time for student discourse and the highest items dealt with the confidence in balancing direct instruction and student talk. Lower scores were seen
with distributive practice as well, with both multiple exposures and student practice over several days scoring 3.85 and 3.83 respectively.

Table 12

*Component (vii): Student Discourse and Distributed Practice*

<table>
<thead>
<tr>
<th>Component (vii)</th>
<th>1- No expertise at all</th>
<th>2- Novice</th>
<th>3- Advanced beginner</th>
<th>4- Competent</th>
<th>5- Highly proficient</th>
<th>6- Regional expert</th>
<th>Mean</th>
<th>S.D.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confident balance between direct instruction and student talk</td>
<td>0 (0)</td>
<td>9.1 (15)</td>
<td>24.2 (40)</td>
<td>33.9 (56)</td>
<td>30.9 (51)</td>
<td>1.8 (3)</td>
<td>3.92</td>
<td>0.99</td>
</tr>
<tr>
<td>Multiple exposures to an idea over several days</td>
<td>1.8 (3)</td>
<td>7.9 (13)</td>
<td>24.2 (40)</td>
<td>37.0 (61)</td>
<td>27.3 (45)</td>
<td>1.8 (3)</td>
<td>3.85</td>
<td>1.03</td>
</tr>
<tr>
<td>Students practice a few problems over several days</td>
<td>3.0 (5)</td>
<td>12.1 (20)</td>
<td>18.8 (31)</td>
<td>36.4 (60)</td>
<td>24.2 (44)</td>
<td>3.0 (5)</td>
<td>3.81</td>
<td>1.14</td>
</tr>
<tr>
<td>Students exchange ideas, agree/disagree, including ways of representation</td>
<td>2.4 (4)</td>
<td>16.4 (27)</td>
<td>25.5 (42)</td>
<td>34.5 (57)</td>
<td>20.6 (34)</td>
<td>0.6 (1)</td>
<td>3.56</td>
<td>1.08</td>
</tr>
<tr>
<td>At least 25% of class time for student discourse</td>
<td>3.0 (5)</td>
<td>22.4 (37)</td>
<td>21.8 (36)</td>
<td>30.9 (51)</td>
<td>19.4 (32)</td>
<td>2.4 (4)</td>
<td>3.48</td>
<td>1.19</td>
</tr>
</tbody>
</table>

Note: survey items re-ordered to be from highest to lowest mean.

With some variance with relatively lower items within components (ii), (iii), (v), and (vi), it was important to test for the internal consistency. Each of the seven components of the Interim Instructional Assessment (IIA) Process contained four to six statements that were tested for internal consistency using Cronbach’s alpha. The number of items within each component had internal consistency as noted by Cronbach’s Alpha in Table 13 and aggregate values seem to have a nice range on the one to six Likert scales. As seen in Table 13, Cronbach’s Alpha was
acceptable for all seven components with only component (iii) having a relatively low score of 0.58. Two factors likely contribute to this lower, yet acceptable score; there were only four items and one relatively low score based on teachers feeling a sense of autonomy around professional learning.

Table 13

Survey Reliability Summary by Component and IIA Process Composite

<table>
<thead>
<tr>
<th>Survey Component</th>
<th>Number of items</th>
<th>Cronbach’s Alpha</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i) Prioritization</td>
<td>5</td>
<td>0.77</td>
<td>4.86</td>
<td>1.4 – 6.0</td>
</tr>
<tr>
<td>(ii) Supportive</td>
<td>6</td>
<td>0.75</td>
<td>4.52</td>
<td>1.7 – 6.0</td>
</tr>
<tr>
<td>(iii) Assessment</td>
<td>4</td>
<td>0.58</td>
<td>4.20</td>
<td>1.8 – 5.8</td>
</tr>
<tr>
<td>(iv) Design</td>
<td>5</td>
<td>0.76</td>
<td>4.32</td>
<td>2.4 – 5.8</td>
</tr>
<tr>
<td>(v) Analysis</td>
<td>6</td>
<td>0.83</td>
<td>3.93</td>
<td>2.2 – 5.7</td>
</tr>
<tr>
<td>(vi) Learning Goal</td>
<td>6</td>
<td>0.85</td>
<td>3.79</td>
<td>1.7 – 5.5</td>
</tr>
<tr>
<td>(vii) Engagement</td>
<td>5</td>
<td>0.87</td>
<td>3.73</td>
<td>1.6 – 5.6</td>
</tr>
<tr>
<td>IIA Process Composite</td>
<td>37</td>
<td>0.91</td>
<td>4.18</td>
<td>2.8 – 5.4</td>
</tr>
</tbody>
</table>

Based on 165 teacher surveys

The IIA Process composite score had a reasonably normal distribution with expected values near zero for kurtosis (0.028) and skewness (-0.373). In fact, six of the seven components had a small kurtosis between negative one and one, as well as a small skewness between zero and a negative one. The only component to differ seems to be negatively skewed (-1.223) and slightly leptokurtic based on the kurtosis (2.065) (Lomax & Hahs-Vaughn, 2012). A visual inspection of each histogram, Q-Q plot, and box plots showed that each of the seven components and the composite score was approximately normally distributed. A probable reason for the
skewed component of prioritization was due to one low outlier (ID #8782) with an aggregate score of 1.4 when the component had the highest mean of 4.86. The mean component value ranged from the lowest score of 3.73 for engagement of students and the highest aggregate score of 4.86 for the prioritization of standards. All 165 teachers were included, as outliers were not removed, after examining each component score and the IIA Process composite score to confirm valid responses.

Therefore, the levels of teacher implementation of the seven research-based components regarding the utilization of interim instructional assessments within a collaborative environment varied across the six-point Likert scale. Teachers responded more confidently with the first four components regarding the prioritization of standards, supportive structures, assessment literacy and assessment design with mean scores ranging from 4.2 to 4.9. Teachers were less confident with the final three components involving the analysis of assessment data and specific instructional strategies involving the learning goals and student engagement around student discourse and distributive practice with scores ranging from 3.7 to 3.9.

In addition to the teacher survey data, certain teacher demographics were also collected that were considered for other research questions, primarily, number of years teaching mathematics, major in mathematics, and level of professional learning around interim instructional assessments. Teachers ranged from one year to 42 years of teaching mathematics with the average number of 14.7 years (SD = 8.81) as a continuous variable. The majority of teachers had at least a major in mathematics, with only 15% only having a minor with no additional course work in mathematics, this variable was treated as dichotomous. Finally, the level of professional learning was given a point value to represent four different exposures to training around interim instructional assessments. Table 14 describes each of the four different
levels, the point value assigned for computational reasons, and distribution of teachers.

Approximately one-third of the teachers surveyed had some training from the regional center and the vast majority did not. However, 15.8% did have extensive training that did not involve the regional center. The fact that 52.7% of teachers did not have training around assessments and data analysis may have contributed to the lower score from item (iii-D) found Table 10.

Table 14

**Professional Learning Levels and Distributions**

<table>
<thead>
<tr>
<th>Description of Professional Learning</th>
<th>Point Value</th>
<th>Number of Teachers</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participated in the creation of HS or MS math interim assessments at the ISD</td>
<td>1.0</td>
<td>42</td>
<td>25.5%</td>
</tr>
<tr>
<td>Participated in the PLC meetings at the ISD and did not create interim assessments</td>
<td>0.7</td>
<td>10</td>
<td>6.1%</td>
</tr>
<tr>
<td>Extensive training in PLC or assessment utilization but not involved with ISD</td>
<td>0.3</td>
<td>26</td>
<td>15.8%</td>
</tr>
<tr>
<td>No extensive training around assessment utilization or PLCs in recent years</td>
<td>0.0</td>
<td>87</td>
<td>52.7%</td>
</tr>
</tbody>
</table>

n = 165 teachers

**Building Level Statistics**

The 165 teachers belonged to secondary school buildings across one particular region in the mid-west. Thirty-eight middle and high schools were invited to participate in this research study, which included 14 traditional high schools, 16 traditional middle schools, and eight public school academies (PSA). Three of the PSAs chose not to participate. Therefore, there were a total of 35 schools studied in the region. High schools represented 46% of schools, and 14% were either a middle school or high school PSA. The average school building had 34% (SD = 0.16) of students who qualified as economically disadvantaged (ED) and schools ranged from
15% to 71% of students qualifying ED. In addition, the average school had 7% (SD = 0.04) of students with disabilities (SWD), 3% (SD = 0.04) of English Learners (EL), and 19% (SD = 0.15) who neither Asian nor Caucasian (NonCA). There were two schools that only had one teacher respond to the survey: one PSA and one traditional middle school. The average number of teachers surveyed per school was 4.7 (SD = 2.50) with the largest school containing 13 teachers who participated in the survey.

**Student Level Statistics**

Students ranged from 6th grade to 11th grade in the 2017-2018 school year for this research study. Each of the 13,494 students spent a majority of the school year with one of the 165 teachers who completed a survey around interim instructional assessment within a collaborative environment. In addition, the students had a valid state assessment score (M-STEP or PSAT) from the previous school year, 2016-2017, and a valid state assessment score (M-STEP, PSAT, or SAT) from the 2017-18 school year. Students with missing data were excluded. The state assessment scores were standardized as z-scores using the state mean and standard deviation within the respective grade. Table 15 shows the descriptive statistics of the student level variables used within this study. Both the pre-test and post-test were normally distributed as continuous variables and all other variables were dichotomous. The 33% of students who qualified as economically disadvantaged because the student qualifies as free or reduced lunch based on poverty level or qualification for programs such as migrant education, homeless assistance, foster care, temporary assistance to needy families (TANF), or supplemental nutrition assistance program (SNAP). Students with Disabilities (SWD) represent students who have an Individual Education Plan (IEP) and are assigned the standard state assessment rather than an alternative assessment, which accounts for the 7% in this study.
### Table 15

**Student Level Descriptive Statistics**

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Mean z-score or % of Students</th>
<th>Range of z-scores</th>
<th>Standard Deviation</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math Pre-test</td>
<td>0.58</td>
<td>-3.6 – 4.8</td>
<td>0.93</td>
<td>-0.16</td>
<td>-0.18</td>
</tr>
<tr>
<td>Math Post-test</td>
<td>0.34</td>
<td>-3.1 – 3.2</td>
<td>0.88</td>
<td>-0.010</td>
<td>-0.10</td>
</tr>
<tr>
<td>Econ. Disadvantaged</td>
<td>33%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student w Disability</td>
<td>7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English Learner</td>
<td>4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Asian or White</td>
<td>20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>49%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Aligned</td>
<td>64%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

n = 13,494 students

In order to treat ethnicity as a dichotomous variable, Asian students and Caucasian students were grouped together and all others were assigned a status of NonCA, which constituted 20% of students. The course that each student was enrolled with the teacher of record was coded as having a strong alignment (Test Aligned) to the state test for example, most middle school courses aligned with the M-STEP and Algebra 1 aligned with the PSAT. As Table 15 illustrates, the majority of students (64%) were enrolled in a math course that tightly aligned with the state test. Finally, the region did perform collectively above the state average with a mean standardized math score of 0.34 in 2018 and 0.58 in 2017. Test alignment might attribute to the slight difference in standardized scores, as there exists a stronger alignment in 2017 using 5th through 10th-grade scores compared to 2018 using 6th through 11th-grade scores. In other words, the course content does not align as well to the state assessment in higher grades. Overall, this study examined 13,494 students nested within 165 teachers contained within 35 schools.
Research Question 2: Training across the Region Related to Survey Results

Another purpose of my study was to ascertain the role of training with teachers to utilize interim instructional assessments had on the levels of utilization within these schools. The next question groups the schools by those who participated to some degree in regional teacher training and those who did not, to see if the levels of utilization differ between the groups, specifically:

2. To what degree are there differences in these components between teachers who participated to some degree in training around interim instructional assessments and those teachers who did not? To what extent did buildings differ based on their decision to send teachers to participate in some degree of training?

Teachers self-reported their participation with professional learning based on four levels as illustrated in Table 14 earlier in this chapter. In summary, point values were assigned to teachers who created regional interim assessments (1.0), teachers who participated in regional data analysis (0.7), teachers who had some other training around data utilization (0.3), and teachers who did not have any recent training. The majority of schools, 57.1%, sent at least one teacher to the professional learning around the creation and utilization of interim instructional assessments and an additional 14.3% of schools participated in the regional data analysis. Therefore, 28.6% of schools in this study did not send any teachers to regional training around the utilization of interim instructional assessments in a collaborative environment.

The hierarchical linear model for this research question had two levels: the teacher level and the building level. As described in Chapter 3, the variable of interest was the level of professional development, and there may exist potential covariants at both the teacher level and school level. In order to assess the impact of professional learning, a baseline was established using the unconditional model, and then the covariants were added to create a parsimonious
model.

**Unconditional Model HLM Level 2**

The unconditional model (2.0) found in Chapter 3 provided the baseline for subsequent models when adding both covariants and the survey variables of interest. Based on the unconditional model, the greatest amount of variance occurred at level one as expected. Table 16 summarizes the analysis results of the unconditional two-level HLM.

**Table 16**

*Two Level HLM Unconditional Model Results*

<table>
<thead>
<tr>
<th>Description</th>
<th>Variance Component</th>
<th>Unconditional ICC</th>
<th>Reliability estimate</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1, $\sigma^2$ (teachers)</td>
<td>0.2236</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2, $\tau_\beta$ (schools)</td>
<td>0.0496</td>
<td>18.22%</td>
<td>0.479</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Level 1 n = 165; Level 2 n = 35

The reliability at level two was relatively low, possibly due to the average school having fewer than five teachers who took the survey and seven buildings having only one or two teachers who were surveyed. As the table above shows, the level-two variance component was significant, whereas, over 81% of the variation occurs between teachers and less than 19% occurs between schools. Some of this variation might be explained by the covariants within this study and the levels of professional learning around the seven research-based components regarding the utilization of interim instructional assessments within a collaborative environment.

**Parsimonious Covariant Model**

Prior to adding the IIA Process composite as the variable of interest, two blocks of models were analyzed to determine the relationship with both teacher level covariants and school
level covariants. When the two covariants at level two were added, both number of years teaching math \((p > .1)\) and the teacher’s level of education at least has a major in mathematics \((p > .5)\) were not significant. Likewise, all of the level three covariants were not significant with the exception of NonCA \((p < .01)\). In fact, one of the level three variables of interest, the school effect of sending teachers to professional learning, had a \(p\)-value greater than .5. In addition, the variance-covariance components test was conducted \((\chi^2 = 1.5938, p > .5)\) comparing the fullest model to the suggested parsimonious model and concluded that the simpler model was more appropriate. The majority of schools (29 out of 35) had sent at least one teacher the regional training and the average building sent 41.4% (s.d. = .362) of teachers to either the creation of interim assessments or the regional professional learning communities. Therefore, the parsimonious covariant model has two levels and controls for the covariant of non-Caucasian/non-Asian students (NonCA) at the school level. Table 17 summarizes the results of the parsimonious model; the mixed model uses the IIA Process composite as the dependent variable and grand mean centers NonCA, such that:

\[
IIA\_Process\_Composite_{ij} = \gamma_{00} + \gamma_{01}(cNonCA) + r_{ij} + u_{0j}
\]  
(2.3)
Table 17

*Results of the Parsimonious Level 2 Survey Model (with robust standard errors)*

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Coefficient</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\gamma_{00}$</td>
<td>4.1877</td>
<td>0.0474</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>cNonCA, $\gamma_{01}$</td>
<td>-0.8966</td>
<td>0.2082</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>$R^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1, $\sigma^2$ (teachers)</td>
<td>0.22322</td>
<td>0.0016</td>
</tr>
<tr>
<td>Level 2, $\tau_\beta$ (schools)</td>
<td>0.03125</td>
<td>0.3702</td>
</tr>
</tbody>
</table>

Level 1 n = 165; Level 2 n = 35

Therefore, 37% of the between school variation was explained by the covariant. However, the majority of the variation occurred between teachers. The parsimonious model allows for the next step to focus on the variable of interest and addressing the question of the relationship for different levels of professional learning regarding interim instructional assessments.

**Professional Learning Levels Related to Levels of Implementation**

The professional learning levels have point values that vary from no exposure (zero) to full training on the creation, implementation, and utilization of interim assessment in a collaborative environment (one) (Table 14). However, professional learning was not a dichotomous variable since some teachers only participated in the regional data analysis (0.7) and other teachers had some other exposure to data analysis or assessment utilization (0.3). By inserting the professional learning levels at level one, the teacher level, I hoped to explain some of the between-teacher variance. I also hoped that the interpretation of the coefficient would be both significant and meaningful to conclude a positive relationship between professional learning and the implementation of the seven components captured in the composite score. Table 18
captures the results of the HLM analysis, both the coefficient and the changes in the variance.

Table 18

*Results of the Professional Learning Model (with robust standard errors)*

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Previous Coefficient</th>
<th>Coefficient</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, ( \gamma_{00} )</td>
<td>4.1877</td>
<td>4.0946</td>
<td>0.0558</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>cNonCA, ( \gamma_{01} )</td>
<td>-0.8966</td>
<td>-0.7514</td>
<td>0.2088</td>
<td>.001</td>
</tr>
<tr>
<td>PD Level, ( \gamma_{10} )</td>
<td>n/a</td>
<td>0.2460</td>
<td>0.0979</td>
<td>.013</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Previous Variance</th>
<th>Variance Component</th>
<th>Change in ( R^2 )</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1, ( \sigma^2 ) (teachers)</td>
<td>0.22322</td>
<td>0.21133</td>
<td>0.0533</td>
</tr>
<tr>
<td>Level 2, ( \tau_\beta ) (schools)</td>
<td>0.03125</td>
<td>0.03980</td>
<td>-0.2736</td>
</tr>
</tbody>
</table>

Level 1 n = 165; Level 2 n = 35

Therefore, the professional learning levels explain over 5% of the variation at the teacher level even though the \( R^2 \) decreased at the school level. In addition, the coefficient associated with professional learning was significant (\( p < .05 \)) and shows a positive relationship.

Specifically, for a teacher who attended the training offered at the regional center, the teacher has a higher composite survey score, +0.25, compared to another teacher who had no training and controlling for the building environment containing an average number of Non-Caucasian/Non-Asian students. In other words, in a typical school, a teacher without any training would be expected to have scored 4.09 on a six-point scale, regarding their level of implementation of the seven research components around the utilization of interim instructional assessments in a collaborative environment. Likewise, another teacher who participated fully in the regional training, also in a typical school, would be expected to have a higher average score of 4.34.

The IIA Process composite score consisted of all 37 items in the teacher survey around
implementation levels of interim assessment utilization in a collaborative environment. An additional question that might be raised would inquire into the relationship of professional learning and the seven research components that make up the composite score. Table 19 provides a summary of key results when the composite score, as the outcome variable, was decomposed into the subcomponents that were significant.

Table 19

Results of the Professional Learning Model on the Components (with robust standard errors)

<table>
<thead>
<tr>
<th>Outcome Variable</th>
<th>Variables</th>
<th>Coefficient</th>
<th>SE</th>
<th>( p )</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIA Process Composite</td>
<td>Intercept, ( \gamma_{00} )</td>
<td>4.0946</td>
<td>0.056</td>
<td>&lt; .001</td>
<td>.44</td>
</tr>
<tr>
<td></td>
<td>PD Level, ( \gamma_{10} )</td>
<td>0.2460</td>
<td>0.098</td>
<td>.013</td>
<td></td>
</tr>
<tr>
<td>Design</td>
<td>Intercept, ( \gamma_{00} )</td>
<td>4.1963</td>
<td>0.077</td>
<td>&lt; .001</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>PD Level, ( \gamma_{10} )</td>
<td>0.3621</td>
<td>0.122</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Intercept, ( \gamma_{00} )</td>
<td>3.7742</td>
<td>0.091</td>
<td>&lt; .001</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>PD Level, ( \gamma_{10} )</td>
<td>0.3823</td>
<td>0.159</td>
<td>.017</td>
<td></td>
</tr>
<tr>
<td>Learning Target</td>
<td>Intercept, ( \gamma_{00} )</td>
<td>3.6371</td>
<td>0.095</td>
<td>&lt; .001</td>
<td>.40</td>
</tr>
<tr>
<td></td>
<td>PD Level, ( \gamma_{10} )</td>
<td>0.4069</td>
<td>0.145</td>
<td>.006</td>
<td></td>
</tr>
<tr>
<td>Engage</td>
<td>Intercept, ( \gamma_{00} )</td>
<td>3.5699</td>
<td>0.106</td>
<td>&lt; .001</td>
<td>.52</td>
</tr>
<tr>
<td></td>
<td>PD Level, ( \gamma_{10} )</td>
<td>0.4390</td>
<td>0.155</td>
<td>.005</td>
<td></td>
</tr>
</tbody>
</table>

Level 1 \( n = 165 \); Level 2 \( n = 35 \)

The table only shows the four subcomponents that professional learnings levels were significant in the model. The prioritization of standards \( (p > .2) \), supportive structures \( (p > .2) \), and assessment literacy \( (p > .5) \) were not significantly related to the level of professional development.

Similar to the interpretation of the coefficient when the IIA Process composite score was influenced by the professional learning levels, each of the remaining four subcomponents had a positive relationship with the implementation levels captured by the subcomponent. For
instance, on a scale of one to six, with six being the best, the typical teacher in a typical school rated themselves at a 4.2 for assessment design when the teacher did not report any professional development, whereas, a teacher who participated in the regional training would have an expected value of 4.6. Likewise, the typical teacher without training rated themselves at a 3.8 for data analysis, and teachers with regional training averaged a rating of 4.2. Similarly, self-reported implementation of learning targets and engagement strategies, such as student discourse, both increased from 3.6 to 4.0. With this positive relationship to implementation levels, the theory of action for professional learning holds that greater implementation will also result in greater student learning gains as investigated by research question three.

**Research Question 3: Relationship between Survey Components and Math Achievement**

The final question accounts for the student, teacher, and school level demographics to determine if at least one of the components will predict improved student scores in mathematics on high-stakes, state-level tests linked directly to teachers.

3. To what extent was there a relationship between the level of teacher implementation of these seven components involved in the utilization of interim instructional assessments and math achievement score improvement over time, controlling for certain student, teacher, and school building factors (i.e., poverty, ethnicity, special education, and math pre-test)?

In order to address this question, a three-level hierarchical linear model was created placing 13,494 students at level one, 165 teachers at level two, and 35 schools at level three.

**Unconditional Model HLM Level 3**

The unconditional model (3.0) found in Chapter 3 provided the baseline for subsequent models when adding both covariants and the survey variables of interest. There have been
numerous studies in education that suggest the range of intraclass correlations (ICC) between schools typically range from .15 to .25 for student math achievement (Hedges & Hedberg, 2007). However, I was unable to find suggested ICC values for a three-level model in education likely due to the difficulty in obtaining teacher roster data. Based on the unconditional model, the greatest amount of variance occurred at level one as expected. Table 20 summarizes the analysis results of the unconditional three level HLM.

Table 20

<table>
<thead>
<tr>
<th>Description</th>
<th>Variance Component</th>
<th>SE</th>
<th>Unconditional ICC</th>
<th>Reliability estimate</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1, ( \sigma^2 ) (students)</td>
<td>0.6032</td>
<td>0.0074</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2, ( \tau_\pi ) (teachers)</td>
<td>0.1550</td>
<td>0.0202</td>
<td>19.55%</td>
<td>.940</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Level 3, ( \tau_\beta ) (schools)</td>
<td>0.0346</td>
<td>0.0176</td>
<td>4.36%</td>
<td>.466</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Level 1 \( n = 13,494 \); Level 2 \( n = 165 \); Level 3 \( n = 35 \)

The reliability was fairly high at level two; however, the reliability at level three was fairly low possibly due to the average school having less than five teachers who took the survey and seven buildings having only one or two teachers who surveyed. As the table above shows, both the level 2 and level 3 variance components were significant, however, nearly 20% of the variation occurs between teachers and less than 5% occurs between schools. This may suggest that the typical school variation of 15-25% would be better explained if the teacher level were made available. Some of this variation might be explained by the covariants within this study and the levels of teacher implementation of seven research-based components regarding the utilization of interim instructional assessments within a collaborative environment.
Parsimonious Covariant Model

In order to explore how the variation might be explained, covariants were added by specific blocks by levels. The first block added seven uncentered covariants at level one with fixed effects where the estimated mean math achievement, 0.2387, and each of the covariants were significant at \( p < .001 \), with the exception of English Learners which was also significant at \( p < .01 \). Table 21 shows how the variance changes when adding the following covariants at level one: pre-test, economically disadvantaged, English Learner, students with disabilities, non-Caucasian/non-Asian, female, and course alignment to the state test.

Table 21

Three Level HLM Block 1 Covariants Model Results

<table>
<thead>
<tr>
<th>Description</th>
<th>Variance Component</th>
<th>SE</th>
<th>( R^2 )</th>
<th>Reliability estimate</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1, ( \sigma^2 ) (students)</td>
<td>0.2931</td>
<td>0.0036</td>
<td>0.5141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2, ( \tau_\pi ) (teachers)</td>
<td>0.0460</td>
<td>0.0062</td>
<td>0.7032</td>
<td>.907</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Level 3, ( \tau_\beta ) (schools)</td>
<td>0.070</td>
<td>0.0045</td>
<td>0.7981</td>
<td>.370</td>
<td>.005</td>
</tr>
</tbody>
</table>

\( Level 1 n = 13,494; Level 2 n = 165; Level 3 n = 35 \)

Although the variance at both level two and three remains significant, the level one covariants explained a large portion of teacher variance (70.3%) and a larger portion of school variance (79.8%). Additional covariants at the teacher and school level may explain the variation at these two levels even further.

The second block added the teacher level covariants of years teaching math centered on the grand mean and the dichotomous effect of teachers having at least a major in mathematics or only a minor in their college degree. This second iteration revealed the dichotomous covariant of
major was not significant ($p = .48$). Therefore, major was removed as a covariant before adding level 3 covariants. The third block added school level covariants, which considered the same seven variables at level 1 aggregated to the school level and grand mean centered. However, six of the seven covariants were not significant ($p > .1$) except English Learners ($p < .05$). Also, the variance-covariance components test was conducted ($\chi^2 = 9.404, d.f. = 7, p$-value = .224) comparing the fullest model to the suggested parsimonious model and concluded that the simpler model was more appropriate. Therefore, those level two and three covariants that were not significant were eliminated from the model to create the parsimonious covariant model with values summarized in the following mixed model:

$$Math_{ijk} = \gamma_{000} + \gamma_{001}(cEL_k) + \gamma_{010}(cYears_{jk}) + \gamma_{100}(PreTest_{ijk}) + \gamma_{200}(Aligned_{ijk}) + \gamma_{300}(ED_{ijk}) + \gamma_{400}(SWD_{ijk}) + \gamma_{500}(EL_{ijk}) + \gamma_{600}(NonCA_{ijk}) + \gamma_{700}(Female_{ijk}) + e_{ijk} + r_{0jk} + u_{00k}$$

Each of the terms were explained in Chapter 3 with the exception of the meaning of the intercept based on this parsimonious model. The intercept, $\gamma_{000}$, represents the expected mean math achievement for an Asian or Caucasian, male student when his school has the average population of English Learners, his teacher has taught math the average number of years, and the student scored at the state average on the pre-test. Table 22 summarizes the change in variance components from the unconditional model, and Table 23 summarizes the results of the HLM analysis for the parsimonious covariant model above.
Three Level HLM Parsimonious Covariants Model Variations

<table>
<thead>
<tr>
<th>Description</th>
<th>Variance Component</th>
<th>SE</th>
<th>$R^2$</th>
<th>Reliability estimate</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1, $\sigma^2$ (students)</td>
<td>0.2931</td>
<td>0.0036</td>
<td>.5141</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 2, $\tau_\pi$ (teachers)</td>
<td>0.0423</td>
<td>0.0057</td>
<td>.7273</td>
<td>.900</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Level 3, $\tau_\beta$ (schools)</td>
<td>0.0026</td>
<td>0.0031</td>
<td>.9109</td>
<td>.197</td>
<td>.077</td>
</tr>
</tbody>
</table>

Level 1 $n = 13,494$; Level 2 $n = 165$; Level 3 $n = 35$

Results of the Parsimonious Covariant Model (with robust standard errors)

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Coefficient</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\gamma_{000}$</td>
<td>0.2397</td>
<td>0.0448</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>cEL, $\gamma_{001}$</td>
<td>-1.6175</td>
<td>0.2928</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>cYears, $\gamma_{010}$</td>
<td>0.0063</td>
<td>0.0022</td>
<td>.006</td>
</tr>
<tr>
<td>Math Pre-test, $\gamma_{100}$</td>
<td>0.6001</td>
<td>0.0133</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>TestAligned, $\gamma_{200}$</td>
<td>-0.1792</td>
<td>0.0532</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Econ. Disadvantaged, $\gamma_{300}$</td>
<td>-0.0828</td>
<td>0.0139</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Student w Disability, $\gamma_{400}$</td>
<td>-0.2825</td>
<td>0.0301</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>English Learner, $\gamma_{500}$</td>
<td>-0.1130</td>
<td>0.0380</td>
<td>.003</td>
</tr>
<tr>
<td>Non Caucasian/Non Asian, $\gamma_{600}$</td>
<td>-0.1165</td>
<td>0.0182</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Female, $\gamma_{700}$</td>
<td>-0.1190</td>
<td>0.0133</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Level 1 $n = 13,494$; Level 2 $n = 165$; Level 3 $n = 35$

With much of the variance already explained at level two and three, I expected to see very little change in $R^2$ when adding the variables of interest around the seven research components surveyed, and I hoped to see significant coefficients. Notice that with 91% of the variance explained at level three, the variance was no longer significant. Fortunately, the variables of
interest are found at the teacher level, where 73% of the variance has already been explained.

**Survey Components in Relation to Math Achievement**

In order to introduce the variables of interest, there are eight models built in Chapter 3 and Appendix E that introduce one of the seven research-based components and the IIA Process composite score regarding the utilization of interim instructional assessments within a collaborative environment, specifically:

i. prioritization of standards for a guaranteed and viable curriculum;

ii. supportive relationships and structures;

iii. assessment literacy and autonomy;

iv. quality interim assessment design;

v. an analysis that leads to action planning;

vi. high impact instruction: learning goals; and

vii. high impact instruction: engagement around discourse and practice?

The models 3.1 through 3.8 built on the parsimonious covariant model by inserting the IIA Process composite score from the survey and then inserting one component at a time. Introducing these variables of interest at level 2 had little to no impact on the covariants at level 1. However, the covariants at level 2 and level 3 may change. Therefore, level 1 covariants were not reported in this section.

The first variable tested was the IIA Process composite score from all 37 items on the survey. Teachers ranged from 2.8 to 5.4 on the six-point Likert scale, and the average teacher had an IIA Process composite score of 4.2. As the model in Chapter 3 suggested, when the composite score was added, it was allowed to vary randomly. However, the random effect for the composite score was not significant ($p > .5$). In addition, the variance-covariance
components test was conducted ($\chi^2 = 0.3877, d.f. = 2, p > .5$) and concluded that the simpler model was more appropriate. Therefore, the fixed effect for the IIA Process composite score is seen in Table 24 that highlights the changes that occurred with the introduction of the composite score. The composite score reduced the level two variation to 0.0391 which increased the $R^2$ from 72.7% to 74.8% of the variation explained.

Table 24

*Results of the Composite Survey Model (with robust standard errors)*

<table>
<thead>
<tr>
<th>Variable Description</th>
<th>Previous Coefficient</th>
<th>Coefficient</th>
<th>SE</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept, $\gamma_{000}$</td>
<td>0.2397</td>
<td>0.2357</td>
<td>0.0443</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>cEL, $\gamma_{001}$</td>
<td>-1.6175</td>
<td>-1.3836</td>
<td>0.2814</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>cYears, $\gamma_{010}$</td>
<td>0.0063</td>
<td>0.0063</td>
<td>0.0022</td>
<td>.006</td>
</tr>
<tr>
<td>IIA Process Composite, $\gamma_{020}$</td>
<td>n/a</td>
<td>0.1085</td>
<td>0.0373</td>
<td>.004</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Previous Variance</th>
<th>Variance Component</th>
<th>Change in $R^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 2, $\tau_\pi$ (teachers)</td>
<td>0.0423</td>
<td>0.0391</td>
<td>0.0204</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Level 3, $\tau_\beta$ (schools)</td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.0158</td>
<td>.053</td>
</tr>
</tbody>
</table>

*Level 1 n = 13,494; Level 2 n = 165; Level 3 n = 35*

Therefore, the IIA Process composite score explains 2.04% of the variation at level two and 1.58% of the variation at level three. In addition, the coefficient was a significant factor in the overall model and suggests that a teacher who self-reports one unit higher on the composite score than another teacher will likely have an average classroom $z$-score that was 0.11 higher. When combined with the findings from research question 2, professional learning had a positive relationship with implementation levels self-reported on the survey, and those higher survey scores have a positive relationship to student learning.
The IIA Process composite score consists of all 37 items in the teacher survey around implementation levels of interim assessment utilization in a collaborative environment. The models in Chapter 3 sought to decompose the composite score into the seven research components that make up the composite score. Table 25 provides a summary of key results when the composite score, as the primary independent variable of interest, was decomposed into the subcomponents that were significant when added to the parsimonious model.

Table 25

Results of the Student Achievement Model and Components (with robust standard errors)

<table>
<thead>
<tr>
<th>Variable of Interest</th>
<th>Variables</th>
<th>Coefficient</th>
<th>SE</th>
<th>p</th>
<th>Reliability</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIA Process Composite</td>
<td>Intercept, ( \gamma_{000} )</td>
<td>0.2357</td>
<td>0.0443</td>
<td>&lt; .001</td>
<td>.893</td>
</tr>
<tr>
<td></td>
<td>Composite, ( \gamma_{020} )</td>
<td>0.1085</td>
<td>0.0373</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>Support</td>
<td>Intercept, ( \gamma_{000} )</td>
<td>0.2373</td>
<td>0.0445</td>
<td>&lt; .001</td>
<td>.898</td>
</tr>
<tr>
<td></td>
<td>Support, ( \gamma_{020} )</td>
<td>0.0443</td>
<td>0.0229</td>
<td>.055</td>
<td></td>
</tr>
<tr>
<td>Assess. Design</td>
<td>Intercept, ( \gamma_{000} )</td>
<td>0.2378</td>
<td>0.0442</td>
<td>&lt; .001</td>
<td>.897</td>
</tr>
<tr>
<td></td>
<td>Design, ( \gamma_{020} )</td>
<td>0.0649</td>
<td>0.0279</td>
<td>.022</td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Intercept, ( \gamma_{000} )</td>
<td>0.2410</td>
<td>0.0237</td>
<td>&lt; .001</td>
<td>.895</td>
</tr>
<tr>
<td></td>
<td>Analysis, ( \gamma_{020} )</td>
<td>0.0646</td>
<td>0.0236</td>
<td>.007</td>
<td></td>
</tr>
<tr>
<td>Engage</td>
<td>Intercept, ( \gamma_{000} )</td>
<td>0.2366</td>
<td>0.0440</td>
<td>&lt; .001</td>
<td>.894</td>
</tr>
<tr>
<td></td>
<td>Engage, ( \gamma_{020} )</td>
<td>0.0648</td>
<td>0.0185</td>
<td>&lt; .001</td>
<td></td>
</tr>
</tbody>
</table>

Level 1 n = 13,494; Level 2 n = 165; Level 3 n = 35

The table only shows the four components that were relatively significant when added to the parsimonious model including supportive structures/relationships, assessment design, analysis, and engagement. The prioritization of standards \( (p > .2) \), learning targets \( (p > .3) \), and assessment literacy \( (p > .2) \) were not significantly related to student math achievement.

However, the support structures and relationships were left in the table based on the fact that the \( p \)-value indicated significance \( (p = .047) \) with the final estimation of fixed effects without robust
standard errors.

Similar to the interpretation of the coefficient when student math achievement was influenced by the self-reported implementation captured by the IIA Process composite score, each of the remaining four subcomponents in Table 25 had a positive relationship with the implementation levels captured by the subcomponent. For instance, the typical teacher in a typical school who rated themselves one point higher for assessment design than the average teacher in the study would expect a classroom math achievement \( z \)-score of 0.303, whereas, the average teacher would expect an average achievement score of 0.238. Likewise, the typical teacher with an average rating for data analysis would expect an average \( z \)-score of 0.241 and a teacher who reported one unit higher would expect an average classroom \( z \)-score of 0.306. Similarly, self-reported supportive structures and engagement strategies such as math discourse both increased from 0.237 to 0.282 and 0.301 respectively. Of course, as previously stated, the combined process of these seven components raised scores from 0.236 to 0.344 when a teacher self-reported one unit higher than the average rating of teachers in the study.

**Chapter 4 Summary**

In this chapter, a detailed statistical analysis presented findings to address the three research questions. First, each of the seven research components around interim instructional assessments (IIA) within a collaborative environment and the IIA Process composite score varied across the six-point Likert scale. Research Question 1 inquired about the distribution of these results and the seven components of the IIA Process and IIA Process composite score had a relatively normal distribution across the six-point Likert scale. Scores tended to range from the lowest range of one or two to the highest range of five or six on the Likert scale. In fact, the survey components, teacher demographics, and school demographics met assumptions needed
for the methodology used.

Second, the results concerning the second research question indicated a positive relationship between professional learning levels and implementation levels, however, only a small amount of variance was explained. The coefficient for professional learning levels suggested that a teacher without any training on the IIA Process would expect to have a composite score 4.09 whereas a teacher who fully participated in the training would expect a composite score of 4.34. Therefore, the model suggested a significant and positive relationship between the levels of professional learning and the levels of implementation around the interim instructional assessment process.

Finally, Research Question 3 inquired if such increased levels of implementation would relate to increased levels of student achievement. The results addressing the third research question also indicated a positive relationship between self-reported implementation levels and student math achievement. In other words, a teacher with the lowest IIA Process composite score of 2.8 would expect to have a lower average student achievement score compared to the teacher with the highest composite score of 5.4. Specifically, the teacher with the lowest survey score would expect a math achievement z-score of 0.09 whereas the teachers with the highest survey score would expect a math achievement z-score of 0.37. Table 26 and 27 provides a summary of the results discussed in this section. In chapter 5, I took a closer look at the implications of these results and examined each of the seven components that make up the interim instructional assessment process.
Table 26

Summary of Greer (2019) Research Results on Research Question and IIA Process

<table>
<thead>
<tr>
<th>Research Question or Topic</th>
<th>Greer (2019) Research Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. What were the levels of teacher implementation of seven research-based components around the IIA Process?</td>
<td>Survey items varied across a six-point scale (Tables 6-13), had a relatively normal distribution, and had internal consistency.</td>
</tr>
<tr>
<td>2a. To what extent did buildings differ based on sending teachers to regional training?</td>
<td>The majority of buildings (29 of 35) sent some teachers to regional training and the results do not show this decision to be statistically significant for this population.</td>
</tr>
<tr>
<td>2b. To what degree are there differences between teachers who participated to some degree in training around the IIA Process?</td>
<td>The IIA Process composite score had a significant and positive relationship (0.25*, Table 19) with teacher levels of training such that teachers with no training had an expected average score of 4.09 and those with training had an average expected score of 4.34 on a scale from 1 to 6.</td>
</tr>
<tr>
<td>3a. The unconditional 3-level HLM model for Research Question 3 examined the variation of student achievement across the student, teacher, and school level.</td>
<td>Demonstrated nearly 20% of the variation occurred at the teacher level and less than 5% occurred between schools (Table 20).</td>
</tr>
<tr>
<td>3b. To what extent was there a relationship between the level of teacher implementation of the IIA Process and math achievement score improvement when controlling for certain factors?</td>
<td>The IIA Process composite score had a significant and positive relationship with student math achievement levels such that the teacher with a lower IIA Process score of 2.8 would expect the students to have an average z-score of 0.09 on the state math test. Whereas a similar teacher with a higher IIA Process score of 5.4 would expect the students to have an average z-score of 0.37 on the state math test.</td>
</tr>
</tbody>
</table>
Table 27

Summary of Greer (2019) Research Results on the Components of the IIA Process

<table>
<thead>
<tr>
<th>Components of IIA Process</th>
<th>Greer (2019) Research Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>IIA Process composite score (components (i)-(vii))</td>
<td>The composite score for all components was found to have a significant, positive relationship (0.25*, Table 19) with professional learning and a significant, positive relationship (0.11**, Table 25) with student math achievement.</td>
</tr>
<tr>
<td>(i) Prioritization of standards for a guaranteed and viable curriculum</td>
<td>Independently not significant when tested in either the relationship with training nor with math achievement, likely due to the population and higher scores.</td>
</tr>
<tr>
<td>(ii) Supportive relationships and structures</td>
<td>Significant, positive relationship (0.04*, Table 25) with secondary math achievement when tested independently, however, not significant independently with training.</td>
</tr>
<tr>
<td>(iii) Assessment literacy and autonomy</td>
<td>Independently not significant when tested with either the relationship with training nor with math achievement.</td>
</tr>
<tr>
<td>(iv) Quality interim assessment design</td>
<td>Significant, positive relationship (0.36**, Table 19) with professional learning and a significant, positive relationship (0.06*, Table 25) with secondary math achievement when tested independently.</td>
</tr>
<tr>
<td>(v) An analysis that leads to action planning</td>
<td>Significant, positive relationship (0.38*, Table 19) with professional learning and a significant, positive relationship (0.06**, Table 25) with secondary math achievement when tested independently.</td>
</tr>
<tr>
<td>(vi) high impact instruction: learning goals and feedback</td>
<td>Significant, positive relationship (0.41**, Table 19) with professional learning but not a significant relationship with secondary math achievement when tested independently.</td>
</tr>
<tr>
<td>(vii) high impact instruction: student discourse and distributive practice</td>
<td>Significant, positive relationship (0.44**, Table 19) with professional learning and a significant, positive relationship (0.06***, Table 25) with secondary math achievement when tested independently.</td>
</tr>
</tbody>
</table>
CHAPTER 5
DISCUSSION

The purpose of my study was to ascertain the extent to which math teachers utilized an interim instructional assessment (IIA) process within middle and high schools, and how such utilization levels connected to student achievement. Another purpose was to ascertain the role that the intensity of teacher training had on the levels of utilization within these schools. To address these relationships, hierarchical linear modeling (HLM) was used to analyze 35 schools containing 165 math teachers who completed a survey and taught 13,494 secondary students.

The format of this chapter includes a discussion of the major results, the implications to school leaders, the relationship to existing studies, and recommendations for future research.

Discussion of Major Results

The seven components around the utilization of interim assessments in a collaborative environment and the two relationships seen in the research questions are illustrated in Figure 7 with updated results from Chapter 4, specifically from summary of Table 27. The first relationship sought whether regional teacher training influenced the effective utilization of the components around interim assessments, as seen in research question two. This positive relationship (0.25*, Table 27) is illustrated in Figure 7 by the black arrow between training and components around the IIA Process. In addition, there were four components that had a significant, positive relationship when tested independently as noted by the black circles around components (iv) – (vii). The second relationship examines whether the levels of teacher utilization of these seven components impacted math achievement on the state tests, as seen in research question three, by tying student results directly to teachers. This second positive relationship (0.11**, Table 27) is illustrated in Figure 7 by the blue arrow between the
components around the IIA Process and student math achievement on state tests. In addition, there were four components that had a significant, positive relationship when tested independently as noted by the blue, rounded rectangles around components (ii), (iv), (v), and (vii).

**Figure 7.** Updated Conceptual Framework with Results around the IIA Process.

**Levels of Training had a Positive Relationship with Levels of Utilization**

In order to address the first relationship between regional teacher training and the levels of utilization around interim instructional assessments as reported on the teacher survey, a two-
level HLM model was created (model 2.1 and 2.2, Chapter 3). The analysis began with the
unconditional model 2.0 (Chapter 3) and found the most parsimonious model of covariants
(model 2.3, Chapter 4) prior to testing the survey levels. Surprisingly, the parsimonious model
found no significance in the effect of a school that sent a teacher to the regional training, and the
variance-covariance component test confirmed the simpler model was more appropriate.
Therefore, buildings did not differ based on their decision to send teachers to the regional
training as research question two inquired. Research question two also inquired about the
relationship between the regional training and the individual level of teacher implementation of
the seven components and the IIA Process composite score.

The IIA Process composite score from the teacher survey regarding the levels of
utilization around interim instructional assessment was used as the outcome variable and the
level of professional learning for teachers was found to be significant as seen in Table 18
(Chapter 4). Although only about 5% of the variation at the teacher level was explained by
professional learning, the coefficient implied a positive relationship. In fact, according to Table
18, a teacher who attended the regional training would expect a higher composite survey score,
approximately 0.25 units higher on a six-point Likert scale than another teacher who had no
training. In other words, the expected composite score of a teacher who attended the regional
training was 4.34 (out of 6) compared to a score of 4.09 of a teacher who did not attend training.
Therefore, regional training had a positive relationship with the overall utilization around interim
instructional assessments in a collaborative environment.

Next, the individual components that comprised the IIA Process composite score were
tested by replacing the composite score as the dependent variable with each of the seven
research-based components. The results indicated that not all of the seven components had a
significant relationship with the regional training. Specifically, the prioritization of standards for a guaranteed and viable curriculum was not significant. However, it was also the highest rated component by teachers with a mean of 4.86 (out of 6) as seen in Table 13 (Chapter 4). Therefore, it stands to reason the prioritization of standards was a common practice across the region regardless of training. Likewise, the other two components that did not have a significant relationship with regional training also had some of the highest average scores of 4.52 for supportive structures and 4.20 for assessment literacy. However, the four remaining components each had a statistically significant relationship with the regional training.

The component for engagement around student discourse and distributive practice yielded the lowest average response of 3.73 (out of 6) from the teacher survey as seen in Table 13 of Chapter 4. Hattie et al. (2017) defined student discourse as moving beyond discussion and about the “exchange of ideas, including ways of representing, thinking, talking, agreeing, and disagreeing” (p. 136). In addition, Hattie et al. found that distributed practice of mathematical skills versus mass practice had an effect size of 0.71 as discussed in Chapter 2. Although the mean was relatively low, teachers who engaged in the regional training had an expected score of 4.00 compared to the lower score of 3.57 expected of those who did not have any training, an expected increase of 0.43. Likewise, the other three components also showed an expected increase for those involved in regional training such that clear learning targets increased 0.41, assessment design increased 0.36, and data analysis increased 0.38 (Table 19, Chapter 4). Therefore, the regional training had a positive relationship with the overall utilization around interim instructional assessments and specifically on the components of assessment design, data analysis, clear learning targets and effective engagement through student discourse and distributive practice. Research question three found a similar positive relationship with the IIA
Process composite score and some of the same components.

**Levels of Utilization had a Positive Relationship with Student Achievement**

Research question three addressed the second relationship between student achievement in mathematics and the utilization of the seven research-based components (prioritization, support, literacy, design, analysis, learning targets, and engagement) around interim instructional assessment in a collaborative environment. Five of the components coincided with the findings from research question two such that prioritization of standards and assessment literacy did not have a significant effect from professional learning or on student achievement. Whereas, assessment design, analysis, and engagement had a positive relationship connected to both professional learning and student achievement. However, before a deeper discussion of these relationships, it seemed important to start with lessons learned from the unconditional model and parsimonious covariant model.

There seemed to be very few research studies that connect secondary teachers at the middle school and high school to individual student achievement data to create a three-level hierarchical linear model (HLM). Hedges and Hedberg (2007) noted there had been numerous studies in education using a two-level HLM model where the unconditional intraclass correlations (ICC) between school buildings typically range between 15 and 25%. However, in this three-level HLM model, the unconditional ICC at the school level (4%) was substantially smaller than the unconditional ICC at the teacher level (20%) based on Table 20 (Chapter 4). In fact, when establishing a parsimonious model of covariants, 91% of the school level variation was explained such that the level three variance component was no longer significant ($p = .077$, Table 22, Chapter 4). This might suggest the importance of teacher-level data rather than school-level data. In a similar finding from this study, research question two found that building
level participation was not significant \((p > .5)\) whereas teacher level participation in regional training was significant \((p < .05)\). Fortunately, the purpose of this study focused on the relationships at the teacher level and did not set out to dispute any research established at the school level.

The IIA Process composite score from the teacher survey regarding the utilization levels of interim instructional assessment was introduced into the parsimonious covariant model as the primary variable of interest and was found to be significant as seen in Table 23 (Chapter 4). However, only about 2\% of the variation at the teacher level and 1.6\% of the variation at the school level was explained by the IIA Process composite score; the coefficient implied a positive relationship. In fact, according to Table 23, for a teacher who rated themselves one point higher than the average teacher on a six-point Likert scale would expect to have an average classroom achievement \(z\)-score that was 0.11 units higher. In other words, the expected math \(z\)-score of a classroom with an average composite score 0.24 units above the state average compared to an expected classroom score of 0.35 for a teacher who rated themselves one unit higher on the composite teacher survey (Table 23, Chapter 4). According to Table 13 of Chapter 4, the composite teacher score ranged from the lowest rating of 2.8 to the highest rating of 5.4 where the average was 4.2. Therefore, the self-reported level of overall utilization around interim instructional assessments in a collaborative environment had a positive relationship with student achievement for secondary mathematics. Combined with the positive relationship between professional learning and levels of utilization as suggested by the conceptual framework (Figure 4, Chapter 1) correctly suggested a positive relationship with the upward arrows between the three boxes illustrating this research study.

To further explore how each of the seven research-based components related to student
achievement, the IIA Process composite score was substituted for each of the seven components, as seen in Appendix E. Similar to the first relationship seen with research question 2, prioritization of standards for a guaranteed and viable curriculum did not have a significant relationship with student achievement. This finding likely has more to do with the distribution of the teachers in this region and may have been limited by the construction of the survey instrument. First, prioritization had the highest component mean of 4.86 on a six-point Likert scale, the distribution was both leptokurtic and negatively skewed based on the results for research question one found in Chapter 4. Based on this distribution of 165 teachers, and the fact that prioritization had a \( p \)-value between .2 and .3, there does not seem to be evidence to contradict the established research of Marzano (2003) on the importance of prioritizing standards for a guaranteed and viable curriculum. However, for this research study, prioritization of standards did not have a significant relationship with student achievement. Similarly, learning targets also did not have a significant relationship with student achievement although professional learning did have a positive relationship with learning targets. It was not the position of this research study to discredit the mass amount of research on learning targets nor prioritization of standards, as this study was limited to the distribution across the region and a researcher-created survey tool.

Assessment literacy did not have a significant relationship with student achievement nor professional learning. As discussed in Chapter 2, understanding the difference between assessments that follow scope and sequence and those that do not may be an attributing factor to the inconsistency of results regarding interim assessments. In this research study, participants clearly understood the difference, as it was explained to participants before the completion of the survey. It also stands to reason that a teacher knowing the difference would not have an impact
on student achievement. However, a future study might consider a meta-analysis that compares the utilization of interim benchmark assessments (i.e. pre/post-tests) compared to interim instructional assessments, those that follow scope and sequence of instruction. For this research study, the assessment design questions focused on the utilization of assessments that followed the scope and sequence of instruction.

There existed encouraging results that showed both assessment design and data analysis had a positive relationship with both regional training and student achievement. As seen in Chapter 2, several researchers and distinguished practitioners had articulated the need for quality assessment design and effective data analysis as part of a collaborative process that improves teaching and learning. Both components had a significant and positive relationship with achievement, such that the average teacher would expect a math achievement $z$-score of 0.24 and a teacher who reported one unit higher on either component would expect a math achievement $z$-score of 0.30. Likewise, engagement strategies such as student discourse and dispersed practice, also had a significant and positive relationship that increased the expected achievement score by 0.06 (Table 25, Chapter 4).

Finally, the component for supportive structures and supportive relationships seemed significant enough to include in Table 25 of Chapter 4. Depending on the use of robust standard errors, the $p$-value was either .047 or .055, and the coefficient again suggested a positive relationship such that the expected achievement score would increase from 0.24 to 0.28 if the component score was one unit higher than the average. However, the support component was not significantly related to the levels of training, which might indicate the regional center needed to do a better job teaching the vital behaviors of a professional learning community as described in Chapter 2. Of course, this component also addressed the supportive structures, such as
proximity to other math teachers and designated time for collaboration. This study treated the support component as a broad category within a thorough process around interim assessments, and there exist other studies that attempt to more clearly define this type of collaboration and the results to student achievement.

**Relationship of Results to Existing Studies**

This research study proposed that if a process was taught to build interim instructional assessments (Bambrick-Santoyo, 2010; Marshall, 2006; Stiggins & Dufour, 2009) that prioritized a guaranteed and viable curriculum (Frey & Fisher, 2009; Marzano, 2003), teachers might be better equipped to provide effective feedback (Hattie, 2009; Marzano, 2003) that would allow students to track the essential standards (Hattie, 2009). If the right type of assessments was built and an implementation process was taught, those teachers who experienced the process might be able to incorporate other research-based elements into their instruction, such as student self-reported grades (Hattie, 2009), student discourse, and distributive practice (Hattie, Fisher, & Frey, 2017). The positive relationship found in this research study confirmed the importance of the collective components of the process supported by numerous research studies (Clark, 2010; Goertz et al., 2010; Heritage, 2010; Stiggins & DuFour, 2009) detailed in Chapter 2.

Of the studies examined in Chapter 2, there were numerous studies that showed interim assessments of different types or purposes that had a positive result in one school, one grade, or one content area but not in another (Alkhus, 2012; Goren, 2010; Konstantopoulos & Miller, 2013; Marshall, 2006). In fact, none of these studies found a positive relationship with high school mathematics. Also, they tend to show predictability with the assessment only but fail to show a relationship between teacher fidelity of a process around assessment data and student impact on high-stakes tests. However, Oliver (2013) did inquire about the process and found
counter-intuitive results with benchmark assessments that suggested higher self-reported utilization resulted in the smaller gains in student achievement. On the contrary, by using a three-level HLM model and standardizing math achievement scores across 6th through 11th grade, my research study showed a positive relationship between teacher utilization around interim instructional assessments and student achievement in mathematics across secondary schools.

Therefore, this research study should add to what will hopefully be a growing amount of research using 3-level HLM with secondary mathematics. Although some research currently exists for math using HLM, the vast majority avoids the secondary level and tends to be limited to a two-level model. One plausible reason might be the difficulty of obtaining teacher rosters, the complexity of multiple memberships, and the difficulty of cleaning large data sets where students move within a school two to three times per year. There have been numerous studies in education, for both ELA and math at the lower level, that suggest the range of intraclass correlations (ICC) between schools typically range from 0.15 to 0.25 for student achievement (Hedges & Hedberg, 2007). Based on the unconditional model for research question three, the greatest amount of variance occurred at level one as expected. As Table 20 from Chapter 4 shows, both the level 2 and level 3 variance components were significant; however, nearly 20% of the variation occurs between teachers and less than 5% occurs between schools. These results might suggest that the typical school variation of 15-25% would be better explained if the teacher level were made available. This might be one of the major findings to add to the existing research; Table 28 represents a summary of how the major findings from this study relate to previous studies featured in Chapter 2 and how this study might contribute to future studies suggested by the literature review in Chapter 2.
Comparison of Greer Research with Previous Research Findings

<table>
<thead>
<tr>
<th>Previous Research Cited</th>
<th>Greer (2019) Research Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goertz et al. (2010) found the process around interim assessment impacts student achievement at various grades for elementary and middle school math.</td>
<td>Adds to these findings by showing the IIA Process composite score had a positive relationship with both professional learning (0.25*, Table 19) and student achievement (0.11**, Table 25) for all grades across middle school and high school math.</td>
</tr>
<tr>
<td>Fisher and Frey (2009) ethnography on the CFA or IIA Process included prioritization of standards.</td>
<td>Adds quantitative research that supports prioritization of standards, component (i), as part of the IIA Process.</td>
</tr>
<tr>
<td>Supportive relationships and structures are necessary for teachers to diagnose assessment data that will lead to high-quality instruction (DuFour et al., 2010; Olivier &amp; Hipp, 2010).</td>
<td>Supports by adding quantitative research with high school math that supports a collaborative environment, component (ii), as part of the IIA Process.</td>
</tr>
<tr>
<td>Conflicting studies around interim benchmark and interim instructional assessments, none of which showed an impact on high school math (Alkhas, 2012; Carlson et al., 2011; Goertz et al., 2010; Heritage, 2010; Konstantopoulos &amp; Miller, 2013; Oliver, 2013; Pon, 2013)</td>
<td>Adds quantitative research with HLM that supports the design of interim instructional assessments, component (iv), at the middle school and high schools level has a significant and positive relationship (0.06**, Table 25) to student achievement in mathematics.</td>
</tr>
<tr>
<td>Interim assessments must be accompanied by analysis that leads to teacher action (Black &amp; William, 1998; Goertz et al, 2010; Nichols et al., 2009; Stiggins &amp; DuFour, 2009).</td>
<td>Supports by adding quantitative research using HLM that supports analysis that leads to action, component (v), as part of the IIA Process.</td>
</tr>
<tr>
<td>High impact instruction includes student discourse and distributed practice (Hattie et al., 2017)</td>
<td>Supports by adding research using HLM that supports this type of student engagement, component (vii), as part of the IIA Process.</td>
</tr>
<tr>
<td>Intraclass correlations (ICC) between schools typically range from 0.15 to 0.25 for student achievement (Hedges &amp; Hedberg, 2007).</td>
<td>Adds to the HLM research by adding a third level that demonstrated nearly 20% of the variation occurred at the teacher level and less than 5% occurred between schools.</td>
</tr>
</tbody>
</table>
Implications for Educational Leaders

My study postulated there would be differences among teachers across the region regarding the seven components for the IIA Process, that regional training would increase the utilization of seven components among teachers, and that teachers who excel in one or more of the seven components would correspond with higher student achievement results on the state assessment. The results concluded that differences did exist and that regional training increased the utilization of interim instructional assessments, specifically with the IIA Process composite score (0.25*, Table 19) and four components. Furthermore, the results concluded that higher utilization of interim assessments in a collaborative environment corresponded with higher student achievement (0.11**, Table 25). Educational leaders, both teachers and administrators, who want to continue to improve teaching and learning for secondary mathematics might draw some inferences from these positive results.

First, there seems to be cause for celebration in this particular region that most teachers had prioritized a guaranteed and viable curriculum (Marzano, 2003) and had a reasonable understanding of assessment literacy (Bambrick-Santoyo, 2010). Specifically, the prioritization of standards for a guaranteed and viable curriculum was not significant; however, it was the highest rated component by teachers with a mean of 4.86 (out of 6). Therefore, it stands to reason, the prioritization of standards is a common practice across the region regardless of training. There were only a few teachers in this region who may need to address the prioritization of standards, as seen by the distribution as both leptokurtic, and negatively skewed illustrated in Figure 8 below. With the number of teachers who rated themselves high in this area, it may suggest one reason why this region tends to outperform the state as seen in Tables 2 and 4 (Chapter 3). However, this may prove to be the first step for leaders outside of this region.
Educational leaders outside of the region and a limited number of educators inside this region should also investigate the importance of utilizing assessments that follow the scope and sequence of instruction. This study confirmed that the design of these interim assessments (Bambrick-Santoyo, 2010; Chappulus et al., 2012) had a statistically significant and positive relationship with student achievement for secondary mathematics.

Second, the difficulty to move from a quality assessment design to a meaningful data analysis was articulated by researchers (Goertz et al., 2010; Nichols et al., 2009) in Chapter 2. However, the results of this study showed that regional training was related to the increased utilization of teachers in these two areas (.36**, .38* respectively, Table 19) and that the increased utilization had a positive relationship with higher student achievement (.06*, .06** respectively, Table 25). Both components had a significant and positive relationship with achievement such that the average teacher would expect a math achievement $z$-score of 0.24 and a teacher who reported one unit higher on either component would expect a math achievement $z$-score of 0.30. When these findings coupled with the positive relationship found between supportive structures/relationships and student achievement (.04*, Table 25), most educators
would draw an inference to *Professional Learning Communities* (PLC) as described in Chapter 2. Therefore, it may behoove administrators to learn the current status of PLCs in the building and to listen to teachers on how they might support increasing their craft around assessment design, analysis, and collaboration with others. It might be a structural issue where teachers do not have easy access to other professionals (3.79 out of 6, Table 7), or it might be honing teacher practice around identifying the causal theory within the data analysis that leads to more effective instruction (4.24 out of 6, Table 10). The results of this study merely identified the positive relationship when teachers feel supported and have opportunities for collaboration specific to assessment analysis.

Finally, the heart of this process around interim assessments must lie in the effectiveness of classroom instruction. The component that rated the lowest among the seven was titled, “VII High Impact Instruction: Engagement around student discourse and distributive practice” and Figure 8 shows the distribution. Hattie et al. (2017) defined student discourse as moving beyond discussion and about the “exchange of ideas, including ways of representing, thinking, talking, agreeing, and disagreeing” (p. 136). In addition, Hattie et. al. found that distributed practice of mathematical skills versus mass practice had an effect size of 0.71 as discussed in Chapter 2. The overall average for engagement was the lowest of the seven components at 3.73 (Table 13, Chapter 4), and each of the five indicators ranged from 3.5 to 3.9 on a one to six scale (Table 12, Chapter 4). Teachers self-reported component score (vii) ranged from the lowest score of 1.6 to the highest score of 5.6, as seen in Table 13 of Chapter 4. Therefore, student discourse and distributive practice might be two primary areas of focus for this region based on the low mean and the statistically significant and positive relationship to student achievement. What seems evident from this research study is that there exist teachers within this region who feel confident
in these two areas and who have likely seen higher student achievement results. The question then arises, how might this region or other regions learn from the highest performing teachers around the practices of these two engagement strategies? This question and others are addressed in the recommendations for future research and address one possible approach to the overall school improvement process.

**Recommendations for Future Research**

Of course, there are certain shortcomings to strictly using a quantitative approach such as my study and not seeking meaning through a qualitative approach. Creswell (2014) explained the benefits of qualitative research as exploring the meaning of something like the use of interim instructional assessments for teachers. A qualitative approach would have allowed for observing how teachers utilize assessment data in a collaborative environment. My study followed a deductive approach by examining the relationship between teachers utilizing a research-based set of seven components and the student math achievement on the state assessment. Other factors might have attributed to an increase or decrease in achievement scores. Therefore my study was not attempting to explain causation of any of the seven components.

However, a likely next step might be to take a more exploratory approach, such as Designed-Based Research (DBR). Such an approach would lend itself well to the overall school improvement process for improving teacher and learning secondary mathematics. According to Cobb (2003), DBR differs from traditional experimental research by its intention to “inquire more broadly into the nature of learning in a complex system and to refine generative or predictive theories of learning” (p. 12). Heath and Heath (2010) reference the incredible story of Jerry Sternin who helped resolve malnutrition in Vietnam, not by importing an American solution; rather, Jerry studied the ‘bright spots’ in Vietnam to discover the solution that a small
number of mothers had found to maintain the health of their children. Sparks (2004) made a connection between the work of Jerry Sternin and school reform by breaking the process into four basic steps: define, determine, discover and design. Cobb goes on to explain the iteration process of DBR that parallels Sparks’ four basic steps, including the determination of positive deviants, or bright spots. Therefore, data is only the starting point; the true power is in the stories of transformation behind the data.

The first step would be to define the problem, such as math achievement for students at the secondary level. Next, the study must determine if any positive deviants exist. Although it was not the purpose of my study, the parsimonious covariant model for research question three could be used to identify positive deviants using the level two residual file. By using the parsimonious covariant model, the model would control for the influence of factors that are outside the control of teachers. The residuals at level two would report how well the class average achieved on the state math test compared to how well the model predicted based on certain pre-test and demographic factors. Once the positive deviants, or bright spots, are identified, further qualitative studies could be conducted to discover vital behaviors that lead to such higher achievement. When more information would likely be discovered about how teachers are impacting student achievement, then the final step would be to design an appropriate support system to scale up the implementation of such vital behaviors. For example, this study suggested a focus on student discourse as the lowest rated component of the IIA Process, yet the distribution of teacher responses suggest there exists teachers in the region who have found a way to engage more students in math talk. Also, the data suggested that higher levels of utilization were strongly related to higher levels of math achievement. This school improvement process would be similar to the work of the Reading Now Network in West Michigan and might
be adopted as a Secondary Math Network in the region.

One benefit of my study was the use of hierarchical linear models (HLM) as the most appropriate methodology for education when hierarchical structures exist from the student to teacher to building level and the inadequacy of traditional statistical techniques for modeling (Raudenbush & Bryk, 2002). Specifically, a 3-level HLM model was applied in which students represented the first level, teachers at the second level, and buildings at the third level. The use of traditional models to study hierarchical structures, such as multiple regression and ANCOVA, often has misleading results. Using HLM allows for the separation of between-building, between-teacher and within classroom error terms, which allows for interpretations of both the teacher effect and the building effect (Raudenbush & Bryk, 2002).

However, this study had a nonexperimental ex-post facto design that had notable limitations, such as the lack of random assignment of a treatment variable. The study was limited to one geographical location since the aim was to link student data to teachers and teachers to buildings. The survey was limited to 37 questions and did not address every assessment practice, such as the emergence of the possible benefits of performance-based assessments. Since the survey was given to a relatively small number of schools (35) based on convenience sampling, generalizations are limited to the region with similar demographics. In 2017/18, the State of Michigan tested approximately 699,000 secondary students whereas my study used a specific region that tested approximately 22,500 secondary students or 3.2% of the state. The specific region in this study had fewer students who qualify for economically disadvantaged status, more Hispanic students, and fewer African American students as cited in the population section. As the data hubs in Michigan develop, additional studies may be able to research a larger population that might include the entire state.
Concluding Thoughts

According to MI School Data (2018), only about one-third of students in Michigan met the benchmark for college readiness in high school or met proficiency on the M-STEP in middle school. According to the College Board (2017), Michigan ranked 48th in the nation on the SAT Math sub-score. Likewise, over the past twenty years, the NAEP (2015) reported that Michigan results have remained stagnant, near 278, on the NAEP whereas the national average had increased from 272 to 282. According to recent national and state level assessments, only about one-third of Michigan students are proficient in secondary math. Both practitioners and researchers have asserted that a process around interim assessments would impact student achievement. Previous studies have been inconsistent in demonstrating the impact of an interim assessment process on student achievement, especially with high school mathematics. Moreover, previous studies were not found to utilize Heiarchical Linear Models (HLM) to test such a relationship. Therefore, the purpose of my study was to ascertain the extent to which math teachers utilized an interim instructional assessment (IIA) process within middle and high schools, and how such utilization levels connected to student achievement. Another purpose was to ascertain the role that the intensity of teacher training had on the levels of utilization within these schools.

The data analysis revealed differences at the teacher level across the region regarding their utilization of seven components of the IIA Process, as well as differences among the student-level math achievement data provided by the state. Such differences were found to have a positive relationship between regional training and higher, self-reported practices of utilizing interim assessments in the classroom. The analysis also showed a positive relationship between growth in student achievement on the state assessment and interim assessment utilization. Likely
the greatest implication to educational leaders would be the positive relationship between the growth in student math achievement and some of the individual components, in particular, those related to professional learning communities (assessment design, data analysis, supportive structures and relationships) and those related to high impact instruction (student discourse and distributive practice). Educational leaders would want to assess how their school utilizes interim instructional assessments that follow the scope and sequence of instruction and the structures in place to allow for data analysis within a collaborative environment. In other words, those who collaborate around assessment results seem to have a greater impact on student learning. In addition to teachers talking with each other, this research supports the notion that students learn better when talking with each other as well. I hope that greater use of these strategies will result in improved achievement statewide and we will continue to look for bright spots in education because teachers make such a profound difference. The data is only the starting point; the true power is in the stories of transformation behind the data.
REFERENCES


England: SAGE Publications Ltd.


Fenske, M. S. (2013). Educator perceptions of low-income elementary students and their effects on reading (Order No. 3602854). ProQuest Dissertations & Theses Global. (1468722435).


findings from an evaluation of the Formative Assessments of Students Thinking in Reading (FAST-R) Program in Boston Elementary Schools. New York, NY: Manpower Demonstration Research Corporation.


Kappan, 90(9), 640-644.

Appendix A

Research Project Invitation to Administrators
Research Project Invitation to Administrators

August 17, 2018

Subject: Secondary Math Research Project Invitation

To: ________ Building Principals, Curriculum Directors, and Superintendents

From: Doug Greer, ________ and WMU doctoral student

Dear Administrator,

As we have learned from the power of collaboration from the Reading Now Network, you are invited to participate in a research project that will hopefully spawn a collaboration among secondary mathematics instruction. The research study will examine the relationship between teacher’s instructional strategies in second mathematics and the impact on student achievement. Dr. Jianping Shen, Professor at Western Michigan University, and I are conducting the study titled *A Relationship between Student Math Achievement and Teachers Utilizing a Process involving Interim instructional Assessments*. I intend to use only de-identified information gathered from secondary math teachers across the area as part of my dissertation.

As a part of the research study, a survey instrument was created to ascertain teacher engagement around standards, instruction, and assessment. The survey was based on the work of Hattie, Fisher, & Frey (2017) and Marzano (2003) and was grouped into seven components:

- Prioritization of standards for a guaranteed and viable curriculum
- Supportive relationships and structures
- Assessment literacy and autonomy
- Quality interim assessment design
- Analysis that leads to action
- High impact instruction: learning goals
- High impact instruction: engagement

The survey would likely take your middle school and high school math teachers approximately 20 minutes to complete. It consists of 35 questions that are responded to on a six-point Likert scale. Teacher identity will be kept strictly confidential and coded for research purposes. The reason for assigning each teacher a random, four-digit number will be to connect students to their responses to instructional strategies. Before any analysis is performed, all identifiable information will be removed. Teachers may choose not to participate at all, and can simply not complete the survey. Moreover, the information in my dissertation will mask the identity of every student, teacher and school building. In short, I guarantee the utmost confidentiality.
My hope is to have a chance to meet with you prior to any teacher participating in the survey. Should you choose to participate as a district or building, I would like to work with you on how to best connect with the secondary teachers about the research study and the survey. It would be my pleasure to meet with teachers to explain the hope of contributing to a secondary math network, the basis of the research study, and answer any questions regarding the survey.

In closing, the true value of this study comes from the potential that certain instructional practices may be found to strongly relate to higher student achievement and beyond this research study, that we might share such evidence among teachers. In addition, I hope that teachers would be willing to come forward, not as a part of the study, and offer best practices around the strategies a similar study may suggest that have the greatest impact on learning. Essentially, teachers may have an incentive through the Marshal Plan and through this ISD to be identified as having the greatest impact with certain subgroups of our population. If I have not already spoken with you about the benefits of our regional plan for a math network, please inquire.

Please feel free to contact me with any questions, concerns, or suggestions. You may also contact Dr. Shen (269-387-3887), Human Subjects Institutional Review Board (269-387-8293) or the Vice President of Research (269-387-8298) if questions or problems arise during the course of the study.

Thank you,

Doug Greer
Cell:
Email:

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped data and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.
Appendix B

Research Project Invitation to Teachers
Research Project Invitation to Teachers

August 17, 2018

Subject: Secondary Math Research Project Invitation

To: Math Teachers at «District»

From: Doug Greer, _____ and WMU doctoral student

Dear «First» «Last»:

You are invited to participate in a research project that will hopefully spawn a collaboration among secondary mathematics instructors across the region. The research study will examine the relationship between teacher’s instructional strategies in second mathematics and the impact on student achievement. Dr. Jianping Shen, Professor at Western Michigan University, and I are conducting the study titled A Relationship between Student Math Achievement and Teachers Utilizing a Process involving Interim instructional Assessments. I intend to use only de-identified information gathered from secondary math teachers, like yourself, across the area as part of my dissertation. Your randomly generated code is: «Code» (You will need this for the survey). As a part of the research study, a survey instrument was created to ascertain teacher engagement around standards, instruction, and assessment. The survey was based on the work of Hattie, Fisher, & Frey (2017) and Marzano (2003) and was grouped into seven components:

- Prioritization of standards for a guaranteed and viable curriculum
- Supportive relationships and structures
- Assessment literacy and autonomy
- Quality interim assessment design
- Analysis that leads to action
- High impact instruction, learning goals
- High impact instruction, engagement

The survey would likely take you approximately 20 minutes to complete. It consists of 35 questions that are responded to on a six-point Likert scale. For the questions focused on a collective effort of your math department, the scale will be based on your level of agreement, as follows:

- Strongly disagree
- Moderately disagree
- Slightly disagree
- Slightly agree
- Moderately agree
- Strongly agree

The other questions will be based on your perceived level of expertise within your classroom
during the 2017/18 school year, as follows:

- No expertise at all
- Novice
- Advanced beginner
- Competent
- Highly proficient
- Regional expert

Teacher identity will be kept strictly confidential, as evident by the random, four-digit code. The reason for the code is to connect the de-identified students who spent two semesters with you to your responses around instructional strategies. Before any analysis is performed, districts, buildings, and teachers will be coded to hide their identity even from the researcher. Whereas student identifiable information will be completely removed. You may choose not to participate at all, and can simply not complete the survey. In short, I guarantee the utmost confidentiality.

The true value of this study comes from the potential that certain instructional practices may be found to strongly relate to higher student achievement and that we might share such evidence among teachers. Beyond this research study, I hope that teachers would be willing to come forward and offer best practices around the strategies this study may suggest that have the greatest impact on learning. This would be for the ISD research only and never published. The only exception would be by teacher permission to apply for the $5,000 stipend associated with the Marshall Plan. The only way this would be possible will be if teachers indicate on the survey that I have permission to decode their identity if and only if it corresponds to the highest achievement and I promise to talk only to them about sharing any information with others.

I’m hoping to be available personally to help administer the survey. Please feel free to contact me with any questions, concerns, or suggestions. You may also contact the Dr. Shen (269-387-3887), Human Subjects Institutional Review Board (269-387-8293) or the Vice President of Research (269-387-8298) if questions or problems arise during the course of the study.

Thank you,

Doug Greer
Cell:
Email:

This consent document has been approved for use for one year by the Human Subjects Institutional Review Board (HSIRB) as indicated by the stamped data and signature of the board chair in the upper right corner. Do not participate in this study if the stamped date is older than one year.
Appendix C

HSIRB Approval Letter
Date: October 16, 2018

To: Jianping Shen, Principal Investigator
   Doug Greer, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: IRB Project Number 18-08-19

This letter will serve as confirmation that your research project titled “A Relationship between Student Math Achievement and Teachers Utilizing a Process Involving Instructional Interim Assessments” has been approved under the exempt category of review by the Western Michigan University Institutional Review Board (IRB). The conditions and duration of this approval are specified in the policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may only be conducted exactly in the form it was approved. You must seek specific board approval for any changes to this project (e.g., you must request a post-approval change to enroll subjects beyond the number stated in your application under “Number of subjects you want to complete the study”). Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the IRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: October 15, 2019
Appendix D

Teacher Survey Regarding Interim Assessments
Survey: Teaching and Learning Math

Purpose of this survey
This survey will connect instructional strategies with student achievement on the state assessment from last year (M-STEP, PSAT, SAT) in order to discern which strategies seem to result in higher achievement on high-stakes testing. Please reflect back on last year and complete the survey based on what strategies and practices you used during 2017/18. The study considers student demographics such as free/reduced lunch, students with disabilities, and English learners.

Research Consent
Your responses are kept strictly confidential. In fact, you should receive a random generated code to place on this survey that will help to link your responses concerning standards, assessment, and instruction with de-identified students who spent at least two semesters with you. For the purposes of the research study, teacher and student identity will never be disclosed and will be kept strictly confidential using random codes. By recording your random number below and completing the survey you acknowledge the confidentiality of this survey and consent to the research study.

RECORD Randomly Assigned Number: ________________________________

1) What type of classes did you teach last year (2017/18)?

Check all that apply

☐ 6th – 8th grade math

☐ Pre-algebra or Algebra I

☐ Geometry, Algebra II, Pre-calculus or trigonometry

☐ Calculus or AP Stats

☐ Integrated HS math course 1 or 2

☐ Integrated HS math course 3 or 4

☐ Other ____________________________________________________________
2) How many years have you taught secondary mathematics (6th – 12th grade)? __________

3) In terms of your college coursework, which one statement best describes you:
   - Minor in mathematics with no additional college-level courses in mathematics
   - Minor in mathematics plus additional college-level courses in mathematics
   - Major in mathematics
   - Masters or specialist degree in teaching mathematics

4) In terms of professional learning, please check the box that best describes you:
   - I participated in the creation of HS or MS math interim assessments at the ISD
   - I did not create the ISD math interim assessments but I participated in the PLC meetings at the ISD examining the assessment data
   - I have not been involved in PLC or interim assessments at the ISD but I have had extensive training in PLC or assessment utilization (such as Delta Math)
   - I have not had extensive training around assessment utilization or PLCs in recent years
My View of My Math Department: Strongly disagree ↔ Strongly agree

For each of the following statements, select your level of agreement based on your math department during the 2017/18 school year.

1) Prioritization of Standards for a Guaranteed and Viable Curriculum

Please respond to each question with your level of agreement regarding your math department.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly disagree</th>
<th>Moderately Disagree</th>
<th>Slightly disagree</th>
<th>Slightly agree</th>
<th>Moderately Agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) My math department had clearly identified knowledge and skills essential for all students to master.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) The essential knowledge and skills of each course were prioritized so students had ample opportunity to learn.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) The depth and breadth of the standards expected to be covered were addressed within the time available.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) We made a clear distinction between knowledge and skills that students must master versus mere exposure in each unit.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) We frequently engaged in formal or informal discussions about specific content issues that hinder student learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## II) Supportive Relationships and Structures

Please respond to each question with your level of agreement regarding your math department during the past school year, 2017/18.

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Moderately disagree</th>
<th>Slightly disagree</th>
<th>Slightly agree</th>
<th>Moderately agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) A safe culture to take risks existed within my department.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) My department exhibited a unified and sustained commitment to improving teaching and learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) We engaged in vulnerable and supportive dialogue while examining student data to enhance teaching and learning.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) The proximity of my department personnel allowed for ease in collaborating with colleagues.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) The school schedule promoted collective learning and shared practice.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f) Adequate time was dedicated during the school year to facilitate collaborative work within my department.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### III) Assessment literacy and autonomy
Please respond to each question with your level of agreement regarding your math department during the past school year, 2017/18.

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Moderately disagree</th>
<th>Slightly disagree</th>
<th>Slightly agree</th>
<th>Moderately agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) When using formative assessments, we frequently shared promising practices around instruction.</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>b) Departmentally we supported allowing students to make up points on interim assessments (chapter tests, projects, etc.) to better assess mastery.</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>c) As a department, we spent more meaningful time analyzing interim rather than summative assessments.</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
<tr>
<td>d) We had the autonomy to personalize any training around assessment utilization in recent years.</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
<td>〇</td>
</tr>
</tbody>
</table>
The focus has changed from your department to you as an individual and the scale has changed to your perceived level of expertise.

**My Perception of My Classroom:**  Novice → Regional Expert

**IV) Quality Assessment Design for Instruction.**
Please respond to each question with your individual level of expertise this past school year, 2017/18.

<table>
<thead>
<tr>
<th>Question</th>
<th>Novice</th>
<th>Advanced beginner</th>
<th>Competent</th>
<th>Highly proficient</th>
<th>Regional expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Establishing a proficiency standard (or success criteria) that I expect every motivated student to achieve on assessments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Ensuring interim instructional assessments contain sufficient evidence for each essential skill or knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Providing timely feedback on essential knowledge and skills for individual students following interim instructional assessments.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) Ensuring items on interim instructional assessments vary in both complexity and difficulty.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) Ensuring items on interim instructional assessments align with the content and rigor of the state assessment.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
V) **Analysis that Leads to Action**

Please respond to each question with your individual level of perceived expertise this past school year, 2017/18.

<table>
<thead>
<tr>
<th></th>
<th>No expertise at all</th>
<th>Novice</th>
<th>Advanced beginner</th>
<th>Competent</th>
<th>Highly proficient</th>
<th>Regional expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Initiating or leading the analysis of interim instructional assessments with other educators frequently, at least every 6-9 weeks.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>b) Using formative and interim assessment data to build on student strengths and address student weaknesses.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>c) Regularly examining assessment items to determine why students are struggling.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>d) Using interim instructional assessment data to identify students who need additional time.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>e) Creatively finding ways to provide additional time for students who struggle.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>f) Creatively finding ways to provide other accommodations for students who struggle in addition to time.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
VI) High Impact Instruction: Learning Goals

Please respond to each question with your individual level of perceived expertise this past school year, 2017/18.

<table>
<thead>
<tr>
<th>No expertise at all</th>
<th>Novice</th>
<th>Advanced beginner</th>
<th>Competent</th>
<th>Highly proficient</th>
<th>Regional expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Knowing how to begin each unit by presenting students with clear learning targets.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Providing timely feedback on the extent to which students are achieving learning goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Effectively guiding students to keep track of their own progress of learning goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) Ensuring that students were involved in complex projects that required them to address the content in unique ways.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) Knowing how to end each unit by providing students with clear feedback on learning goals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>f) Knowing how to prepare students for the skill set of the next unit based on foundations of the previous unit(s).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### VII) High Impact Instruction: Engagement

Please respond to each question with your individual level of expertise this past school year, 2017/18.

<table>
<thead>
<tr>
<th>Question</th>
<th>No expertise at all</th>
<th>Novice</th>
<th>Advanced beginner</th>
<th>Competent</th>
<th>Highly proficient</th>
<th>Regional expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>a) Confidently choosing a balance between direct instruction and student math talk throughout each unit.</td>
<td>○ ○ ○ ○ ○ ○ ○ ○ ○ ○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b) Utilizing at least 25% of class time confidently for effective student discourse or math talk.</td>
<td>○ ○ ○ ○ ○ ○ ○ ○ ○ ○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c) Confidently organizing students for the exchange of ideas, including ways of representation, thinking aloud, agreeing and disagreeing.</td>
<td>○ ○ ○ ○ ○ ○ ○ ○ ○ ○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d) Delivering lessons so students do a few problems on a given concept each day over several days rather than many problems for only a couple of days.</td>
<td>○ ○ ○ ○ ○ ○ ○ ○ ○ ○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>e) Having multiple exposure to an idea over several days and spacing the practice of skills an extended period.</td>
<td>○ ○ ○ ○ ○ ○ ○ ○ ○ ○</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5) In this research study, all names were coded to prevent identification of students, teachers, and school buildings since the purpose is to identify instructional strategies that have the greatest impact. The intent of this survey is to measure student growth on state assessment after controlling for certain demographics.

Only for the purposes of a secondary math network across our region, you may choose to allow your name to be decoded. In the event that significant student growth will be attributed to a handful of coded teachers, decoding the teacher id will not be performed without the teacher’s permission and will never be published for research. Only for the ISD purposes of a secondary math network and the ability to apply for $5,000 of Marshal Plan funds. Please select one of the following:

- Only if I am one of the teachers with the highest amount of student achievement growth, please decode my identity and contact me prior to anyone else learning my identity so that I can share some of my best practices and decide if I want to share or apply for the $5,000 stipend.

- Even if I am one of the teachers the research suggests identifying based on the highest achievement, under no circumstances are you to decode my identity.

**Optional final thoughts before submitting**

6) What strategies or structures in place in your school do you believe have led to higher student achievement in your school that may not have been sufficiently covered on this survey?

7) What hurdles or structures do you believe have hindered higher student achievement in your school?

8) Any additional comments, suggestions, or questions for the researcher?
Appendix E

Additional Models for Research Questions 2 and 3
Models 2.3 through 2.9 for Research Question 2

Model 2.3: Prioritization Component 1

\[ \text{Prioritize}_{jk} = \beta_{0k} + \beta_{1k} (\text{ProfLearn}_{jk}) + \beta_{2k} (\text{Years}_{jk} - 1) \]

\[ + \beta_{3k} (\text{Degree}_{jk}) + r_{jk} \]

\[ \beta_{0k} = \gamma_{00} + \gamma_{01} (\text{BuildPD}_{0k}) + \gamma_{01} (\text{HS}_{0k}) + \gamma_{01} (\text{cED}_{0k}) + \gamma_{01} (\text{cNonCA}_{0k}) \]

\[ + \gamma_{01} (\text{cEL}_{0k}) + \gamma_{01} (\text{cSWD}_{0k}) + \gamma_{01} (\text{cFemale}_{0k}) + u_{0k} \]

\[ \beta_{1k} = \gamma_{10} \]

\[ \beta_{2k} = \gamma_{20} \]

\[ \beta_{3k} = \gamma_{30} \]

where \( \text{Prioritize}_{jk} \) is the average survey result for the first survey component prioritizing standards for teacher \( j \) in building \( k \), \( \beta_{0k} \) is the adjusted mean survey score for component one in building \( k \) for a first year teacher with a minor in mathematics who did not experience any professional learning around assessment utilization. \( \beta_{1k} \) is the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building \( k \) and \( \beta_{2k} \) to \( \beta_{3k} \) are the covariate effects for building \( k \). The level-1 model has one random effects component, \( r_{jk} \), which is the random effect associated with each teacher \( j \) in building \( k \). The building level or level-2 model is: where \( \gamma_{00} \) is the overall adjusted mean survey score for component one, \( \gamma_{01} \) is the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and \( \gamma_{10} \) to \( \gamma_{30} \) are controlled covariate effects. The level-2 model has one random effect components; \( u_{0k} \) is the random effect associated with the component one survey score for each school after controlling for the participation of some teachers in a professional learning series within a school.
Model 2.4: Supports Component 2

\[ \text{Supports}_{jk} = \beta_{0k} + \beta_{1k}(\text{ProfLearn}_{jk}) + \beta_{2k}(\text{Years}_{jk} - 1) + \beta_{3k}(\text{Degree}_{jk}) \] (2.4)

\[ + \eta_{jk} \]

\[ \beta_{0k} = \gamma_{00} + \gamma_{01}(\text{BuildPD}_{0k}) + \gamma_{01}(\text{HS}_{0k}) + \gamma_{01}(\text{cED}_{0k}) + \gamma_{01}(\text{cNonCA}_{0k}) \]

\[ + \gamma_{01}(\text{cEL}_{0k}) + \gamma_{01}(\text{cSWD}_{0k}) + \gamma_{01}(\text{cFemale}_{0k}) + u_{0k} \]

\[ \beta_{1k} = \gamma_{10} \]

\[ \beta_{2k} = \gamma_{20} \]

\[ \beta_{3k} = \gamma_{30} \]

where \( \text{Supports}_{jk} \) is the average survey result for the second survey component about supportive structures and relationships for teacher \( j \) in building \( k \), \( \beta_{0k} \) is the adjusted mean survey score for component two in building \( k \) for a first year teacher with a minor in mathematics who did not experience any professional learning around assessment utilization. \( \beta_{1k} \) is the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building \( k \) and \( \beta_{2k} \) to \( \beta_{3k} \) are the covariate effects for building \( k \). The level-1 model has one random effects component, \( \eta_{jk} \), which is the random effect associated with each teacher \( j \) in building \( k \). The building level or level-2 model is:

where \( \gamma_{00} \) is the overall adjusted mean survey score for component two, \( \gamma_{01} \) is the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and \( \gamma_{10} \) to \( \gamma_{30} \) are controlled covariate effects. The level-2 model has one random effect components; \( u_{0k} \) is the random effect associated with the component two survey score for each school after controlling for the participation of some teachers in a professional learning series within a school.
Model 2.5: Assessment Literacy Component 3

\[ \text{AssessLit} = \beta_{0k} + \beta_{1k}(\text{ProfLearn}_{jk}) + \beta_{2k}(\text{Years}_{jk} - 1) + \beta_{3k}(\text{Degree}_{jk}) \]  

+ \eta_{jk}

\[ \beta_{0k} = \gamma_{00} + \gamma_{01}(\text{BuildPD}_{0k}) + \gamma_{01}(\text{HS}_{0k}) + \gamma_{01}(\text{cED}_{0k}) + \gamma_{01}(\text{cNonCA}_{0k}) \]

+ \gamma_{01}(\text{cEL}_{0k}) + \gamma_{01}(\text{cSWD}_{0k}) + \gamma_{01}(\text{cFemale}_{0k}) + u_{0k}

\[ \beta_{1k} = \gamma_{10} \]

\[ \beta_{2k} = \gamma_{20} \]

\[ \beta_{3k} = \gamma_{30} \]

where \( \text{AssessLit}_{jk} \) is the average survey result for the third survey component about assessment literacy for teacher \( j \) in building \( k \), \( \beta_{0k} \) is the adjusted mean survey score for component three in building \( k \) for a first year teacher with a minor in mathematics who did not experience any professional learning around assessment utilization. \( \beta_{1k} \) is the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building \( k \) and \( \beta_{2k} \) to \( \beta_{3k} \) are the covariate effects for building \( k \). The level-1 model has one random effects component, \( \eta_{jk} \), which is the random effect associated with each teacher \( j \) in building \( k \). The building level or level-2 model is: where \( \gamma_{00} \) is the overall adjusted mean survey score for component three, \( \gamma_{01} \) is the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and \( \gamma_{10} \) to \( \gamma_{30} \) are controlled covariate effects. The level-2 model has one random effect components; \( u_{0k} \) is the random effect associated with the component three survey score for each school after controlling for the participation of some teachers in a professional learning series within a school.
Model 2.6: Quality Assessment Design Component 4

\[
Design_{jk} = \beta_{0k} + \beta_{1k}(ProfLearn_{jk}) + \beta_{2k}(Years_{jk} - 1) + \beta_{3k}(Degree_{jk}) + \eta_{jk}
\]  

\[
\beta_{0k} = \gamma_{00} + \gamma_{01}(BuildPD_{0k}) + \gamma_{01}(HS_{0k}) + \gamma_{01}(cED_{0k}) + \gamma_{01}(cNonCA_{0k}) + \gamma_{01}(cEL_{0k}) + \gamma_{01}(cSWD_{0k}) + \gamma_{01}(cFemale_{0k}) + u_{0k}
\]

\[
\beta_{1k} = \gamma_{10}
\]

\[
\beta_{2k} = \gamma_{20}
\]

\[
\beta_{3k} = \gamma_{30}
\]

where \(Design_{jk}\) is the average survey result for the fourth survey component about quality assessment design for teacher \(j\) in building \(k\), \(\beta_{0k}\) is the adjusted mean survey score for component four in building \(k\) for a first year teacher with a minor in mathematics who did not experience any professional learning around assessment utilization. \(\beta_{1k}\) is the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building \(k\) and \(\beta_{2k}\) to \(\beta_{3k}\) are the covariate effects for building \(k\). The level-1 model has one random effects component, \(\eta_{jk}\), which is the random effect associated with each teacher \(j\) in building \(k\). The building level or level-2 model is: where \(\gamma_{00}\) is the overall adjusted mean survey score for component four, \(\gamma_{01}\) is the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and \(\gamma_{10}\) to \(\gamma_{30}\) are controlled covariate effects. The level-2 model has one random effect components; \(u_{0k}\) is the random effect associated with the component four survey score for each school after controlling for the participation of some teachers in a professional learning series within a school.
Model 2.7: Analysis Component 5

$$Analysis_{jk} = \beta_{0k} + \beta_{1k}(ProfLearn_{jk}) + \beta_{2k}(Years_{jk} - 1) + \beta_{3k}(Degree_{jk})$$ (2.7) + \eta_{jk}

$$\beta_{0k} = \gamma_{00} + \gamma_{01}(BuildPD_{0k}) + \gamma_{01}(HS_{0k}) + \gamma_{01}(cED_{0k}) + \gamma_{01}(cNonCA_{0k}) + \gamma_{01}(cELD_{0k}) + \gamma_{01}(cSWD_{0k}) + \gamma_{01}(cFemale_{0k}) + u_{0k}$$

$$\beta_{1k} = \gamma_{10}$$

$$\beta_{2k} = \gamma_{20}$$

$$\beta_{3k} = \gamma_{30}$$

where $Analysis_{jk}$ is the average survey result for the fifth survey component about data analysis that leads to action for teacher $j$ in building $k$, $\beta_{0k}$ is the adjusted mean survey score for component five in building $k$ for a first year teacher with a minor in mathematics who did not experience any professional learning around assessment utilization. $\beta_{1k}$ is the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building $k$ and $\beta_{2k}$ to $\beta_{3k}$ are the covariate effects for building $k$. The level-1 model has one random effects component, $\eta_{jk}$, which is the random effect associated with each teacher $j$ in building $k$. The building level or level-2 model is: where $\gamma_{00}$ is the overall adjusted mean survey score for component five, $\gamma_{01}$ is the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and $\gamma_{10}$ to $\gamma_{30}$ are controlled covariate effects. The level-2 model has one random effect components; $u_{0k}$ is the random effect associated with the component five survey score for each school after controlling for the participation of some teachers in a professional learning series within a school.
Model 2.8: Instruction Learning Goals Component 6

\[ \text{LearnGoal}_{jk} = \beta_{0k} + \beta_{1k}(\text{ProfLearn}_{jk}) + \beta_{2k}(\text{Years}_{jk} - 1) + \beta_{3k}(\text{Degree}_{jk}) + r_{jk} \]

\[ \beta_{0k} = \gamma_{00} + \gamma_{01}(\text{BuildPD}_{0k}) + \gamma_{01}(\text{HS}_{0k}) + \gamma_{01}(\text{cED}_{0k}) + \gamma_{01}(\text{cNonCA}_{0k}) + \gamma_{01}(\text{cEL}_{0k}) + \gamma_{01}(\text{cSWD}_{0k}) + \gamma_{01}(\text{cFemale}_{0k}) + u_{0k} \]

\[ \beta_{1k} = \gamma_{10} \]

\[ \beta_{2k} = \gamma_{20} \]

\[ \beta_{3k} = \gamma_{30} \]

where \( \text{LearnGoal}_{jk} \) is the average survey result for the sixth survey component about high impact instruction: learning goals for teacher \( j \) in building \( k \), \( \beta_{0k} \) is the adjusted mean survey score for component six in building \( k \) for a first year teacher with a minor in mathematics who did not experience any professional learning around assessment utilization. \( \beta_{1k} \) is the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building \( k \) and \( \beta_{2k} \) to \( \beta_{3k} \) are the covariate effects for building \( k \). The level-1 model has one random effects component, \( r_{jk} \), which is the random effect associated with each teacher \( j \) in building \( k \). The building level or level-2 model is: where \( \gamma_{00} \) is the overall adjusted mean survey score for component six, \( \gamma_{01} \) is the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and \( \gamma_{10} \) to \( \gamma_{30} \) are controlled covariate effects. The level-2 model has one random effect components; \( u_{0k} \) is the random effect associated with the component six survey score for each school after controlling for the participation of some teachers in a professional learning series within a school.
**Model 2.9: Instruction Engagement Component 7**

\[
Engage_{jk} = \beta_{0k} + \beta_{1k}(Prof\text{Learn}_{jk}) + \beta_{2k}(Years_{jk} - 1) + \beta_{3k}(Degree_{jk}) + \eta_{jk}
\]

\[
\beta_{0k} = \gamma_{00} + \gamma_{01}(BuildPD_{0k}) + \gamma_{01}(HS_{0k}) + \gamma_{01}(cED_{0k}) + \gamma_{01}(cNonCA_{0k}) + \gamma_{01}(cEL_{0k}) + \gamma_{01}(cSWD_{0k}) + \gamma_{01}(cFemale_{0k}) + u_{0k}
\]

\[
\beta_{1k} = \gamma_{10}
\]

\[
\beta_{2k} = \gamma_{20}
\]

\[
\beta_{3k} = \gamma_{30}
\]

where \(Engage_{jk}\) is the average survey result for the final survey component high impact instruction: engagement for teacher \(j\) in building \(k\), \(\beta_{0k}\) is the adjusted mean survey score for component seven in building \(k\) for a first year teacher with a minor in mathematics who did not experience any professional learning around assessment utilization. \(\beta_{1k}\) is the average difference in survey scores between teachers who participated to some degree in the professional learning and teachers who did not in building \(k\) and \(\beta_{2k}\) to \(\beta_{3k}\) are the covariate effects for building \(k\). The level-1 model has one random effects component, \(\eta_{jk}\), which is the random effect associated with each teacher \(j\) in building \(k\). The building level or level-2 model is: where \(\gamma_{00}\) is the overall adjusted mean survey score for component seven, \(\gamma_{01}\) is the average difference between means of two schools that differ by participation in sending teachers to some degree of professional learning around assessment utilization, and \(\gamma_{10}\) to \(\gamma_{30}\) are controlled covariate effects. The level-2 model has one random effect components; \(u_{0k}\) is the random effect associated with the component seven survey score for each school after controlling for the participation of some teachers in a professional learning series within a school.
Models 3.2 through 3.8 for Research Question 3

Model 3.2: Supportive Relationships and Structures

Level 1:
\[ MathZ_{ijk} = \pi_{0jk} + \pi_{1jk}(PreTz_{ijk}) + \pi_{2jk}(ED_{ijk}) + \pi_{3jk}(SWD_{ijk}) + \pi_{4jk}(EL_{ijk}) + \pi_{5jk}(NonCA_{ijk}) + \pi_{6jk}(Fem_{ijk}) + e_{ijk} \]

Level 2:
\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(Supportive_{0jk}) + \beta_{02k}(Course_{0jk}) + \beta_{03k}(Years_{0jk} - 1) + \beta_{04k}(Degree_{0jk}) + r_{0jk} \]

\[ \pi_{pjk} = \beta_{p0k} \text{ for all covariates where } p = 1 \text{ to } 6 \]

Level 3:
\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(cPreTz_{00k}) + \gamma_{002}(cED_{00k}) + \gamma_{003}(cSWD_{00k}) + \gamma_{004}(cEL_{00k}) + \gamma_{005}(cNonCA_{00k}) + \gamma_{006}(HS_{00k}) + u_{00k} \]

\[ \beta_{p0k} = \gamma_{p00} \text{ for all covariates where } p = 1 \text{ to } 6 \]

where \( MathZ_{ijk} \) is the standardized math \( z \)-score for a student \( i \) taught by teacher \( j \) in building \( k \), \( \pi_{0jk} \) is the mean math achievement of teacher \( j \) in building \( k \), \( \beta_{00k} \) is the mean math achievement in building \( k \), \( \beta_{01k} \) is the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \( Supportive_{0jk} \) is the survey component score for supportive relationships and structures for teacher \( j \) in building \( k \), and \( \gamma_{000} \) is the grand mean for the state which is zero based on \( z \)-scores. The model contains covariates at the student level, teacher level and grand centered at the building. The model also contains three random effect components such that \( e_{ijk} \) is the random effect associated with each student \( i \) connected to teacher \( j \) in building \( k \), \( r_{0jk} \) is the random effect associated for each teacher \( j \), and \( u_{00k} \) is the random effect associated with the average math achievement score for building \( k \).
Model 3.3: Assessment Literacy and Autonomy

Level 1:
\[
\text{Math}Z_{ijk} = \pi_{0jk} + \pi_{1jk}(\text{PreT}z_{ijk}) + \pi_{2jk}(E_D_{ijk}) + \pi_{3jk}(\text{SWD}_{ijk}) + \pi_{4jk}(E_L_{ijk}) + \\
\pi_{5jk}(\text{NonCA}_{ijk}) + \pi_{6jk}(\text{Fem}_{ijk}) + e_{ijk}
\]

Level 2:
\[
\pi_{0jk} = \beta_{00k} + \beta_{01k}(\text{AssessLit}_{0jk}) + \beta_{02k}(\text{Course}_{0jk}) + \beta_{03k}(\text{Years}_{0jk} - 1) \\
+ \beta_{04k}(\text{Degree}_{0jk}) + r_{0jk}
\]
\[
\pi_{pjk} = \beta_{p0k} \quad \text{for all covariates where } p = 1 \text{ to } 6
\]

Level 3:
\[
\beta_{00k} = \gamma_{000} + \gamma_{001}(\text{PreT}z_{00k}) + \gamma_{002}(E_D_{00k}) + \gamma_{003}(\text{SWD}_{00k}) + \gamma_{004}(E_L_{00k}) + \\
\gamma_{005}(\text{NonCA}_{00k}) + \gamma_{006}(\text{HS}_{00k}) + u_{00k}
\]
\[
\beta_{p0k} = \gamma_{p00} \quad \text{for all covariates where } p = 1 \text{ to } 6
\]

where \(\text{Math}Z_{ijk}\) is the standardized math z-score for a student \(i\) taught by teacher \(j\) in building \(k\), \(\pi_{0jk}\) is the mean math achievement of teacher \(j\) in building \(k\), \(\beta_{00k}\) is the mean math achievement in building \(k\), \(\beta_{01k}\) is the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \(\text{AssessLit}_{0jk}\) is the survey component score for assessment literacy and autonomy for teacher \(j\) in building \(k\), and \(\gamma_{000}\) is the grand mean for the state which is zero based on z-scores. The model contains covariates at the student level, teacher level and grand centered at the building. The model also contains three random effect components such that \(e_{ijk}\) is the random effect associated with each student \(i\) connected to teacher \(j\) in building \(k\), \(r_{0jk}\) is the random effect associated for each teacher \(j\), and \(u_{00k}\) is the random effect associated with the average math achievement score for building \(k\).
Model 3.4: Quality Assessment Design

Level 1:

\[ MathZ_{ijk} = \pi_{0jk} + \pi_{1jk}(PreTz_{ijk}) + \pi_{2jk}(ED_{ijk}) + \pi_{3jk}(SWD_{ijk}) + \pi_{4jk}(EL_{ijk}) + \pi_{5jk}(NonCA_{ijk}) + \pi_{6jk}(Fem_{ijk}) + e_{ijk} \]

Level 2:

\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(Design_{0jk}) + \beta_{02k}(Course_{0jk}) + \beta_{03k}(Years_{0jk} - 1) + \beta_{04k}(Degree_{0jk}) + r_{0jk} \]

\[ \pi_{pjk} = \beta_{p0k} \text{ for all covariates where } p = 1 \text{ to } 6 \]

Level 3:

\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(cPreTz_{00k}) + \gamma_{002}(cED_{00k}) + \gamma_{003}(cSWD_{00k}) + \gamma_{004}(cEL_{00k}) + \gamma_{005}(cNonCA_{00k}) + \gamma_{006}(HS_{00k}) + u_{00k} \]

\[ \beta_{p0k} = \gamma_{p00} \text{ for all covariates where } p = 1 \text{ to } 6 \]

where \( MathZ_{ijk} \) is the standardized math z-score for a student \( i \) taught by teacher \( j \) in building \( k \), \( \pi_{0jk} \) is the mean math achievement of teacher \( j \) in building \( k \), \( \beta_{00k} \) is the mean math achievement in building \( k \), \( \beta_{01k} \) is the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \( Design_{0jk} \) is the survey component score for quality assessment design for teacher \( j \) in building \( k \), and \( \gamma_{000} \) is the grand mean for the state which is zero based on z-scores. The model contains covariates at the student level, teacher level and grand centered at the building. The model also contains three random effect components such that \( e_{ijk} \) is the random effect associated with each student \( i \) connected to teacher \( j \) in building \( k \), \( r_{0jk} \) is the random effect associated for each teacher \( j \), and \( u_{00k} \) is the random effect associated with the average math achievement score for building \( k \).
Model 3.5: Analysis that leads to Action

Level 1:

\[ MathZ_{ijk} = \pi_{0jk} + \pi_{1jk}(PreTz_{ijk}) + \pi_{2jk}(ED_{ijk}) + \pi_{3jk}(SWD_{ijk}) + \pi_{4jk}(EL_{ijk}) + \pi_{5jk}(NonCA_{ijk}) + \pi_{6jk}(Fem_{ijk}) + e_{ijk} \]

Level 2:

\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(Analysis_{0jk}) + \beta_{02k}(Course_{0jk}) + \beta_{03k}(Years_{0jk} - 1) + \beta_{04k}(Degree_{0jk}) + r_{0jk} \]

\[ \pi_{pjk} = \beta_{p0k} \text{ for all covariates where } p = 1 \text{ to } 6 \]

Level 3:

\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(cPreTz_{00k}) + \gamma_{002}(cED_{00k}) + \gamma_{003}(cSWD_{00k}) + \gamma_{004}(cEL_{00k}) + \gamma_{005}(cNonCA_{00k}) + \gamma_{006}(HS_{00k}) + u_{00k} \]

\[ \beta_{p0k} = \gamma_{p00} \text{ for all covariates where } p = 1 \text{ to } 6 \]

where \( MathZ_{ijk} \) is the standardized math z-score for a student \( i \) taught by teacher \( j \) in building \( k \), \( \pi_{0jk} \) is the mean math achievement of teacher \( j \) in building \( k \), \( \beta_{00k} \) is the mean math achievement in building \( k \), \( \beta_{01k} \) is the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \( Analysis_{0jk} \) is the survey component score for analysis that leads to action for teacher \( j \) in building \( k \), and \( \gamma_{000} \) is the grand mean for the state which is zero based on z-scores. The model contains covariates at the student level, teacher level and grand centered at the building. The model also contains three random effect components such that \( e_{ijk} \) is the random effect associated with each student \( i \) connected to teacher \( j \) in building \( k \), \( r_{0jk} \) is the random effect associated for each teacher \( j \), and \( u_{00k} \) is the random effect associated with the average math achievement score for building \( k \).
Model 3.6: High Impact Instruction: Learning Goals

Level 1:

\[ MathZ_{ijk} = \pi_{0jk} + \pi_{1jk}(PreTz_{ijk}) + \pi_{2jk}(ED_{ijk}) + \pi_{3jk}(SWD_{ijk}) + \pi_{4jk}(EL_{ijk}) + \pi_{5jk}(NonCA_{ijk}) + \pi_{6jk}(Fem_{ijk}) + e_{ijk} \]

Level 2:

\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(LearnGoal_{0jk}) + \beta_{02k}(Course_{0jk}) + \beta_{03k}(Years_{0jk} - 1) + \beta_{04k}(Degree_{0jk}) + r_{0jk} \]

\[ \pi_{pjk} = \beta_{p0k} \text{ for all covariates where } p = 1 \text{ to } 6 \]

Level 3:

\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(cPreTz_{00k}) + \gamma_{002}(cED_{00k}) + \gamma_{003}(cSWD_{00k}) + \gamma_{004}(cEL_{00k}) + \gamma_{005}(cNonCA_{00k}) + \gamma_{006}(HS_{00k}) + u_{00k} \]

\[ \beta_{p0k} = \gamma_{p00} \text{ for all covariates where } p = 1 \text{ to } 6 \]

where \( MathZ_{ijk} \) is the standardized math z-score for a student \( i \) taught by teacher \( j \) in building \( k \), \( \pi_{0jk} \) is the mean math achievement of teacher \( j \) in building \( k \), \( \beta_{00k} \) is the mean math achievement in building \( k \), \( \beta_{01k} \) is the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \( LearnGoal_{0jk} \) is the survey component score for high impact instruction: learning goals, for teacher \( j \) in building \( k \), and \( \gamma_{000} \) is the grand mean for the state which is zero based on z-scores. The model contains covariates at the student level, teacher level and grand centered at the building. The model also contains three random effect components such that \( e_{ijk} \) is the random effect associated with each student \( i \) connected to teacher \( j \) in building \( k \), \( r_{0jk} \) is the random effect associated for each teacher \( j \), and \( u_{00k} \) is the random effect associated with the average math achievement score for building \( k \).
Model 3.7: High Impact Instruction: Engagement

Level 1:
\[ MathZ_{ijk} = \pi_{0jk} + \pi_{1jk}(PreTz_{ijk}) + \pi_{2jk}(ED_{ijk}) + \pi_{3jk}(SWD_{ijk}) + \pi_{4jk}(EL_{ijk}) + \pi_{5jk}(NonCA_{ijk}) + \pi_{6jk}(Fem_{ijk}) + e_{ijk} \]

Level 2:
\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(Engage_{0jk}) + \beta_{02k}(Course_{0jk}) + \beta_{03k}(Years_{0jk} - 1) + \beta_{04k}(Degree_{0jk}) + r_{0jk} \]
\[ \pi_{pjk} = \beta_{p0k} \text{ for all covariates where } p = 1 \text{ to } 6 \]

Level 3:
\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(cPreTz_{00k}) + \gamma_{002}(cED_{00k}) + \gamma_{003}(cSWD_{00k}) + \gamma_{004}(cEL_{00k}) + \gamma_{005}(cNonCA_{00k}) + \gamma_{006}(HS_{00k}) + u_{00k} \]
\[ \beta_{p0k} = \gamma_{p00} \text{ for all covariates where } p = 1 \text{ to } 6 \]

where \( MathZ_{ijk} \) is the standardized math z-score for a student \( i \) taught by teacher \( j \) in building \( k \), \( \pi_{0jk} \) is the mean math achievement of teacher \( j \) in building \( k \), \( \beta_{00k} \) is the mean math achievement in building \( k \), \( \beta_{01k} \) is the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \( Engage_{0jk} \) is the survey component score for high impact instruction: engagement, for teacher \( j \) in building \( k \), and \( \gamma_{000} \) is the grand mean for the state which is zero based on z-scores. The model contains covariates at the student level, teacher level and grand centered at the building. The model also contains three random effect components such that \( e_{ijk} \) is the random effect associated with each student \( i \) connected to teacher \( j \) in building \( k \), \( r_{0jk} \) is the random effect associated for each teacher \( j \), and \( u_{00k} \) is the random effect associated with the average math achievement score for building \( k \).
Model 3.7: Prioritization of Standards

Level 1:

\[ MathZ_{ijk} = \pi_{0jk} + \pi_{1jk}(PreTZ_{ijk}) + \pi_{2jk}(ED_{ijk}) + \pi_{3jk}(SWD_{ijk}) + \pi_{4jk}(EL_{ijk}) + \pi_{5jk}(NonCA_{ijk}) + \pi_{6jk}(Fem_{ijk}) + e_{ijk} \]

Level 2:

\[ \pi_{0jk} = \beta_{00k} + \beta_{01k}(Prioritize_{0jk}) + \beta_{02k}(Course_{0jk}) + \beta_{03k}(Years_{0jk} - 1) + \beta_{04k}(Degree_{0jk}) + r_{0jk} \]

\[ \pi_{pjk} = \beta_{p0k} \text{ for all covariates where } p = 1 \text{ to } 6 \]

Level 3:

\[ \beta_{00k} = \gamma_{000} + \gamma_{001}(cPreTZ_{00k}) + \gamma_{002}(cED_{00k}) + \gamma_{003}(cSWD_{00k}) + \gamma_{004}(cEL_{00k}) + \gamma_{005}(cNonCA_{00k}) + \gamma_{006}(HS_{00k}) + u_{00k} \]

\[ \beta_{p0k} = \gamma_{p00} \text{ for all covariates where } p = 1 \text{ to } 6 \]

where \( MathZ_{ijk} \) is the standardized math \( z \)-score for a student \( i \) taught by teacher \( j \) in building \( k \), \( \pi_{0jk} \) is the mean math achievement of teacher \( j \) in building \( k \), \( \beta_{00k} \) is the mean math achievement in building \( k \), \( \beta_{01k} \) is the rate of change between teachers when the survey component increases by one unit on a six-point Likert scale, \( Prioritize_{0jk} \) is the survey component score for the prioritization of standards for teacher \( j \) in building \( k \), and \( \gamma_{000} \) is the grand mean for the state which is zero based on \( z \)-scores. The model contains covariates at the student level, teacher level and grand centered at the building. The model also contains three random effect components such that \( e_{ijk} \) is the random effect associated with each student \( i \) connected to teacher \( j \) in building \( k \), \( r_{0jk} \) is the random effect associated for each teacher \( j \), and \( u_{00k} \) is the random effect associated with the average math achievement score for building \( k \).
Appendix F

Administrator Overview of Training
Administrator overview of process for building interim assessments


- Primary purpose of assessments is to be instructional, secondary purposes TBD
- Teacher trust and buy-in established throughout the process as assessments are built by teachers with guidance from the ISD
- Standards prioritized in order to guarantee learning in a viable time frame (essential, important and nice to know)
- Assessments to follow the sequence of instruction common across districts involved
- Assessments divided into 4 – 6 units of study per HS course (resulting in time intervals between 6-9 weeks supported by research)
- Districts will set aside time following assessments for analysis and re-teaching

Assessments — “Standards are meaningless until you define how to assess them” Bambrick-Santoyo (2010)

- Assessments are transparent and are the starting point of instruction
- Interim assessment covers all essential standards with the option to address some important standards provided all standards have sufficient evidence
- High quality assessment items, multiple formats, levels, and assessable components of each standard are attended to in building each assessment
- Depth of Knowledge varies for each standard and used to help define success criteria
- Aligned to state assessments and college ready expectations in terms of complexity and difficulty
- Aligned to state assessment in format and length when possible (consider online delivery and performance tasks)
- Spirals by revisiting essential standards on subsequent assessments (about 1/3 of test)
- Diagnostic assessment: moving from who did what, to why

Analysis — Moving from “who did what wrong” to “why,” which then leads to “what next”

- The OAISD 3-step analysis process combines multiple research recommendations
- Utilizing inQwizIT and/or other user-friendly data displays
- Teachers invest in scoring free response (FR) section and verifying short answer(s)
- Selected response Qs scored automatically, teachers scoring FR section, and analysis to be finalized ideally within 48 hours but no longer than one week
- Analysis NOT to be done in isolation, collaboration and dialogue recommended


- Data connected to standards and instruction
- Effective follow-up attached to re-teaching action plan, supported by administration
- Provide effective feedback for individual students
- Enable students to track their own improvement on the essential standards
- Monitoring, support and follow-up by the building level administrator(s)
**Analysis** — Moving from “who did what wrong” to why, which leads to instructional strategies

**Part 1 - Global Impressions**
Using the “Course Overview” inQwizIT report, make observations about the performance of students and the standards. Simply make observations without making excuses, drawing inferences or conclusions based on global data. Record on template provided (paper or electronic) or record on the ‘Intervention Group’ inQwizIT report once printed.

- How well did the class do as a whole?
- Which standards seem to be the strongest? Should these be retaught to only some?
- Which standards seem to be the weakest? Should these be retaught to the whole group?
- Who seems to be the strongest students? Is this expected?
- Who seems to be the most struggling students? Is this expected?
- Do you see any trends with students with disabilities, English Learners or a specific ethnicity or gender?

**Part 2 - Item Analysis … WHY?**
At this time, it is appropriate to make conjectures, inferences and establish causal theory. You may record this directly on the assessment or on the template provided (paper or electronic).

- Look at each question, particularly those that were ‘bombed,’ did students all choose the same wrong answer? Why was a specific distractor chosen or why was a common mistake made? (NOTE: you may want to look at questions grouped by standard)
  - Did students make a simple mistake?
  - Does it seem as though the students missed an important concept?
  - Was something about the test format difficult?
  - Did students possibly struggle with any vocabulary?
- Break down each standard, did students do similarly on every question or were some questions more complex? Why might students struggle on some questions but not others?
- Compare similar standards, do the results of one standard influence the other standard? If so, what impact might this have on their results?

**Part 3 – Action Planning**
You should record this on the blank ‘Action Planning’ template, by hand or electronically.

- What will be done to help students meet benchmark?
  - Focus on small details and/or test format, reteach academic vocabulary and/or conceptual understandings
- When will students be retaught?
  - During classroom warm-up's (bellringer), independent practice time, reteaching days...
- How will students be retaught?
  - Whole group or small group instruction, teacher-lead instruction, student-lead instruction, video instructional
- How will students demonstrate mastery of the standard?

Keep the analysis focused on standards rather than an overall grade!
Appendix G

Permission from J. Hattie
Permission from J. Hattie

From: John Hattie [mailto:jhattie@unimelb.edu.au]
Sent: Saturday, March 31, 2018 8:00 PM
To: Doug Greer <dgreer@oaisd.org>
Subject: RE: Seeking permission concerning "Barometer of Influence"

More than happy for you to use

Best wishes for the Ph.D

John Hattie
Laureate Professor
Melbourne Graduate School of Education
100 Leicester St, University of Melbourne, Carlton, Victoria, AUSTRALIA 3010

Chair, Board of the Australian Institute for Teaching and School Leadership
Associate Director of the ARC-SRI: Science of Learning Research Centre:
http://www.slrc.org.au
orcid.org/0000-0003-3873-4854

From: Doug Greer [mailto:dgreer@oaisd.org]
Sent: Sunday, 1 April 2018 2:48 AM
To: John Hattie <jhattie@unimelb.edu.au>
Subject: Seeking permission concerning "Barometer of Influence"

Dr. Hattie,

I would like to thank you for your contributions to education and your continual updates to improve teaching and visible learning. I have seen you present in Michigan in the past and we sited your research when building interim instructional assessments across our west Michigan region. I am now pursuing my Ph.D. and I have been working on my dissertation proposal. In the literature review, I cite your work around desired effects that align with the utilization of interim instructional assessments. I believe the reader would benefit from the figure commonly referred to as the Barometer of Influence. I am wondering, may I have your permission to include this figure with a proper citation?
Thank you for your consideration and I hope you enjoyed the Easter weekend.

Take care,

Doug Greer
Director of School Improvement
Ottawa Area ISD
877-702-8600 x4109

Practical SI Timeline
Twitter: @Doug_Greer4
Appendix H

Permission to Use De-identified Data for Research
August 20, 2018

Office of the Vice President for Research
Western Michigan University
1903 W Michigan Ave
Kalamazoo, MI 49008

To Whom It May Concern:

It is an honor to offer this letter of assurance on behalf of Doug Greer. In his time at OAISD, Doug has been instrumental in facilitating a systems approach to educational practice. He, more than most, has drawn connections between previously isolated programs and departments and has helped to facilitate coordination of these formerly independent efforts into much more effective delivery. He bridges the work of career and technical education, special education, assessment and improvement, across our intermediate school district as well as with our local school districts he serves. His service mindset has driven him, with our blessing, to impact educational programming far beyond the Ottawa area.

As Mr. Greer approaches candidacy status in his Ph.D. program, it has been a pleasure to watch him marry theory, meaningful research, and operational practice on a daily basis. The connections are seamless, unlike many such efforts. Presently, Doug is leading a study around secondary mathematics with the blessing of our local superintendents. His position as Director of School Improvement allows him access to student achievement data and teacher roster files needed for his study to inform a secondary math network. In addition, Doug has received permission to de-identify the data to use as a part of his dissertation study, titled A Relationship between Student Math Achievement and Teachers Utilizing a Process involving Instructional Interim Assessments. We look forward to learning from the study.

Please do not hesitate to contact me directly if you have any questions.

Sincerely,

Peter Haines, superintendent
616-738-8940 ext. 4001
PHaines@oaisd.org