24 KARAT OR FOOL'S GOLD? AN ANALYSIS OF THE IMPACT OF INTERIM ASSESSMENT DATA ON CLASSROOM TEACHER INSTRUCTIONAL DECISIONS

A Dissertation presented to

the Faculty of the Graduate School at the University of Missouri-Columbia

---

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Educational Leadership & Policy Analysis

---

by STEVEN SNEAD

Dr. Se Woong Lee, Dissertation Supervisor

July, 2024

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

24 KARAT OR FOOL'S GOLD? AN ANALYSIS OF THE IMPACT OF INTERIM ASSESSMENT DATA ON CLASSROOM TEACHER INSTRUCTIONAL DECISIONS

presented by Steven Snead,

a candidate for the degree of Doctor of Educational Leadership & Policy Analysis and hereby certify that, in their opinion, it is worthy of acceptance.

_____

Dr. Se Woong Lee

_____

Dr. Lisa Dorner

_____

Dr. Beth Whitaker

_____

Dr. Sarah Marriott

ACKNOWLEDGMENTS

I would first like to thank the hardworking teachers of Oakland County who participated in this study. I have the daily privilege of working with teachers, principals, and a wide variety of instructional leaders across our beautiful, diverse county, and I can assure you that you will not find a more passionate and dedicated group of educational professionals. The general public has only a small window into the daily work of teachers and might not fully understand the complexity of the work that teachers bring into the classroom each and every day. We cannot thank our teachers enough. Yet, I hope this dissertation is one small contribution that helps to elevate the voice of teachers within the educational research community, honoring their experiences, expertise, and professionalism.

I would also like to acknowledge my wife, Yolanda Snead, for supporting me in all my educational and professional endeavors. I could not have possibly begun this educational journey without her support, from late-night cups of coffee to power me up during extended writing sessions, or even the occasional word of affirmation, pushing me to keep going. There is perhaps no more precious gift to a person than a strong, wise, and steadfast partner in life. I thank God that He blessed me with her.

Finally, I would like to acknowledge the many friends, family, and professional associates who inspired me and pushed me in this doctoral journey. In particular, I would like to thank Dr. David Snead, Dr. Brandon Snead, and Dr. Marcus Collins for their counsel and words of support.

TABLE OF CONTENTS

ABSTRACT

Data-based decision-making has been a frequently used policy and practice intervention used in schools to help inform the decision-making processes of educational practitioners, with the aim of improving student outcomes. Interim benchmark assessments are designed by commercial test developers to support educators in this framework. In fact, the state of Michigan has legally mandated the use of interim benchmarks in public schools to accelerate learning as a result of COVID-19 pandemic-related school closures. However, while student assessment instruments are core tools of data-based decision-making, not all assessments are equally useful to educators, particularly classroom teachers. There is a lack of evidence that shows how teachers use data from interim benchmark assessments in their practice and if there are differences in perceptions of the usefulness of this data across the various assessment tools that exist in the market. This study seeks to explore teacher perceptions about the usefulness of the data interim benchmark assessments provide and how frequently teachers use this data in their educational decisions. Using survey data from classroom teachers, this study employs quantitative methods to provide descriptive statistics on the degree of usefulness of interim benchmark assessments and the frequency of teachers' use of these data in their educational decisions. Though this study is limited to teachers in Michigan, the results of the study have national implications for how policymakers can support classroom-level educational decisions through the use of assessment data.

Keywords: Interim Benchmark Assessment, Student Testing, Data-Based

SECTION 1 – INTRODUCTION TO THE DISSERTATION-IN-PRACTICE

Background

In Michigan, as in every state across the United States, the COVID-19 pandemic significantly disrupted nearly every aspect of life. These disruptions were perhaps no more pronounced, prolonged, and profound than in the K-12 schools (Huck & Zang, 2021). In Michigan, Governor Gretchen Whitmer used the emergency powers of the governor's office to order schools to close buildings, shifting from in-person learning to virtual learning for three months. Many schools were closed to in-person learning for up to one full school year (Kilbride et al., 2021). As a result of these disruptions to in-person learning, the Michigan Department of Education paused the administration of its state assessment system due to the inability of schools to administer these tests in a remote environment. The data these state tests provide are designed to be used by educators and policymakers to make critical decisions to support student learning, including identifying students for additional support, improving school curriculum, staffing, and even budget priorities (Michigan Department of Education, 2022). Key educational decision- makers, including teachers, were left without a necessary tool to support student learning without these data.

In response to the need for more data from the canceled state assessment system, the Michigan Legislature required schools to administer interim assessments to students in grades K through 8 in Reading and Mathematics. Although several commercial test developers publish and sell interim benchmark assessments to schools, they share common features (Marion et al., 2007; Perie et al., 2007). They all use a select response

testing methodology to determine a student's level of knowledge and skill in an academic domain, providing teachers and school leaders with student performance data to support a variety of educational decisions. These tests are designed to be administered to students online via computer and do not require in-person administration. Unlike state assessments, interim benchmark assessments produce several instantaneous performance results available to teachers, school leaders, students, and their families. Many schools in Michigan have used interim benchmark assessments before this legislation for various functions, but this represented the first time that these assessments were codified into law (Kilbride et al., 2021).

Although pandemic-related school closures have ended and schools have resumed normal operations, including the resumption of the state assessment system, the Michigan legislature has continued to require schools to administer these assessments as a condition of receiving school funding (Michigan Legislature, 2022). The stated purpose of this continuation is to study student learning loss and provide educators and policymakers with information to support key educational decisions in a broader effort to accelerate student learning (Kilbride et al., 2022).

Though this well-intended law received bi-partisan support from Michigan elected officials, there is little to no research evidence to suggest that school implementation of interim benchmark assessments has any significant impact on teacher and school practices, and resultingly, no significant gains in student outcomes have been attributed to school implementation of benchmark assessment data (Marion et al., 2007; Chojnack et al., 2013; Slavin et al., 2013).

There may be an assumption underlying the law and the pre-pandemic

implementation of interim benchmark assessments that simply having academic data enables teachers to make better decisions to support improved teaching and learning, or so-called data-based making. (Marion et al., 2007; Buckley, et al., 2010). Data-based data-based-making has been a popular policy and practice used by schools to inform and modify educational decisions on an ongoing basis to improve educational outcomes for students (Hargreaves et al., 2013). However, prior research that has attempted to study education practitioner implementation of data-based decision- making, as well as this strategy's effect on student achievement, has shown mixed results (Carlson et al., 2017; Visscher, 2021; Peters et al., 2021; Ikemoto & Marsh, 2007). Furthermore, there is a lack of evidence that shows how teachers use interim benchmark assessments, in particular, in an overall data-based decision framework applied within the classroom. Despite this lack of evidence, Michigan has continued to allocate millions of dollars to reimburse districts for the cost of these tests. Additionally, to administer these assessments, schools must allocate an indeterminate level of resources and funding to run these assessments, including technology, internet infrastructure, human capital, and time (Michigan House Fiscal Agency, 2021; Michigan Senate Fiscal Agency, 2021). Given the significant allocation of resources dedicated to interim benchmark assessments, the need for robust research evidence of their impact on teacher practice represents an opportunity to study the various types of instructional decisions that teachers make using benchmark assessment data.

Previous researchers and interim benchmark assessment publishers suggest that teachers can use interim benchmark data in various ways. Researchers have theorized that these data can help teachers select the professional development they need in specific

areas of their own instructional practice and measure their own instructional effectiveness (Buckley et al., 2010; Olah et al., 2010). They also suggest that teachers can use interim benchmark assessment data to create instructional groups, plan instruction, identify students for intervention, and set learning goals for individual students (Goertz et al., 2009). However, though researchers have suggested that these instructional decisions are theoretically possible, there needs to be more evidence to show what teachers actually do with these data. This presents a compelling case to engage with classroom teachers directly about the ways in which they use these data in their instructional decisions. With this knowledge, policymakers can improve policies aimed at providing support to teachers to improve student educational outcomes.

Additionally, as numerous interim benchmark assessments are available on the market, each assessment tool provides teachers with a unique and proprietary reporting system of student assessment results. In Michigan, the most widely used assessments are MAP, published by Northwest Education Association (NWEA), iReady, published by Curriculum Associates, and STAR, published by Renaissance Education (Kilbride et al., 2022). Each of these publishers provides academic information to teachers on students in different ways, thus potentially leading to differences in teacher perception of the usability of the reporting these systems provide.

Teacher perception of the usability of reports is a compelling phenomenon to study because there is an emerging body of evidence that shows that teacher perception of the usability of instructional resources impacts the degree to which they use that resource in instruction (Tuma et al., 2021; van Leeuwen, 2019). Moreover, specifically regarding assessment tools, researchers have found that teachers' affective attitudes

toward using assessment data have been shown to influence how they use that data (Prenger & Schildkamp, 2016). Although this study will be limited to researching how Michigan teachers use these tests under Michigan policy enactment, these tests are used across the United States and, in NWEA's case, even across the globe. Therefore, the results of this study can inform a broad array of data-based decision-making- related policies enacted by a variety of governments and institutions.

Research Questions

The purpose of this study is to investigate how teachers are utilizing interim benchmark assessment data to inform their instructional decisions regarding professional practice in the classroom. The essential questions that this dissertation will attempt to answer are as follows:

1. To what degree do teachers believe their district's interim benchmark assessment tool provides useful student information in instructional decision-making?

2. What kinds of instructional decisions do teachers make using data from interim benchmark assessments?

3. How frequently do teachers use data from interim benchmark assessments to make those decisions?

4. What is the correlation between teachers' perception of the usefulness of their district's interim benchmark assessment and the frequency by which they reference those data in making those decisions?

5. What is the difference in teacher perception of usefulness across interim benchmark assessment publishers?

Significance of this Research

This dissertation would contribute to the available research on school district use of interim benchmark assessment data at the classroom level across a large population of teachers from multiple districts. The data gleaned from this study would represent a unique opportunity to research how the Michigan Benchmark assessment law has impacted instructional decision- making at the classroom level and the degree of difference in instructional decisions teachers make based on the individual interim benchmark assessment their districts use. This project would represent the only research of this kind that has been conducted in the state of Michigan. Additionally, a review of available literature suggests that previous research has yet to study how teachers use interim benchmark assessments as they currently exist in the market. This study would broadly contribute to the literature on data-based decision-making, particularly as it relates to classroom instruction, with potential implications for state and local policy. Though one study of note found no change in teacher practice after an initial year of schoolwide implementation of interim benchmark assessment data, researchers suggested that this lack of impact was due to insufficient experience with the school's selected assessment tool (Chojnacki et al., 2013). Therefore, the timing of this study will provide the research community with useful data on the decisions teachers make using interim benchmark assessment data after several years of implementation. In Michigan's case, at the time of this study, all schools serving students in grades K-8 have administered a benchmark assessment for at least three years.

The results of this study are intended to support policymakers at the state and local levels in understanding the impact of this policy on classroom instruction. The

amount of state funding allocated towards interim benchmark assessment, in addition to significant technology and human resources, represents a significant effort on the part of the state to support teachers in instructional delivery post-pandemic. This study would elucidate the degree to which this policy is, or is not, impacting practice and uncover potential differences in teachers' perceptions of the usefulness of interim benchmark assessment data across the state's most popular publishers. This would help state and local policymakers understand the return on investment of the benchmark assessment law while also helping local practitioners understand the differences between the various interim benchmark assessment publishers. While some research would suggest that, in a very broad context, assessment data can influence instructional practices (Chojnacki et al., 2013; May & Robinson, 2007; Quint & Smith, 2008), no other research of the scope of this study exists on interim assessments as they exist in the educational market today. If policies requiring the purchase and use of interim benchmark assessments continue to persist, it is critical for the research community to understand how teachers view, value, and use these tools in their instruction.

SECTION 2 - LITERATURE REVIEW

Interim Assessment – A Contested Term

Any examination of the use of interim assessments in K-12 school systems must first begin with an exploration of the term interim assessment and how the field has applied this term to various psychometric tests used in schools. This is a contested term, and there is a lack of consensus from previous research about precisely what the term interim/benchmark assessment actually means. The terms benchmark assessment and

interim assessment may be used interchangeably throughout research literature and in practice. Perie, Marion, Gong, and Wurtzel (2007), collectively representing The Center for Assessment, Achieve Inc., and The Aspen Institution define an interim assessment as an assessment used to evaluate “students' knowledge and skills relative to a specific set of academic goals typically within a limited time frame” (p.1). In a separate paper, Perie, Marion, and Gong (2009) define interim assessment as " medium- scale, medium-cycle assessments, falling between summative and formative assessments and usually administered at the school or district level…typically given several times a year" (p.6). Herman, Osmundson, and Dietl (2010), representing The Assessment and Accountability Comprehensive Center, provide a similar definition, stating that "Benchmark assessments are assessments administered periodically throughout the school year at specified times during a curriculum sequence to evaluate student knowledge and skills relative to an explicit set of longer-term learning goals" (p. 1). Additionally, the Michigan Assessment Consortium (2017) provides a more recent, yet simple definition, "an assessment program that is administered periodically to students, such as at the conclusion of a marking period" (p.17).

Reviewing these definitions to anchor one's examination of the research literature is essential, as various bodies of research may examine the use of interim and benchmark assessments in contexts that may vary distinctly from one another. In some research studies examined throughout the literature, interim benchmark assessments were district-created paper- pencil tests or tests developed to a specific set of state or district standards and learning targets by an external educational consultation firm or district central office (Slavin et al., 2011; Olah et al., 2010). In other studies, the assessments under review

were commercially developed tests from a curriculum publisher or a professional psychometric company implemented voluntarily by individual schools or as a part of a state assessment system (Brown et al., 2009; Chojnacki et al., 2013). Others have suggested an expansive view of interim assessments, from commercially developed assessments to locally created performance tasks (Perie et al., 2009). For this paper, and in alignment with current Michigan law, the aim of this research study will focus exclusively on commercially developed computer adaptive interim assessments, designed to national standards and performance norms, in Reading and Math administered to students online in the Fall, Winter, and Spring of a typical school year.

Functions of Interim Assessments in Schools

Interim assessments have a variety of uses and functions within a comprehensive assessment system. These tests can be used to communicate expectations, plan curriculum and instruction, monitor and evaluate learning, and predict future performance (Herman et al., 2010). However, though these tests may serve these purposes, some researchers suggest that there needs to be more evidence to suggest that these tests provide useful information to educators to improve learning (Perie et al., 2007; Perie et al., 2009). Other researchers have used the term "benchmark" assessment to refer to a district or state-developed assessment administered quarterly. With support to teachers, researchers found that teachers altered their practices based on the data (Carlson et al., 2011). A critical difference between the assessment tools studied by previous researchers and the interim assessments that will be examined in this research study is the nature by which the assessments are designed and implemented (Herman et al., 2010). Two of the most popular interim assessments used by districts in Michigan are the NWEA MAP

assessment and the Curriculum Associates i-Ready assessment (Kilbride et al., 2021). These two assessments are proprietary, commercially developed assessments published by their respective organizations.

Both publishers claim that their assessments are aligned to Common Core State Standards, and while there may be some general applicability to state curriculum standards, by definition, they are not built to any one state's specific set of curriculum standards. They are also not built to one district's locally adopted curriculum. Additionally, performance on these assessments is normed across a national population of test-takers and is not respective to any one state's population of test-takers. Indeed, a selling point for these assessments from their publishers is that student performance is nationally normed so that school stakeholders can understand how students perform against a national pool of peers, not simply peers within a school, district, or state.

Interim Assessments are not Formative Assessment

Additionally, some researchers have studied so-called formative assessments in an era when the broader educational community did not universally understand the term "formative" assessment (Perie et al., 2009; Perie et al., 2007; Shepard, 2009). Since the publication dates of much of the research in this literature review, it is essential to note that understanding of certain terms within the field of K12 Education has shifted, with some being dated or outright misaligned with current understanding in the field. For example, some researchers and benchmark assessment developers themselves refer to interim benchmark assessments as formative assessments, as researched in the landmark study Assessment and Classroom Learning'' by Black and Wiliam (Black & Wiliam,

1998; Perie et al., 2007; Goren, 2010). Since that time, the Council of Chief School State Officers defines formative assessment as "a process used by teachers and students during instruction that provides feedback to adjust ongoing teaching and learning to help students improve their achievement on intended learning outcomes" (FAST-SCASS, 2017). Inherent in this definition is the word process, as in an instructional process embedded within the teaching and learning experience. Benchmark assessments therefore cannot be accurately deemed as formative ostensibly because they are tests administered outside of instruction. Indeed, Perie, Marion, and Gong suggested that "many so-called formative assessments are not at all similar to the types of assessment and strategies students by Black and Wiliam (1998) but instead are interim assessments." (2009).

Data-Based Decision-Making in Schools

A misconception among practitioners is that the purpose of assessment data is solely to support accountability and drive compliance. (Mandinach & Schildkamp, 2021). Because of these misconceptions, assessment can potentially drive teachers to frame student learning in a deficit framework rather than a focus on growth and continuous development. However, a variety of studies show how teachers use student data from various sources and assessment methods in classroom-level instructional decision-making (Buckley et al., 2010; Olah et al., 2010; Tsai & Tosh, 2020). Teachers use data in various instructional decisions, but those decisions can be unpredictable and are highly influenced by several factors, including training, teacher assessment literacy, and technology access (Horne et al., 2015; Buckley et al., 2010; May & Robinson, 2007). In one study with a small group of mathematics instructors, researchers found that two groups of math teachers within the same district came to different conclusions and

instructional implications based on the same results from a district-developed assessment (Horne et al., 2015). In Ohio, a small study of 42 teachers revealed that some used a newly developed state system of assessment reports on state data to make decisions on what to teach and focus on specific students at risk of not graduating (May & Robinson, 2007). In another study, researchers structured a quasi-experimental study to compare the difference in patterns of teachers' instruction using self-completed instructional checklists over a school year. They found that there was an insignificant difference between the practices of teachers who had access to a commercially developed interim assessment and those who did not. (Chojnacki et al., 2013). The researchers suggested that the low impact on teacher practices might be due to teachers having only one year of experience using the assessment tool. They also suggest that the small sample size of teachers completing checklists might harm the study's power.

Therefore, this dissertation would build upon the previous body of research by studying a large sample of teachers across several schools and districts after several years of mandatory implementation driven by state policy. It would also focus on nationally normed commercially developed interim assessment tools as they are currently understood and used in the K-12 market today. This focus and anticipated sample size sets this dissertation study apart from previous research conducted on interim benchmark assessments and would contribute to policymaker understanding, as well as practitioner understanding, of how data from interim benchmark assessments influence teacher instructional decisions.

SECTION 3 - THEORETICAL FRAMEWORK

Utility Theory

The framework of this study relies on a Utility Theory in the context of how teachers use the educational resources provided to them by their districts to support their instructional needs. Evidence suggests teachers have broad, independent decision-making authority in classroom pedagogy (Wang et al., 2021; Usma-Wilches, 2006; Sparks et al., 2012). Therefore, each teacher makes a set of choices regarding what resources they use and how they use them in classroom instruction. Utility theory may explain the differences in how teachers use the educational resources afforded to them.

Utility Theory emerges from the field of Economics to explain how consumers make choices in a free market. Utility Theory was first offered by Swiss mathematician Daniel Bernoulli in 1738 as a way to explain how consumers make decisions based on the degree of satisfaction or utility they gain from a given item (Fishburn, 1968). The more utility that an item provides, the more the consumer is likely to value or prefer that item. However, individuals each have different beliefs and circumstances that influence how much utility they derive from an item or choice. Cardinal Utility, a derivative of Utility Theory, postulates that utility can be measured by imaginary or theoretical units called “utils,” thus giving every choice a specific utility value unique to each individual.

Considering interim benchmark assessment data as a resource available for teachers to use in the market of instructional decision-making, teachers may likely assign varied amounts of utils to interim benchmark assessment data when they make choices about professional practice in the classroom. Although state laws and districts may

mandate the administration of interim benchmark assessments, there are no laws that mandate teachers' actual use of these data in classrooms. Local district policies on instructional programs may vary, but generally speaking, teachers are afforded wide autonomy in instructional decision-making. In one study, most teachers reported that they have a great deal of autonomy in determining what teaching methods they use and how they plan lessons, yet conversely, only 12% reported that they have autonomy on what data they collect on student achievement (Worth & Van den Brande, 2020). This suggests that while districts may mandate assessment administration and data collection, teachers make individual choices on how and why they use data in classroom-level instructional decisions. Additionally, while data-based decision-making and assessment administration may be required by district policies and/or encouraged practices, the actual use of data by teachers in the classroom may significantly depart from district guidance (Melucci, 2013).

Usability is a key dimension that researchers have found influences how and why teachers use instructional resources (Wang et al., 2021). Teachers are more likely to use instructional resources that they find easy to use and will often adapt district-provided resources to make them easier to use (Wang et al., 2021; Bugler et al., 2017). Utility theory, therefore, may explain why teachers have a greater level of uptake of certain resources in their instructional decision-making process. Many interim benchmark assessment vendors market their materials to classroom teachers, highlighting usability features such as enabling teachers to easily understand student capabilities, providing actionable insights for instructional change, and generating easy- to-interpret reports (Curriculum Associates, 2023; Renaissance Education, 2023; NWEA, 2023).

Additionally, as mentioned previously, teachers' affective attitude towards assessment has been shown to be a psychological factor that influences how teachers use assessment data in their practice (Prenger & Schildkamp, 2018). This suggests that teachers individually assign varying degrees of value to assessment data in their practice. Because teachers make many choices in their practice throughout the course of instruction, there is likely a degree and difference in utils or usefulness that interim benchmark assessment data plays in each of those choices. Are data from interim benchmark assessments 24 karat gold that teachers view as highly useful and frequently used to support classroom instruction? Or, are these data fools' gold that teachers simply do not value and find useless? This study will attempt to illuminate this phenomenon.

SECTION 4 – RESEARCH METHODOLOGY

Context of the Study

The community that served as the focus of this study is the Oakland Schools Intermediate School District (ISD) constituency, which contains all public schools in Oakland County, Michigan, and a small number of schools in several neighboring counties. Oakland County is located northwest of the city of Detroit and is considered a part of the greater Metro-Detroit area. Although many communities in Oakland County are commonly referred to as the suburbs of Detroit, Oakland County has urban, rural, exurban, and suburban communities within its county borders. Oakland County is the second largest populated county in the state of Michigan, with 1,259,431 residents as of the 2020 Census living across 62 local municipalities. (U.S. Census Bureau, 2022). Oakland County, on average, is a particularly affluent community with a median housing

value of $268,600 compared to $172,100 across Michigan and a per capita income of $49,012 compared to $34,768 across Michigan. However, there is significant economic diversity within the county, ranging from a median household income of $20,398 in the city of Pontiac, an urban community, to Novi Township, with a median household income of $175,625 (U.S. Census Bureau, 2022).

The Oakland Schools ISD constituency contains 28 local education agencies (LEAs), commonly known as school districts, and 22 public school academies (PSAs), commonly known as charter schools. Each LEA and PSA is its own independently operated educational entity, each of which has an elected or appointed school board. LEAs in Oakland Schools range in size from 15,000 students in its largest LEA to just under 1,000 students in its smallest LEA. PSA range in size and in school site configuration. Most PSAs in the county are single school site entities, with only two that operate more than one school site. Oakland County PSAs range in size from 38 students in its smallest PSA to 2,896 students in its largest PSA, which is a statewide Virtual school charter by an Oakland County LEA. In total, Oakland Schools LEAs and PSAs combined serve 172,972 students in grades K through 12 across 322 public schools. 58.58% of students in Oakland Schools are White, 19.5% are African American, 9.35% are Asian, 7.43% are Hispanic, 4.79% are multi-racial, and .28% are American Indian or Alaska Native. 7.8% of students are English Learners. 12.22% of students are identified as having a disability. 36.93% of students are Economically Disadvantaged (Michigan's Center for Educational Performance and Information, 2023).

The Oakland Schools ISD student population varies in several ways from the overall statewide student population. Oakland County is more racially diverse than the

state, with more than double the percentage of Asian students (9.35% in Oakland County vs. 3.54% in Michigan), slightly more African American students (19.5% in Oakland County vs. 18.13% in Michigan), and fewer White students (58.58% in Oakland County vs. 63.65% in Michigan). The Oakland Schools ISD population also typically outperforms the statewide average on a variety of school performance metrics. For example, in 2022, 51.4% of 3rd-grade students were at or above proficiency in English Language Arts (ELA) on the state's accountability test, the Michigan Student Test of Educational Progress (M-STEP), compared to the statewide average proficiency rate of 41.6% of 3rd-grade students on the 3rd Grade ELA M-STEP. Also in 2022, 59% of Oakland County 11th-grade students met the college and career readiness benchmark on the Evidence-Based Reading and Writing (EBRW) portion of the SAT, Michigan's high school accountability test, compared to the statewide rate of 51.6%.

Data from the state's primary public education database, MISchoolData (maintained by Michigan's Center for Education Performance and Information), shows that there are approximately 12,320 full-time K-12 teachers in Oakland Schools ISD that are employed across the 28 LEAs and 22 PSAs (2023). As of spring 2023, 22.5% have been employed by their district for one year or less. Yet, in the spring of 2021, just 15% of staff had one year or less longevity. This is a marked increase in the percentage of staff who have one year of longevity or less in their district. These data may reflect alignment with national trends that suggest that districts across the United States are facing challenges filling open positions and hiring an increasing number of new staff post-pandemic. (Bleiberg & Kraft, 2022). However, 21.3% of Oakland Schools ISD teachers have been employed by their district between 2 and 5 years, and 56% have been

employed by their district for six or more years. So, while there has been an increase in new hiring across Oakland County, the county's teacher workforce largely has a substantial amount of experience in their current district.

In addition to longevity, 63.7% of Oakland Schools ISD teachers have a Master's degree, which is a higher rate than the national average of 51% among public school teachers (National Center for Education Statistics, 2023). Approximately 80% of teachers in the county are female, and 20% are male, following similar national gender patterns across all public schools that reflect that the vast majority of teachers are female. Also, in alignment with national trends, most Oakland County teachers are White. Over 89% of teachers in the county are White, 7.8% are African American, 1.1 are Asian, and 1.0% are Hispanic (Michigan's Center for Educational Performance and Information, 2023).

Public school operation in Michigan is largely funded by state tax revenues through a state law that is passed annually called the School Aid Act, first created in 1994 by Proposal A. Previous to Proposal A, each school district relied largely on local property taxes to fund schools. Proposal A and the School Aid Act shifted the burden of school taxation and funding from local municipalities to the state through a combination of an increase in the state sales tax rate and a significant reduction in local property tax rates. Using this statewide revenue pool, the School Aid Act provides a basic per-pupil foundation allowance that apportions money to LEAs and PSAs based on their respective annual student enrollment counts. In Fiscal year 2022, the basic per-pupil foundation allowance was $8,700 per student. However, some districts were allowed to keep higher property tax rates to support the funding of their schools. The rationale behind "grandfathering" these districts into keeping higher property tax rates was that Proposal

A's cap on school-allocated local property tax rates would be too low for very affluent communities to maintain their higher-than-average school spending. This is particularly apropos to Oakland County, wherein 10 of the county's 28 LEAs receive more than the basic per-pupil foundation allowance, more than any other county in Michigan.

Author Positionality

As of the writing of this dissertation, I am set to complete my 21st year in public K-12 Education. I began my career in public education as a high school English teacher at the Detroit Public Schools in the state of Michigan. I am a product of the Detroit Public Schools, and I am dedicated to improving educational systems that I believe will enhance the quality of life and opportunity for all students, particularly students from historically marginalized communities. I have a Bachelor's degree in Secondary Education from the University of Michigan and a Master's degree in Educational Leadership from Wayne State University. Throughout my career, I have tried to use my educational experiences to enhance those of the students I have been honored to serve.

As a Black male who was born and raised in Detroit, I have witnessed the power of education to transform the lives of individuals and communities. Therefore, my research interests are improving education systems, resulting in better opportunities and achievement for all students. I now serve as the Supervisor for the Curriculum & Assessment Unit at Oakland Schools. Research, evaluation, and assessment are the focus of my work within the organization, as I also lead our organization's assessment and research team. I believe that assessment and educational data, in general, under the right conditions, can lead to better decisions by educational professionals in their efforts to

support students. This belief is shaped by my experience in working with teachers and school administrators who have been able to improve student educational programs, informed by access to quality, relevant student data.

As these experiences have shaped my belief about K -12 educational systems, I acknowledge my bias toward quality data and assessment and their role in educational decision- making processes. This informs and compels my interest in this study, especially given the current policy landscape in Michigan that requires school professionals to administer interim benchmark assessments. I recognize that not all school personnel hold similar views towards assessment, yet I also recognize that not all available data are high-quality and appropriate to use in a professional educational setting. Having the right data at the right time with the right people can support high-quality educational decision-making. Therefore, my goal is to contribute to the available body of research that informs the research community about how teachers perceive the usefulness of interim-benchmark assessments so that policymakers and practitioners can improve the role of assessment in continuous improvement. It is prudent for me to disclose that through my current employment, I have established strong professional relationships with most district assessment coordinators in Oakland County. As their local county ISD, I provide these staff and teachers in their districts with professional learning, consultation, and resources to improve their educational programs through educational data. My relationships with these direct-level leaders and my established body of work in the county helped in the recruitment and participation in this study.

Research Design

This research utilized an online survey designed to be completed by classroom teachers in Oakland County ISD who have been assigned to teach students in Kindergarten through grade 8. The survey utilized single-select and Likert-scale type items to measure the degree to which teachers find data from interim benchmark assessments useful to their practice, as well as an open-ended response to maximize the collection of unique and idiosyncratic participant perspectives. This methodology allowed for statistical analysis of data, particularly a correlational analysis between the usefulness of interim benchmark assessments and the frequency of use. This methodology also allowed for the use of ANOVA analysis to determine the statistical difference in the mean rates of perception of usefulness across the interim benchmark assessment vendors used in the county.

Furthermore, an online survey allowed for the most efficient way to collect data from the largest sample possible from the intended population, given that the population is situated in the second most populous county in the state of Michigan. Additionally, there may be a significant variation in perspectives across individual teachers, thus requiring the collection of perspectives from a significant number of participants to conduct a thorough analysis and form valid conclusions. A survey allowed for the most efficient collection of data without presenting a significant burden of participation on districts, schools, and individual teacher participants. Conducting observations and personal interviews would not have been a practical way to collect this data, nor would it have allowed for statistical calculation. Furthermore, the volume of observations and personal interviews that would be necessary to collect across multiple school districts

implementing various interim benchmark assessment tools was beyond the nature and length of this study.

Participants and Recruitment

The specific population that this research project studied were K-8 classroom teachers employed in the LEAs and PSAs within the Oakland Schools ISD. This is the population that was most directly impacted by Michigan's interim benchmark assessment law and was most likely to routinely administer interim benchmark assessments. All LEAs and PSAs in Oakland Schools ISD serving students in kindergarten through 8th grade with a student enrollment count below 9,000 students were invited to participate in this study through coordination with each district's assessment coordinator. It was beyond the feasibility of this study to conduct and coordinate the survey instrument's distribution across many teachers and schools in a single, large school district. Districts of this size have complicated and layered communication channels between central offices, schools, and classroom teachers that make coordinating survey participation challenging and time-consuming. Broad system communications in districts across leadership levels, such as in large districts with many schools and programs, can be underdeveloped (Norqvist & Ärlestig, 2021). Central office-driven change across a school system, therefore, can be very challenging and requires significant time to implement. The decision to use interim benchmark assessments in the manner prescribed by Michigan law is the responsibility of the LEA, not individual schools, and thus, getting a clear picture of the implementation and use of an IBA within a large district presents significant challenges to researchers. Therefore, this study focused on analyzing school districts of comparable size and enrollment to yield more robust empirical evidence. Excluding these larger districts from

recruitment and participation in this study meant the removal of four out of the 28 LEAs. No PSAs servicing students in grades K-8 were excluded from the recruitment and participation of this study.

Within this targeted population of schools, this study attempted to recruit all K-8 teachers, not simply just teachers who teach Reading and/or Math. A typical elementary teacher assigned to students in Kindergarten through grade five was likely responsible for teaching multiple content areas (Math, Reading, Science, Social Studies, etc.) to the same students for a significant portion of a school day. Including these teachers in the study made intuitive sense because they teach subjects assessed by interim benchmark assessments. Additionally, this study also included Reading and Math teachers serving students in grades six through eight, such as in a typical middle school setting, who were assigned to teach one content area to a group of students for a single period within a school day.

However, this study also recruited teachers assigned to teach K-8 students who teach a subject other than Reading or Math. Given the varied nature of teacher working assignments, it could be reasonable to suggest that some teachers might inherently find data from a Reading and Math interim benchmark assessment less useful because they teach a content area that these assessments do not directly measure. However, there is evidence to suggest that districts adopt policies that require all teachers within a given school to reference interim benchmark data in their classroom practice. Teachers in schools face schoolwide accountability pressures in Reading and Math (Erichsen & Reynolds, 2020), and schools adopt schoolwide reforms that impact all staff across virtually all content areas (Stevens et al., 2022; Sherwood et al., 2021). Therefore, all teachers assigned to teach

students in kindergarten through 8th grade working in districts with 9,000 or fewer students were encouraged to engage with the study's survey.

Determining the number of K-8 classroom teachers within these districts was challenging. The Michigan Center for Education Performance and Improvement (CEPI) maintains a database of teacher FTE at the district and school level through its registry of educational personnel, which is publicly available on MiSchoolData.org. However, these FTE counts are irrespective of specific grade level assignments or content area assignments. This study specifically focused on K-8 classroom teachers only and did not include other ancillary instructional staff such as instructional coaches, specialists, or interventionists. These staff are often coded as teachers in Michigan's registry of educational personnel, which therefore made using this registry as the basis for staffing FTE inappropriate for this study.

The lack of specific grade-level teacher staffing data required the study to calculate an estimated number of classroom teachers serving students in grades K through 8. I calculated these estimates for each participating district based on average class size and K-8 student count. However, this was further complicated by the fact that CEPI does not publish an official average class size calculation. Student-teacher ratios calculated by the department are notoriously misleading since all staff who serve in an instructional role are included in the count of teachers within a district, even though they may not actually serve in classroom teaching roles (Jacob et al., 2016). However, there is evidence to suggest that the average class size in Michigan is 27 students (Schramm, 2016). Therefore, this study calculated an estimate of K-8 teachers in the target population based on an assumption of 27 students per classroom teacher. For each district in the target population, I calculated

the sum of enrollment of students in kindergarten through 8th grade and divided this total number by 27. Using this calculation, this study estimated that there were 2,981 classroom teachers serving students in grades K-8 in the districts targeted for recruitment in this study, and this estimate served as the study population count.

As an incentive for districts to participate in this study, each district was provided with an aggregate anonymized summary of responses collected in their district, provided that there was a minimum response rate of 10% from their district, with a minimum N size of 10 responses. Aggregate countywide results were also shared with assessment coordinators at the conclusion of this study. District assessment coordinators will be able to use these data from this study to support continuous improvement planning in their districts, as well as future cycles of assessment tool review and adoption.

An initial invitation to participate in this study was sent via e-mail to 43 assessment coordinators from eligible LEAs and PSAs, with follow-up emails sent to assessment coordinators who did not initially respond within seven days. As mentioned previously, PSAs do not typically participate in many services offered by Oakland Schools. I do not have many relationships with assessment leaders in PSAs to reference in recruitment. Therefore, recruiting PSAs to participate in this study was a challenge. Fifteen local education authorities and one public school academy agreed to be a part of this study, based on the initial invitation e-mail and follow-up reminder emails. Among these districts, this study estimated that there were 1,434 teachers serving students in grades K through 8.

Each district assessment coordinator in a participating district served as the primary contact for district participation. I did not have direct communication with any

study participants. Instead, I provided each assessment coordinator with emails and communication templates that they forwarded to teachers within their district. Because Interim benchmark assessments are designed to be administered multiple times per year, with the first administration date occurring between September and October, October was an ideal time for teachers to start participating in this study. In most cases, classroom teachers administer this assessment to their students and have instantaneous access to student data. Additionally, districts are required by Michigan law to share individual student data with parents no later than 30 days after the start of the school year. Therefore, it was reasonable to conclude that teachers would have access to student data by mid- October, again making October an ideal time for teachers to start participating in this study. Assessment Coordinators were asked to forward the study's survey to teachers in their districts during the week of October 23rd, with additional follow-up emails sent during the weeks of October 30th and November 6th. District assessment leaders were sent weekly totals of the number of participants from their district as a means of transparency and encouraging participation. This three-week data collection period, combined with weekly updates to district assessment leaders, was a reasonable length of time to collect a suitable number of participants.

This study attempted to obtain between 248 (8%) and 787 (26%) participants or more of the population of approximately 2,981 teachers in Oakland County serving students in kindergarten through 8th grade in districts with less than 9,000 students. Having an absolute minimum number of 200 responses is sufficient for correlation analysis (Guilford, 1954). This study's research questions required regression analysis, which is further explained in the Analysis section of this dissertation. Therefore,

obtaining a minimum of 200 responses was necessary to yield valid interpretations of the data. Additionally, based on a population size of 2,981 teachers, a sample size of at least 248 would yield a margin of error of 5% within a confidence level of 90% (Taherdoost, 2017).

The survey instrument recorded a total of 398 responses, which is about 13% of the target population. However, not all survey respondents completed the entire battery of survey items. Among the recorded participants, 256 respondents completed all select-response items within the survey, and those 256 responses served as the focus for conducting analysis in this study. This sample of 256 teachers from the total population of 2,981 teachers represents 8.5% of the total population. This is less than 10% of the total population, and thus the findings should be interpreted with caution. Nonetheless, Taherdoost (2017) stated that a sample of at least 248 is sufficient for analysis of this population size.

Instrument

As there is no single industry-recognized survey instrument that is used by schools to measure teacher perceptions of the utility of interim benchmark assessments, this study required the creation of a survey instrument to answer the study's research questions. I created this instrument using the recommendations of previous research, policy brief recommendations, as well as recommendations directly from interim benchmark assessment publishers themselves (Buckley et al., 2010; Olah et al., 2010. Perie et al., 2007; Wang et al., 2017; Curriculum Associates, 2021; Nordengren, 2022). Additionally, given that the intended participants are classroom teachers and that time is

not of abundant quality in a typical teacher's workday (Nahmais, 2015), the survey was designed to be completed in under 5 minutes and could have been completed via a computer, tablet, or cell phone. The short nature of this survey, combined with its accessible multi-platform design, presented a minimal burden of participation for busy teachers. See *Appendix A* for the text of the entire survey instruction. Below, I describe a summary and rationale for each of the question blocks in the survey instrument.

Participant Identification Block

It is necessary to collect a minimum level of identifying information from participants to ensure that the survey is completed by the intended population. In three single select items, the survey asked participants to identify their district, primary content area, and assigned grade levels. This survey did not collect personally identifiable information and participants should have had a level of anonymity that allowed them to feel comfortable answering honestly and candidly. No data from the identification block were shared with districts.

Interim Assessment Identification Block

Teachers were asked to share the name of their district's interim assessment tool from a list of available interim benchmark assessment tools used in Michigan. It was necessary to ask teachers about this because there was not a publicly available database that discloses what interim benchmark assessment each district uses. Furthermore, the length of time that a district has implemented an interim benchmark assessment and the length of time that individual teachers have to implement this tool may influence the degree to which it is used in classroom- level instructional decision-making (Chojnacki et al., 2013). Therefore, it was necessary to ask how many years of experience teachers have

with this tool, and this was asked in a single select item.

Interim Benchmark Assessment Usefulness Block

The Data collected in the Interim Benchmark Assessment Usefulness Block will help to directly answer Research Questions 1, 2, and 3.

Research Question # 1 - To what degree do teachers believe their district's interim benchmark assessment tool provides useful student information in instructional decision- making? Herman et al. recommended, among several factors, that districts select interim benchmark assessments based on a sense of high utility and how much an interim benchmark assessment helps the school meet its intended purposes (2010). Therefore, if the central purpose of data-based decision-making strategies is to provide information to school practitioners to improve educational services to students, then it is logical to ask teachers about their perceptions of the utility of an interim benchmark assessment in helping them improve instruction for students. This question was asked in a Descriptive Likert-scale item [Not Useful, Somewhat Useful, Useful, Very Useful], with each of those descriptors being coded to a numeric value: [Not Useful=1, Somewhat Useful =2, Useful=3, Very Useful=4].

Instructional Decisions Block

The data collected in the Instructional Decisions Block helped to directly answer Research Questions 2 and 3. Additionally, the data collected in this block was used to conduct an analysis that will answer Research Questions 4 and 5, explained in further detail in the Analysis section of this proposal.

Research Question #2 - What kinds of instructional decisions do teachers make using data from interim benchmark assessments? Teachers were asked to select

up to 7 instructional decisions that the research literature suggests can be informed by interim benchmark assessment data. Previous researchers have suggested that interim benchmark assessments can be used to help teachers identify their own professional learning (Buckley et al.,2010), evaluate their own instructional effectiveness (Buckley et al., 2010; Olah et al., 2010), adapt instruction and curriculum (Perie et al., 2007), identify students for in-class intervention (Tsai & Tosh, 2020), and set student learning goals (Nordengren, 2022; Curriculum Associates, 2023). Additionally, the IBA publishers themselves recommend that teachers use their assessments to create instructional groups and plan individualized instruction (Nordengren, 2022; Curriculum Associates, 2023). Teachers were asked how useful they believe their district interim benchmark assessment is in helping them make these decisions in descriptive Likert scale items [Not Useful, Somewhat Useful, Useful, Very Useful] with each of those descriptors being coded to a numeric value: [Not Useful=1, Somewhat Useful =2, Useful=3, Very Useful=4].

Teachers were also asked how frequently they used data from their district's IBA to inform these seven classroom instructional decisions. Teachers could select from four options, Never, Rarely, Occasionally, or Frequently, with each of those responses coded to a numerical value: [Never=1, Rarely=2, Occasionally=3, Frequently=4]

Additionally, teachers had the option of providing additional ways in which they use interim benchmark assessment data that may not have been discussed in previous research. This was asked for in an open-ended response item. Having an open-ended item was necessary to capture all of the various ways in which teachers use interim benchmark data in their instructional decision-making processes while also providing for the expression of individualized teacher-participant voice.

Analysis

Descriptive Analysis

To help illustrate the degree, frequency, and kinds of decisions teachers make, I provided a described analysis of the findings of the survey instrument. Descriptive analysis is an efficient way of characterizing a phenomenon and helping researchers understand relationships between variables (Loeb et al., 2017). Additionally, the descriptive analysis provides an easily understandable summary of research findings to an audience of practitioners and policymakers (Dynarski & Kisker, 2017), a group of educational decision-makers this study hopes to inform. Using SPSS, I provided several data tables summarizing responses with descriptive statistics showing n sizes of responses, percentages of responses, and mean calculations. The descriptive analysis will help answer Research Question 1 – To what degree do teachers believe their district's interim benchmark assessment tool provides useful student information in instructional decision-making Research Question 3 – How frequently do teachers use data to support these decisions? For Research Question 2 – What kinds of instructional decisions do teachers make using data from interim benchmark assessments - I also provided a frequencies table of the data from item 8 in the survey. The open-ended response data was coded using an inductive coding approach to develop themes and was summarized in one combined theme frequencies table. I read each response, created codes based on the text within those responses, re-read the responses a second time, and applied the codes to each response as applicable. This should help the research community gain a deeper understanding of which factors are most strongly associated with the usefulness of an interim benchmark assessment.

Regression Analysis

The key theory behind his study - Utility Theory – would suggest that teachers assign varying levels of utility or usefulness to interim benchmark assessment and that this may be expressed in terms of frequency of use. Therefore, it stands to reason that there was a relationship between teachers' perception of the usefulness of IBAs and the frequency by which they use them in classroom instructional decisions. To show this relationship, I conducted a regression analysis to show the relationship between the usefulness of functions of IBA and the frequency by which teachers use those functions in classroom-level decision-making. This analysis helps to answer Research Question 3 – What is the correlation between teachers' perception of the usefulness of IBA and the frequency by which they reference those data in making those decisions?

In this regression analysis, I calculated the relationship between each pairing of IBA usefulness functions and the frequency of those functions. For example, the usefulness responses for the setting student learning goals function of an IBA were paired with the setting student learning goals frequency responses. The usefulness of responses in the survey served as the independent variable, and the frequency of use of responses will serve as the dependent variable. There are seven pairings of usefulness and frequency in total. This will also help support the theory that the more useful teachers perceive an assessment function to be, the more likely they will utilize that function in instructional decision-making. The inverse should also be true. The less useful an assessment function is, the less frequently it will be used. The formula below explains how this research question will be calculated.

$$\boldsymbol{Y} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} \cdot \boldsymbol{X} + \boldsymbol{\varepsilon}$$

Y serves as the dependent variable, in this case, the frequency by which teachers use data from IBAs in their instructional decision-making. **X** represents the independent variables and the perception of the usefulness of each discreet function of the IBA, respectively. The paired variables from the survey instrument were 1) creating instructional groups in your classroom, 2) planning whole group instruction, 3) planning individualized instruction, setting learning goals with students, 4) identifying students for in-class intervention, 5) measuring your own effectiveness, and 6) identifying your own professional development needs. I hypothesized that there would be a strong correlation between the usefulness and frequency of use of IBAs for in- class decision-making for each of these IBA functions.

Analysis of Variance

Lastly, I conducted a one-way analysis of the variance of the overall mean rates of the usefulness of each of the IBA tools. Because there are multiple IBAs that teachers utilized across the sample, it was possible that there could be statistically significant differences in the usefulness of the various IBA publishers. IBAs are adopted by districts in a free-market economy, and IBA vendors compete with one another for customers. Within this competitive market, IBA vendors naturally claim that their products are superior to those of their competitors. Conducting a one-way ANOVA test, therefore, would show if there would be a statistically significant difference between two or more groups by testing for the difference of means of those groups (Simkus, 2023). ANOVA also helps to avoid Type I errors that might emerge through conducting multiple t-tests (Wilcox, 2002).

SECTION 5 - FINDINGS

Characteristics of the Respondents

Fifteen districts agreed to participate in this study, including 14 LEAs and one PSA. This represents half of the 28 LEAs in Oakland County. As anticipated, this study netted very little participation from PSAs. Only one out of 19 eligible PSAs participated, likely due to the researcher's lack of strong relationships and professional familiarity with key PSA assessment leader personnel employed by PSAs. The districts in this study represent a diverse array of geographic localities, consisting of districts serving city, rural, and suburban communities, albeit city and rural districts are underrepresented in this study. See **Table 1**.

Table 1

Characteristics of LEAs & PSAs Participating in this Study

	K-8 Teaching Population	Number of LEAs	Number of PSAs	% Suburban	% City	% Rural
Sample	1434	14	1	86.66%	6.67%	6.67%
Population	2981	43	20	53.49%	37.21%	9.30%

The student enrollment sizes of these districts range from 300 students to over 8,000 students. Collectively, the districts in this study serve a combined 38,733 K-8 students, serving approximately 48% of Oakland County's total 80,488 K-8 student population. Among the participating districts, 59.22% of students are White, which is comparable to the county's overall 58.58% White student population. The participating districts severed a larger proportion of African American students, 27.29%, compared to the county's overall 19.50% African American student population. Conversely, the

participating districts served a smaller proportion of Asian students, 3.30%, when compared to the county's 5.61% Asian student population. This study's participating districts also served a slightly smaller proportion of Hispanic students, 5.50%, compared to the county's 7.95% Hispanic student population. See **Table 2**.

Table 2

Student Demographics of LEAs and PSAs Participating in this Study

	K-8 Student Population	% White	% Black	% Hispanic	% Asian
Sample	38,733	59.22%	27.29%	5.50%	3.29%
Population	80,488	58.50%	22.80%	7.95%	5.61%

Additionally, there appeared to be a somewhat even distribution of teachers assigned to teach each grade level. Note that teachers were able to report being assigned to multiple grades. See **Table 3**.

Table 3

Number and Percentage of Teachers Assigned to Teach Each Grade Level

	K	1	2	3	4	5	6	7	8
n	51	60	63	69	65	68	47	45	54
%	9.77	11.49	12.07	13.22	12.45	13.03	9.00	8.62	10.34

**Teachers may be assigned to teach multiple grade levels at the same time.*

Most respondents in the study, 56.64%, teach both Reading/ELA and Math, while an overwhelming majority of respondents, 88.67%, teach Reading, Math, or some combination of both. Given that this study targets K-8 teachers, and the typical teaching assignment of grade level teachers in grades K through 5 includes responsibility for teaching both Reading and Math in a self-contained setting, it is reasonable to conclude that teachers of both Reading and Math constitute a majority of the K-5 study population. Only a small percentage of respondents, 17.9%, taught neither Reading/ELA nor Math. These would likely be specials teachers in grades K-5 or other specialized content area

teachers in grades 6-8. It is reasonable to assume that teachers who are assigned to teach subjects other than Reading or Math were less likely to complete the survey, as interim benchmark assessments assess student achievement in subjects that they are not primarily assigned to teach. See **Table 4**.

Table 4

Number and Percentage of Teacher's Teaching Assignments

Teaching Assignment	n	%
Reading/ELA	49	19.14
Math	33	12.89
Both Reading/ELA and Math	145	56.64
Neither Reading/ELA or Math	29	11.33

In reviewing the findings of this study, it is important to consider the subject matter that teachers are primarily responsible for teaching. Recall that the Michigan Benchmark Assessment law required schools to assess students in Reading and Math in grades Kindergarten through 8th Grade. Additionally, the benchmark assessment vendors discussed in this study provide assessment solutions for Reading/ELA and Math but not for other subject areas. Yet, many schools often have schoolwide Reading or Math improvement goals that involve the cooperation of all teachers in the school (Stevens et al., 2022; Sherwood et al., 2021). In some schools, all teachers are evaluated on schoolwide reading data, even though there may be teachers assigned to teach courses that don't explicitly have alignment with reading. For example, in one of the participating districts, 5% of an individual teacher's evaluation rating is derived from a schoolwide Interim Benchmark score in reading. This means that teachers who do not necessarily teach Reading as a part of their job assignment, say a 6th-grade Science teacher or a 4th-grade gym teacher, would be evaluated in part using schoolwide reading performance.

This practice is commonly referred to as *all teachers are reading teachers* mantra (Draper, 2002).

Most respondents had significant experience with their district's interim benchmark assessment. 73.13% of respondents had three or more years of experience with administering their district's selected interim benchmark assessment. 10.88% of respondents had two years of experience, 6.12% had one year of experience, and 9.86% were within their first year of experience with their district's interim benchmark assessment. A previous study that examined change in teacher practice after an initial year of using interim benchmark assessment data observed that the data these assessments provided did not significantly impact instruction (Chojnacki et al., 2013). The researchers suggested that additional years of experience may have been necessary to see more substantial changes in instruction. The experience of this study's respondents would suggest that these teachers would theoretically be more positioned to incorporate interim benchmark assessment data into their practice. See **Table 5**.

Table 5

Years of experience with the district's selected interim benchmark assessment

Years of Experience	n	%
In 1st year	19	7.42
1-2 years	42	16.40
3 years or more	195	76.17

Findings Summary

Overall, this study found that most teachers believed that interim benchmark assessments, in general, can help them in their instructional practice. However, the actual tests their districts adopted were only somewhat useful in their instructional decision-making process. The specific use cases of the interim-benchmark assessments matter greatly, as there are significant differences in the perception of the usefulness of the interim-benchmark assessment depending on the specific instructional decisions that teachers make. In this case, the three most useful functions of interim benchmark assessments were using them to identify students for in-class intervention, planning individualized instruction, and creating instructional groups. Additionally, this study found that teachers use interim benchmark assessments in combination with other types of data to make wholistically informed instructional decisions. Interim benchmark assessments serve as one piece of evidence that informs teachers' instructional decisions. Finally, while there were some differences in the perception of usefulness across the specific benchmark assessment vendors, these differences were not statistically significant. These findings are further explained in the sections below.

Teachers generally believe IBAs can be helpful, but there is disagreement.

The first research question in this study asked to what degree teachers believed their district's interim benchmark assessment tool provides useful information in instructional decision-making. However, prior to examining the primary research question, it is prudent to examine the beliefs of the teachers in this study about the overall utility of assessment in instructional practice. On its surface, this is a logical premise. Teachers who believe that assessments are useful may tend to use them more frequently, and teachers who do not believe that assessments are useful may be less likely to use

them in their practice. Prior research has suggested that teacher beliefs may influence the degree to which they use and find usefulness in data derived from commercially developed tests (Datnow, 2015). Overall, this study found that teachers generally agreed that interim benchmark assessments can help improve their instructional practice. 68.36% of all teachers rated that they either agreed or strongly agreed that IBAs can help them in their classroom. Conversely, 31.34% of teachers strongly disagreed or disagreed that IBAs could help them in their classroom practice. An interesting note, however, is that teachers in this study were less likely to strongly agree that interim benchmark assessment data can help them improve their practice. Only 8.98% percent of teachers rated that they strongly agreed that IBAs can help improve their practice. See **Table 6**.

Furthermore, this study found a difference in levels of how much teachers in tested subjects vs. non-tested subjects agreed that IBAs were helpful to their practice. For example, among teachers who are assigned to teach math, reading, or both math and reading, 69.55% believed that data from interim benchmark assessment can help them improve their practice. 58.62% of teachers who taught neither Reading nor Math strongly agreed or agreed that IBAs can help them in the classroom. Although the participation of teachers of non-tested subject areas in this study is small (n=29), the data shows that both teachers who are assigned to teach a content area tested by IBAs and teachers who are not assigned to teach a content area tested by IBAs generally agreed that IBAs can help them improve their instructional practice. All in all, these data suggest that teachers in this study have a generally positive disposition towards the usefulness of IBA instructional practice. However, more than a third of teachers in tested subjects disagreed that IBAs were helpful, and over 40% of teachers in non-tested subjects disagreed that these tests

were helpful. See **Table 6**.

Table 6

The degree to which teachers believe IBAs can be helpful to their practice

	All Teachers		Teachers of Tested Subjects		Teachers of Non-Tested Subjects
Rating	n	%	n	%	%
Strongly Agree	23	8.98	22	9.69	3.45
Agree	152	59.38	136	59.91	55.17
Disagree	64	25.00	55	24.23	31.03
Strongly Disagree	17	6.64	14	6.17	10.34

Given that the data suggests that most teachers in this study have a favorable disposition towards Interim Benchmark Assessments in general, how did the teachers rate the usefulness of their district's specific interim benchmark assessment data? Districts in Michigan have local authority to choose the IBA among a range of assessment products available on the commercial market. So, while it may be true that teachers, in general, may have a positive belief in the usefulness of IBAs, teachers perceive that their district's selected interim benchmark assessment may be useful to their classroom practice. The perception of the usefulness of IBAs in their practice may be influenced by their beliefs and by the level of experience they've had with IBAs (Datnow, 2015). However, this study has shown that the teachers in this study have generally positive beliefs towards IBAs and have multiple years of experience with IBAs. **See Table 5**.

Teachers don't have very high perceptions of the usefulness of their district's IBA.

This study found that 37.89% of teachers reported that their district's interim benchmark assessment was useful or very useful. The plurality of teachers, 45.31%, rated that their district's IBA was somewhat useful, and 16.80% of teachers reported that their district's IBA was not at all useful to their instructional decision-making. This shows that teachers generally do not have very high perceptions of the usefulness of their district's interim benchmark assessment—well over half saw it as only somewhat useful or not at all useful. See **Table 7**.

Table 7

Teacher perception of the usefulness of their district's interim benchmark assessment

Rating	n	%
Very Useful	24	9.37
Useful	73	28.52
Somewhat Useful	116	45.31
Not at all useful	43	16.80

Additionally, there were clear differences in the level of usefulness of a district's selected IBA between teachers who teach a tested subject area and those who do not. Among teachers in tested subject areas, 38.76% reported their district's IBAs were useful or very useful, compared to 31.03% of teachers who do not teach a tested subject area who reported the same perception of usefulness. What is a striking but not surprising difference is within the "Not at all useful" rating. Only 13.22% of teachers in tested subject areas rated their district's IBA as not at all useful compared to 44.83%, which is more than triple the rate of teachers in non-tested subject areas.

Table 9

Teachers of tested subjects vs teachers of non-tested subjects – perception of usefulness of IBAs

	Teachers of Tested Subjects		Teachers of Non-Tested Subjects	
Rating	n	%	n	%
Very Useful	22	9.69	2	6.89
Useful	66	29.07	7	24.14
Somewhat Useful	109	48.02	7	24.14
Not at all useful	30	13.22	13	44.83

Although this study found that teachers across the board found their district's selected interim benchmark assessment tools somewhat useful to their classroom instructional decisions, there was a difference in the perceptions of usefulness across the

different IBA tools utilized in Oakland County. The three most used IBA tools in the county were NWEA Map, i-Ready, and FastBridge. Only a small fraction of respondents reported using S.T.A.R. or a district-developed Interim Benchmark Assessment. Among the three most used IBA tools in the county, i-Ready users reported the most utility with their assessment solution; however, in no case did more than half of teachers rate any tool useful or very useful. i-Ready was the closest at 48% of i-Ready-using teachers rating it as useful or very useful. See **Table 10**.

Table 10

Overall, how useful do you believe your district's interim benchmark assessment is in your instructional decision-making? Disaggregated by IBA tool.

	i-Ready		NWEA MAP		FastBridge	
Rating	n	%	n	%	n	%
Very Useful	11	14.67	2	9.09	1	2.22
Useful	25	33.33	4	25.76	2	26.67
Somewhat Useful	26	34.67	4	48.48	4	53.33
Not at all Useful	13	17.33	2	16.67	8	17.78

Teachers Find Identifying Students and Planning Individualized Instruction Essential in IBAs.

The second research question asks what kinds of instructional decisions teachers make using data from interim benchmark assessments. This study found that two functions were the highest-rated uses of an IBA. 59.38% of teachers rated that identifying students for in-class intervention was either a useful or very useful function of IBAs. 50.98% of teachers rated that planning individualized instruction was a useful or very useful function of IBAs. More than half of respondents did not rate the other functions as useful or very useful. See **Table 11**.

Table 11

The usefulness of IBAs in 7 discreet classroom functions (N=256)

	Mean	Very Useful		Useful		Somewhat Useful		Not Useful	
Function		n	%	n	%	n	%	n	%
Identifying students for in-class intervention	2.72	62	24.22	90	35.16	74	28.91	30	11.72
Planning individualized instruction	2.52	51	20.00	79	30.98	77	30.20	48	18.82
Creating instructional groups in your classroom	2.44	39	15.23	81	31.64	90	35.16	46	17.97
Setting learning goals with students	2.36	42	16.47	69	27.06	82	32.16	62	24.31
Planning whole group instruction	2.03	22	8.63	53	20.78	91	35.69	89	34.90
Measuring your own instructional effectiveness	1.88	16	6.25	44	17.19	88	34.38	108	42.19
Identifying your own professional	1.76	12	4.69	37	14.45	84	32.81	123	48.05

Interestingly, there appeared to be a very strong plurality of responses for the not useful option for measuring your own instructional effectiveness function and identifying your own professional development needs function. More teachers rated those two functions as not useful than any other function rating. 76.56% of teachers reported that IBAs were either not useful or somewhat useful for measuring their own instructional effectiveness. 80.86% of teachers reported that IBAs were either not useful or only somewhat useful for identifying their own professional development needs.

While not necessarily a researched practice, many of the IBA vendors recommend that teachers use their assessment tools to set learning goals with students. In this practice, teachers share performance reports with students and their families and attempt to set a performance goal, say increasing a RIT score by 10 points, by the next assessment window. In fact, both NWEA and iReady have pre-built reports that are intended to be used with students to set goals for themselves to improve performance in future

assessment cycles. However, only 43.75% of teachers reported that using IBAs to set learning goals with students was very useful or useful.

Additionally, while these vendors and some practitioner guidance (Demchak & Sutter, 2019) may recommend that teachers use these data to look for trends and patterns to plan whole group instruction, only 29.29% of teachers reported that planning whole group instruction was a useful or very useful function of IBAs. Furthermore, small group instruction, particularly among elementary teachers, is an evidence-based practice strongly recommended by the Essential Instructional Practices produced by the General Education Leadership Network of the Michigan Association of Intermediate Schools Administrators (Michigan Association of Intermediate School Administrators General Education Leadership Network Early Literacy Task Force, 2018). Planning for small group instruction has been a significant focus of the Oakland Schools, particularly in literacy and math. In fact, reports for creating ideal groups of students with similar learning achievement levels can be found in some of the pre-built reports in IBAs, such as NWEA. Yet, just 47.09% of teachers reported that using IBAs to create instructional groups was either very useful or useful.

While this study asked teachers about the above seven distinct common uses of interim benchmark assessment data, the survey also asked teachers to respond to an open-ended question: *What other types of instructional decisions do you make using data from your district's interim benchmark assessment that haven't been listed?* Eighty-seven teachers responded to this open-ended survey item; however, not all of these participants yielded responses relevant to the question. Sixteen respondents to the open-ended item wrote “none” or “nothing,” which may perhaps suggest that these seven functions have

captured the most common uses that teachers were likely to report. One respondent even reported, “Your list of uses is fairly thorough.”

Eleven respondents included responses to this item that did not necessarily reveal an instructional decision and simply listed some other assessment methods or assessment tools that they use to support their work. A strong theme of referencing or using other assessments in combination with data from IBAs emerged throughout the analysis of these qualitative responses. Ten respondents mentioned using IBA data to compare or triangulate patterns of student responses with other assessments, including diagnostic assessments, summative assessments such as M-STEP, or formative assessments within their own classrooms. One respondent said, “Interim benchmark assessments are only a snapshot of a student’s potential strengths and weaknesses and should never be utilized alone. A combination of a screener like the NWEA with more diagnostics assessments like Accidence Math and Reading or Fountas and Pinnell reading assessments must be used to have a complete vision of a student.” Another respondent said, “This assessment is a “flag” for further diagnostics such as a reading inventory or a spelling inventory,” while another reported, “Information from our benchmarks also, in conjunction with other data points, helps us determine which students will receive Tier 2, pull out literacy support.” These responses show that teachers use IBA data with other data to support a variety of decisions, including both within their own classrooms, while also leveraging broader supports available across the school, like MTSS or Special Education Services.

Indeed, while these responses did not necessarily contain themselves to in-classroom instructional decisions, teachers nonetheless participate in many levels of decision-making within a school to support individual students. Supporting individual

students was a strong theme found in the open-ended responses. Fourteen respondents said they used IBAs to identify students for intervention, nine mentioned using IBAs in course placement either for advanced coursework or remediation intervention services, and six reported that they used IBA data as a part of the Individualized Educational Planning process within Special Education services.

Among these responses, respondents reported using IBAs for "identifying students for tutoring," to "pick a focus area for MTSS," and to make "student groups for collaborative work." See **Table 12** for a full listing of themes found within the open-ended responses.

Table 12

Most common themes are mentioned in open-ended responses.

Themes	Count
Identifying students for intervention	14
Comparison with other assessments	10
Identifying students for course placement	9
Creating instructional or intervention groups	8
Identifying learning gaps needs	7
Assisting in the IEP process	6
Lesson planning & curriculum adjustments	5

Teacher Frequency of Use Varies Greatly on Specific Use Cases

The third research question asked teachers how frequently they used IBAs to make specific instructional decisions. Teachers in this study used their district's interim benchmark assessment tool across a broad range of frequencies for the distinct purposes found in research Question #2. 75.29% of teachers reported that they frequently or occasionally used IBAs to identify students for in-class intervention (mean response rate

of 2.97). 61.39% of teachers reported the same for planning individualized instruction (mean response rate of 2.65). This mirrors the pattern observed in the usefulness ratings. A majority of teachers found identifying students for intervention and planning individualized instruction were the two most useful functions of an IBA. Additionally, 63.29% of teachers frequently or occasionally used IBAs for creating instructional groups (mean response rate of 2.63), even though less than half of teachers report that using IBAs for this function was very useful or useful.

There were also very strong patterns of infrequency of use of IBAs for certain functions. 43.53% of teachers reported that they never use their district's IBA to identify their own professional development needs. When combined with the teachers who rarely use IBAs for this function, 71.77% of teachers rarely or never use IBA data to inform their own professional development needs. This function had a mean frequency of use response mean of 1.91, making it the least frequently used function of an IBA. Furthermore, 65.89% of teachers rarely or never used IBAs to measure their own instructional effectiveness. At a response mean of 2.08, using IBAs to measure one's own instructional effectiveness is the second least used function of an IBA.

54.51% of teachers, just over half, frequently or occasionally use IBAs for setting student goals, even though IBA publishers develop specific pre-built reports for this very purpose. This function had a mean response rate of 2.51. In a similar fashion, despite their being specific reports for examining whole class or whole grade-level achievement patterns, only 40.79% of teachers frequently or occasionally used IBAs for planning whole group instruction, a mean response rate of 2.19. See **Table 13**.

Table 13

Frequency of use of IBAs by classroom instructional decision

		Never		Rarely		Occasionally		Frequently	
Function	Mean	%	n	%	n	%	n	%	n
Identifying students for in-class intervention	2.97	9.80	25	14.90	38	43.53	111	31.76	81
Planning individualized instruction	2.65	17.25	44	22.35	57	38.04	97	22.35	57
Setting learning goals with students	2.51	22.35	57	23.14	59	35.69	91	18.82	48
Creating instructional groups in your classroom	2.63	17.19	44	19.53	50	46.88	120	16.41	42
Planning whole group instruction	2.19	32.16	82	27.06	69	30.59	78	10.20	26
Measuring your own instructional effectiveness	2.08	33.73	86	32.16	82	26.67	68	7.45	19
Identifying your own professional development	1.91	43.53	111	28.24	72	22.35	57	5.88	15

A strong correlation between usefulness and frequency of use.

The fourth research question studied the correlation between teachers' perception of the usefulness of their district's interim benchmark assessment and the frequency by which they reference those data in making those decisions. This study found very strong relationships between the distinct usefulness of IBA functions and the frequency by which teachers use those functions in their work. For example, in the previous sections of this dissertation, it was found that teachers did not have a high perception of the utility of IBAs for identifying their professional learning needs and that most teachers ever or rarely used IBAs for that purpose.

This study found a significant relationship, an Unstandardized B Coefficient of .846, for the identifying professional needs function ($p.<.001$). Even in functions that demonstrated lower correlation strength to the frequency of use, the relationship is still very

strong. For example, the identification of students for in-class intervention function is still strong at an unstandardized B coefficient of .641. These data suggest that the more useful a function, the more frequently that function is used, and the less useful a function is, the less that function is used. See **Table 14**.

Table 14.

Relationship between usefulness (IV) and the frequency of use (DV). (N=255)

Function	Unstandardized B	Coefficients Std. Error	Standard Coefficients Beta	Sig.
Identifying students for in-class intervention	.641	.046	.663	<.001
Planning individualized instruction	.749	.041	.752	<.001
Setting learning goals with students	.769	.042	.755	<.001
Creating instructional groups in your classroom	.701	.045	.703	<.001
Planning whole group instruction	.816	.042	.775	<.001
Measuring your own instructional effectiveness	.789	.043	.752	<.001
Identifying your own professional	.846	.044	.772	<.001

No Differences in Usefulness across IBAs

Earlier analysis in this dissertation showed that there were differences in the level of perception of usefulness across the three most commonly used interim benchmark assessment tools in the county. Collectively, the mean rate of usefulness of all interim benchmark assessments on a 4.0 scale was 2.28, or slightly above the somewhat useful label. The IBA rated the most useful, iReady, was at a mean rate of 2.43. NWEA was rated at 2.25. FastBridge, 2.14. While iReady may have been rated the highest at 2.43, it should be noted across all interim benchmark assessments and disaggregating data by assessment publisher that no interim benchmark assessment was rated at or above a 3.0. This suggests that teachers do not find interim benchmark assessments to be very useful

tools for classroom instructional decisions, regardless of which tool is used. See Table 15.

Table 15

Difference in mean rates of usefulness across most common interim benchmark assessments

IBA	N	Mean	Std Deviation	Lower Bound	Upper Bound	Min	Max
All	277	2.28	.840	2.18	2.38	1	4
NWEA	145	2.25	.829	2.11	2.38	1	4
iReady	82	2.43	.930	2.22	2.63	1	4
FastBridge	50	2.14	.700	1.94	2.34	1	4

To measure the statistical significance of these differences, a one-way ANOVA analysis was performed to determine whether or not the mean rates of usefulness across all of the benchmark assessment publishers were statistically different from one another. This test revealed a p-value of .282 between the mean rates of usefulness across vendors. Therefore, this test did not show a statistically significant difference in rates of usefulness across IBA vendors. These findings give further credence to Utility Theory. If it is true that teachers do not have very high perceptions of the usefulness of their district's IBAs and that most see IBAs as either somewhat useful or not at all useful, then Utility Theory would suggest that, regardless of the IBA, none of them would be useful to classroom teachers. Given that there is not a statistically significant difference among mean rates of IBA usefulness, no matter how different IBA vendors may claim that their assessments are from those of their competitors, teachers largely see IBAs en masse in a similar fashion. See **Table 16**.

Table 16

ANOVA analysis of mean rates of usefulness of IBAs

	Sum of Squares	df	Mean Square	F	Sig
Between Groups	3.568	4	.892	1.270	.282
Within Groups	193.143	275	.702		
Total	196.711	279			

SECTION 6 – DISCUSSION

Districts should focus use of IBAs on identifying students for intervention and planning individualized support.

This study used a utility theory to hypothesize that the more useful a teacher perceives a function of an interim benchmark assessment, the more frequently they would use that function in their in-classroom instructional decision-making. The inverse would also theoretically be true. The less the perception of a function's usefulness, the less frequently that function would be used. Therefore, there should be a strong relationship between utility and frequency of use. The findings of this study suggest that for each function studied, there was a strong correlation between its usefulness and the frequency of use by teachers. This is the central heart of utility theory, closely related to rational choice theory. Utility theory within a professional social context postulate that individuals assign different levels of value to items within their environment and make use of those items based on those varying levels of utility (Fishburn, 1968). For example, a mathematics teacher would find a pencil a more useful tool in classroom instruction than a permanent mark. Even though the cost of a permanent marker may be more expensive and more financially valuable than a lead-based pencil, the context in which the math teachers operate makes the use of a pencil all the more valuable and, therefore,

useful.

Within the context of teaching and learning, supporting individual students is of high value, and that value may be driving the preference for teachers to use interim benchmark assessments to inform individual student learning needs (Huck, 2021). As all schools across the globe are still wrestling with learning recovery from COVID-19 pandemic-related learning loss, supporting students who need additional support remains a high priority. Broad-based analysis of student academic performance data would still suggest that schools in the United States have not fully recovered from learning loss (Huck, 2021). And furthermore, there are a number of recently passed legislative acts in the state of Michigan that push schools to continue to provide individualized support for students, including Michigan's Read by Grade Three law, Extended COVID-19 Learning Plans laws, and Michigan's most recent School Aid Act which provides additional funding for high dosage tutoring programs. These factors combined present a compelling policy environment that pushes teachers to continue to focus on individualized learning and to make full use of tools that help inform individualized learning needs. IBAs appear to be moderately suited for that function.

Improving alignment of district policies and expectations of IBA use to the utility that teachers place on IBAs may improve the use of IBAs in classroom-level decision-making and thereby increase their value to teachers. Despite the fact that teachers largely only somewhat believe that their district's IBA is useful, this study does suggest that teachers do indeed believe that IBAs, in general, can be valuable tools to support them in the classroom. While prior research has shown that teachers use data to inform instructional decisions in a generalized way (Tsai & Tosh, 2020), there is little research to

show what specific use cases teachers say they utilize for interim benchmark assessments. While IBA providers may recommend certain practices to their clients, there is not a large body of evidence that shows how district clients, particularly teachers, use IBAs within the classroom. This study suggests, however, in both quantitative responses and qualitative responses, that teachers use IBAs in a screening capacity to identify students who may need additional intervention. Guidance on the use of data from organizations like the Michigan Assessment Consortium and the National Association for School Psychologists recommends that schools screen students for academic deficiency throughout a given school year (National Association of School Psychologists, 2016). Furthermore, Michigan law specifically requires that all students in grades Kindergarten through 3rd grade must assessed for reading deficiencies within the first 90 days of school. Given that screening is such a strongly advised practice, it makes sense that teachers perceive that identifying students for intervention is among the most useful functions of an interim benchmark assessment.

Additionally, districts should consider emphasizing policy and guidance for teachers to use IBAs to plan individualized support for students. Both iReady and NWEA provide specific curriculum resources for students based on their individualized performance. These features may contribute to teachers' strong preference for using IBAs to support individualized instruction. NWEA, prior to its acquisition by Houghton Mifflin Harcourt, provided free online instructional content to students through Khan Academy. In fact, iReady specifically recommends to its clients that students should spend 45 minutes or more engaged in its proprietary online software suite, which provides customized instructional content tailored to the concrete learning needs of each student

(Curriculum Associates, 2021). iReady presents evidence to its clients that students who spend 45 minutes or more in its online software suite are more likely to meet proficiency on state standards (Curriculum Associates, 2021). In some districts, providing students with 45 minutes of iReady computerized lessons during class time is a requirement for teacher lesson planning. All three IBA vendors included in this study provide training and guidance to teachers on specific ways that teachers can plan individualized instruction for students, particularly for students who fall beneath certain performance targets. Providing supplementary instructional content to students, as well as creating instructional groups based on those supplementary instructional resources, appears to have a strong value proposition for a significant proportion of teachers.

Overall, the frequency of use data from this study suggests that teachers, in large measure, use data from interim benchmark assessments to assist them in meeting the needs of individual students. They frequently or occasionally use these data to identify students who may need intervention, use the data to inform small group interventions, help them plan those interventions, and help inform the setting of student learning goals. This aligns with prior research that suggests that these are ideal functions of interim benchmark assessments (Herman, 2010; Perie et al., 2009). Although research has not definitely proven that having access to IBA data shifts instructional practice or improves student outcomes (Perie et al., 2009), if teachers are to use IBA data at the classroom level, focusing on the learning needs of individual students appears to be the most ideal and frequently used function.

Districts should shift focus away from functions that teachers don't find useful.

Prior research and guidance suggest that establishing the purpose of use is a

necessary first step in using assessment data (Michigan Assessment Consortium, 2017; Perie et al., 2007). This may suggest that not all tests, including IBAs, can serve all purposes equally. Certain assessments that are well suited for a particular set of purposes might not be well suited for others. For example, a formative assessment may be well-suited to help a teacher determine the next steps in a particular daily lesson, but it would not be well-suited to help a college admissions officer determine college admission. Therefore, we would reasonably expect a college admissions officer never to use formative assessment to help them determine college admission.

In this same light, teachers rarely or never used interim benchmark assessments to help them determine their own professional learning needs. This study showed that teachers did not find this function useful, and they never or rarely use IBAs for this function. This finding stands in conflict with earlier research, which showed that district administrators sought to have IBAs inform professional learning decisions (Buckley et al., 2010). On its surface, this can be seen as a logical conclusion: the academic performance results of students should drive the professional learning direction of staff. For example, if an interim benchmark assessment shows that students are below grade level in Numbers and Operations in Math, then a teacher should seek professional learning opportunities to help them improve their knowledge and professional capacity to teach mathematical operations. However, if teachers believe that IBAs have low utility for informing their professional learning needs, then district and school leaders should investigate other sources of data to inform professional learning needs. If district adoption, usage, and support of IBAs are in alignment with teacher utility theory, then it is possible that teacher usage of IBAs for specific functions may improve.

When we examine the findings for measuring the instructional effectiveness function of IBAs, we observe that, in large measure, teachers do not find this to be a useful function of IBAs, and teachers do not frequently use IBAs for this purpose. It is interesting to observe that this is a less frequently used function of IBAs. Many districts in this study require school administrators to use data from interim benchmark assessments in their teacher evaluation frameworks.

Michigan educator evaluation law, until recently, required districts to base teacher evaluation ratings, in part, on assessment and growth data derived from valid and comparable assessments across grades. Many districts have local interpretations of this law that drive them to use data from IBAs to measure the instructional effectiveness of teaching and administrative staff. Yet, despite these policy requirements, teachers did not frequently use IBAs to measure their own instructional effectiveness.

Planning whole group instruction is another function of IBAs that was not frequently used by most teachers. In general, data-based decision-making frameworks recommend that teachers use data to make informed decisions about whole group and whole class instruction (Peters, 2021; Prenger, 2018). For example, within the commonly adopted Multi-Tiered System of Student Support Frameworks, there is a strong insistence on improving Tier 1 instruction for all students (National Association of School Psychologists, 2016). Data reviews then become critical components of how schools can improve core instruction for all students. For example, if a review of IBA data suggests that students may struggle with identifying the author's purpose in informational text, then this might inform whole school efforts to increase informational text- based curricular adoption and learning activities. This also might inform how teachers use

classroom curriculum resources to plan instruction for all students to improve performance in this target area. Yet, the findings in this study would suggest that using data from IBAs might not be a preferred method for teachers to measure and inform whole group planning needs.

As discussed earlier, the qualitative analysis of the open-ended responses in the survey revealed, in part, that many teachers reported using multiple sources of assessment data, including diagnostic assessments, formative assessments, and classroom observations, to determine student intervention. This use of multiple data sources by teachers may explain why the identifying students for in-class intervention coefficient might be slightly lower than the other functions. Teachers may use multiple data sources to inform student intervention and, therefore, have varied perceptions of utility and frequency of use for that function, but teachers nearly universally see little value in IBAs for identifying their own professional development needs.

IBAs aren't that much different from each other, but more research is needed.

While teachers in this study appeared to believe that Interim Benchmark assessments were somewhat useful, it is reasonable to assume that there was a difference in perception of usefulness across various interim benchmark assessments; each IBA publisher in this study presents its teachers with different information in different vendors. We have previously established and discussed that there is a broad interpretation of the term interim benchmark assessment in the K -12 education field. Each IBA publisher in this study provides teachers with different information in different ways. What these tools do have in common is that they are administered online directly to students, they are administered three times a school year, and they assess students in

Reading and Math. However, there are many distinct differences between these IBA tools. NWEA aggregates performance across grade levels into what it calls “RIT scores.” iReady’s achievement measure is a “scaled score,” and FastBridge’s is simply labeled a “score.” All three tools place those scores along a performance percentile range, spanning from

1-99, against a national database of student scores respective to each vendor. However, different performance labels are associated with performance percentiles for student scores. iReady labels student scores as being on or below grade level, FastBridge labels scores according to high or low risk, and NWEA labels students with low and high achievement ranges.

It makes sense then that teachers could have different perceptions of the usefulness of IBAs that may be driven by the type of IBA they use. Because each district typically utilizes one IBA across its entire collection of schools, teachers only have experience with one tool and do not use a range of tools within a single year. This study presents a unique opportunity to have some level of standard comparability across the three tools. EdReports, an organization that rates K12 curriculum materials, had plans to provide a review of these tools based on the expertise within their organization. This would have provided the K-12 Education field with a standardized comparison across many interim benchmark assessment tools. However, as of the writing of this dissertation, those plans were abandoned as NWEA and other vendors not in this study withdrew consent to have their assessment undergo review (Fittes, 2022). The purpose of this dissertation was to study teacher perceptions of the usefulness of interim benchmark assessment data in classroom-level instructional decisions and not necessarily to provide

a comprehensive review of each tool. Yet this study does contribute to the field's knowledge about how these tools compare to one another and presents a standard by which to compare different IBA vendors.

Even though there were no statistically significant differences across any of the tools in this study, it is interesting to note that slightly more teachers reported that they found iReady more useful than other tools. However, these results might have been influenced by the n size of iReady users in this study. iReady is a tool used by a minority of school districts in Oakland County. NWEA is the most widely used interim benchmark assessment vendor in the county. FastBridge is used by just a handful of districts. Further research across a much larger population of users, perhaps a statewide study conducted by a state education agency or even a national study, is advised.

Districts should capitalize on experienced teachers to improve IBA utility and usage.

In order to improve teacher perceptions of the utility of IBAs and thereby improve the frequency of their use in the classroom, experienced teachers might be an untapped resource in improving teacher practice within a school. Prior research has shown that professional learning and guidance in the use of data-based decision-making protocols support teacher practice (Datnow, 2015; Tsai & Tosh, 2020). However, research has also shown that personal guidance from veteran teachers to less veteran teachers on classroom instructional decisions was more useful than general guidance on data use alone. This may suggest that to improve the utility of IBAs within the context of instructional decisions, districts may need to incorporate more teacher voices and specific use cases from experienced teacher peers in professional learning design. Simply providing access

to data has not been shown to shift instructional practices (Tyler, 2013), and ultimately, shifting instructional practices is the central goal of data-based decision-making.

While prior research has shown that teachers use data to inform instructional decisions in a generalized way (Tsai & Tosh, 2020), there is little research to show what specific use cases teachers say they utilize for interim benchmark assessments. While IBA providers may recommend certain practices to their clients, there is not a large body of evidence that shows how district clients, particularly teachers, use IBAs within the classroom. This study suggests, however, in both quantitative responses and qualitative responses, that teachers use IBAs in a screening capacity to identify students who may need additional intervention. Guidance on the use of data from organizations like the Michigan Assessment Consortium and the National Association for School Psychologists recommends that schools screen students for academic deficiency throughout a given school year (National Association of School Psychologists, 2016). Furthermore, Michigan law specifically requires that all students in grades Kindergarten through 3rd grade must assessed for reading deficiencies within the first 90 days of school. Given that screening is such a strongly advised practices, it makes sense that teachers perceive that identifying students for intervention is among the most useful functions of an interim benchmark assessment.

Utility studies on other assessment formats may be needed.

Prior research has shown that teachers use a variety of student data to support several classroom decisions (Datnow, 2015; Buckley et al., 2010; Tsai & Tosh, 2020). Indeed, this study has suggested that teachers use a variety of multiple sources of data in their decision-making process; they do not simply rely on one sole source of data to

support their classroom instructional decisions. Given that teachers have multiple sources of data to support their instructional decisions, like the Michigan Student Test of Educational Progress (M-STEP), curriculum-based measures, locally developed assessments, diagnostic assessments, progress monitoring assessments, and information gleaned from the formative assessment process, it is reasonable to suggest that teachers may hold different perceptions of the usability of different types of assessments. For example, it may be reasonable to hypothesize that teachers find the formative assessment process extremely useful to their instructional decisions, while M-STEP, given once a year in the spring of each school year, might not be useful at all. From this study, it is clear that teachers do find IBAs useful, but only to a limited degree. There may be other assessments that they find more useful to the context of the classroom instructional decisions.

Additionally, this study limited its scope to the utility of interim benchmark assessments within classroom instructional decision-making. However, as previously discussed, assessments have many different uses and functions (Michigan Assessment Consortium, 2017), and those functions are not limited to classroom-level teacher-made decisions alone. For example, school districts routinely determine the placement of students in specialized educational services.

Teachers, along with family caregivers, school administrators, and special education staff, collectively determine student placement into specialized educational services. In that context, it may be possible that interim benchmark assessments are very useful because they provide standardized reporting of student achievement compared to state and national standards. However, within the context of the classroom, this study

suggests that teachers do not find IBA very useful for supporting day-to-day classroom instruction. Additional research is needed to discover what is likely to be various levels of utility that teachers place in other assessment methods.

Limitations

Charter Schools, entities that receive authorization from external agencies like a university or school district, often have charter agreements that require schools to meet certain performance metrics. Many times, Charter schools use IBA performance as an accountability measure, and therefore, there may be a different professional attitude towards IBA performance in charter schools than in traditional public schools. There was only one charter school that agreed to be a part of this study, making charter school teachers underrepresented in this study.

Additionally, large school districts, districts with 9,000 students or more, were not included in this study, as it was beyond the feasibility of this study to coordinate the distribution of the survey instrument across such a large body of schools and teachers within a single district. More research is needed across a larger body of teachers, perhaps even a statewide or national study. If interim benchmark assessments continue to be a part of the everyday fabric of K-12 Education, more research should be conducted on how schools use these data to support student learning.

Conclusion

The central inquiry of this dissertation was centered on the utility value that teachers place on interim benchmark assessments within their classroom instructional decision-making process. Are these tests indeed 24 Kart gold that take a front and center

place in teacher instructional decisions? Or, are they fool's gold, promising valuable sources of student information but ultimately ending up being worthless? This study would suggest that the answer is somewhere in between. Teachers find interim benchmark assessment data to be somewhat useful in their decision-making process and use IBA data, along with many other types of data, to support a variety of instructional decisions. IBAs might not be the panacea of useful information that policymakers and IBA vendors claim, but teachers do use these data, to a degree, to inform their practice.

Now, the key question remains: Why? While further research is needed to explain why teachers do not fully and frequently utilize IBA data in their instructional decision-making process, the answer may lie in the scholarship of noted marketing professor Dr. Marcus Collins of the University of Michigan. In his signature text, *For the Culture: The Power Behind What We Buy, What We Do and Who We Want to Be*, Dr. Collins theorizes that the adoption of behaviors is influenced by people, not necessarily by advertisements, product quality, and features, or humbly speaking, publications from university researchers. "People, not marketing communications, convert people" (Collins, 2023, p. 116). Therefore, if policymakers and school leaders want to convert teachers into full and frequent users of IBA data, the path forward may lie in uplifting the practices of their fellow teachers who successfully utilize IBA data within their classrooms. This is akin to finding a congregation of powerful IBA users and uplifting them by preaching the gospel of good data use in classroom practice. The state of Michigan has invested significant funds in districts to support purchasing these IBA tools, which have only modest impacts on classroom practices. It may be high time that policymakers redirect these funds from investment in products and, instead, into people.

SECTION 7 - SCHOLARLY PRACTITIONER REFLECTION

Engaging in this dissertation process, indeed participating in the entire doctoral journey, has deepened my appreciation for research and fueled my intellectual curiosity for the unknown. Unfortunately, it has been my experience in K-12 schools, perhaps largely due to the custodial nature of schools' responsibility for the care of minor children, that there often is little interest or time devoted to deeply exploring and conducting research. Decisions, though well-intentioned, are often rushed and made without sufficient complexity of dialogue. When school practitioners are engaged in self-reflection, I often find that many cannot name the evidence base that supports the strategies they use. Sometimes, those strategies have no evidence base at all. Yet many school stakeholders hold strong convictions about what schools should do to support student learning and well-being. Therefore, through the completion of this dissertation, it is my earnest desire to help shape and shift dispositions toward quality research within the ranks of K-12 educators. It is my desire to make all schools powerful and wonderful places for humanity to grow and thrive.

This first begins with the dissemination of this study. As I have mentioned in the Acknowledgement sections of this dissertation, I am profoundly grateful for the teachers and educational leaders who have supported me in this project. As promised, this spring and summer, I will be sharing the individual results of each district's data with their district-level assessment leader. I will also share the aggregate findings of the study with all Oakland County assessment leaders and assistant superintendents through my role as the Supervisor of the Curriculum and Assessment unit at Oakland Schools. Furthermore,

as I write this very statement, I am pleased to report that this study has been selected by the Michigan School Testing Conference Executive Planning Committee to be delivered in a presentation at its upcoming annual conference in February 2025. This conference is attended by nearly 400 educators across the great state of Michigan and will be the only presentation that addresses how educators use data from interim benchmark assessments to inform their practice. I also plan to pursue opportunities to share the findings from this study at the American Education Research Association (AERA) Annual Meeting in 2025.

I have previously had the privilege of attending AERA in 2023 and 2024, and I can say without a doubt that it is an experience that I have absolutely loved. In combination with my doctoral journey and this dissertation experience, they collectively have given me opportunities to engage in and witness rigorous, thoughtful dialogue on models, methodologies, and projects that vary in scope but are united in one goal: advancing our collective knowledge about how we can best invest in children. Through education, both through policy and practice, we have an opportunity to create futures for our children that can shatter the barriers of the unknown. I welcome that journey, and I hope that this dissertation is but one footstep into that journey towards the unknown.

As I begin this journey, I take with me an appreciation for rigorous research methodology and research design. I have poured many hours into the writing and revision of this dissertation's methodology section, with the careful and astute guidance of Professor Se Woong Lee. This process has been intellectually challenging for me, yet also rewarding. In fact, I specifically chose a quantitatively focused dissertation project because I believe the methodological rigor would be challenging (and fun). This process has developed me as a scholar to not only deeply explore topics of interest but also to

take on new challenges that are new and novel to who I am as an educator. I am a high school English teacher by trade, so statistics and models do not come naturally to me. However, I do fancy myself as a storyteller of sorts, and through this dissertation, I have learned that good research allows us to tell powerful stories. Even in quantitative data, the numbers are indeed speech, and there are stories to be told through the data if we have the skills and ears to listen. In this project, I hope I have told the stories of the hard- working teachers of Oakland County who skillfully incorporate data into their classroom decision-making processes. And I hope to continue telling the stories of many school community members whose voices go unheard in educational policy decision-making.

Through the skills and dispositions I have gained through this dissertation process, I believe I have enhanced my skill in crafting powerful stories to influence critical decision- making – influencing them for the better. This is what I hope to carry forward in my path as an educational leader, both formally in my role as a county-wide administrator and informally as a colleague united with my educational peers to critically examine and re-examine our practices in schools. Collectively, there is so much we know, yet there is also so much we don't know. This process has taught me to embrace the humility of the unknown, chasing the various and varied ways of knowing and bringing people along for the ride. So, in many ways, this dissertation is not the end but rather the beginning of a new phase in my work and identity as a scholar and a servant in helping our schools be the thriving heartbeats of our community.

REFERENCES

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education*, 5, 7-74. http://dx.doi.org/10.1080/0969595980050102

Bugler, D., Marple, S., Burr, E., Chen-Gaddini, M., & Finkelstein, N. (2017). How teachers judge the quality of instructional materials. San Francisco, CA: WestEd.

Buckley, K., Christman, J., Goertz, M., & Lawrence, N. (2010). Building With Benchmarks: The Role of the District in Philadelphia's Benchmark Assessment System. *Peabody Journal of Education, 85:2*, 186-204, DOI: 10.1080/01619561003685346

Carlson, D., Borman, G. D., & Robinson, M. (2011). A Multistate District-Level Cluster Randomized Trial of the Impact of Data-Driven Reform on Reading and Mathematics Achievement. *Educational Evaluation and Policy Analysis, 33*(3), 378–398. https://doi.org/10.3102/0162373711412765

Chojnacki, G.J., Eno, J., Liu, F., Meyers, C.V., Konstantopoulos, S., Miller, S.R., & Ploeg, A.V. (2013). Do Interim Assessments Influence Instructional Practice in Year One? Evidence from Indiana Elementary School Teachers.

Collins, M. (2023). *For The Culture: The Power Behind What We Buy, What We Do, and Who We Want to Be*. Public Affairs.

Curriculum Associates. (2021). I-Ready Efficacy Research Summary. *Curriculum Associates.* Retrieved on March 17, 2024. https://www.curriculumassociates.com/-/media/mainsite/files/i-ready/iready-efficacy-research-summary-2021.pdf

Curriculum Associates. (2023). I-Ready Success in Action. *Curriculum Associates*. Retrieved on August 12, 2023. https://www.curriculumassociates.com/programs/implementing-i-ready

Datnow, A., Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven

decision making: A literature review of international research. *J Educ Change* **17**, 7–28. https://doi.org/10.1007/s10833-015-9264-2

Demchak, M., & Sutter, C. (2019). Teachers' Perception of Use and Actual Use of a Data-Based Decision-Making Process. *Education and Training in Autism and Developmental Disabilities*, *54*(2), 175–185. https://www.jstor.org/stable/26663975

Draper, R. (2002). Every Teacher a Literacy Teacher? An Analysis of the Literacy-related Messages in Secondary Methods Textbooks. *Journal of Literacy Research*. 24(3):357-384

Erichsen, K., & Reynolds, J. (2020). Public school accountability, workplace culture, and teacher morale. *Social Science Research, 85*. https://doi.org/10.1016/j.ssresearch.2019.102347

Fishburn, P. (1968). Utility Theory. *Management Science* 14(5):335-378.

Fittes, E. (2022). Are Interim Assessments Living Up to Their Billing? New Review Aims to Find Out. *EdWeek Market Brief*. Retrieved March 17, 2024. https://marketbrief. edweek.org/marketplace-k-12/interim-assessments-living-billing-new-review-aims-find/

Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards-FAST SCASS. (2018). *Revising the Definition of Formative Assessment*. Washington, DC: Council of Chief School State Officers.

Formative Assessment for Students and Teachers State Collaborative on Assessment and Student Standards. (2018). *Revising the Definition of Formative Assessment*. Washington, DC: Council of Chief School State Officers.

Guiklford, J.P. (1954). Psychometric methods (2nd ed.). McGraw-Hill.

Goren, P.N. (2010). Interim Assessments as a Strategy for Improvement: Easier Said Than Done. *Peabody Journal of Education, 85*, 125 – 129.

Herman, L. Osmundson, E., Dietel, R. (2010). *Benchmark Assessment for Improved Learning* (AACC Policy Brief). Los Angeles, CA: University of California.

Huck, C. Zang, J. (2021). Effects of COVID-19 Pandemic on K-12 Education: Systemic Literature Review. *Educational Research and Development Journal*, 24(1) pp. 53-84

Ikemoto, G.S., & Marsh, J.A. (2007). Cutting through the "Data-Driven" Mantra: Different Conceptions of Data-Driven Decision Making. *Teachers College Record: The Voice of Scholarship in Education, 109*, 105 - 131.

Jacob, B., Crespin, R., Libassi, C. Dynarski, S. (2016). Class Size in Michigan: Investigating the Risk of Being in Very Large Classes. *Education Policy Initiative Policy Brief #5*. https://edpolicy.umich.edu/sites/epi/files/2021-07/class-size-policy-brief-revised.pdf

Kilbre, T., Hopkins, B., Strunk, K., Yu, D. (2022). *Michigan's Fall 2021 Benchmark Assessments – April 2022*. Education Policy Innovation Collaborative. https://epicedpolicy.org/wp-content/uploads/2022/04/ Benchmark_Report_April2022.pdf

Kilbre, T., Hopkins, B., Strunk, K. (2021). *Michigan's 2020-21 Benchmark Assessments – August 2021*. Education Policy Innovation Collaborative.

https://epicedpolicy.org/wp-content/uploads/2022/02/Benchmark-

RptvI_Aug2021.pdf

Loeb, S., Dynarski, S., McFarland, D., Morris, P., Reardon, S., & Reber, S. (2017). Descriptive analysis in education: A guide for researchers. (NCEE 2017–4023). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance.

Mandinach, E., Schildkamp, K. (2021). Misconceptions about data-based decision making in education: An exploration of the literature, *Studies in Educational Evaluation* 69. https://doi.org/10.1016/j.stueduc.2020.100842.

May, H. & Robinson, M. (2007). A Randomized Evaluation of Ohio's Personalized Assessment Reporting System (PARS).

Melucci, L. (2013). Teacher Perceptions and Use of Data-Driven Instruction: A Qualitative Study. [Doctoral Dissertation, Capella University]. https://www.proquest.com/docview/ 1449198403

Michigan Assessment Consortium. (2017). *Assessment Literacy Standards*. Michigan Assessment Consortium. https://www.michiganassessment consortium.org/ assessment-literacy-standards/

Michigan Association of Intermediate School Administrators General Education Leadership Network Early Literacy Task Force (2018). Essential instructional practices in language and emergent literacy: Birth to age 3. Lansing, MI: Authors. https://www.gomaisa.org/downloads/literacy_essentials/9.2022_updated_birth_to_age_3_literacy_essentials.pdf

Michigan's Center for Educational Performance and Information. (2023). *Staffing Count*

for Oakland Schools ISD. https://mischooldata.org/staffing-count?Common_Locations=1I,106,0,0~2A,0,0,0&Common_SchoolYear=24&Common_LocationIncludeComparison=False&Portal_InquiryDisplayType=Snapshot&Common_StaffingReportCategory=Longevity&Common_StaffingGroup=Teachers&Common_StaffingCountType=FullTimeEquivalency. Retrieved August 12, 2023.

Michigan Department of Education (2022). Education Assessment: What It Is, What It Means, And What It Offers. https://www.michigan.gov/mde/-/media/Project/Websites/mde/Year/2021/03/24/MI_Education_Assessment_System.pdf?r

Nahmias, C. K. (2015). *Content area teachers' perceptions of the factors that promote or inhibit infusion of content area reading strategies into instruction*). ProQuest, LLC.

National Association of School Psychologists. (2016). *Integrated Model of Academic and Behavioral Supports* [Position statement]. Bethesda, MD.

National Center for Education Statistics. (2023). *Characteristics of Public School Teachers. Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved April 17, 2023, from https://nces.ed.gov//programs/coe/indicator/clr.

Nordengren, C. (2022). *Read the latest in student goal-setting guidance*. NWEA. https://www.nwea.org/blog/2022/read-the-latest-in-student-goal-setting-guidance/

Norqvist, L., & Ärlestig, H. (2021). *Systems Thinking in School Organizations – Perspectives From Various Leadership Levels*. Journal of Educational Administration, 59(1), 77–93. https://doi.org/10.1108/JEA-02-2020-0031

Perie, M., Marion, S., Gong, B. (2007). Moving Toward a Comprehensive Assessment System: A Framework for Considering Interim Assessment. Dover, NH: The National Center for the Improvement of Education Assessment, Inc.

Peters, M. T., Förster, N., Hebbecker, K., Forthmann, B., & Souvignier, E. (2021). Effects of Data-Based Decision-Making on Low-Performing Readers in General Education Classrooms: Cumulative Evidence From Six Intervention Studies. Journal of learning disabilities, 54(5), 334–348. https://doi.org/10.1177/00222194211011580

Perie, M., Marion, S., Gong, B., Wurtzel, J. (2009). The Role of Interim Assessments in a Comprehensive Assessment System. Dover, NH: The National Center for the Improvement of Educational Assessment, Inc.

Prenger, R. & Schildkamp, K. (2018). *Data-based decision making for teacher and student learning: A psychological perspective on the role of the teacher*. Educational Psychology, 38:6, 734-752. https://doi.org/10.1080 /01443410.2018.1426834

Schramm, M. (2016). What is the average class size in Michigan's schools? Michigan Public Radio. https://www.michiganpublic.org/education/2016-09-15/what-is-the-average-class- size-in-michigans-schools

Shepard, L.A. (2009). Commentary: Evaluating the Validity of Formative and Interim Assessment. Educational Measurement: Issues and Practice, 28: 32-37. https://doi.org/10.1111/j.1745-3992.2009.00152.x

Sherwood, E., Domingues, M., Elsen, G., Tagessen, J., Wang, H., Xie, W. S., Kane, P., Maldonado, A., & Cyrus, M. (2021). Collaboration for Literacy Improvement: The Experiences of Teacher-Researchers in the First Stages of an Inquiry Process. Clearing House, 94(3), 128–136. https://doi.org/10.1080/00098655.2021.1907147

Slavin, R.E., Cheung, A.C., Holmes, G., Madden, N.A., & Chamberlain, A. (2013). Effects of a Data-Driven District Reform Model on State Assessment Outcomes. American Educational Research Journal, 50, 371 - 396.

Sparks, D., Ralph, J., Malkus, N. (2012). Public School Teacher Autonomy in the Classroom Across School Years 2003- 04,2007-08 and 2011-12. Stats in Brief. NCES 2015-089. National Center for Education Statistics.

Stevens, E. A., Murray, C. S., Scammacca, N., Haager, D., & Vaughn, S. (2022). Middle school matters: examining the effects of a schoolwide professional development model to improve reading comprehension. Reading & Writing, 35(8), 1839–1864. https://doi.org/10.1007/s11145-022-10271-9

Taherdoost, H. (2017). Determining Sample Size; How to Calculate Survey Sample Size. International Journal of Economics and Management Systems. 2, 237-239.

U.S. Census Bureau. (2022). Quick Facts Michigan Oakland County United States. Retrieved August 13, 2023. from https://www.census.gov/quickfacts/fact/table/MI,oaklandcounty Michigan, US/PST045222

Usma-Wilches, J. (2006). "Teacher autonomy: a review of the research literature." Graduate Research Papers. 1634. https://scholarworks.uni.edu/grp/1634

Leeuwen, A.V. (2019). Teachers' perceptions of the usability of learning analytics reports in a flipped university course: When and how does information become actionable knowledge? Educational Technology Research and Development, 67, 1043-1064.

Tuma, A.P., Doan, S., Henry, D., Lawrence, R., Woo, A., & Kaufman, J.H. (2021). Teachers Perceptions of What Makes Instructional Materials Engaging, Appropriately Challenging, and Usable: A Survey and Interview Study.

Tsai, T., Tosh, K. (2020). Educator Access to and Use of Data Systems. RAND Corporation. https://www.rand.org/pubs/research_reports/RR2575z8-1.html

Tyler, J. (2013). If you Build It, Will They Come? Teachers' Online Use of Student Performance Data. Education Finance and Policy, 8(2), 168–207.

Wang, E., Tuma, A., Doan, s., Henry, D., Lawrence, R., Woo, A., Kaufman, J. (2021). Teachers Perceptions of What Makes Instructional Materials Engaging, Appropriately Challenging, and Usable: A Survey and Interview Study, RAND Corporation. United States. Retrieved from https://policycommons.net/artifacts/4836728/teachers-perceptions- of-what-makes-instructional-materials-engaging-appropriately-challenging-and-usable/5673296/ on 05 Oct 2023. CID: 20.500.12592/t0knwp.

Worth, J., Van den Brande, J. (2020). *Teacher autonomy: How does it relate to job satisfaction And retention*? Slough: National Foundation for Educational Research.

APPENDIX A

Survey Items Framework

Participant Identification Block

1. Which district or charter school do you work in? [Drop Down List of Districts & Charter Schools]
2. Which grade level(s) do you primarily teach [Multi-Select Multiple choice of K-8]
3. Which content areas(s) do you primarily teach?
    a. Reading
    b. Math
    c. Reading and Math
    d. Neither Reading or Math

Interim Benchmark Assessment Identification Block

Note: The State of Michigan provides funding for schools to administer an approved interim benchmark assessment to students in grades K-8 in Reading and Math.

4. Which of the following Interim Benchmark Assessments does your district administer to your students?
    a. FastBridge
    b. iReady
    c. NWEA
    d. STAR
    e. District Developed Assessment
5. How many years of experience do you have in administering or using data from this assessment? [Multiple choice, 0-5 years or more]

Interim Assessment Usefulness Block

6. How strongly do you agree or disagree that interim benchmark assessments can help you improve your instructional practice? [Strongly Disagree, Disagree, Agree, Strongly Agree]

7. Overall, how useful do you believe your district's interim benchmark assessment is inyour instructional decision-making? [Not Useful, Somewhat Useful, Useful, Very Useful]

Instructional Decisions Block

8. How useful do you believe the data from your interim benchmark assessment is in making the following instructional decisions? [Not Useful, Somewhat Useful, Useful, Very Useful]

a. Creating instructional groups in my classroom
b. Planning whole group instruction
c. Planning instruction for individual students
d. Setting learning goals with students
e. Identifying students for in-class intervention
f. Measuring my instructional effectiveness
g. Identifying my own professional development needs
h. None of the above

9. How frequently do you use the data from your interim benchmark assessment in the following instructional decisions [Never, Rarely, Occasionally, Frequently]

a. Creating instructional groups in my classroom
b. Planning whole group instruction
c. Planning instruction for individual students
d. Setting learning goals with students
e. Identifying students for in-class intervention
f. Measuring my instructional effectiveness
g. Identifying my own professional development needs

10. What other ways do you use data from this assessment in your instruction? [open text box field]

Survey items based on research (Bulkley et al., 2010; Olah et al., 2010; Herman et al., 2010; Perie et al., 2007)

VITA

Steven Snead serves as the Supervisor of Curriculum & Assessment at Oakland Schools. His team is responsible for providing educators across 28 districts and 22 charter schools with professional learning, resource development, and consultation to support deep learning for all students. He is also a member of the Board of Directors for the Michigan Assessment Consortium, helping to promote assessment literacy across the state of Michigan. He's given numerous local, statewide, and national presentations on assessment-focused practices, including district assessment design, formative assessment, MTSS, educator evaluation, and equitable assessment practices.

Steven is a proud graduate of Cass Technical High School in Detroit. He holds a B.A. in Secondary Education from the University of Michigan, a M.Ed. from Wayne State University, and a Doctorate in Educational Leadership and Policy Analysis from the University of Missouri. The focus of his doctoral research was exploring the utility of interim benchmark data in classroom teacher instructional decision-making.