DIFFERENCE IN STUDENT ACHIEVEMENT BETWEEN MATH CURRICULUM CHOICE IN THE STATE OF LOUISIANA

The undersigned, approved by the Department Chair of Graduate Studies in Education, have examined a dissertation entitled:

DIFFERENCE IN STUDENT ACHIEVEMENT BETWEEN MATH CURRICULUM CHOICE IN THE STATE OF LOUISIANA

Presented by Brian D. Neugebauer, a candidate for the degree of Doctor of Education and hereby certify that in their opinion it is worthy of acceptance.

Dr. Kevin Patterson, Committee Member
Assistant Professor of Education

Dr. Benny Fong, Committee Member
Assistant Professor of Education

DIFFERENCE IN STUDENT ACHIEVEMENT BETWEEN MATH CURRICULUM CHOICE IN THE STATE OF LOUISIANA

A Dissertation
Presented to
The Faculty of the Graduate Education Department
Southwest Baptist University

In Partial Fulfillment
of the Requirements for the Degree

Doctor of Education

By

Brian D. Neugebauer, B.S., M.S.
Dr. Kevin Patterson, Dissertation Advisor

July 2023

ACKNOWLEDGMENT PAGE

To my wife and kids, I want to acknowledge the cost you paid and the support you provided to help me complete this journey. You supported me through this process with compassion, endured me missing more events and memories than we ever planned, and found reservoirs of sacrificial love that sustained us. With true excitement and joy toward our future, thank you. Explicitly, to my wife Rachel, your sacrifice for this accomplishment was likely the only one to match, and possibly even exceed, my own. There is only one of you, and I am both thrilled and blessed to be the one with you. This accomplishment is not mine; it is ours. My daughter Audrey, your life is an answer to prayers and a blessing that reaches beyond your grasp. You possess a spirit that is both rare and unmatched in its zeal for life; I am immensely excited to share in more of it and see where your life takes you. Drew, my son, you joined our lives during a rather crazy time for us and the world and are so thankful you did despite vast uncertainty. Your peaceful joy and inquisitive spirit provide a unique blend of stability and entertainment for everyone near you. I am thankful to be a larger part of both your present and your future as I watch you develop into the man you will become.

To the rest of our family, thank you. I would like to directly thank my parents, along with Rachel's parents and grandparents, for your contributions to this process and being a part of our team. You stepped in with support, meals, childcare, trips of encouragement and relaxation, and conversations that lifted our spirits. Connie, your mastery of the English language was a tremendous help; thank you for the time you spent helping me craft key segments of ideas with a constant willingness to brainstorm how I could best communicate my ideas. For the rest of our family, we appreciate both the big and small gestures of support, patience, and understanding during these busy years.

To my committee, thank you for your sustained efforts over the past few years to guide my thoughts, provide continual feedback, and help me stay focused on the completion of the task. I realize all dissertations have their own flow and process, but you unknowingly took on the challenge of guiding someone who is both dyslexic and ADHD, with strong tendencies to work vigorously off and on for weeks, or sometimes months, and then bombard you with either questions or several new pages of writing that developed during that time, presenting a unique experience for all involved. Specifically, Dr. Patterson, know that the constant encouragement to move me forward paired with the time you spent responding to my detailed emails, on the phone with me in conversation, or through Zoom meetings helped this task reach completion. Likewise, Dr. Fong, you enhanced both my process and research overall by pushing my thinking, making sure no idea or concept was missed, and ensuring everything I did was thoughtful and with purpose. Again, thank you for your contributions toward helping me achieve this goal.

To my support network, thank you. Cohort 8, I could not have been a part of a better group of smart, encouraging, lively colleagues, and you helped make this process fun. To my friends, I appreciate your constant invitations to events I could rarely make but knew I needed, the fun and laughter when I was able to participate, and your understanding and patience during times I wish I could have supported you more. For those who participated in my research, thank you for taking the time to respond; this project could not have been possible without you. Last, but foremost, is the provision of purpose, energy, and grace supplied by God through this process, and my life as a whole, that helped me not only complete this task but continues to help me develop as a person to positively impact those I encounter.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory to compare math curricula choice to Louisiana Education Assessment Program (LEAP) math achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. The independent variable of interest, math curricula choice, was generally defined as curriculum materials. The dependent variable of interest was third, fourth, and fifth grade math LEAP achievement scores. Studies from diverse researchers, both recent and less recent, suggest different curricula materials can have varying results on student learning (Bhatt & Koedel, 2012; Bhatt et al., 2013; Koedel et al., 2017; Lein Authement, 2022; Polikoff, Petrilli, et al., 2020; Walsh, 2009; White, 2018). Considering the potential value of curriculum materials, the LDOE attempted to leverage curriculum material influence, partially through the aid of a tiered ranking system of Tier 1 through Tier 3, for school use to help inform curriculum material selection to facilitate improvement in student achievement (Kaufman et al., 2018).

Information for this study was gathered through LDOE provided data, survey results, and school parishes within Louisiana. A one-way ANOVA was used to compare the means from the different tiers. The findings from this study produced mixed results and support the need for continued persistence within this area of study. Explicitly, more investigation should determine whether the Tier 1 curriculum material's high performance in maintaining the highest average scores for all comparisons with reliable sample sizes for each grade and year of this study, and statistically significant results for higher achievement during the 2017-2018 school year, was a consistent pattern with Tier 1 curriculum materials for other grades and years, other curriculum materials nationwide, or if these results are an irregularity.

CHAPTER ONE

Introduction

Albert Einstein once stated, “Do not worry about your difficulties in mathematics. I can assure you mine are still greater” (Hawking, 2009, p. 337). While the accuracy of Einstein’s self-deprecating statement might be up for debate, the reality is many American students are experiencing difficulties in mathematics. A quick review of both international and national achievement of American students will confirm this problem. Using the Program for International Student Assessment (PISA) for international comparison, the United States ranked 37th out of about 80 countries in mathematics in 2018 (Schleicher, 2019), which was only a marginal improvement from being ranked 39th in 2015 to a similar listing of countries (Organization for Economic Co-operation and Development [OECD], 2018). However, the National Center for Educational Statistics (NCES) painted a better, but not completely reassuring, picture from the 2019 Trends in International Mathematics and Science Study (TIMSS) ranking fourth grade United States students at 15 out of 64 and eighth grade students as 11 out of 46 (NCES, 2019).

In the United States, the National Assessment of Educational Progress (NAEP) test in 2017 showed the math gap between high-performing and low-achieving math students continued to grow and has been growing since 2007 (Oglesby-Phelps, 2022; Sparks, 2018). A major factor influencing the achievement gap and the growth in the achievement gap is the expanding skill gap between students from different socioeconomic backgrounds (Hanushek et al., 2019; Owens 2018; Whilby, 2020). Furthermore, the overall scores on the NAEP test for fourth graders from 2007 to 2017 were stagnant, ranging from 240 to 242 on a 500-point scale (U.S. Department of Education, National Center for Educational Statistics [NCES], 2019). While these scores are higher than the results during the 1990s, ranging from 213 to 226, they still reflected a decade of scores that lack growth.

The combination of stagnant growth and a widening achievement gap in mathematics has educators, administrators, and policy makers searching for new ways to improve student learning and achievement. While several factors can influence student achievement, one of the prevailing avenues school districts have pursued to increase student achievement is using effective curricula materials (Koedel et al., 2017; Reys et al., 2003; Solomon et al., 2019; Superfine et al., 2010). To aid school districts, or parishes as they are referred to in Louisiana, in selecting quality curriculum materials, the Department of Education in Louisiana (LDOE) has developed a tier system to evaluate curriculum materials' alignment to the state standards and grade level learning (Louisiana Department of Education [LDOE], 2022).

The researcher examined what curriculum materials were used for third, fourth, and fifth grade mathematics within the state of Louisiana to see if student achievement differed based on curriculum material choice between each tier. Typically, the small amount of research related to what curriculum materials are used in schools is a major roadblock when it comes to evaluating their use (Blazar et al., 2020; Cummins-Colburn, 2007; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018), but the state of Louisiana's tiered system with preexisting textbook groups alleviated some of this challenge by creating population groups. This allowed the researcher to attempt to discover the differences in student achievement and curriculum material choice. To achieve this goal, the researcher used the existing Louisiana Department of Education textbook tiers (LDOE, 2022) as a curriculum material differentiator. To gain a large data set, the researcher sent out surveys to all the elementary school parish administrators in Louisiana to determine what curriculum materials were used at each school and utilized the LDOE data to review student achievement.

In Chapter One, the researcher presents the problem statement by briefly discussing the consequences of not providing proper math education and then addressing the general lack of literature around curriculum material choice and student achievement in mathematics. Robert Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) formed the theoretical framework of this study and is utilized within the purpose statement to examine curriculum material choice and mathematical student achievement in the state of Louisiana. The significance of the study detailed how this research adds to the body of literature in this area while also inspecting whether a statistically significant difference between curriculum material choice and student achievement in mathematics exists. Other foundational aspects of Chapter One include the research questions, null hypotheses, definition of terms, limitations, delimitation, and assumptions, along with what designs and controls were used to help this study remain valid and reliable.

Problem Statement

A strong mathematical understanding is a key component in ensuring an individual's success in academics and future career, therefore effective math instruction is crucial for building a healthy workforce (Daro & Asturias, 2019; Denton, 2021; Hiser, 2016; National Council of Teachers of Mathematics [NCTM], 2018; White, 2018). Logically, failure to properly educate students in mathematics will not only negatively impact each individual, but collectively have an adverse effect on the society to which individuals belong. As a result, the best learning outcomes will only be achieved when educators promote equity in education by providing schooling that prepares all students for their future in academics or the workforce regardless of students' initial abilities (Lambert 2018). Analyzing curricula materials based on student achievement scores has been done before and is an area where further research is needed (Blazar et al., 2020; Bhatt & Koedel, 2012; Chingos & Whitehurts, 2012; Hiser, 2016; Kane, 2016; Koedel et al., 2017; Polikoff, Campbell, et al., 2020; Nargi, 2018; Ruggeri, 2021; Solomon et al., 2019; Steiner, 2017).

Some recent studies suggested different curricula materials have varying impacts on student learning (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Cress, 2019; Koedel et al., 2017; Lein Authement, 2022; Polikoff, Petrilli, et al., 2020; Walsh, 2009; White, 2018). Yet, administrators are often unable to make better decisions related to instructional materials due to a lack of evidence on the impact of the materials being used (Blazar et al., 2020; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Koedel et al., 2017; Polikoff, Campbell, et al., 2020; Ruggeri, 2021). Koedel et al. (2017, p. 3) described the deficiency of research as a "frustrating lack of information," while Steiner (2017, p. 1) viewed the shortage of research in this area as a "matter of deep concern and urgent need." Of the research that has been done on the effectiveness of curricula, few research methods stand up to scrutiny (Chingos & Whitehurts, 2012; Solomon et al., 2019). As school districts and policy makers continue to recognize the importance of math education and measure student achievement based on standardized tests, school districts are searching for the best approaches for not only teaching students, but also demonstrating student achievement on standardized tests. Because research aimed at curriculum materials has received significantly less focus than the research and theory of teaching, opportunity for discovery in this area is plentiful (Remillard, 2005). Logically then, a proper analysis of curricula material choice, which often requires an initial investigation into what curriculum materials are being used (Kane, 2016; Koedel et al., 2017; Polikoff, 2018), is a key part of the educational structure that needs to be scrutinized. The state of Louisiana attempted to help teachers and school parishes partially overcome these difficulties with their curriculum materials tier system.

To help school districts find curriculum materials that reliably align with the Louisiana state standards, the Louisiana Department of Education provided a tiered ranking system for school districts to optionally use to help guide curriculum material selection. The tiered system consisted of curriculum materials rated as Tier 1 through Tier 3. Curriculum materials categorized as a Tier 1 material have been reviewed and shown to best align to the Louisiana state standards with a Tier 2 and then a Tier 3 material designation for material choice less aligned to the standards than the previous (LDOE, 2022). This process will be discussed in more detail in Chapter Two. This study utilized Louisiana Department of Education's tier groupings of curriculum materials, along with any non-tiered materials not listed, to determine whether a difference exists between curriculum materials and student achievement between each tier.

Purpose Statement

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) to compare math curricula choice to Louisiana Education Assessment Program (LEAP) math achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. The independent variable of interest, math curricula choice, will be generally defined as curriculum materials. The dependent variable of interest is third, fourth, and fifth grade math LEAP achievement scores. Both recent and historical research show a strong connection in mathematics education between the written curriculum within the textbook or program choice and the curriculum and instruction that takes place in the classroom (Arican, 2018; Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Monaghan, 2013; Remillard, 2005; Ruggeri, 2021; Tyson & Woodward, 1989; Walsh, 2009; Westbury 1990).

Specifically, within the math classroom, Ball and Cohen (1996) and Bellens et al. (2020) argued textbooks were the leading factor for what takes place during instruction, with Ball and Cohen (1996) and Walsh (2009) asserting few teachers detour from textbook outlines and information. Additionally, Arican (2018) and Ruggeri (2021) both stated the use of textbooks or curriculum materials to implement instruction was heightened further when teachers were not comfortable teaching mathematics. Solomon et al. (2019) went on to note approximately 85 percent of elementary classrooms within the United States rely on at least one curriculum material for science and mathematics. Elliott (1990) states more than any other subject area, math textbooks are more prone to be used as taught curriculum at the elementary level. Then additional research conducted by Tyson and Woodward (1989) showed 75 percent to 95 percent of the total classroom instruction at the elementary level comes from the textbook structure. This led Monaghan (2013) to proclaim math textbooks as the “de facto” curriculum and the best textbooks should be viewed as the ones that get the best results, which Polikoff, Petrilli, et al. (2020) also contended. As a result, studies in this area often use textbook, curriculum, curricula, and curriculum materials interchangeably to represent the written curriculum from the textbook or program of choice (Blazar et al., 2020; Koedel et al., 2017; Rahman, 2018; Suppa, 2018), which this study has also done to examine the differences between curriculum material choice and student achievement.

Significance of Study

The significance of this study was found by examining whether a statistically significant difference exists comparing student achievement between groups of curriculum material choices. As discussed previously, analysis related to the differences in student achievement and curriculum choices is lacking (Blazar et al., 2020; Bhatt & Koedel, 2012; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Kane, 2016; Koedel et al., 2017; Polikoff, 2018; Steiner, 2017), thus significance from this study is found by adding to this area of need. Achievement scores on high stakes tests were a main source of measuring student success for years, but it was not until the No Child Left Behind (NCLB) Act in 2001 that it became national policy (Ford, 2018). This emphasis on using an accountability system utilizing high stakes testing was carried into the Common Core State Standards (CCSS) era in the fields of mathematics and English Language Arts (ELA). For students and schools to achieve greater success on these standard tests, it is important for both best practices from research and practical application to complement each other within curriculum designs (Schoenfeld, 2014). Since a curriculum material change requires some level of both time and money, it is important to know if there is a statistically significant difference between the available choices for schools to best allocate their time and money. But lack of research in this area means educators and policy makers are merely beginning to recognize the degree to which math curriculum materials influence math learning (Solomon et al., 2019), and the need for more research in this was an area this study helps address.

This issue is further complicated by the fact there is only a small amount of research related to what curriculum materials are being used in schools (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). Discovering what curriculum materials are used is time consuming since only a handful of states gather information on curriculum material adoption use (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). Thus, determining what differences exist between different curriculum choices is inherently problematic and starts with an investigation related to what materials are utilized within the research population. The state of Louisiana provided a unique opportunity to partially overcome this challenge since the state has already separated curriculum materials into tiers based on how closely they are aligned to the state standards. This allowed for analysis of curriculum materials within each preexisting tier rather than individual analysis of each curriculum material choice.

Research Questions

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) to compare math curricula choice to math Louisiana Education Assessment Program (LEAP) achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. Several questions were examined within this study related to the difference of math curriculum material choice and Louisiana student achievement on the LEAP:

RQ1. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ1a. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ1b. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ1c. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

RQ2. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ2a. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ2b. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ2c. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

RQ3. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ3a. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ3b. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ3c. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

Null Hypotheses

The researcher utilized Louisiana Department of Education (LDOE) data to examine the math achievement scores of all public Louisiana school districts for grades three through five to scrutinize the null hypotheses:

$H_0 1$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP.

$H_0 1a$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2016-2017.

H_o1b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2017-2018.

H_o1c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2018-2019.

H_o2: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP.

H_o2a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2016-2017.

H_o2b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2017-2018.

H_o2c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2018-2019.

H_o3: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP.

H_o3a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2016-2017.

H_o3b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2017-2018.

H_o3c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of

Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2018-2019.

Theoretical Framework

Cumulative learning theory has been used as a framework for developing curriculum in schools over the past several decades (Beatty & Pritchett, 2012; Okey & Gagne, 1970; Spaull & Kotze, 2015). Gagne's cumulative learning theory was of particular interest to this study because it has been used as a framework within education-based studies before (Chen, 2021; Udensi, 2019), applied to examinations of mathematical learning within studies (McBride, 1996; Walsh, 2009), and considered in similar studies to specifically compare the difference of curriculum material selection and student achievement (Walsh, 2009). Gagne's cumulative theory can be summarized by three basic tenets. The first is to deliver instruction on component tasks that scaffold to the desired task. Second, to confirm each component task is mastered. Third, to sequence component tasks for optimum transference to the desired final task (Gagne, 1962a, 1962b, 1965; Walsh 2009). These three tenets essentially summarize what curriculum materials overall are designed to do, as they help the teacher scaffold learning through instruction, monitored progress through learning activities, and provide sequenced learning with clear assessment over the established learning goals (Machalow, 2020; Remillard & Kim, 2020; Stein et al., 2007; Suppa 2018). Additional frameworks used to study the difference of curriculum material selection and student achievement that support Gagne's cumulative learning theory were the behaviorist stimulus-response works of Pavlov, Watson, and Skinner. This is discussed in more detail within Chapter Two.

Definition of Terms

Accountability systems: systems that evaluate a school's performance based on predetermined performance measures (Hudak, 2022).

Achievement Gap: the differences between the test scores of minority students and/or low-income students and the test scores of their White peers (Pardini, 2022).

Behaviorism: theory of learning based upon the notion all behaviors are acquired through one's conditioning. Behaviorist theory posits a person's conditioning occurs through interaction with the environment while responses to environmental stimuli shape our actions (Uibelhoer, 2020).

Casual-comparative design: a research design that seeks to find relationships between independent and dependent variables after an action or event has already occurred (Rivero, 2022).

Common Core Standards: a set of high-quality academic standards in mathematics and English language arts/literacy (ELA). These learning goals outline what a student should know and be able to do at the end of each grade. The standards were created to ensure all students graduate from high school with the skills and knowledge necessary to succeed in college, career, and life, regardless of where they live. (Common Core State Standards Initiative [CCSSI], 2022a.).

COVID-19: 'CO' stands for corona, 'VI' for virus, and 'D' for disease. Formerly, this disease was referred to as '2019 novel coronavirus' or '2019-nCoV.' The COVID-19 virus is a new virus linked to the same family of viruses as the Severe Acute Respiratory Syndrome (SARS) and some types of common cold (Barthelemy, 2022).

Criterion-Referenced: Determined how an individual is performing in relation to a criterion, or educational objectives (Conrad, 2021).

Curriculum: the content taught to students, the sequencing structure of that content, and the learning goals associated with that content (Suppa, 2018).

Data-driven decision-making: a process using results of summative and formative assessment data to improve instructional planning across schools (Mai-Huntley, 2022).

Educational Standards (Standards): what students should know and be able to do to demonstrate proficiency by the end of a grade level or course (Robertson, 2022).

Enacted Curriculum: refers to what actually takes place in the classroom related to the curriculum including the teacher's implementation of the official curriculum and the ways in which learners experienced the official curriculum (Roeder, 2020).

Fidelity: refers to the provision or delivery of instruction in the way it was designed or prescribed (Joyner, 2022).

High Stakes Testing: standardized testing used as a data source to determine distribution of funds/resources and used as a means of evaluation of students/teachers (Benton, 2022).

Intended Curriculum: The intended curriculum refers to the teacher's aims, learning goals, and intentions for students (Suppa, 2018).

LEAP or LEAP 2025 Test: the standardized test given in the state of Louisiana to students in grades three through eight. The acronym LEAP stands for Louisiana Educational Assessment Program. The Louisiana Educational Assessment Program (LEAP 2025) test is used to assess how well students perform (Waller, 2021).

Parish: the equivalent of a school district in the state of Louisiana.

Student Achievement: the measurement of academic content a student learns in a given timeframe (Barthelemy, 2022).

Written Curriculum: refers to the written content, learning goals, activities, and resources the teacher uses to plan lessons, such as a textbook or teacher's guide (Suppa, 2018).

Limitations

Limitations within this study have been identified and include:

1. District and school curriculum choice: The researcher was not able to control school district curriculum material selection.
2. Fidelity: The researcher did not account for the degree to which teachers implemented the curriculum materials as intended by the developer.
3. Implementation dip: The researcher did not categorize results based on the number of years curriculum materials had been utilized within each school district.
4. Participant drop-out: The researcher is not able to account for student migration between different curriculum materials within a school year and can only evaluate the student school and curriculum material choices at the time of the LEAP.
5. Pedagogy: The researcher did not examine pedagogical differences between teachers who used different curriculum materials. Differences in pedagogy within different socioeconomic groups were also not examined.
6. Professional development: The researcher did not include analysis based on either the quality or quantity of professional development on curriculum materials by participants within the study.

7. Purposive Sampling: The researcher sent a survey to each public elementary school building administrator in the state of Louisiana with some combination of grades three, four, or five within the building.
8. Reliability and validity of test: The researcher examined the reliability and validity of the LEAP data presented by LDOE.
9. Socioeconomic levels: The research was not able to control which schools or school districts utilized different curriculum materials based on socioeconomic levels or account for differences in educational achievement based on socioeconomic levels.
10. Teacher effectiveness: The researcher did not attempt to gauge the effectiveness of individuals or groups of teachers within the LDOE tier groups.
11. Time period: The researcher used data from the 2016-2017, 2017-2018, and 2018-2019 school years. Due to COVID-19 interference with 2019-2020 testing, the earlier year of 2016-2017 was added to the data set to replace the 2019-2020 school year.

Delimitations

Delimitations within this study have been identified and include:

1. Curriculum: The researcher only analyzed curriculum material choice between Tier 1, Tier 2, Tier 3, and non-tiered curriculum material choices for 2016-2017, 2017-2018, and 2018-2019 returned from the survey.
2. Location: The research only used public school districts within the state of Louisiana.
3. Sample size and selection: LEAP data was only analyzed for elementary schools, specifically grades three, four, and five.

4. Research questions: The researcher limited the research questions to only include questions related to the null hypotheses and not all questions that might relate to this area of study.
5. Theoretical objectives: Gagne's cumulative theory provided the framework for analyzing differences in student achievement and curriculum material choice; other theoretical concepts related to achievement were not included as an objective.
6. Time frame: Data was analyzed from a three-year period consisting of LEAP data from 2016-2017, 2017-2018, and 2018-2019 school years.
7. Variables: The researcher limited the variables of interest to those mentioned in the purpose statement, which were student achievement and curriculum material choice.

Assumptions

Assumptions within this study have been identified and include:

1. Curriculum: This study assumed principals or school leaders had accurate and up-to-date information related to which school districts were or were not using specific curriculum materials during the 2016-2017, 2017-2018, and 2018-2019 LEAP test years.
2. Data: Both the data gathered from the researcher survey and data compiled by the LDOE will be accurate.
3. Generalizability: This study assumed the use of purposive sampling of each school in the state of Louisiana provided a large enough and broad enough data sample for the results of this study to be valid for the state of Louisiana and applicable to other related research, and therefore be generalizable.

4. LDOE validity and reliability related to LEAP data: This study assumed the data posted by the Louisiana Department of Education (LDOE) and gathered by the LEAP assessment is valid and reliable. Support for this is discussed in Chapter Three.

Design and Controls

A causal-comparative research design was used to compare student achievement across school parishes within the state of Louisiana based on curriculum material choice. This quantitative study utilized LEAP achievement data from all school parishes in the state of Louisiana found on the LDOE website. The purpose of this type of design was to discover what curriculum materials are commonly used in schools within each curriculum tier and what correlation and impact exists, if any, between the dependent and independent variables.

Several different limitations, delimitations, and assumptions within this study need to be acknowledged and are addressed below to maintain academic integrity. While the researcher was not able to select what curriculum materials a school district used, the researcher was able to group the curriculum materials into tiers based on LDOE information. The researcher utilized the LDOE curriculum materials tier system which rates curriculum materials as Tier 1, Tier 2, or Tier 3 curriculum depending on how well the curriculum materials align to the Louisiana state standards. Louisiana was selected for this study because of the existence of this unique tier system resource. The LDOE tier systems are discussed in more detail in Chapter Two. Although the LDOE does provide a tier system to assist educators in evaluating how well curriculum materials align to the state standards, there is no public record outlining specific textbook or curriculum material use for each school within the state of Louisiana. Therefore, the researcher sent out a survey to each public school in Louisiana to determine which schools used the varying curriculum materials. As a result of all Louisiana schools having the opportunity to complete the survey, the data should have generalizability since the data reflects the state as a whole despite no random sampling being a limitation of this study. More specifically, the survey was sent out by the researcher to all public schools and their administrative leaders within the state of Louisiana to determine the specific curriculum materials used during the 2016-2017, 2017-2018, and 2018-2019 school years.

Since the time frame of the study is recent, the researcher assumed administrators either had an awareness of what curriculum materials were used during those years or had the ability to access information to complete the survey. It should be noted, the 2019-2020 school year was projected to be a part of this study. However, it was not included because the LEAP was not administered in the 2019-2020 school year due to COVID-19 related issues and schools not participating in the LEAP that year. As a result, the 2016-2017 school year was used as a replacement. Other curriculum and student achievement related concepts such as professional development of teachers, teacher experience, student migration, student socioeconomic status, and the number of years a curriculum material had been implemented were not included in the survey and viewed as limitations or delimitations of this study. While gathering that type of data could be possible and beneficial in some cases, the researcher determined it would be overly burdensome to gather and analyze that additional information along with the fact it would distract from the main objective and framework of the study.

The survey the researcher sent out only asked for curriculum materials for third, fourth, and fifth grade since those are the first three elementary years students in Louisiana take the year-end LEAP assessment. A list of all public schools in Louisiana was gained by using the LDOE website which listed each public school in the state and therefore provided the necessary information to send the surveys. The LEAP data was attained by accessing the LDOE website, which displayed the achievement results for every public school by grade in Louisiana. Researched and utilized data was ex post facto nonexperimental since it was collected after the LEAP assessments had already taken place each year and the schools had already selected their curriculum materials.

Summary

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) to compare math curricula choice to Louisiana Education Assessment Program (LEAP) math achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. Gagne's cumulative learning theory can be broken down into three main components: providing scaffolded instruction of supporting skills or tasks, confirming supporting skills or tasks are mastered, and finally, sequencing supporting tasks in a way to help students best master the desired final task(s) (Gagne, 1962a, 1962b, 1965). This essentially outlined what a curriculum, and curriculum materials, provided (Walsh, 2009). As stated above, textbooks are a driving force behind curriculum and instruction within the classroom (Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Monaghan, 2013; Remillard, 2005; Tyson & Woodward, 1989; Walsh, 2009; Westbury 1990). According to Gagne's theory, a curriculum that scaffolds instruction, builds supporting skills, sequences tasks, and builds to the final task and skills should have the most beneficial results. Therefore, the researcher decided to investigate if there was a difference between student achievement and curriculum materials associated with the LDOE tier system.

Since records on what curriculum materials used by each school are scarce (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018), the state of Louisiana provided a unique opportunity for investigation since the LDOE provided a tiered system outlining numerous textbooks based on each textbook's alignment to the Louisiana state standards. The LDOE tiered system provided two things. First, it provided a starting point to conduct research on curriculum materials by listing several of the curriculum materials utilized by schools, which most states do not have in place (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). Second, the tiered system provided categories to determine how well each curriculum material is aligned to state standards, or what Gagne (Gagne, 1962a, 1962b) would view as the support tasks and final tasks. Using Gagne's cumulative theory as the lens through which to view curriculum materials, the above-mentioned research questions were developed with the aim to investigate what difference existed, if any, between student achievement and curriculum materials from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum materials.

Chapter Two of this study provides a literature review examining topics directly connected to the current state of education within the United States and the state of Louisiana, the importance of math education as a whole, and other achievement related factors such as socioeconomic differences and curriculum choice in education. Topics are arranged thematically, and peer reviewed findings are utilized to guide the researcher and support the theoretical framework of the study. Chapter Three outlines the methods for gathering and analyzing data within this study, the selection and categorization of school districts based on specific curriculum material usage, and the data collection techniques utilized. Chapter Four focuses on communicating the results from this study. Chapter Five consists of a summary review of the entire scope and sequence of this study and outlines the implications of the findings along with recommendations for future studies.

CHAPTER TWO

REVIEW OF RELATED LITERATURE

Introduction

Former University of California, Los Angeles professor James Caballero stressed the importance of math education when he said, “I advise my students to listen carefully the moment they decide to take no more mathematics courses. They might be able to hear the sound of doors closing” (Caballero, 1989). Dr. Caballero was not alone in his thinking as many researchers have stated the importance of mathematical knowledge to optimize opportunity individually and collectively at the societal level (Daro & Asturias, 2019; Denton, 2021; Hiser, 2016; NCTM, 2018; White, 2018). The opportunity to examine whether curriculum materials could help keep doors open for more students within mathematics provided the backdrop for this causal-comparative study, the purpose of which was to use Gagne’s cumulative learning theory (Gagne, 1962a, 1962b, 1965) to determine if there is a relationship between math curricula choice to Louisiana Education Assessment Program (LEAP) math achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. To support that purpose, Chapter Two provides a thematic review of literature related to math curriculum materials.

Chapter Two begins with a brief history of the standardized testing movement, to examine legislation related to both education and math education movements within the United States. The legislation review is followed by discussion associated with textbook and curriculum material choice within education. The rationale for starting with the standardized testing, legislation, and curriculum material topics was to provide context for how curriculum materials connect with the current Common Core State Standards (CCSS), the LEAP assessment, and the standardized testing atmosphere within the United States. Chapter Two ends with analysis of learning theories associated with Gagne's cumulative learning theory and curriculum materials use, which were previously outlined in Chapter One.

Review of Math Standardized Testing and Educational Policy in the United States

The genesis of the contemporary accountability movement within the United States started nearly 50 years ago with the publication of *A Nation at Risk* (Hudak, 2022; Kemler, 2022; Mai-Huntley, 2022; Monaghan, 2013). It was at this point the United States was confronted with the fact other countries were not only equal to the United States education system but surpassing it, and the education system within the United Stated needed adjustment (Horne Wooley, 2022; Mitra, 2018; Pardini, 2022). From that point forward, the creation of state level commissions, along with emphasis on rigorous and career ready education and standardized testing, received increased levels of scrutiny and importance to improve education within the United States (Horne Wooley, 2022; Hudak, 2022). Math education was one of the key areas of concern since math performance was one of the primary factors used to rank schools (Burnette, 2018). However, schools failing to properly educate students in math literacy continued to be a major problem in the United States despite years of effort to improve in this area (Berrett & Carter, 2018).

Assessment over the last hundred years has mostly been rooted in behaviorist ideology as instruction; curriculum materials and classroom practice focused on what skills could be seen and what information could be memorized to show student mastery of content (Hageman, 2020). Essentially, from an accountability standpoint, the focus by education stakeholders was on Gagne's second and third tenets of cumulative learning theory as they attempted to determine if students had accurately learned component skills, tenet two, for the student to reach mastery of the final task, the expected learning for the grade level, which is tenet three (Gagne, 1962a, 1962b, 1965). The combination of the above events led to an increase in school accountability through standardized tests, which were a key element of the No Child Left Behind (NCLB) Act and remain a prevalent focus today (Cage, 2021; Davis, 2022; Walker, 2022).

The United States has both recently and historically gone through several different educational transitions and political alignments (Horne Wooley, 2022; Lein Authement, 2022; Stralek, 2018). The expanded use of standardized testing as a measurement of student learning and accountability starting in the 1980s (Clemons, 2022), the No Child Left Behind (NCLB) Act along with previous legislation such as the G.I. Bill, the Elementary and Secondary Education Act (ESEA) of 1965, and ideas put forth in *A Nation at Risk* have aimed to increase equity in education within the United States (Horne Wooley, 2022). Recently math and science education has received even more attention as the United States attempts to stay internationally competitive in math and science career fields (Cruze, 2022) and in the global marketplace as a whole (Horne Wooley, 2022). The following review of education legislation within the United States focuses primarily on events since the year 2001, prudent to the current educational environment related to standardized testing, learning standards, and math education.

No Child Left Behind ACT and Standardized Testing

Starting in the early 2000s, The No Child Left Behind (NCLB) policy was an attempt by the federal government to increase academic readiness of students within the United States, as a continuation of the Elementary and Secondary Education Act (ESEA) of the 1960s (Cruze, 2022; Pennington, 2022; Perry, 2021). As a result, education reform from the last two decades, starting with NCLB, had focused on promoting equity and equal opportunity within the United States education system (Hanselman, 2018). In terms of Gagne's first tenet of cumulative learning theory, delivery of instruction on component tasks to reach the final task was lacking equity across the country and adjustments to instructional practices were needed. The NCLB Act specifically brought with it a renewed focus on observable and tangible learning outcomes based on standardized test results to measure student progress. This focus on learning outcomes connected to both the second tenet of cumulative learning theory, confirmation of component tasks or knowledge, and the third tenet, students reaching the desired final result through properly sequenced instruction. The direct use, even if not directly stated, of Gagne's cumulative learning theory to guide education policy should not be viewed as a surprise since it has been used to guide curriculum frameworks before (Beatty & Pritchett, 2012; Okey & Gagne, 1970; Spaull & Kotze, 2015).

Standardized testing ultimately dominated the operationalization of NCLB despite the Act's structure, which focused on student achievement along with state flexibility, research-based practices for instruction, and more school choice for families (Cruze, 2022; Horne Wooley, 2022). Horne Wooley (2022) and Walker (2022) stated one of the main goals of NCLB was to reduce the achievement gap that existed within public education which was most visible between minority or special populations and non-minority students. Additionally, NCLB intended to raise achievement for all students through a focus on curriculum connected to state standards and outlining the curriculum teachers should focus on within their classrooms (Horne Wooley, 2022). Thus, the NCBL effort to view student achievement to some extent as a byproduct of curriculum, represented a need to have instruction properly scaffolded to each desired task, tenet one of cumulative learning theory

(Gagne, 1962a, 1962b, 1965). The NCLB era's use of standardized testing to measure what students have actually learned, accountability for schools by requiring schools make Adequate Yearly Progress (AYP), and for schools to develop a plan if they did not meet the desired AYP goals on high stakes tests (Horne Wooley, 2022), reflected the focus on tenets two and three of cumulative learning theory, as previously stated. The widening achievement gap and inequities in education were believed to be a main reason the United States had fallen behind other countries and NCLB aimed to correct that issue utilizing the above methods (Cage, 2021). In fact, NCLB had the stated objective, "all children will reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments" (Regenstein et al., 2018, p. 3) by the year 2014 (Perry, 2021; Williams, 2019).

Since math and English were the focus of the standardized tests within NCLB, other content areas such as science, social studies, and the fine arts suffered because they were not tested areas (Heise, 2017; Cruze, 2022). Generally, the results of NCLB impacted low-income school districts the most with changes seen through the adoption of new curriculum, longer school days and more emphasis on standardized assessments (Troppe et al., 2017). The heavy focus on academic standards and testing within NCLB was a major concern as the policy approached reauthorization, therefore preliminary work on adjustments to improve the legislation began in 2006 (Fertile, 2022). From the perspective of cumulative learning theory, too much emphasis has been placed on measuring the final task, through standardized testing, without enough focus on the actual scaffolding and delivery of instruction to the final task (Cruze, 2022; Fertile, 2022; Horne Wooley, 2022; Pennington, 2022). Pennington (2022) went on to highlight the lack of national or common standards across the United States as a concern and most likely having a negative impact on the educational system overall.

Every Student Succeeds Act and Common Core State Standards

The issues associated with NCLB standardized testing and varying curriculums led to adjustments in the subsequent years. Two of the more recent reforms to math education in the United States that followed NCLB were the implementation of the Common Core State Standards (CCSS) in 2010 and Every Student Succeeds Act (ESSA) of 2015. The ESSA was the replacement to NCLB and further continuation of the ESEA of 1965 (Barthelemy, 2022; Cage, 2021; Koenig, 2022). The ESSA attempted to reconcile the shortcomings of the NCLB Act which consisted of states lowering standards to more easily meet high stakes testing goals, the predisposition of penalizing failure instead of celebrating achievement, evaluating student achievement scores in isolation rather than looking at student progress and finally NCLB's tendency to apply a one-size-fits-all type of correction process for under-achieving schools (Barthelemy, 2022). To avoid the one-size-fits-all approach, the federally mandated AYP goals from NCLB were substituted for state designated goals that simply required states to include student achievement goals for all students and student subgroups (Cage, 2021; Koenig, 2022; Perry, 2021). While the ESSA still utilized high stakes testing, the focus somewhat shifted as the ESSA goals consisted of first helping states align curriculum to better match college and career readiness standards and second, to focus on equity in education (Barthelemy, 2022; Clemons, 2022), which represented a more holistic view of cumulative learning theory principles with a focus on both the process toward the final goal as well as the final goal itself. Chu (2019) summarized the change between NCLB and the shift to ESSA by stating there was higher devotion to "equity and excellence (that can be represented globally) with a focus on closing the achievement and opportunity gaps among students within and between schools and districts, especially students who have been historically underserved in terms of educational achievement" (p. 3).

Clemons (2022) also stated graduation rates and tests scores are the main criteria used to evaluate schools to designate them as either low- or high-performing within ESSA. Part of the ESSA effort was to provide education departments within each state the means to improve school performance and toughen accountability within each state (Kannam & Weiss, 2019). To address accountability required by ESSA, states provided a School Performance Grade (SPG) and categorized low-performing schools into one of three categories: Comprehensive Support and Improvement (CSI), Targeted Support and Improvement (TSI), or Additional Targeted Support and Improvement (ATSI) as of the 2017-2018 school year. This was part of the schools' report card (Clemons, 2022). Additionally, ESSA mandated states provide comprehensive information regarding the performance of all public schools within the state (Clemons, 2022). The continued use of high stakes standardized testing and reporting of student data for accountability purposes within ESSA, which was also used by NCLB, meant the ESSA was essentially a revision and expansion of NCLB rather than a complete fundamental shift in testing and accountability practices (Barthelemy, 2022; Clemons, 2022; Williams, 2019).

The second recent event following NCLB that influenced math education within the United States was the CCSS. The CCSS not only brought about common national standards but also heightened expectations for math education and a renewed focus on skills needed for college and career readiness with practices aligned with high-performing international countries (CCSSI, 2021b; Fertile, 2022; Mangram & Solis Metz, 2018; Pennington, 2022).

Allensworth et al. (2021) stated, “the adoption of the Common Core State Standards in Mathematics (CCSS-M) by most states across the country brought ambitious new targets for what students should learn and be able to demonstrate in school” (p. 1). Additionally, standards across grade levels were built with more vertical harmony, intentionally connecting previous knowledge to current learning from year to year (Pennington, 2022). This correlated to the three cumulative learning theory tenets by examining how to properly scaffold instruction, evaluate component tasks to reach the final task, and examine if the sequence properly led to the desired final outcome. The ambitious task outlined by the CCSS did cause substantial adjustments for what was taught at each grade level in most states (Letwinsky & Cavender, 2018). In all, 41 states along with the District of Columbia and four United States territories implemented the CCSS (CCSSI, 2022b). Even with the political pushback and repeal efforts, most states are still utilizing the CCSS (Lein Authement, 2022; Petrilli, 2017) or a close representation of them even if the state shifted from the official CCSS (Friedberg et al., 2018). This was significant since the LEAP 2025, simply referred to as the LEAP test, was built on the CCSS which was the measurement of student achievement within this study and other recent research within Louisiana (Lein Authement, 2022).

Planning for the CCSS started when state commissioners of education and governors from 48 states, along with other United States districts and territories, gathered in 2009 to develop the CCSS (CCSSI, 2022b). “Core math standards emphasize increased communication about math (purportedly, to demonstrate understanding) while somewhat de-emphasizing the performance of mathematical procedures (such as performing computations, solving equations, or solving word problems)” (Rebarber & McCluskey, 2018, p. 10). Because of this shift, Fertile (2022) detailed how a lack of training or professional development opportunities for teachers related to the shift, and continuation of CCSS instruction, hindered the potential of the CCSS; this could be viewed as a lack of prioritization of tenet one, effectively scaffolding instruction, of cumulative learning theory. While it is true public scrutiny caused several states to shy away from the CCSS over the past few years, most states still use the CCSS through slight modification, or renaming them, leaving the core of the standards in place despite some setbacks (Allensworth et al., 2021; Friedberg et al., 2020; Petrilli, 2017).

Since the concepts within the CCSS had already been shown to be successful by other countries, there had been optimism the CCSS would have a positive impact on the United States education system as well. (CCSSI, 2021a; Pennington, 2022). In fact, the CCSS were envisioned to "serve as an equalizer for high poverty students and the related achievement disparity" (Buzick et al., 2019; Davis, 2019, p. 1). However, despite the success other countries were having, concerns related to student performance on standardized mathematics assessments persisted and were seen through low ACT scores (Gewertz, 2018), a decline in the math ranking of the United States (OECD, 2020), an overall lack of student achievement growth (Denton, 2021; Oglesby-Phelps, 2022; Sparks, 2018) and an expanding achievement gap (Rotermund & Burke, 2021; Sparks, 2018) despite the efforts made through the CCSS. As stated in Chapter One, the combination of stagnant growth and a widening achievement gap in mathematics has educators, administrators, and policy makers searching for new ways to improve student learning and achievement. While several factors can influence student achievement, one of the prevailing avenues school districts have pursued for an increase in student achievement is the use of effective curricula materials (Koedel et al., 2017; Reys et al., 2003; Solomon et al., 2019; Superfine et al., 2010).

Factors Influencing Elementary Students' Math Achievement

The above educational policies and practices have directly impacted elementary math students and shaped the educational experience of those students; however, the results of these policies left room for more improvement. In fact, recent national and international testing revealed progress in mathematical achievement from students within the United States yet showed little or no progress with lower achieving students performing worse, which has widened the achievement gap in recent years (Carunungan, 2022; Oglesby-Phelps, 2022; Rotermund & Burke, 2021; Sparks, 2018). This little to no progress was seen through only incremental growth on the NAEP over the last decade (Denton, 2021; Hussar et al., 2020; NCES, 2019), where approximately 59 percent of fourth grade students performed below NAEP proficiency levels (NAEP, 2019). The NAEP test results were particularly concerning because the design of the NAEP test aimed to specifically represent characteristics from students across the United States and best approximate the ability of all fourth-grade students within the population (NCES, 2020). Internationally on the PISA, the United States ranked of 37 out of 80 countries in 2018, which only slightly improved on the rank of 39 in 2015 (OECD, 2018; Schleicher, 2019). Additionally, the TIMSS results from 2019 were marginally positive and showed fourth-grade United States students ranked 15 out of 64 and 8th grade students ranked 11 out of 46 (Carunungan, 2022; NCES, 2019). Despite those somewhat positive results on the TIMSS and PISA, the United States Department of Education (2021) still expressed concern over the current math literacy level of elementary students.

Educators have had difficulty with teaching mathematics not only because the concepts have been a challenge for some elementary students, but also because it has been hard to always determine where learning gaps exist (Gillespie, 2021), highlighting the importance of proper examination of component tasks that lead to the final task from tenet two of cumulative learning theory. Another gap experienced, but not always addressed by students, was the motivation and engagement gap which has also been connected to student achievement (Ager-Sharrieff, 2022). The lack of success related to properly motivating students during mathematical learning opportunities has continued to be a concern (Lee et al., 2021). Both Goodwin (2018) and Olgesby-Phelps (2022) discussed how important it was for students, especially struggling students on the lower end of the achievement gap, to make a connection between the subject and themselves, which is referred to as math identity. In fact, positive math identity has been shown to impact student learning (Nelson, 2021). A common hindrance to developing a positive math identity experienced by students, and possibly their parents and teachers before them, was the myth that some people were naturally good "math people" and other people were simply not good "math people" (Boaler et al., 2018). Clark (2021) pointed out the development of positive math identity for students included the initial belief they could actually learn and improve in mathematics and develop a growth mindset, which was why the myth that people either are or are not "math people" and similar beliefs about mathematics by students hindered student growth. Furthermore, Carunungan (2022) addressed both student and teacher math identity and noted part of the issue elementary students encountered with mathematics was the elementary teacher, since elementary teachers were often uneasy teaching mathematics and therefore resorted to teaching in ways that were not always best for students. Based on Gagne's cumulative learning theory, using teaching methods that do not allow students to make mathematical connections has the potential to not only hinder scaffolding rigorous course content, (tenet one), but also, impact the efficiency at which students learn and progress through component tasks and formative assessments (tenet two).

Despite these obstacles, Hatcher (2018) encouraged teachers to understand and help students work through mindset issues and reframe negative feelings toward mathematics to empower students to learn better. In fact, Clark (2021) discussed how teachers who helped students develop positive math mindset actually helped them engage with mathematics in a more meaningful way. While the issue might seem complicated, Brodesky et al. (2021) stated part of the solution was as simple as educators showing students they believed the students had the ability to master math content and placed high expectations on each student's mathematical growth. Meanwhile, Carunungan (2022) saw the variance in math mindset by students, and teachers, and called for more research-based, teacher-friendly curriculum materials to support teacher instruction. Teacher-friendly resources could be especially beneficial since teachers who struggle in mathematics tend to heavily rely on the provided curriculum materials (Arican, 2018; Ruggeri, 2021). To help struggling students, Sterner et al. (2019) outlined a more detailed approach summarized by the need for systemic curriculum materials along with scaffolded learning with discussion, the application of problem-solving strategies, and cumulative reviews at the end of tasks. These ideas from Sterner et al. (2019), while not directly stated, clearly connected to the cumulative learning theory concepts of instruction that effectively scaffolded learning and monitored progress as students progressed toward their learning goals in the most effective way possible. The more detailed systems approach outlined by Sterner et al. (2019) was echoed by Polikoff (2021) who viewed curriculum materials as one possibility to improve both teacher instruction, through curriculum material provision of correct learning goals and standards, along with the potential increase in student achievement.

Textbooks and Curriculum Material Choice within Math Education

The assertion by Polikoff (2021) that curriculum materials could be leveraged to both improve teacher instruction and increase student achievement aligned with the purpose of this causal-comparative study. Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) was used by the researcher to determine whether a relationship existed between math curricula choice and Louisiana Education Assessment Program (LEAP) math achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. An examination of Polikoff's assertion, as well as the purpose of this study, required an investigation into several other related topics connected to curriculum materials and student achievement to thoroughly scrutinize the issue. First and foremost, was the association between math curriculum materials and the math curriculum taught within the classroom. If so, to what degree did math curriculum materials influence the actual instructional practices of teachers? If there was little to no association between math curriculum materials and teacher utilization of those materials, then any investigation into curriculum effectiveness might be seen as fundamentally flawed because of curriculum materials' limited association to either what was taught or how it was taught. Additionally, the researcher found it prudent to examine how curriculum materials had adjusted to the CCSS and potential gaps in research related to the CCSS, curriculum materials, and student achievement within mathematics. To empower schools to make informed curriculum materials decisions (through the transition to the CCSS and address the lack of efficacy studies on curriculum materials and student achievement), the state of Louisiana's unique tiered ranking system was also a key element of the literature review within this section given its central role within both the purpose statement and study as a whole.

Curriculum Materials as Curriculum

Marzano (2003) stressed the need for a guaranteed and viable curriculum within schools. But have curriculum materials influenced the curriculum and educational process within schools, and if so, in what way? Despite teachers having their own unique opinions and teaching styles, teachers more often than not addressed curriculum topics based on the prescribed materials provided by the school district for which they work rather than cover topics not within those materials (Ball & Cohen, 1996; Bellens et al., 2020; Chingos & Whitehurts, 2012; Machalow, 2020). Furthermore, teachers tend to follow the sequence of topics within the provided curriculum resources rather than determining their own order (Reyes et al. 2003; Ruggeri, 2021; Walsh, 2009) and do so even more when they are not comfortable with the content (Arican, 2018; Ruggeri 2021). Therefore, curriculum materials have often provided both the scope and sequence for classroom curriculum; the pedagogical approach teachers utilized during instruction has also been shown to be influenced by those same curriculum materials (Machalow, 2020; Remillard & Kim, 2020; Stein et al., 2007; Suppa 2018). When viewed from the perspective of Gagne's cumulative learning theory, the curriculum materials used by teachers within the classroom address all three tenets of the theory as they influence the scaffolded component task instruction of teachers, formative assessment, and learning opportunities for teachers to monitor student progress, as well as the scope and sequence utilized by the teacher intended to lead to mastery of the final task.

Both recent and historical research have shown a strong connection in mathematics education between the written curriculum within the textbook, program choice or curriculum materials, and the enacted curriculum and instruction that takes place within the classroom (Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Koedel et al., 2017; Machalow, 2020; Monaghan, 2013; Remillard, 2005; Solomon et al., 2019; Tyson & Woodward, 1989; Walsh, 2009; Westbury, 1990). Eisner (1987) surmised textbooks were the leading stimulus for teachers to determine what was taught, which was supported by Monaghan (2013) years later and specified math curriculum is significantly dependent on textbooks. Solomon et al. (2019), nearly 30 years after Eisner (1987), confirmed Eisner's analysis, stating curriculum materials still have a prominent place within the math classroom, followed by Machalow (2020) who asserted textbooks are what help transition a neutral standard into an actual classroom learning experience.

This longitudinal consistency presented by Eisner (1987), Monaghan (2013), Solomon et al. (2019), and Machalow (2020) within the math classroom was not unique. Ball and Cohen (1996) and Bellens et al. (2020) argued textbooks were the leading factor for what took place during instruction, while Ball and Cohen (1996) and Walsh (2009) asserted few teachers detour from textbook outlines and information. Arican (2018) and Ruggeri (2021) both stated the use of textbooks or curriculum materials to implement instruction was heightened further when teachers were not comfortable teaching mathematics. Solomon et al. (2019) went on to note approximately 85 percent of elementary classrooms within the United States relied on at least one curriculum material for science and mathematics. Elliott (1990) stated more than any other subject area, math textbooks are more prone to be used as taught curriculum at the elementary level, while research conducted by Tyson and Woodward (1989) showed 75 percent to 95 percent of the total classroom instruction at the elementary level comes from the textbook structure. Blazar et al. (2020) found across six different demographic and geographic areas approximately 94 percent of elementary teachers used district-adopted curriculum materials on more than half of their lessons. This led Monaghan (2013) to proclaim math textbooks are the “de facto” curriculum and the best textbooks should be viewed as the ones that get the best results, which Polikoff, Campbell, et al. (2020) also contended. Given the previously stated relationship between cumulative learning theory and curriculum materials, the theory’s core tenets of scaffolded instruction of smaller tasks or skills that build to the final task, confirmation or formative assessment to aid in mastery of those small tasks, and the sequence of instruction and task acquisition to reach mastery of final tasks would also appear to have a prominent place within elementary math education.

Research has also shown teachers in more literature-based subjects use trade books, authentic resources, primary documents, and other non-textbook related items to shape curriculum, yet mathematical instruction tended to utilize only the provided curriculum materials, such as textbooks (Remillard, 2005). In fact, Remillard (2005) also pointed out teachers who taught both mathematics and more literature-based subjects would enhance the textbook or provided curriculum materials in language arts and reading but not do so when they taught mathematics. A decade before Remillard (2005), Ball and Cohen (1996) concluded any endeavor to introduce change in mathematics teaching needed to be rooted in textbook or curriculum materials change due to the reliance on curriculum materials by teachers when teaching mathematics, echoed again recently by Polikoff (2021). The influence curriculum materials have had on the educational process within mathematics and elementary classrooms has been foundational, both in terms of teacher lesson preparation and instruction (Ball & Cohen, 1996; Land et al., 2019; Polikoff, 2021; Remillard, 2005, 2018; Stein et al., 2007). Concerning math curricula specifically, Polikoff (2015) stated curriculum materials were one of the most important influencers on teacher instruction, which was supported both recently and historically through the work of Ball and Cohen (1996), Bellens et al., (2020), Eisner (1987), Elliott (1990), Machalow (2020), Monaghan (2013), Polikoff (2021), Remillard (2005, 2018), Solomon et al. (2019), Tyson and Woodward (1989), Walsh (2009), and Westbury (1990).

Curriculum Materials and Instruction

While the above section focused on the affiliation between curriculum materials as curriculum, instruction was mentioned due to the interdependent nature of curriculum, curriculum materials, and instruction within the educational process. Machalow (2020) summarized this dynamic suggesting teachers similarly understood and operationalized standards within classrooms principally through textbook information, since textbooks exist in the middle ground between a neutral standard and classroom enactment. However, a closer examination of how curriculum materials specifically transformed instruction was warranted and is included below. The curriculum enactment process from Stein et al. (2007) was a relatively recent expansion of how curriculum materials, teacher instruction, and student learning interact and has been utilized in recent studies (Harvey, 2021; Remillard, 2018b; Remillard & Kim, 2020; Richman, 2021; Suppa, 2018; Tanck, 2021).

The curriculum enactment process essentially outlined in three steps how curriculum materials, teacher planning, and classroom instruction interacted in the teaching and student learning process, which were similar to concepts within the three tenets of Gagne's cumulative learning theory. The textbook or curriculum materials provided the written curriculum which led to an intended learning opportunity created by the curricula author. The teacher then operationalized the written curriculum or intended learning opportunity for their classroom which, after operationalization, became the intended curriculum. While the curriculum materials, the written curriculum, directly interacted with the teacher to develop the intended curriculum, the instruction during the lesson the teacher actually provided the students was the enacted curriculum. What the students experienced within instruction may or may not have had degrees of variance from the written curriculum and intended curriculum (Stein et al., 2007). Tanck (2021) reflected on that process and concluded that while teachers did possess leverage on how curriculum was delivered, both the importance of objectives within curriculum materials and their entanglement with teacher curriculum design cannot be overlooked.

Both the degree to which the student took advantage of the learning opportunity and the teacher's fidelity in how he or she implemented the written curriculum can impact student learning (Suppa, 2018). While recent research (Rahman, 2018; Remillard & Kim, 2020; Suppa, 2018) has focused on this process, the idea is not new. Approximately 25 years ago as Ball and Cohen (1996) studied the importance of textbooks and other curriculum materials, they noted how the enacted curriculum was something the curriculum materials, teachers, and students helped co-construct. Through an examination of curriculum materials, Rahman (2018) inferred curriculum materials within the mathematics classroom had a substantial influence on both what was taught by the teacher and learned by the students, which is supported by the works of Monaghan (2013), Remillard and Kim (2020), Schoenfeld (2014), and Wang and McDougall (2019), who drew similar conclusions.

While Rahman (2018) and Suppa (2018) mostly focused on the interaction between what the teacher planned (intended curriculum), and what actually took place during the teacher's lesson (enacted curriculum), Suppa made two important observations related to curriculum materials as the written curriculum and instruction. First, Suppa (2018) stated most student learning studies focused on teacher effectiveness and the strategies utilized by teachers and overlooked the impact the intended learning opportunities supplied through the written curriculum in the form of curriculum materials.

Second, Suppa (2018) expressed how the significance of curriculum materials were two-fold when viewed through the curriculum enactment process. Traditionally, curriculum materials were seen as a student support and something that provided the students with a learning opportunity. But curriculum materials have been shown to provide the teacher an opportunity to learn since teachers interacted with the written curriculum or curriculum materials as the teacher developed the intended learning opportunities for students. This was not only supported by Suppa (2018) but also by Monaghan (2013), Prigodich (2021), Rahman (2018), and historically by Ball and Cohen (1996). This led Suppa (2018) to surmise one of the ways to partially provide more equity in math education was to study how written curriculum could support improvements in how mathematics was taught. Suppa (2018) in this regard was not alone, as other researchers from the past few decades came to the shared opinion the learning opportunities students obtained during instruction related to mathematical achievement (Monaghan, 2013; Remillard & Kim, 2020; Stein & Kim, 2009). These learning opportunities, based on cumulative learning theory, represent the importance of delivery of instruction scaffolded to the desired task and examining the sequence of instruction to optimize transference.

Curriculum Materials and Alignment to State Standards

Due to the influence curriculum materials have on curriculum and instruction, Steiner (2017) stated curriculum choice can noticeably influence student learning, but the lack of research in this area hinders quality instruction, purchasing, and policy choices related to curriculum materials. In fact, curriculum material alignment to the CCSS has been one key area of concern (Polikoff, 2021). If viewed through the lens of cumulative learning theory, the proper alignment of standards, curriculum materials, and instruction would represent the importance of instruction correctly scaffolded to the desired task, along with investigating whether the overall sequence of instruction was optimized for transference. Polikoff (2015) investigated fourth grade mathematics textbook alignment and noticed the degree to which textbooks advertised alignment to the new CCSS and how well they are actually aligned were not in sync. This is deeply concerning since "the ways that textbooks present standards may stand in for the standards themselves from teachers' and students' perspectives. This gives curriculum developers great power in determining or guiding how standards are translated into action in classrooms" (Machalow, 2020, p. 21). While many problems existed, issues related to how curriculum material aligned to the CCSS were echoed by Polikoff (2015, 2021) and Polikoff, Petrilli, et al. (2020) who stated textbooks overemphasized memorization, procedures, and other content not connected with the standards and further research in this area was needed. Other researchers noticed similar issues and claimed efforts to raise teacher effectiveness within mathematics will consistently fail to improve achievement if the curriculum materials they used were not high-quality and aligned to the CCSS (Chingos & Whitehurts, 2012; Polikoff, Petrilli, et al. 2020). According to Machalow (2020), the earliest fully aligned curriculum to the CCSS was in 2013, despite the introduction of the CCSS in 2010, with two of the other largest publishers within the United States not introducing a fully aligned curriculum to the CCSS until the year 2020.

Issues Tracking Curriculum Material Usage and Efficacy

While statistics have shown substantial curriculum material usage during the current era of CCSS (Blazar at al., 2020; Koedel et al., 2017; Solomon et al., 2019), research related to what curriculum has been utilized by schools and the influence of curriculum materials on student learning has been hindered because nearly all the states do not track or record what curriculum materials have been used (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). Of the few states that have gathered and reported data on curriculum materials, the data has suffered from misspellings or a general lack of organization and made the data unviable or immensely difficult to use (Kane, 2016). Even in California, which required schools to report what curriculum materials were used, there appeared to be a lack of accountability as the data appeared to have a significant number of schools that did not report curriculum materials or lacked the necessary detail to distinguish between similar editions of related curriculum materials (Koedel et al., 2017). A proper analysis of curricula material choice has required an initial investigation into what curriculum materials were used (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018), which has been burdensome, and a key part of the educational structure that needs to be scrutinized. Building on Koedel's already mentioned frustration (Koedel et al, 2017) and Steiner's concern related to the lack of knowledge associated with curriculum materials (Steiner, 2017), Blazar et al. (2020) investigated this issue in a collaborative effort that involved ten co-authors. Since 2010, Blazar et al. (2020) found only one multi-textbook randomized trial, two randomized trials assessing the effectiveness of a single textbook, and a few non-experimental studies that relied on matching techniques to estimate textbook effects had been completed to examine curriculum material effectiveness.

Following that summary of available research, or lack thereof, Blazar et al. (2020) went on to state most curriculum materials used currently have never gone through a rigorous test of efficacy.

In addition to the above issues, deficient research in this area might be the result of a combination of lack of commitment or lack of interest in this area of research. Remillard (2005) described how an ample amount of research had been dedicated to the research and theory of teaching, yet little data is available on curriculum resources or guides. This view is supported by Suppa (2018) who stated most student learning studies focus on teacher effectiveness and the strategies utilized by teachers, while often overlooking the impact the intended learning opportunities supplied through curriculum materials. Whether there was one dominant reason or several reasons, experts in the field agree there is a gap in research related to the relationship between the choice in curriculum material and its impact on student achievement (Blazar et al., 2020; Bhatt & Koedel, 2012; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Kane, 2016; Koedel et al., 2017; Polikoff, 2018; Polikoff, Campbell, et al., 2020; Steiner, 2017).

Curriculum Materials and Student Achievement

While research related to curriculum materials and student achievement was limited, as shown above, there was still a body of research related to the topic available for discussion even if it was scarce. To highlight the importance of proper curriculum selection, Chingos and Whitehurts (2012) discovered through their research the choice of curriculum materials had a comparable effect on student achievement to teacher effectiveness. Given the previously mentioned connection curriculum materials had to cumulative learning theory as a tool that helped guide the scaffolded delivery of instruction, provide formative assessment and learning activities for progress monitoring, and helped form the scope and sequence of learning within classrooms, the importance of proper curriculum materials selection was also examined by the researcher. Chingos and Whitehurts (2012) noted improving teacher effectiveness was complicated and expensive, however another avenue to improve student achievement would be through the selection of better curriculum materials. Chingos and Whitehurts (2012) argued selecting better materials would be easier, less expensive, and quicker to implement. The utilization of curricula to increase student achievement was not a new concept and has been used by stakeholders in the educational community for many years (Polikoff, 2021; Reys et al. 2003; Solomon et al. 2019; Superfine et al. 2010). Curriculum material efficacy research was important not only because of the potential to enhance student learning but also because of its potential to increase student learning with little to no added expense to school budgets since curriculum materials are already part of school budgets (Koedel et al., 2017; Solomon et al., 2019).

Within this study, the researcher specifically aligned the research questions for grades three, four, and five for three reasons. First, the grade range aligned this study with previous investigations that were similar (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Kendrick et al., 2020; Koedel et al., 2017; Nargi, 2018). Second, the elementary grades tend to be the focus of improving math achievement since difficulty in the elementary years cannot only be detected as early as kindergarten but are also used to predict later success or difficulty within mathematics (Metzger, 2018; Solomon et al., 2019). Third, the LDOE did not start LEAP testing until the third grade (LDOE, 2018), which would make research of other grade levels related to student achievement problematic because they would lack alignment to grades three, four, and five within this study.

As mentioned before, both Steiner (2017) and Chingos and Whitehurts (2012) stated curriculum choice can noticeably influence student learning, with Ruggeri (2021) adding curriculum materials greatly influenced a student's opportunity to learn and, as a result, different curriculum material created different opportunities. This would appear logical given Ruk's (2021) review of teaching practices, which found textbooks were the most recurrently utilized resource within math classrooms by teachers to determine student tasks. When viewed through the lens of Gagne's cumulative learning theory, this would correlate to different curriculum materials potentially providing different scaffolding opportunities, different progress monitoring opportunities, and different sequences of tasks toward the desired learning goals that might naturally result in different student achievement based on different curriculum materials. Additional research in this area, both recently and historically, and through a variety of research designs showed curriculum materials had a significant educational effect on student achievement or the potential to increase student achievement (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Cress, 2019; Koedel et al., 2017; Lein Authement, 2022; Polikoff, Petrilli, et al., 2020; Walsh, 2009; White, 2018). Koedel et al. (2017), with later affirmation from Polikoff (2021), stated even minor increases in student achievement related to curriculum materials could be viewed as significant because implementation would be as simple as picking one curriculum material instead of another, requiring little cost and effort, and even more importantly the curriculum material choice would have a school-wide effect and not be constrained to a specific population within the school. Polikoff (2021) also described curriculum improvement as a high leverage improvement since it had the ability to scale and reach the entire population, and went on to recommend individual state agencies increase their role and help evaluate materials to ensure schools are more knowledgeable and empowered to select quality materials, much like the LDOE in Louisiana.

Despite the above research which detailed the significance of curriculum materials, it was prudent to discuss related research that provided contradictory findings. Even as Monaghan (2013) and Blazar et al. (2020) outlined the prominent place curriculum materials have within the math classroom, their research showed mixed results and concluded there was no consistent difference between a student's achievement who used one type of curriculum material compared to another. Similar conclusions were made by other researchers who investigated differences in curriculum materials and student achievement ranging from single school comparisons (Kendrick et al., 2020; Womersley, 2020) to those who utilized larger samples (Blazar et al., 2020; Nargi, 2018). Furthermore, both Monaghan (2013) and Nargi (2018) expressed concern that a direct comparison between curriculum material choice and student achievement is too simplistic when evaluating curriculum materials, given the complex nature of student learning and classroom instruction. However, Nargi (2018) went on to say those concerns should not limit further research in this area. Instead, Nargi (2018) stated more causal and correlational studies were needed to provide more clarity.

When discussing the mixed results, both Blazar et al. (2020) and Koedel et al. (2017) acknowledged the confines in this area of research due to the limited number of states that gather curriculum material information along with the limited number of research studies in this area which most likely contribute to the mixed results. The limited research related to what curriculum materials schools utilize presents a major roadblock when it comes to evaluating curriculum material use (Blazar et al., 2020; Cummins-Colburn, 2007; Hutt &

Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). Often administrators have been unable to make better decisions related to instructional materials due to the lack of evidence on the impact of the materials used (Blazar et al., 2020; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Koedel et al., 2017; Polikoff, Campbell, et al., 2020; Ruggeri, 2021). Studies in this area were a relatively recent focus and the lack of research meant educators and policy makers have only started to recognize the degree to which math curriculum materials did or did not influence math learning (Solomon et al., 2019).

Louisiana Department of Education

As the United States adjusted to NCLB, CCSS, and the ESSA movements, the Louisiana Department of Education (LDOE) made changes at the state level that coincided with the changes that had occurred at the national level (Walker, 2022). In 2014 the LDOE developed district report cards, adjusted teacher certification procedures, and aligned assessments on standardized tests for elementary and high school (Walker, 2022). Kaufman et al. (2018) stated the focus of the LDOE was on "identifying, recommending, and even developing resources and tools intended to make it easier for districts and teachers to teach in ways that are aligned with state standards and assessments" (p. 50). Considering the potential importance of curriculum materials, as detailed above, the LDOE attempted to leverage curriculum material influence within the classroom to help improve student achievement (Kaufman et al., 2018). Through the lens of cumulative learning theory, the LDOE attempted to leverage an increase in the quality of scaffolded instruction, progress monitoring, and the overall sequence of learning opportunities to optimize transference. A byproduct of that thinking led the LDOE to create the tiered list of resources that can be accessed by teachers, school parishes, and other interested parties so they could more easily teach in ways aligned to the state standards (LDOE, 2022). This freedom of curricula choice and LDOE guidance, was brought about by a change in Louisiana textbook adoption laws in 2015 (Kaufman et al., 2018).

The review process for curriculum materials utilized a rubric from an educational nonprofit, which helped the LDOE determine how aligned to the Louisiana State Standards (LSS) each curriculum material was, which after review allowed it to be placed in the corresponding tier (LDOE, 2022). The tier rankings consisted of three tiers: Tier 1, exemplified quality, meant all non-negotiable criteria was covered and the curricula scored the best on indicators of superior quality. Tier 2, approached quality, meant the curricula met all non-negotiable criteria and met some indicators of superior quality. Tier 3, did not represent quality, referred to curriculum that had not met the non-negotiable criteria (LDOE, 2022). For the purposes of this study, the researcher included another tier labeled "non-tiered curriculum" to potentially include curriculum used by schools not contained within one of the three existing LDOE tiers.

Prior to the above listed tiers, the LDOE simply provided a list of acceptable curriculum materials without guidance or bias. Once the LDOE shifted to the new tier system, they discarded the previous list with the hope schools would utilize the new tier system, which was designed to aid school parish selection of quality curriculum materials (Kaufman et al., 2018). However, it should be noted the LDOE did not promote one curriculum material over another and when the LDOE referred to the tiered curricula list the LDOE stated, "Each local school system should determine if their use is appropriate to meet the educational needs of their students" (LDOE, 2022, para. 1.). The process used to review which curricula belongs in each tier did not provide high marks easily, with only four curriculum materials reaching Tier 1, while fourteen were evaluated as Tier 3. (Kaufman et al., 2018). Curriculum material providers, often publishing companies, were provided feedback from the LDOE related to tier ranking with suggested area of improvement if the curricula provider desired a better evaluation (LDOE, 2022). While school parishes were not required to select specific curriculum materials, it should be noted they were incentivized in Louisiana to select Tier 1 curriculum. This was done through an offer of free professional development that corresponded to Tier 1 curriculum material use provided by the LDOE (Kaufman et al., 2018). The efforts of the LDOE to provide support for curriculum material choice to raise student achievement was not only supported by the educational practices previously mentioned within this chapter, but were also based on sound educational theory, including Gagne's cumulative learning theory, which is discussed below.

Learning Theories and Curriculum Materials

To complement the above section related to both historic and current practical research related to curriculum materials, the researcher provides the below theoretical support for curriculum material use and the association of cumulative learning theory with curriculum materials. This was done to detail both the practice, seen above, and theory of curriculum materials, seen below. For centuries, prominent philosophers and theorists debated and have continued to debate the ideas behind knowledge and learning with names such as Bruner, Dewey, Gagne, Pavlov, Piaget, Skinner, Watson, and Vygotsky, all expanding on these ideas within the last hundred years. Despite the expansive research and theory related to learning and knowledge, there was not a singular learning theory or theoretical framework universally associated with the use of curriculum materials to aid in the acquisition and measurement of knowledge. Rather, there were a multitude of different theories that supported their use and established their place in the learning process, which are outlined below. As mentioned in the previous section, Marzano (2003) explained that a guaranteed and viable curriculum was a key factor shown to influence student achievement. The importance of curriculum prompted the researcher to examine curriculum materials and then apply Gagne's cumulative learning theory as the theoretical framework to examine curriculum material choice within this study.

As previously stated in Chapter One, the three tenets of cumulative learning theory essentially summarized what textbooks, curriculum materials, and curriculum overall were designed to do as they helped the teacher scaffold learning through instruction, monitored progress through learning activities, and provided sequenced learning with clear assessment over the established learning goals. While the cumulative learning theory from Gagne (Gagne, 1962a, 1962b, 1965) served as the theoretical foundation of this study, additional details related to this theory and Gagne's thinking provided more context and support for cumulative learning theory. From this, the researcher found it important to review the origins of stimulus-response research and behaviorism to understand how those related to curriculum materials and cumulative learning theory since Gagne's thinking was rooted in behaviorism (Galoyan, 2019; Hunsaker, 2019; Nimtz, 2018) which viewed the introduction of new information as the stimuli (Zhou, 2020). Additional analysis of how Gagne and other researchers interpreted, operationalized, and expanded on these concepts was also included within this section.

As Dellner (1979) pointed out, Gagne's background originated in military practice where external instruction was the stimulus for learning and problem solving. Thus, an understanding of stimulus-response theory aided in the understanding of Gagne's theories, even if Gagne eventually transitioned past basic stimulus-response thinking (Dellner, 1979; Gagne, 1985; Heaster-Ekholm, 2020; Hunsaker, 2019). Over the last hundred and fifty years, some of the earliest theories of learning originated from non-education settings based on stimulus-response behaviorism research (Hageman, 2020; Lofland, 2021; Russell, 2020; Uibelhoer, 2020). Psychological research related to stimulus-response interactions was most notably associated with Pavlov, Watson and Skinner, but it was Watson's social efficiency theory that helped pioneer the term "behaviorism" (Albert, 2020; Hageman, 2020; LaFave, 2020). It was specifically Pavlov and Watson that laid the behaviorism foundation through their work with animals in stimulus-response environments as they worked on what would later be termed conditional reflex and classical conditioning before behavioristic ideas later transitioned to educational settings (Bates, 2019; McConnell et al., 2020; Russell, 2020; Schunk 2018).

Pavlov, Watson, and Stimulus-Response Behaviorism

Pavlov discovered and later experimented with classical conditioning through the routine task of feeding his dogs (Fernandez, 2021; Gyke, 2020). Through informal observation, Pavlov noticed his dogs would salivate after he brought food into the room. His dogs eventually made a connection between Pavlov and the food, which caused the dogs to salivate simply by Pavlov entering the room, even when there was no food present (Fernandez, 2021; Gyke, 2020). This led to one of the best-known psychological experiments when Pavlov introduced the sound of a ringing noise, specifically a metronome, with food when feeding his dogs. The ringing noise acted as a stimulus, which led the dogs to later demonstrate a learned response, salivating, simply by the ringing noise without the presence of food (Clark, 2018; Fernandez, 2021; Gyke, 2020). Much like Pavlov's dogs, research has shown humans respond to stimuli within their environment (Brock, 2021; Gleckler, 2021). Fernandez (2021) affirmed this and stated the key essentials of behaviorism consisted of three things: the stimulus, the response to the stimulus, and how the relationship amongst the two is maintained.

Watson demonstrated the relationships between learning, stimuli-response, and relationship to stimuli through experiments on infants, including his famously controversial experiment with "Little Albert" (Lafave, 2020; Russell, 2020). To best understand Watson's experiment with Little Albert, it was important to recognize Watson believed prearranged stimuli calculated to incite the desired response or behavior would produce the best results (LaFave, 2020). Within the experiment Little Albert was exposed to white rats, which piqued Little Albert's interest, evident by Little Albert's desire to interact with these animals. However, Watson later introduced loud sounds and used a hammer, metal rods, and similar objects as a stimulus to the environment when Watson brought a white rat within Little Albert's view. Over time, it was observed Little Albert learned to not only fear white rats because of the loud noises but also developed a fear of anything white and furry (Albert, 2020; Bates, 2019; Lafave, 2020; Russell, 2020). LaFave (2020) summarized the application of Watson's research to learning, stating both behavior and thought, from Watson's perspective, are learned through a response to stimulus.

Instructors applied concepts of behaviorism to educational practice by focused use of skills that could be observable, performance measured, demonstrated, and tangible within curriculum objectives (McConnell et al., 2020). The approach of behaviorism, as outlined above, focused on concrete observation or measurement of learned skills that were produced as a result of environmental stimuli. The influence of Gagne's behaviorist thought process

(Galoyan, 2019; Hunsaker, 2019; Nimtz, 2018) is seen through the second tenet of Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965), in that the second tenet focused on the importance of the instructor confirming that each component task or stimulus (Zhoa, 2020), was mastered as knowledge builds. Gagne essentially advocated for what Conrad (2021) stated was a criterion-referenced measurement, which used formative assessment to monitor what students learned based on a set of predetermined educational objectives or tasks.

The application of these basic concepts of behaviorism along with other concepts to instruction and learning helped develop what Gagne called the field of instructional design (Reisner, 2001; Uibelhoer, 2020). McConnell et al. (2020) mentioned observable behaviors associated with learning could be seen once the teacher provided the environment with the proper stimuli. The influence of behaviorist learning theories as they related to measurable and tangible learning outcomes not only influenced instructional design within the classroom, as mentioned by McConnell et al. (2020), but laid the foundation for criterion-referenced assessments as a whole (Uibelhoer, 2020). This was a significant connection to this study, and other studies that utilized state standardized tests to evaluate student achievement since most state standardized tests utilized criterion-referenced tests, including the LEAP (Williams, 2019). While complex learning or problem solving related to stimulus-response environmental factors goes beyond the initial work for Pavlov and Watson, Skinner's work on the process of operant conditioning and both Skinner and Gagne's work with chaining helped show how behaviorism principles could be used to aid in problem solving and learning at a more complex level (Lewis-Davis, 2022; Morales, 2021). Hunsaker (2019) stated Gange's ability to translate learning theory into meaningful instructional practices as a major contribution to the field, with Stalbert (2022) recognizing Gagne as the catalyst for contemporary instructional design.

Skinner, Gagne, and Behaviorism in Education

Following the work of Pavlov and Watson, Skinner expanded on the thinking behind behaviorism through his work on operant conditioning and chaining (Lofland, 2021; Morales, 2021). Skinner built what would later be referred to as a "Skinner Box" to study, or observe in terms of behaviorism, the problem-solving abilities of rats to push a lever. Skinner used food as the environmental stimuli and reinforcement to teach the rats to use the lever (Hudson, 2021; Lofland, 2021; Morales, 2021). Ewen (2010) reflected on Skinner's work and noted operant conditioning showed how the method and concentration of reinforcement strongly impacted both behavioral change and learning. Since Skinner viewed thinking as behavior, he concluded behavior change, thinking and problem solving within an educational context, was impacted by the environmental design (Ewen, 2010). Hudson (2021) went further, and surmised behaviorism essentially viewed all behavior, activities, actions, and learning as a response to environmental stimuli and the associated reinforcement.

When applied to an educational context, correct responses by students would reveal successful conditioning (Russell, 2020). The connection to the first tenet of Gagne's cumulative learning theory complimented Skinner's work as Gagne stressed the importance of scaffolding instruction and learning to meet the goals of the final task (Gagne 1962a, 1962b, 1965). Jad-Moussa (2022) noted Skinner supported this concept and outlined the importance of presentation, assessment, and feedback as part of the instructional design process with presentation equating to stimulus and feedback from the assessment as the reinforcement. Additionally, the added reward of acknowledgement, favorable grades, and good status among peer groups or similar educational practices served as the reinforcement which complements instruction as an educational stimulus that leads to learning (Ewen, 2010; Hudson, 2021; Russell, 2020; Lofland, 2021; Uibelhoer, 2020).

McConnell et al. (2020) reinforced that idea and maintained teachers could alter the environment in a variety of ways but specifically mentioned the importance of creating and operationalizing a lesson plan in order to provide the stimuli or environment for learning. This was significant for the context of this study since curriculum materials for teachers have been publicized to commonly be a large part of the classroom instructional design process and shown to be an even larger part of the instructional design process for elementary teachers during mathematics lessons (Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Monaghan, 2013; Remillard, 2005; Tyson & Woodward, 1989; Walsh, 2009; Westbury 1990). Understanding the importance of instruction as an educational stimulus and precursor to learning, Brock (2021) summarized the role of behaviorism within education stating environment impacted learning because it was the teaching that proceeded learning. Additionally, Brock (2021) stated learning within an academic environment transitioned from initial stimuli to response or behavior development, which was learning. To aid in the learning process within the behaviorism model, new behavior or learning was best followed by reinforcement, which was consistent with other recent summaries of the behaviorism model by Albert (2020), McConnell et al. (2020), Fernandez (2021), LaFave (2020), and Lofland, (2021).

As shown above, textbooks and other curriculum materials can be viewed as an educational stimulus. Curriculum materials provided learning objectives to scaffold instruction illustrated by the practice problems and learning activities utilized by teachers to monitor completion of each component task. Additionally, curriculum materials provided both sequenced instruction and assessment through practice problems, quizzes, tests, or activities to ensure transfer. This idea was supported by Gagne (1974) who surmised the set of planned stimuli provided through instruction influences the process of learning and enhances transference. Gagne often referred to learned skills, behaviors, and observable knowledge of students as "transfer," instead of "learning" (Gagne, 1962b, 1974). As teachers produced the environmental and intellectual conditions that satisfied the needs of students, they helped students internalize control of their learning, which aided in transference (Gagne & Deci, 2005; Vareberg, 2021). The genesis of Gagne's environmental stimulus-to-instruction and instruction-to-transfer viewpoint, can be observed from Gagne's early publications and subsequent writings related to cumulative learning theory consisting of *Military Training and Principles of Learning*, *The Acquisition of Knowledge*, and *The Conditions of Learning*, and other related works.

Cumulative Learning Theory and Gagne

As stated in Chapter One, cumulative learning theory has been used as a framework for developing curriculum in schools over the past several decades to help with transference (Beatty & Pritchett, 2012; Okey & Gagne, 1970; Spaull & Kotze, 2015). Gagne's cumulative learning theory was of particular interest to this study because it has been used within education-based studies before (Chen, 2021; Udensi, 2019), and has been applied to the examination of mathematical learning (McBride, 1996; Walsh, 2009). Furthermore, cumulative learning theory has also been used to specifically compare the impact of math curriculum selection on student achievement by Walsh (2009) in a similar study.

In fact, Gagne utilized fractions and arithmetic sequences (Gagne, 1962a) as well as volume (Gagne, 1968) to help illustrate the connection between cumulative learning theory and both mathematics instruction and curriculum materials. While Gagne was often connected with behaviorism, where the process of learning takes place through systematic and stimulus driven constructs (Dellner, 1979; Galoyan, 2019; Hunsaker, 2019; Nimtz, 2018; White 2018), there have been others who viewed his learning theories as more than basic behavioristic constructs (Dellner, 1979; Hoffman, 2021; Kubota, 1991; Mayers, 2021; Simmons, 2021; White, 2018; Zama & Endeley, 2021; Zhou, 2020). Gagne described instruction as a "set of events external to the learner which are designed to support the internal processes of learning" (Gagne & Briggs, 1979, p. 155). Gagne and Briggs (1979) showed how Gagne focused not only on what would be considered the behaviorist observable events or stimuli external to the learner but also on the mental processing of the learner as he evaluated the learning process. Balancing external stimuli and internal processing within learning led some to refer to Gagne as an eclectic-behaviorist, cognitive-behaviorist, or even a cognitivist (Dellner, 1979; Hoffman, 2021; Lewis-Davis, 2022; Maggio 2021; Mayers, 2021; Simmons, 2021; White, 2018; Zama & Endeley, 2021; Zhou, 2020).

Previously, the researcher stated Gagne's cumulative theory could be summarized by three basic tenets connected to the cumulative learning process Gagne (1962b) outlined from the publication, *Military Training and the Principles of Learning* in 1962. During that same year, Gagne published *The Acquisition of Knowledge* and stated,

> In productive learning, we are dealing with two major categories of variables. The first of these is knowledge, that is, the capabilities the individual possesses at any

given state in the learning; while the second is instructions, the content of the communications presented within the frames of a learning program. (Gagne 1962a, p. 355-356)

As Gagne (1962a) discussed within *The Acquisition of Knowledge*, it was important to notice the focus was on both the internal, "capabilities the individual possesses at any given state in the learning" and the external, "communications presented within the frames of a learning program," with the second essentially referencing instruction or teaching. Gagne (1962a) in more basic terms later stated positive transfer was the result of relevant recall of existing knowledge and the effect of instruction to increase knowledge. Aldamen et al. (2018) summarized cumulative learning as, "the gradual increase in knowledge over the period in which the learning process takes place" (p. 21). Cumulative learning, also referred to as sequence learning, has been so prevalent the majority of curriculum development and presentation of curriculum utilized this concept (Aldamen et al., 2018; Anastasiou, 2022, Nimitz, 2018), of which Gagne has been the foremost advocate (McBride, 1996). Even as Lewis-Davis (2022) reviewed Gagne's work through a cognitivist lens with theorists like Vygotsky and Dienes, it was noted Gagne and others showed how the process of learning was cumulative since learning was built on previous knowledge.

Utilizing the backdrop of Gagne's previous work (Gagne, 1962a, 1962b) to understand learning is cumulative, (Aldamen et al., 2018; Anastasiou, 2022; Dellner, 1979; Gagne, 1962a, 1962b) based on a learner's previous knowledge and the environmental stimulus of instruction, it was possible to understand how Gagne operationalized cumulative learning through his later ideas related to instruction. Gagne, in *The Conditions of Learning*, articulated the five domains of learning and nine events of instruction. It was within this initial work (1965) and future expansion (1985) that Gagne emphasized human learning is cumulative and can be enhanced under certain conditions, which provided the basis for the domains of learning and events of instruction. The five domains consisted of verbal information, intellectual skills, psychomotor skills, attitudes, and cognitive strategies (Gagne 1965, 1985). However, it was the domain of intellectual skills Gagne focused on and believed was the area central to the primary content of school learning (Cummins-Colburn, 2007; Fernandez, 2021; Oberhausen, 1975).

To aid in intellectual skill development, Gagne developed the nine events of instruction: "gain attention, identify objective, recall prior learning, present stimulus, guided learning, elicit performance, provide feedback, assess performance, and enhance retention/transfer" (Gagne, 1985, p. 243). Even though Gagne's domains and events of instruction concepts were over 40 years old, they were still utilized within recent research and teaching (Anastasiou, 2022; Polczynski, 2021). These events of instruction were viewed by Gagne as the environmental stimulus to aid in the cognitive learning process (Kubota, 1991, Zhou, 2020). However, it is important to note these steps were flexible and not concrete as Good and Brophy (1977) articulated when they stated,

> Gagne's nine steps are general considerations to be considered when designing instruction. Although some steps might need to be rearranged (or might be unnecessary) for certain types of lessons, the general set of considerations provide a good checklist of key design steps. (p. 200)

Kubota (1991), like Good and Brophy (1977), discussed the adaptable nature of Gagne's cumulative learning theory as one that was not dominated by either the craft and science of teaching or solely focused on the environmental nurturing of students, but rather a systems approach focused on the importance of sequenced events within the learning process to maximize transfer. Thus, Good and Brophy (1977) and Kubota (1991) underscored flexibility in the operationalization of cumulative learning theory without explicit behavioristic, constructivist, cognitivist, or similar learning theory restrictions based on the type of lesson the instructor desired, since continuity within the sequence of instruction was the foundation of cumulative learning theory (Aldamen et al., 2018; Anastasiou, 2022; Dellner, 1979; Gagne, 1962a, 1968). Gagne (1968) supported this contention as he described cumulative learning as the acquisition of skills, typically oriented from basic to more progressively complex, which interacts and builds to increase learning and is cumulative in impact. However, Gagne (1968) finished by stating learning capacity was built gradually and the "magic" is not found in structure, but merely in learning, retention, and transfer.

Additional theoretical support that stressed the importance of curriculum materials following Gagne's work was Charlie Reigeluth's Elaboration Theory (ET), which was rooted in Gagne's concepts (Hunsaker, 2019) and suggested differing design or sequence of learning can be utilized to produce the best learning outcomes (English & Reigeluth, 1996). Design sequence, or scaffolding, was typically but not always maximized with simple to complex sequencing of concepts (English & Reigeluth, 1996). While numerous factors affected student achievement within and outside of the school's control, the sequence of what was taught and in what order concepts were taught was often within a school's control and was seen through a school district's curriculum choice. Bruner (1966) supported this idea and stated the order of instruction impacted students' capacity to understand, master, and ultimately demonstrate what they have learned. This may especially be true in mathematics instruction in which Bruner's concept of spiral curriculum is often seen. The National

Council of Teachers of Mathematics (NCTM, 2014) asserted the critical issue for processing and retaining information within learning theory was the order of instruction. Research by both Bruner (1966) and Reigeluth (1996) supported Gagne's third tenet within cumulative learning theory, which stated the importance of sequencing instruction for optimum transference (Gagne, 1962a, 1962b, 1965). Bruner, Reigeluth, and many others who followed Gagne's early publications (Gagne, 1962a, 1962b, 1965) helped reinforce cumulative learning theory and either directly or indirectly featured the importance of curriculum materials within the learning process since textbooks and other curriculum materials have been shown to greatly influence curriculum, and therefore the sequencing of instruction and learning (Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Monaghan, 2013; Remillard, 2005; Tyson & Woodward, 1989; Walsh, 2009; Westbury 1990).

Summary

The literature review within Chapter Two focuses on the use of curriculum materials and how education legislation, school and classroom practices, and educational theory have influenced and supported curriculum material use. While most states were using the CCSS in some capacity (Lein Authement, 2022; Friedberg et al., 2020; Petrilli, 2017), those standards and other education legislation alone are not enough to improve student achievement and positively impact the achievement gap (Gewertz, 2018; OECD, 2020; Sparks, 2018). As stated in Chapter One, math achievement must be addressed for the United States to have a capable thriving workforce in the future (Cruze, 2022; Hiser, 2016; Horne Wooley, 2022; White, 2018). Improvement will not come easy but is possible. As Chapter Two illustrates, part of the solution to overcome low student achievement and the achievement gap could reside with effective curricula (Chingos & Whitehurts, 2012; Koedel et al., 2017; Polikoff, Petrilli, et al., 2020; Reys et al., 2003; Solomon et al., 2019; Superfine et al., 2010). However, more research is needed in determining whether curriculum materials did in fact impact student achievement. (Blazar et al., 2020; Bhatt & Koedel, 2012; Chingos & Whitehurts, 2012; Hiser, 2016; Kane, 2016; Koedel et al., 2017; Polikoff, Campbell et al., 2020; Nargi, 2018; Ruggeri, 2021; Solomon et al., 2019; Steiner, 2017).

The next segment, Chapter Three, outlines the methods used for analyzing data within this study, the selection and categorization of schools based on curriculum material usage, and the data collection techniques utilized. Chapter Four focuses on communicating the results from this study. Chapter Five consists of a summary reviewing the entire scope and sequence of this study while outlining implications of the findings. The study concludes with recommendations for future studies.

CHAPTER THREE

RESEARCH DESIGN AND METHODOLOGY

Introduction

Chapter Three outlines what the researcher attempted to discover, who was involved in the research, and what research methods were used by the researcher to gather data. The researcher attempted to discover if curriculum material choice made a difference in student achievement within public school parishes in the state of Louisiana. The researcher utilized public school data from third, fourth, and fifth grade for the purpose of this study. To analyze the data, the researcher utilized a one-way analysis of variance (ANOVA) for $H_0 1$, $H_0 2$ and $H_0 3$, and the corresponding sub hypotheses to determine if a difference existed.

Recent studies from a few different researchers suggested different curricula materials have varying results on student learning (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Cress, 2019; Koedel et al., 2017; Lein Authement, 2022; Polikoff, Petrilli, et al., 2020; Walsh, 2009; White, 2018). However, administrators are frequently lacking the needed statistics to make determinations related to instructional materials due to a shortage of evidence on the differences between student achievement and the curriculum materials being used (Blazar et al., 2020; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Koedel et al., 2017; Polikoff, Campbell, et al., 2020; Ruggeri, 2021). This issue was described by Koedel et al. (2017, p. 3) as a “frustrating lack of information” and this position was echoed by Steiner (2017, p. 1), who addressed the problems as a “matter of deep concern and urgent need.” Chingos and Whitehurst (2012), along with Solomon et al. (2019), continued to highlight the trouble in this area by stating most research done on the effectiveness of curricula failed to stand up to scrutiny. Adding to the problem is the fact research aimed at curriculum materials has received significantly less focus than the research on the theory of teaching (Remillard, 2005). Louisiana has partially overcome these difficulties by providing educators with a curriculum materials tier system based on alignment to the state standards, but connections to student achievement are still absent.

The LDOE theory of action views curricula as an important area of focus to improve student learning and classroom instruction (Kaufman et al., 2018). A byproduct of this thinking led the LDOE to create the tiered list of resources that can be accessed by teachers, school parishes, and other interested parties so they can more easily teach in ways aligned to the state standards (Kaufman et al., 2018). This freedom of choice, which was brought about by a change in Louisiana textbook adoption laws in 2015 (Kaufman et al., 2018), has left a gap in knowledge pertaining to what curricula is being used within Louisiana and if achievement among students is different between curricula choices or between curricula tiers. This gap in knowledge formed the foundation of this study.

The researcher organized Chapter Three by revisiting the purpose statement, research questions, and hypotheses to help connect the research objectives with the data gathering methods below. The remaining sections of Chapter Three progress through the variables and measurement, participants, selection and sampling, research setting, and research design. The chapter then finishes with discussion of the instrumentation used, procedures, data analysis, and summary of Chapter Three.

Purpose of the Study

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) to compare math curricula choice to the Louisiana Education Assessment Program (LEAP) math achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. The independent variable of interest, math curricula choice, will be generally defined as curriculum materials. The dependent variable of interest is third, fourth, and fifth grade math LEAP achievement scores. Both recent and historical research showed a strong connection in mathematics education between the written curriculum within the textbook or program choice and the curriculum and instruction that takes place in the classroom (Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Monaghan, 2013; Prigodich, 2021; Remillard, 2005; Tyson & Woodward, 1989; Walsh, 2009; Westbury 1990). Eisner (1987) summarized textbooks were the leading stimulus for teachers to determine what was taught.

Within the math classroom Ball and Cohen (1996) and Bellens et al. (2020) argued textbooks were the leading factor for what takes place during instruction, with Ball (1996) and Walsh (2009) asserting few teachers detour from textbook outlines and information. Arican (2018) and Ruggeri (2021) both stated the use of textbooks or curriculum materials to implement instruction was heightened further when teachers were not comfortable teaching mathematics. Soloman et al. (2019) went on to note approximately 85 percent of elementary classrooms within the United States relied on at least one curriculum material for science and mathematics. Elliott (1990) stated more than any other subject area, math textbooks were more prone to be used as taught curriculum at the elementary level, while research conducted by Tyson and Woodward (1989) showed 75 percent to 95 percent of the total classroom instruction at the elementary level comes from the textbook structure. This led Monaghan (2013) to proclaim math textbooks are the “de facto” curriculum and the best textbooks should be viewed as the ones that get the best results, which Polikoff, Campbell, et al. (2020) also contended. Since studies in this area often use textbook, curriculum, curricula, and curriculum materials interchangeably to represent the written curriculum from the textbook or program of choice (Blazar et al., 2020; Koedel et al., 2017; Rahman, 2018; Suppa, 2018), this study has also done so to examine the differences in curriculum material choice and student achievement.

Research Questions

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) in comparing math curricula choice to math Louisiana Education Assessment Program (LEAP) achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. Several questions were examined within this study related to the difference of math curriculum material choice and Louisiana student achievement on the LEAP:

RQ1. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ1a. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ1b. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ1c. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

RQ2. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ2a. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ2b. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ2c. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

RQ3. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ3a. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ3b. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ3c. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

Null Hypotheses

The researcher utilized Louisiana Department of Education (LDOE) data to examine the math achievement scores of all public Louisiana school districts for grades three through five to scrutinize the null hypotheses:

$H_0 1$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP.

$H_0 1a$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2016-2017.

$H_0 1b$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2017-2018.

$H_0 1c$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of

Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2018-2019.

H_o2: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP.

H_o2a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2016-2017.

H_o2b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2017-2018.

H_o2c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2018-2019.

H_o3: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP.

H_o3a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2016-2017.

H_o3b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2017-2018.

H_o3c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2018-2019.

Variables and Measurement

The two variables within this study consisted of curriculum material choice, the independent variable, and LEAP mathematic achievement scores, the dependent variable. Curriculum material choice was analyzed by separating the different curriculum materials utilized by schools in the state of Louisiana into previously established tier groups provided by the LDOE. Thus, curriculum materials were grouped and analyzed as a tier group, not analyzed individually. The grouping of curriculum materials into tiers provided the researcher with two benefits. First, the curriculum tier groups allowed the researcher to have a clearly defined variable, which was important since the exact composition of curriculum materials utilized in schools is not known and often an obstacle when conducting curriculum related research (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). Second, the tier groups by their very nature include multiple curricula which helped the researcher limit the likelihood of low population totals which might have existed if individual curriculums were solely analyzed. Additional inspiration for selecting curriculum materials as the independent variable was based on the knowledge curriculum material studies are lacking (Blazar et al., 2020; Bhatt & Koedel, 2012; Chingos & Whitehurts, 2012; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018; Steiner, 2017). Thus, the researcher selected curriculum materials as the independent variable to not only explore the topic but ultimately to add to an area of research in need of more attention.

Mathematic LEAP achievement scores served as a natural dependent variable to analyze the differences in curricula since standardized test scores have been used in curriculum materials studies before (Hiser, 2016; Koedel et al., 2017; Walsh, 2009). Additional support for the selection of an achievement-based dependent variable comes from Monaghan (2013) who asserted math textbooks are the “de facto” curricula in schools, and therefore textbook quality should be judged by those that get the best results, which Polikoff, Campbell, et al. (2020) also contended.

Participants

The participants for this study included all public elementary schools in the state of Louisiana found on the LDOE website. More specifically, the participants for this study included public schools with math LEAP achievement data for grades three, four, and five for testing years 2016-2017, 2017-2018, and 2018-2019. If a public elementary school did not have either math achievement data or some combination of grades three, four, and five that could be analyzed, they were not included in the study. According to the Louisiana Department of Education (2019c), there were 717,109 students who attended public schools within the state in 2018-2019, spread across 72 school parishes, districts, charter schools, or similar designations. These students came from a variety of backgrounds that are both racially/ethnically and economically diverse and dispersed across 1,413 school facilities or reporting sites. Of the 717,109 public school students within the state of Louisiana, 399,831 were classified as minority students, which was about 57 percent of the student population. The largest two racial/ethnic groups of those 717,109 students in public schools were white students totaling 317,278 and black students totaling 309,600. Approximately 69.7 percent of the total public school student population was designated as economically disadvantaged by the LDOE. Limited English proficiency was about three percent. Regarding the third, fourth, and fifth grade student populations, third grade totaled 52,896, fourth grade totaled 54,753, and fifth grade totaled 54,887 in 2018-2019. Thus, the total number of students in the population set for 2019 was 162,536 for those three grades combined. This population is spread out across approximately 750 to 790 unique elementary school buildings (LDOE, 2019b). Similar totals for the 2016-2017 and 2017-2018 school years were found on the LDOE website (LDOE, 2019c) and are outlined in Table 1.

Table 1

Student Population per Grade by Year

School Year	3rd Grade	4th Grade	5th Grade	Total
2016-2017	56,584	56,225	53,636	166,445
2017-2018	55,317	55,779	55,387	166,483
2018-2019	52,896	54,753	54,887	162,536

Selection and Sampling

Purposive sampling was utilized to select the schools that were involved in this study since specific criteria within the population needed to be met to provide the researcher with the appropriate data (Creswell & Poth, 2018). Random sampling was not used for this study for a variety of reasons. First, some buildings designated elementary by LDOE only had a partial representation of grades three, four, and five, did not have enough students within certain grades to publicize results, or did not have those grades within the building at all. As a result, doing a random sample of all buildings with those designations would not guarantee the inclusion of schools within the research guidelines. Likewise, some buildings had a designation of "combination" which included some combination of K-12 that required additional investigation to determine building eligibility for the research study. Another factor that led to purposive sampling was the study's ex post facto research design, which required the participation of current principals within eligible buildings to share information from previous years' curricula material choices; their participation could not be guaranteed.

To gather the specific data needed for this study, the researcher sent out a survey to all school administrators within the 72 school parishes, districts, or similar designations in

Louisiana who served in public schools with some combination of math LEAP achievement data for grades three, four, and five for testing years 2016-2017, 2017-2018, and 2018-2019. The survey included the informed consent acknowledgement, the purpose of the survey and the survey questions the principals were to complete. The sample for this research was then built from all the principals who completed the survey.

The goal of the researcher was to use an ANOVA with fixed effects, omnibus, and one-way for data analysis, with a goal of at least 180 returned surveys, or roughly 45 returned surveys from each of the four-curriculum material tiered groups previously mentioned. Those numbers were set through the use of G*Power software with the following input settings: test family as *F* tests, type of power analysis as *A priori*, effect size *f* as medium or .25, error probability at .05, power set at .8, and the number of groups set to four (Faul et al., 2009). Again, it should be noted only public schools with achievement data from the LDOE website were used. Therefore, private schools, charter schools, parochial schools, and others that did not either participate or have public results on the LDOE website for state-issued standardized testing were not included since no public student achievement data was available to the researcher. The data for public schools that met the research criteria were assigned for analysis to the corresponding curriculum material tier group provided by the LDOE.

The researcher assured the schools involved in this study each responding school would not be directly linked to the school's specific curriculum choice. Rather, each school's curriculum material choice would simply place the school's achievement data into a grouping of schools that used the same curriculum and then the corresponding tier that matched the school curriculum material choice. Even though no identifiable student data or interaction took place within this study, the researcher did go through the process of receiving approval from the Research Review Board (RRB) of Southwest Baptist University, to ensure this study was done in a safe and proficient manner. Dissertation data that was not public access data through the LDOE website was collected and examined using software that required password access to protect district information and anonymity.

Research Setting

The research setting was selected based on the large yet manageable population size within the state of Louisiana, the need for research related to math education and curriculum material choice within the elementary classroom, the availability of performance data through the LDOE, and the uniqueness of the LDOE tiered textbook rating system to compare the variables. The research setting included schools in Louisiana that were a part of the public-school system and had math state assessment data that could be analyzed for grades three, four, and five for the 2016-2017, 2017-2018, and 2018-2019 school years. To ensure a large enough sample size would exist, the researcher sent out the survey to all schools within the state that met those guidelines. The elementary students in third grade totaled 52,896, fourth grade totaled 54,753, and fifth grade totaled 54,887 in 2019 for a combined total of 162,536 across 72 school parishes (LDOE, 2019c). Similar totals existed for the 2017-2018 and 2016-2017 school years as seen in Table 1 above. These student totals spread across approximately 750 to 790 elementary buildings which provided a large enough but manageable variation in building and curricula materials choice for the researcher to conduct the needed research.

Cummins-Colburn (2007) stated there has been little research done on the effect of different texts used by teachers, which is supported by Remillard (2005) who acknowledged this area of research has been overlooked compared to others. Meanwhile, several other researchers (Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Monaghan, 2013; Tyson & Woodward, 1989; Walsh, 2009; Westbury 1990) have made connections between the importance of curriculum materials and student achievement or the relationship between math textbooks and classroom instruction. While this area is not void of research, it does have a "frustrating lack" (Koedel et al., 2017, p. 3) of research. In 2015, the state of Louisiana instituted legislative and educational changes related to their textbook adoption process. This brought about a renewed focus on the textbook as a means for student improvement and the development of Louisiana's tiered textbook evaluation system (Kaufman et al., 2018) to help support educators with their curricula decisions. The lack of research on curricula materials combined with the state of Louisiana's focus on curricula materials to help increase student achievement partially led the researcher to select the state of Louisiana for investigation.

As a point of reference for elementary student performance in the state, the LDOE divides LEAP results into five scoring subgroups from highest achievement to lowest listed as: "A" for advanced, "M" for mastery, "B" for basic, "AB" for approaching basics and "U" for unsatisfactory. For the 2018-2019 school year at the third grade level, 7 percent of students scored at the advanced level, 36 percent at mastery, 26 percent at basic, 21 percent at approaching basics, and 10 percent unsatisfactory, which were similar to the 2016-2017 and 2017-2018 reports (LDOE, 2019b). Fourth grade students for 2018-2019 achieved 3 percent at the advanced level, 38 percent at mastery, 27 percent at basic, 20 percent at approaching basics, and 11 percent unsatisfactory, which were similar to 2016-2017 and 2017-2018 reports (LDOE, 2019b). Fifth grade students for 2019 achieved 4 percent at the advanced level, 30 percent at mastery, 28 percent at basic, 27 percent at approaching basics, and 10 percent unsatisfactory, which were similar to 2016-2017 and 2017-2018 reports (LDOE, 2019b). This type of detailed public data provided by the LDOE, along with additional data that allowed for school parish and specific school building analysis in conjunction with the tiered curricula ranking system established in 2015, was another factor that led the researcher to select elementary schools within Louisiana for research.

Research Design

Widely acceptable research methodology for advanced research consists of quantitative, qualitative, and mixed methods design options (Christensen et al., 2014). In this study, the researcher used a quantitative, causal-comparative, ex post facto investigation to compare student achievement and curriculum materials choice. This was done by utilizing the LDOE curriculum material tier system to categorize curriculum material choice for the independent variable and student achievement for the dependent variable. Since the overall focus of this study was based on gathering, analyzing, and reporting numerical data to answer the research questions, a quantitative study was a logical methodology choice (Christensen et al., 2014). With two established variables to scrutinize, curriculum material choice and student achievement, this study met the guidelines for a causal-comparative study (Salkid, 2010). For the researcher to remain consistent with ethical causal-comparative research practices, no adjustment of the variables was done, and analysis was done utilizing the existing curriculum material tiers and student achievement scores from the LDOE to form the research groups (Creswell & Creswell, 2018; Salkid, 2010). Ex post facto research design means the research was done after the variables had interacted, along with typically measuring variables that naturally interact without manipulation (Christensen et al., 2014);

the researcher examined this type of data for the study and provided the reason for selecting an ex post facto design.

The independent variable within this study was the curriculum material choice of each school. This choice then placed the school in the corresponding curriculum tier which consisted of Tier 1, Tier 2, Tier 3, or non-tiered curriculum. The last tier, defined as non-tiered curriculum, was added by the researcher to group all curriculum material choices schools utilized that are not contained within the LDOE tier listing or did not clearly align to a Tier 1, Tier 2, or Tier 3 curriculum for some other reason. Purposive sampling was used by the researcher to establish the sample for the independent variable. This was done to provide a valid estimate of the different curriculum material choices utilized throughout the state of Louisiana and possibly provide insight into curriculum material use on a larger scale. The dependent variable for this study was student achievement. Student achievement was specifically separated into two subgroups for analysis by percentage. The two groups consisted of one with the top two performing groups on the LEAP: "A" for advanced and "M" for mastery, which were the focus of this study and used for the ANOVA. The other group, which was comprised of the lower three performing groups on the LEAP, "B" for basic, "AB" for approaching basics, and "U" for unsatisfactory, was not part of this study. The advanced and mastery groups were the only two groups used for this study since the researcher attempted to discover whether a difference in curricula materials had a positive difference in achievement rather than a negative difference. Other studies approached achievement by viewing differences through the lens of advanced, mastered, proficient, or similarly designated labels for students who performed well on standardized tests rather than those students who did not (Chandler, 1995; Kodel et al., 2017; Monaghan, 2013; Solomon et al., 2019; Walsh, 2009). To discover if differences occurred, the researcher retrieved open access ex post facto LEAP data from the LDOE website for the 2016-2017, 2017-2018, and 2018-2019 spring testing dates (LDOE, 2019b). Salkind (2010) stated two established variables provide the foundation for a study of difference, which is precisely what the researcher was able to do with curriculum material choice and student achievement.

To perform the analysis on the data, the researcher utilized a one-way ANOVA for H_01, H_02, and H_03 to determine whether a difference exists between curriculum choice and student achievement. To determine if a difference exists within the data sample, the researcher utilized a program called Statistical Package for Social Sciences (SPSS) to compute the needed ANOVA. The researcher focused on finding whether a measurable statistically significant difference exists between the dependent variable of the groups, rather than exploratory research to determine how or why a difference may or may not exist.

Instrumentation

The instrument employed to measure the difference between curriculum material choice and student achievement by the researcher was the Louisiana Education Assessment Program (LEAP). Data from the LEAP has been used as an instrument to measure student achievement in other studies recently (Ebanks, 2020; Walker, 2021) and the use of LEAP data within this study was consistent with how other researchers utilized state testing data. The LEAP is used each year by the LDOE as the benchmark test for student achievement within the state of Louisiana to determine student readiness for the next grade (LDOE, 2018). The mathematics section of the LEAP for grades three through eight includes three types of tasks or questions: Type I (conceptual understanding), Type II (procedural skill and fluency), and Type III (application) questions aligned to the Louisiana State Standards for

Mathematics (LSSM) (LDOE, 2019a). Each of these three types of tasks filter into one of four specific reporting categories: Major Content, Additional & Supporting Content, Expressing Mathematical Reasoning, or Modeling & Application. The LEAP question creation was a collaborative effort between the LDOE and the Data Recognition Corporations (DRC). Questions for the LEAP were then reviewed by Louisiana educator committees to make sure they were appropriate for students based on the LSSM for each grade level (LDOE, 2020). Educators within Louisiana also reviewed test items to look for issues involving bias, fairness, difficulty, and cognitive complexity in addition to alignment to the LSSM for each grade level to ensure questions were appropriate for use on the assessment. Any items deemed inappropriate were revised or simply rejected and not included if they could not be adequately revised (LDOE, 2018, 2019, 2020).

Reliability and validity are two key components of the LEAP that must be examined for the results to be used in a meaningful way. Both Jing (2021) and McLaughlin (2021) stated reliability was the ability to measure and accurately reproduce results. Meanwhile, validity related to the precision with which an assessment, the LEAP in this case, measured what it intended to measure (Creswell, 2019; McLaughlin, 2021). However, validity is not possible if reliability does not exist first (McDonald, 2018; McLaughlin, 2021). To provide transparency and evidence of reliability and validity on the LEAP, the LDOE provided a technical report outlining how the test and test items were created and analyzed on the LDOE website (LDOE, 2018, 2019, 2020), which have been similarly created and administered since 2015. Regarding reliability, the LDOE utilized the DRC to scrutinize the reliability of the LEAP in several areas. One of these areas was raw-test score reliability, which summarizes how consistent the test was across participants. For this, Cronbach's coefficient alpha was utilized. The LDOE technical reports for raw score reliability using Cronbach's coefficient alpha were available for each testing year of this study. Reliability coefficients equal to or greater than .8 on a 0 to 1 scale are established as good and acceptable. The closer the score is to 1, the more consistent the test was and scores of .8 or higher are acceptable in most cases for tests of moderate length (LDOE, 2020).

Cronbach's coefficient alpha for the mathematics section of the LEAP ranged from .89-.92 for 2016-2017, from .91-.92 for 2017-2018, and .92-.94 for 2018-2019, which indicated each year the assessment measured the same construct or content domain reliably (LDOE, 2018, 2019, 2020). These results are outlined in Table 2 below.

Table 2

Cronbach's Coefficient Alpha by Year

LEAP Testing Year	Cronbach's Coefficient Alpha
2016-2017	.89-.92
2017-2018	.91-.92
2018-2019	.92-.94

Note. A range equal to or greater than .8 is viewed as acceptable.

While Cronbach's coefficient alpha provided a good measure of reliability, it was not the only reliability measure examined within the LDOE technical reports. The standard error of measurement (SE measurement) was also reported for the LEAP, which was used to help support reliability in other technical reports for state tests (Arizona Department of Education [AZED], 2018; New York Department of Education [NYSED], 2019) and support reliability for the usage of standardized state data sets in other studies (Cox, 2022; Greathouse, 2021).

While standard error of measurement was often referenced as *SE* measurement in research (Frey, 2018), it was abbreviated as SEM within the LEAP technical report, which was common in other research as well (Barthelemy, 2022; Jing, 2021; LDOE, 2020).

The *SE* measurement is closely connected to the idea of reliability through an inverse relationship that demonstrates an increase in test reliability when low *SE* measurement values exist and less reliability if a test has high *SE* measurement values (Frey, 2018; Luck et al., 2021). The LEAP technical report, along with other research, discussed *SE* measurement as the range around a student's test score, termed the observed score, that likely represented the student's true score (Gardner, 2022; LDOE, 2020; McCrae & Mottus, 2019; McLaughlin, 2021). The LEAP technical report, and similarly stated by Fraenkel et al. (2019), referred to the true score range around the observed scores as an index of random variability (LDOE, 2020) that can be calculated once a reliability coefficient is established (AZED, 2018; Hong, 2022), with Cronbach's coefficient alpha often used as the conventional choice (McNeish, 2018); this was the method utilized by the LEAP technical report (LDOE, 2020).

Table 3 below shows the results from the mathematics section of the LEAP for third, fourth, and fifth grade students during the 2016-2017, 2017-2018, and 2018-2019 school years. It should be noted there were no computer LEAP tests administered for third grade for 2016-2017, and fifth grade only takes the computer test each year. For interpretation purposes, the LEAP technical report examines *SE* measurement by stating it was, "expected that the score a student obtains from a single test administration would fall within one SEM [SE measurement] of the student's true score 68 percent of the time and within approximately two SEMs of the true score 95 percent of the time" (LDOE, 2020, p.206). Thus, the use of *SE* measurement allowed the LDOE to examine the reliability of the LEAP through a comparison of observed scores and expected true scores (LDOE, 2020), along with the previously stated relationship between SE measurement and reliability that demonstrated an increase in test reliability when low SE measurement values exist, and less reliability if a test has high SE measurement values (Frey, 2018; Luck et al., 2021).

Table 3

Mathematics SE measurement Score per Grade by Year

School Year	3rd Grade Computer Test *SE* measurement	3rd Grade Pencil Test *SE* measurement	4th Grade Computer Test *SE* measurement	4th Grade Pencil Test *SE* measurement	5th Grade Computer Test *SE* measurement
2016-2017	N/A	3.73 (62-point test)	3.47 (59-point test)	3.67 (59-point test)	3.58 (61-point test)
2017-2018	3.55 (62-point test)	3.63 (62-point test)	3.46 (62-point test)	3.59 (62-point test)	3.66 (62-point test)
2018-2019	3.53 (62-point test)	3.82 (62-point test)	3.52 (62-point test)	3.65 (62-point test)	3.59 (60-point test)

Evidence for validity within the technical reports provided by the LDOE for each year included general evidence for validity throughout each report along with specifically addressing key validity concepts. One of those areas is content validity. Content validity examines how well items on the assessment reflected the standards (LDOE, 2020), which was established by aligning the assessment to the standards and reviewing how well the items on the assessment matched the standards. This process was previously outlined above. Another specific area of validity examined was construct-related validity, which assesses how valid the construction or format of the assessment was, and specifically how well it measured what it was attempting to measure. Construct-related validity was connected to evidence from the Data Recognition Corporations (DRC) studies on test reliability, convergent validity, and divergent validity, also known as discriminant validity (LDOE, 2020). The LDOE demonstrated construct validity on the LEAP through the convergent and divergent validity analysis done by the DRC. Evidence to support convergent validity and divergent validity to support the overall construct validity of the LEAP can be found in Chapter 9 of the LEAP technical reports each year (LDOE, 2018, 2019, 2020).

Convergent validity examines whether assessment items are narrow enough to measure what is intended. This essentially means the assessment was constructed in a way which accurately allows individuals to show their knowledge if they have the knowledge, thus a student's knowledge and ability to show that knowledge on the assessment were correlated. Meanwhile, divergent validity examines whether assessment items include additional variables or outside knowledge not a part of original intent that hinders the assessment's ability to measure what is intended. This would mean the question or assessment would diverge from the intended purpose and produce results not correlated to the original intent. Analysis of student responses to each item on the LEAP and how each student answered similar items compared to other students was a key element of analysis to help establish validity (LDOE, 2018, 2019, 2020).

Technical report evidence also showed LEAP tests each year were unidimensional, which gave the LDOE confidence the LEAP measured the skills it was attempting to measure. Unidimensional essentially means math items measured math ability and not reading ability or some other content area other than mathematics (LDOE, 2018, 2019, 2020). These factors allowed the LDOE to support the claim LEAP items and tasks had dominant dimension, meaning student LEAP performance reflected each student's ability in each specific content area (LDOE, 2018, 2019, 2020). By design, each LEAP test since 2015 has been created and structured in a similar way to provide longitudinal analysis with similar evidence of reliability and validity within each technical report each year (LDOE, 2018, 2019, 2020).

The scoring results for the LEAP are divided into five scoring groups from highest achievement to lowest listed as: “A” for advanced, “M” for mastery, “B” for basic, “AB” for approaching basic, and “U” for unsatisfactory. Questions on the mathematics test consist of the four categories, which were listed above, and have an identical point breakdown for third, fourth, and fifth grades: Major Content totaling 30 points, Additional & Supporting Content totaling 10 points, Expressing Mathematical Reasoning totaling 10 points, and Modeling & Application totaling 12 points for an overall total of 62 points. The types of questions include multiple choice, multiple select, short answer, keypad input, constructed response, and technology enhanced. Assessments are scored both digitally and by hand depending on the test item or specific format for grades three, four, and five. The LEAP results are open access and found on the LDOE website for every school and school parish. As a result, a data request or permission request did not need to be obtained.

Procedures

The Research Review Board (RRB) at Southwest Baptist University is required to approve dissertation research to confirm human participants are protected prior to conducting research. As a result, a request was submitted to the RRB to gather school building level data on student achievement scores and to distribute an electronic survey to all building and parish administrators to collect information on each school’s curriculum choice within the state of Louisiana, along with phone calls to those same administrators to solicit survey completion.

School achievement and curriculum material data for the 2016-2017, 2017-2018, and 2018-2019 school years were specifically outlined within the RRB request. The RRB request was granted in December of 2022 and the researcher was permitted to collect data, analyze the data, and report the findings. Approval located in Appendix A.

Once RRB permission was obtained, the researcher began the data gathering process. This process included gathering administrator names, emails and/or phone numbers within the sample group, sending administrators a survey, collecting the results of the survey, and then running a one-way ANOVA. To do this, the researcher first obtained the list of current school parishes and elementary school buildings within each parish the LDOE showed LEAP math scores for the 2016-2017, 2017-2018, and 2018-2019 school years. The researcher then gathered names, emails, and phone numbers for each school administrator who worked within an elementary school building in the sample group above from available directory information posted on the LDOE website each year. This ensured only administrators within the desired purposive sample group would be part of the survey.

Once the above-mentioned purposive sample group was defined and email addresses were sorted, the researcher sent the initial email, which included the survey link with statement of informed consent, directions for survey completion, and a summary outlining the researcher's purpose for the survey to all school building or parish administrators and educational leaders within the sample group. A copy of this email can be found in Appendix B. This started the initial data gathering period of at least two weeks. Google Forms was used by the researcher to create and administer the survey. Results from the Google Forms survey were organized and stored within a corresponding Google Sheet used for coding purposes.

As stated above, the goal of the researcher was to use an ANOVA for data analysis. Using G*Power software (Faul et al., 2009), it was estimated the researcher would need at least 180 returned surveys, or roughly 45 returned surveys from each of the four-curriculum material tiered groups. Therefore, the researcher decided to start the initial data gathering process over at least a two-week period, starting with the initial email, to gather enough responses. During this two-week timeframe, the researcher sent a reminder email after one week, found in Appendix C, and another reminder email one day before the end of the initial two-week data gathering process, found in Appendix D. If the initial data gathering process did not produce a large enough sample, the researcher planned to send an additional email, found in Appendix E, and randomly call each school administrator until enough responses were gathered. The phone numbers were gathered from the LDOE website, and a copy of the phone call transcript can be found in Appendix F. Principals and parish administrators, along with school phone numbers and emails, were obtained by the researcher from the LDOE website which was published as directory information.

Data Analysis

Once data collection was complete, the data cleaning process began. As previously stated, Google Forms was used by the researcher to create and administer the survey. Results from the Google Forms survey were organized and stored within a corresponding Google Sheet used for coding purposes. The survey within Google Forms utilized drop-boxes with the previously established Tier 1, Tier 2, Tier 3 curriculum groups provided by the LDOE as options, along with a drop-box for non-tiered curriculum, which was added by the researcher to categorize all curriculum the LDOE did not include in their tier rankings. The data cleaning was a simple process because misspellings, extraneous answers, partially completed surveys, and similar issues were very limited since the structure of the survey naturally avoided these issues. Even partially completed survey results were functional since they provided at least one data point if they had a grade level selection and curriculum choice selection associated with that grade level. The efforts for data cleaning centered around two main points: first, to make sure a single school building did not provide duplicate data points for a specific year, grade, and curriculum combination; second, to make sure there were no extraneous answer or data points the Google Forms structure did not naturally eliminate.

The demographic data that was shared as part of the research findings and used for data analysis was identical to the previously mentioned coded data that reflected the grade level, school year, percentage of students at the advanced or master level, and curricula choice of the schools within the sample population. Once the data cleaning process was complete, a one-way ANOVA was performed to determine whether a difference existed in student achievement among the curriculum choices. Data collection utilized Google Forms while statistical analysis was completed using Statistical Package for the Social Sciences (SPSS) to determine if the researcher would reject or fail to reject the null hypotheses. The dependent variables for each research question were the percentage of students who ranked at the advanced and mastery level on the LEAP, also phrased as student achievement, while the independent variable was the curriculum choice utilized by the corresponding school. Research questions one, two, and three were set up similarly in both the data collection and analysis process, with the only operational differences between questions relating to the specific grade level of each research question for third, fourth, or fifth grade, and the school year analyzed, 2016-2017, 2017-2018, and 2018-2019. Thus, the data was coded to reflect grade level, school year, percentage of students at the advanced or mastery level, and curriculum choice within the sample population. The ANOVA allowed the researcher to compare the difference in mean student achievement between Tier 1, Tier 2, Tier 3, and non-tiered curriculum choices to help reject or fail to reject the null hypotheses. An alpha level of .05 was set to evaluate the level of significance, which was a suitable value to reject or fail to reject the null hypotheses (Pelham, 2013). If the alpha level is less than .05, the null hypothesis will be rejected while an alpha level of .05 or higher would mean the null hypothesis would fail to be rejected. The researcher employed an ANOVA for data analysis because the researcher compared the means across four different groups, instead of an independent *t*-test that would only compare two groups at a time (Laerd Statistics, 2017). After the collection and analysis of data, the researcher will report the means and standard deviations of the dependent variable of each of the groups in Chapter Four.

While a *t*-test and ANOVA have differences, they share the same assumptions about the numerical data that must be scrutinized (Pelham, 2013). Laerd Statistics (2017) outlined the six assumptions for one-way ANOVA tests, which are outlined and applied to this specific study below. The first assumption that must be examined is the assumption that the dependent variable is continuous or measured in intervals. The dependent variable within this study, student achievement, was measured on a 0 to 100 percent scale, and therefore meets this assumption. Second, the independent variables of a one-way ANOVA should consist of at least two different groups or categories and often consists of three or more groups. This study had four groups of independent variables representing the different curriculum tiers: Tier 1, Tier 2, Tier 3, and non-tiered curriculum choices, which meet the criteria for assumption two. The third assumption addressed the need for independence of observation, meaning there was no relationship between groups or influence from the observation process on the groups. Each school participant or achievement score was a part of one group and did not participate or influence another group. As a result, this assumption, like the previous two, was met.

The fourth assumption outlined by Laerd Statistics (201) related to outliers. The results from a one-way ANOVA can have reduced validity if outliers in the data are not accounted for and addressed. To potentially find outliers, the researcher examined the results of the LEAP scores, the dependent variable, for outliers within each independent variable grouping, the three LDOE tier groups and the additional non-tiered curriculum group. Boxplots for dependent variable results within each tier group were created and examined utilizing SPSS software to determine the existence of outliers (Laerd Statistics, 2017). Further discussion related to outliers, if they were present, are addressed in Chapter Four when the Google Form survey results are discussed. Assumption number five stated the dependent variable should have normal distribution, or approximately normal distribution, across each independent variable group. Once data was obtained, the Shapiro-Wilk test of normality was applied using the SPSS software. More discussion related to this can be found in Chapter Four with additional discussion related to survey results. The sixth and final assumption stated the data needs to have homogeneity of variances. This means the independent variables, each of the tiered curriculum groups, should have the same variance across the dependent variable, student achievement. This was analyzed by using Levene's Test through the SPSS software. If this assumption is violated, a Welch ANOVA can be utilized using the SPSS software and is addressed again in Chapter Four.

Another consideration the researcher needed to address was effect size. While an ANOVA analysis with an alpha of .05 can help determine if a significant statistical difference occurs (Pelham, 2013), it does not help describe the effect size. The strength of the association between the variables is referred to as effect size (Gay et al., 2009). When examining effect size, Cohen (1988) Partial Eta Squared guidelines for effect size were followed, in which a small effect size had a value near .01, a medium effect size near .06, and a large effect size near .14. The SPSS ANOVA analysis results table calculated the effected size following each ANOVA, which is reported in Chapter Four.

The researcher's goal was to utilize the ANOVA to compare LEAP scores with curriculum choice to determine if a difference in student achievement occurred based on curriculum choice. Nine total ANOVA tests were performed. The first ANOVA examined third grade student achievement for the 2016-2017 school year across the four-tiered curriculum groups: Tier 1, Tier 2, Tier 3, and non-tiered curriculum choices. The second ANOVA examined fourth grade student achievement for the 2016-2017 school year across the four-tiered curriculum groups. The third ANOVA examined fifth grade student achievement for the 2016-2017 school year across the four-tiered curriculum groups. The fourth ANOVA examined third grade student achievement for the 2017-2018 school year across the four-tiered curriculum groups. The fifth ANOVA examined fourth grade student achievement for the 2017-2018 school year across the four-tiered curriculum groups. The sixth ANOVA examined fifth grade student achievement for the 2017-2018 school year across the four-tiered curriculum groups. The seventh ANOVA examined third grade student achievement for the 2018-2019 school year across the four-tiered curriculum groups. The eighth ANOVA examined fourth grade student achievement for the 2018-2019 school year across the four-tiered curriculum groups. The ninth ANOVA examined fifth grade student achievement for the 2018-2019 school year across the four-tiered curriculum groups. Again, all ANOVA tests employed a .05 level of statistical significance to reject or fail to reject the null hypotheses (Pelham, 2013). If a difference exists, a Tukey's Test for post hoc analysis would be utilized by the researcher to determine where the differences between groups specifically existed. It should also be noted if assumption six above is violated, a Games-Howell post hoc test would be utilized instead of a Tukey post hoc test to account for the violation of assumption six, which will be addressed in Chapter Four should the need arise (Laerd Statistics, 2017).

Summary

The purpose of Chapter Three was to convey the methodology utilized to determine whether a difference between curriculum material choice and student achievement exists. This quantitative, causal-comparative study utilized the LDOE curriculum material tier system to categorize curriculum materials along with LDOE student achievement data from the LEAP for the variables. Since multiple populations or tiers of curriculum materials existed, an ANOVA was used to examine whether differences existed. Both curriculum materials and student achievement data were gathered from the 2016-2017, 2017-2018, and 2018-2019 school years through the use of a survey for curriculum material use and utilization of the open access data from the LDOE website for student achievement information. Next, Chapter Four focuses on communicating the results from this study now that the methodology has been outlined. Chapter Five consists of a summary to review the entire scope and sequence of this study, outlines the implications of the findings, and makes recommendations for future studies.

CHAPTER FOUR

FINDINGS

Introduction

The combination of stagnant growth and a widening achievement gap in mathematics has educators, administrators, and policy makers searching for new ways to improve student learning and achievement. While several factors can influence student achievement, one of the prevailing avenues school districts have pursued to increase student achievement is using effective curricula materials (Koedel et al., 2017; Reys et al., 2003; Solomon et al., 2019; Superfine et al., 2010). Some recent studies suggest different curricula materials have varying impacts on student learning (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Cress, 2019; Koedel et al., 2017; Lein Authement, 2022; Polikoff, Petrilli, et al., 2020; Walsh, 2009; White, 2018). Yet, administrators are often unable to make better decisions related to instructional materials due to a lack of evidence on the impact of the materials being used (Blazar et al., 2020; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Koedel et al., 2017; Polikoff, Campbell, et al., 2020; Ruggeri, 2021).

The objective of Chapter Four is to convey the results of this curriculum materials study following the data collection process and subsequent testing of the data utilizing a one-way ANOVA. To do this, the researcher sequenced Chapter Four to first restate the purpose, significance, and selection of Louisiana for this investigation, along with the research questions and null hypotheses to provide context for data analysis. This transitions to a review of the gathered data, presentation of the findings following the one-way ANOVA, sharing the research results, and transition to Chapter Five.

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) and to compare math curricula choice to Louisiana Education Assessment Program (LEAP) math achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. The independent variable of interest, math curricula choice, was generally defined as curriculum materials. The dependent variable of interest was third, fourth, and fifth grade math LEAP achievement scores. Both recent and historical research showed a strong connection in mathematics education between the written curriculum within the textbook or program choice and the curriculum and instruction that takes place in the classroom (Arican, 2018; Ball & Cohen, 1996; Bellens et al., 2020; Eisner, 1987; Elliott, 1990; Monaghan, 2013; Remillard, 2005; Ruggeri, 2021; Tyson & Woodward, 1989; Walsh, 2009; Westbury 1990).

Statistics have shown sustained curriculum material usage in recent years (Blazar at al., 2020; Koedel et al., 2017; Solomon et al., 2019), yet research related to what curriculum materials have been utilized by schools and the influence of curriculum materials on student learning has been hindered because nearly all the states do not track or record what curriculum materials have been used (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). Of the few states that have gathered and reported data on curriculum materials, the data has suffered from misspellings or a general lack of organization and made it unviable or immensely difficult to use (Kane, 2016). The limited research related to what curriculum materials schools utilize presents a major roadblock when it comes to evaluating curriculum material use (Blazar et al., 2020; Cummins-Colburn, 2007; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018).

In 2015, the state of Louisiana instituted legislative and educational changes related to their textbook adoption process. This brought about a renewed focus on the textbook as a means for student improvement and the development of Louisiana's tiered textbook evaluation system (Kaufman et al., 2018) to help support educators with their curricula decisions. The lack of research on curricula materials combined with the state of Louisiana's focus on curricula materials to help increase student achievement was a motivating factor that led the researcher to select the state of Louisiana for investigation. The tiered system consisted of curriculum materials rated as a Tier 1 through Tier 3 (LDOE, 2022). In addition to the Louisiana Department of Education's tier groupings of curriculum materials, the researcher added the non-tiered curriculum materials category to the data to capture curriculum materials that might have been utilized by a school outside of the LDOE tier rankings.

Research Questions

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) to compare math curricula choice to math Louisiana Education Assessment Program (LEAP) achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. Several questions were examined within this study related to the difference of math curriculum material choice and Louisiana student achievement on the LEAP:

RQ1. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ1a. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ1b. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ1c. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

RQ2. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ2a. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ2b. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ2c. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

RQ3. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

RQ3a. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

RQ3b. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

RQ3c. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

Null Hypotheses

The researcher utilized Louisiana Department of Education (LDOE) data to examine the math achievement scores of all public Louisiana school districts for grades three through five to scrutinize the null hypotheses:

$H_0 1$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP.

$H_0 1a$: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2016-2017.

H_o1b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2017-2018.

H_o1c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2018-2019.

H_o2: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP.

H_o2a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2016-2017.

H_o2b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of

Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2017-2018.

H_o2c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2018-2019.

H_o3: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP.

H_o3a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2016-2017.

H_o3b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2017-2018.

H_o3c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2018-2019.

Analysis of the researcher's data and findings are summarized below. Regarding the data, the next section outlines the data sample, the demographics, and the data cleaning process utilized by the researcher. Concerning the research findings, the section includes each null hypothesis, research question, and corresponding tested assumptions. This is followed by the actual research findings, summary of the data analyzed, and interpretation of outcomes and summary of overall results. To conclude, assertions to either reject or fail to reject the null hypotheses are presented based on the ANOVA results and transition to Chapter Five.

Data Analysis and Findings

This study examined what curriculum materials were utilized within the state of Louisiana for third, fourth, and fifth grades to discover whether student achievement differed based on curriculum material choice. To determine if a difference existed between curriculum material choice and student achievement, the researcher used data from the Louisiana Education Assessment Program (LEAP). During this study, the independent variable of interest, math curricula choice, was commonly defined as curriculum materials. The dependent variable of note was third, fourth, and fifth grade math LEAP achievement scores. It was then the aim of the researcher to send out a survey to administrators within Louisiana and gather at least 180 responses, approximately 45 from each of the four-curriculum material tiers: Tier 1, Tier 2, Tier 3, and non-tiered group, with the addition of the non-tiered group by the researcher to gather data from schools that did not utilize a curriculum listed within the existing three LDOE tiers. Once enough surveys were received, SPSS statistical software was used to initiate an ANOVA to test the null hypotheses.

Sample

Participants in this study included all public elementary schools in the state of Louisiana containing LEAP data on the LDOE website with a large enough sample size for analysis. Data for the LEAP required 10 or more participants, student scores, within a group for public disbursement of the results (LDOE, 2019b). Specifically, for this study, the participants included public schools with math LEAP achievement data for grades three, four, and five for testing years 2016-2017, 2017-2018, and 2018-2019. Each school that met the above criteria provided a possible data point and would be considered a qualifying school that met the purposive sampling conditions. To collect the specific data required for this study, the researcher distributed a survey to all school administrators within the 72 school parishes, districts, or similar designations in Louisiana who served in public schools that met the above qualifying conditions. Included within the survey was the purpose of the survey, the survey questions, and informed consent acknowledgment. The data sample for the investigation was built from all the principals who completed the survey, in addition to the LDOE provided data. However, during the data gathering process, several principals mentioned the need for administrative approval at the parish or district level and did not have authorization at the individual building level to participate in the survey. At that point the researcher expanded email distribution beyond the initial survey to building principals and initiated contact with parish administrators and educational leaders to gain the needed information. In some cases, the parish level approval granted principals the authorization to complete the survey and in other cases the parish level administrator provided partial or comprehensive data for the entire school parish. As a result, the final data sample was assembled from what was gathered from the LDOE, surveys to principals, and contact from parish level administrators. The total number of data points collected for each year, for each grade level, are listed in Table 4. Additional details related to data source, data tier distribution, and specific curriculum material within each tier for the gathered data can be found in Appendix G (see Tables G1-G5).

Table 4

Collected Data by Year and Grade

School Year	Grade	School Participation (Collected Data Points)	Total Schools
2016-2017	3	336	786
	4	336	778
	5	332	757
2017-2018	3	586	789
	4	583	782
	5	575	764
2018-2019	3	580	761
	4	585	770
	5	574	751

Demographics

The previously mentioned data sample was gathered from the state of Louisiana, which had a total student enrollment ranging from 717,000 to 721,000 students per year from 2016-2017 to 2018-2019. These students were instructed from 1,420 different building locations, of which between 750 and 790 were elementary buildings (2019b) and were the focus of this study. The most prominent two racial/ethnic groups for each year were white students and black students. Most of the LDOE provided data consisting of percentages for the entire school population of the state, with the only grade specific details consisting of the specific number of students within each grade (LDOE, 2019c).

Table 5

Demographics

School Year	3rd Grade Students	4th Grade Students	5th Grade Students	White	Black	Limited English Proficiency	Economically Disadvantaged
2016-2017	56,584	56,225	53,636	44.98%	43.89%	3.28%	71.6%
2017-2018	55,317	55,779	55,387	44.64%	43.59	3.53%	69.23%
2018-2019	52,896	54,753	54,887	44.24%	43.17%	3.68%	69.78%

Note. Data and column titles were sourced from LDOE (2019c).

Data Cleaning

Upon completion of the data gathering process and organization of the results, the data cleaning process commenced to help ensure accurate results. This process first involved the examination of the data gathered by the researcher and then analysis of the gathered data to determine the degree to which the collected data adhered to or violated one-way ANOVA assumptions four, five, and six previously mentioned in Chapter Three. Regarding the gathered data, the research survey was designed to intentionally limit issues such as misspellings, conflicts with duplicated school names from different parishes, extraneous answers, partially completed surveys, and similar issues. This was done through the utilization of drop-down lists with all possible parish, school, grade level, and LDOE curriculum tier choices entered beforehand. Then, the survey was sequenced in a way that required participation to first acknowledge informed consent, select the school parish, select a school's name with site code identifier, and then start selecting the grade level(s) and curriculum material(s) within the school building for each year. Because of this sequence, partly finished survey results were functional since the essential identifiable information was already completed for the survey. Thus, the two primary efforts related to data cleaning from the survey centered around a review of the data to make sure each school building did not provide duplicate data points by one administrator or two administrators completing information for the same building, which after review did not occur. The second step in the process was to review the survey results to make sure there were no extraneous answers or data points the Google Forms organization did not naturally resolve, which after review did not occur.

Despite the ease of the survey data cleaning process, the researcher still needed to reconcile situations where the principal survey results, district level administrator survey results, and/or shared curriculum material information and LDOE provided information did not align. In some cases, nonalignment was not an issue if the different results consisted of two or more answers within the same LDOE tier, since placement into correct tier instead of identifying the exact curriculum material was the focus of the investigation. When the three sources of information did not fully align, but two of the three sources did, the researcher used the information the two sources aligned provided instead of the source from the one outlier, which only occurred with four school locations and impacted 18 total data points across all three years for all three grades. A conflict where all three sources of information had no commonality did not occur.

This left the researcher to determine how to resolve situations where two sources of information disagreed, which only took place for the 2017-2018 and 2018-2019 school years. The total data point disputes between two sources were limited to 27 individual data points, spread across 10 different school buildings. In those instances, the researcher initially defaulted to using the LDOE provided data since it was gathered closer to the actual school year in 2020 or earlier instead of when the researcher conducted the research in spring of 2023. However, the researcher did use the principal survey results instead of the LDOE data if the principal data accurately aligned with either the LDOE data or parish level administrator data at other data points consistently to account for pilot program migration when curriculum material shifts occurred, slow adopters of new curriculum material shifts, and to reflect the principal's knowledge of what had occurred in the building. Using the above framework as a guide, the 27 data points that were not aligned when two sources were provided resulted in the researcher utilizing the LDOE data for 16 data points and the principal survey for nine data points for curriculum tier choice. For one school location, that impacted three data points, a principal survey and district administration information did not align: 2017-2018. In that instance, the principal information that was provided was inconsistent with other principal surveys within the same school parish and parish level administrator information provided to the researcher, thus the researcher utilized the parish level provided information that was consistent with other schools in the parish at that time.

Regarding the LEAP data provided by the LDOE the researcher needed to address two situations that required data cleaning. First was the adjustment of LEAP data for advanced and mastery school performance reported as "≤ 1" to the value of "1" to allow those fields to be included in the calculations. Second was the removal of data points that had an "NR" in place of a number value, which indicated the sample population was below 10 and not reported and therefore not of use for this study (LDOE, 2019b). Once data cleaning related to the gathered data was complete, the second part of the data cleaning process of evaluating the final three assumptions for utilization of a one-way ANOVA proceeded. The final three assumptions were previously mentioned in Chapter Three along with the first three assumptions and justification for the data passing the initial three assumptions.

Assumptions. The first of the remaining three assumptions the researcher examined was assumption four, outliers. The results from a one-way ANOVA can have reduced validity if outliers in the data are not accounted for and addressed (Laerd Statistics, 2017). Therefore, the researcher examined outliers in three different ways. The of which was to locate outliers within the data and examine the accuracy of the recorded data. Within the recorded data, two types of outliers were found. First were values that showed 100 percent or 101 percent of students scoring in the advanced or mastery range. In both cases the researcher reviewed the data provided by the LDOE to confirm the data records were accurate. After review, the records used for data analysis were consistent with the LDOE records. While cases of 100 percent of the students scoring advanced or mastery were not common, it did occur within most of the nine data samples for this study one or more times.

As for the 101 percent, the LDOE acknowledges that different techniques were used within their reports for privacy reasons, which may impact totals but not impact the usability of the data (LDOE, 2019b). As a result, these data points were used and kept at the LDOE reported totals.

The second step of the three-step process used to examine outliers, involved the utilization of SPSS to locate outliers through descriptive statistics and boxplots, which followed the procedures outlined in Chapter Three. Results from the data gathering process extended across three academic years, comprising three grades per year, with data from Tier 1, Tier 3, and non-tiered groups for each grade. This created 27 individual data subgroups, three per ANOVA, that required examination of outliers for the nine total ANOVA tests. Within SPSS, running the Explore procedure facilitated the creation of boxplots and descriptive statistics to examine outliers, which were indicated by data values one and a half times the interquartile range (IQR) below the first quartile or above the third quartile for each individual subgroup of data. Of the 27 data subgroups, no outliers were found below the minimum of the boxplots for any of the data sets, but outliers were found above the maximum for 13 of the 27 subgroups utilizing SPSS boxplot calculations (see Appendices K, L and M). In four of those 13 instances, the outliers were within five percentage points of the maximum, with outlier values in those data sets originating with values 95, 97, 97 and 99.5 out of the highest potential value limit of 100. Additionally, the number of outliers within each subgroup represented five or fewer data points for 10 of the 13 subgroups. The remaining three subgroups had outlier totals of eight, nine, and nine, but had total samples sizes of 463 or more data points within each subgroup. Related, all subgroups with one outlier had 23 or more data points and all subgroups with two or more outliers had at least 59 data points. Additionally, none of the outlier values within the 13 data sets were extreme outliers, indicated by a value of three times the IQR above the third quartile, instead of one and a half times the interquartile range (IQR) for traditional outliers. While a few outliers can impact results, sample size was important because the degree to which outliers impact outcomes is typically lessened with larger samples sizes (Bloom, 2020; Liao et al., 2016), allowing the research approach to continue as prepared (Burns McOmber, 2020; Field, 2013). The tables below (see Tables 6, 7, and 8) detail key outlier information for each of the three school years of this study.

Table 6

2016-2017 Outlier Information per Tier Group by Grade within Tier

Grade	Tier	Sample Size (*n*)	Boxplot Q3 Value	Maximum Value (Q4)	Number of Outliers	Outlier Values
3rd	1	274	56	98	2	100, 100
	3	59	57	94	0	-
	Non-tiered	3	69	69	0	-
4th	1	273	53	98	0	-
	3	60	50.75	93	0	-
	Non-tiered	3	74	74	0	-
5th	1	271	44	83	5	89, 89, 98, 98, 100
	3	59	40	75	2	87, 91
	Non-tiered	2	N/A	31	0	-

Note. The upper limit of values was 100. A dash designation, (-), indicated not applicable.

Table 7

2017-2018 Outlier Information per Tier Group by Grade within Tier

Grade	Tier	Sample Size (*n*)	Boxplot Q3 Value	Maximum Value (Q4)	Number of Outliers	Outlier Values
3rd	1	471	55	97	4	98, 100, 100, 100
	3	92	47	81	3	90, 98, 100
	Non-tiered	23	44.25	73	1	100
4th	1	468	52	97	3	99, 100, 100
	3	92	47.5	90	0	-
	Non-tiered	23	43	69	1	90
5th	1	463	41	77	9	83, 84, 85, 86, 89, 91, 94, 95, 100
	3	90	38.25	76	3	84, 86, 88
	Non-tiered	22	29.5	46	1	78

Note. The upper limit of values was 100. A dash designation, (-), indicated not applicable.

Table 8

2018-2019 Outlier Information per Tier Group by Grade within Tier

Grade	Tier	Sample Size (*n*)	Boxplot Q3 Value	Maximum Value (Q4)	Number of Outliers	Outlier Values
3rd	1	569	54	95	8	98, 98, 98, 99, 100, 100, 100, 101
	3	3	60	60	0	-
	Non-tiered	8	66.75	97	0	-
4th	1	572	54	100	0	-
	3	4	65	68	0	-
	Non-tiered	8	59.25	94	0	-
5th	1	568	45.75	88	9	89, 91, 92, 93, 94, 96, 96, 98, 100
	3	4	47.75	85	0	-
	Non-tiered	2	N/A	92	0	-

Note. The upper limit of values was 100 and the inclusion of value 101 was previously discussed above. A dash designation, (-), indicated not applicable.

The final step, the researcher used to gauge the consequences of outliers was to perform the ANOVA without the outliers in the data set to determine if those altered the findings (Laerd Statistics, 2017). Once outliers were removed from the data sets containing outliers, assumptions tests were repeated and the subsequent ANOVA was conducted without the outliers. In the 2016-2017 and 2018-2019 school year results, six of the nine ANOVAs mirrored the original results once the outliers were removed. However, all three grade levels for the 2017-2018 year showed important changes after the removal of outliers, including the 2017-2018 third grade ANOVA results shifting from nearly statistically significant, to statistically significant. More discussion related to results with and without outliers for specific subgroups from the ANOVA results will take place with the presentation of findings section later in Chapter Four.

Laerd Statistics (2017) indicated it is reasonable to maintain the inclusion of outliers in data sets when it is believed the outliers had no material effect on the results of the study but did not indicate a protocol for varying and inconsistent effects on the study. In addition to the three methods for examining outliers mentioned above, Laerd Statistics (2017) along with Liao et al. (2016), placed a general emphasis on transparency regarding outliers when there was no valid reason to remove them along with an acknowledgement there is not a universally accepted procedure regarding outliers. The removal of outliers in general should be set aside for measurement and recording errors (Liao et al., 2016) and is typically a last resort accompanied by discussion related to their values or impact on the outcome of the study (Laerd Statistics, 2017). But it is also unwise to ignorantly keep outliers without examining the impact of keeping the outliers (Judd et al., 2017), which supported Laerd Statistics (2017) view and the researcher's effort to explore the data results with and without outliers. Utilizing the above information as a guide, the researcher kept the outliers within the data sample after examining how the inclusion or removal of outliers impacted ANOVA outcomes, determining if outliers exceed three times the IQR above the third quartile to indicated extreme outliers, and inspecting the recorded data from the LDOE for accuracy.

However, additional commentary in the following sections of Chapter Four detail results with and without outliers to provide full transparency and a better overall understanding of the data for how samples with and without outliers impacted results (Liao, et al., 2016).

The next assumption that required examination was the test of normality utilizing a Shapiro-Wilk test, as previously mentioned in Chapter Three, to examine assumption five. Upon completion of the Shapiro-Wilk test, the results showed 21 of the 27 subgroups comprising the nine ANOVA tests rejected the Shapiro-Wilk null hypothesis ($p < .05$), indicating non-normality, compared to a value to not reject the null hypothesis ($p > .05$) which would support normality (see Table 9). However, most of the subgroups either exceeded or were well under the original goal of approximately 45 data points per subgroup, so it was not surprising that the Shapiro-Wilk test failed to support normality with sample sizes either significantly less than or greater than 50 (Demir, 2022; George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). While the robust ability of an ANOVA to withstand violations of normality has been documented (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017), the researcher did not conclude the results of this study should rest on the ANOVA inherent strength against non-normality alone. Further consideration to support or not support the results of the ANOVA involved the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes, typically numbering between 30 and 40 (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), with some suggesting an acceptable as low as 25 (Bloom, 2020). With larger samples size, it is also suggested to use visual tools such as histogram distribution and Q Q-Plots to evaluate normality (Laerd Statistics, 2017; Appendices K, L and M). Additional discussion and tables related to detailed normality examination of the data can be found in Appendix N (see Tables N1-N5).

Table 9

Shapiro-Wilk Test of Normality per Curriculum Tier by Year and Grade

School Year	Grade	Tier 1 Sample Size (*n*)	Tier 1 Shapiro-Wilk (*p*)	Tier 3 Sample Size (*n*)	Tier 3 Shapiro-Wilk (*p*)	Non-tiered Sample Size (*n*)	Non-tiered Shapiro-Wilk (*p*)
2016-2017	3	274	< .001	59	.234	3	.873
	4	273	< .001	60	.022	3	.659
	5	271	< .001	59	< .001	2	-
2017-2018	3	471	< .001	92	< .001	23	.002
	4	468	< .001	92	< .001	23	.101
	5	463	< .001	90	< .001	22	.005
2018-2019	3	569	< .001	3	.503	8	.065
	4	572	< .001	4	.675	8	.037
	5	568	< .001	4	.712	2	-

Note. Shapiro-Wilk test supports normality if $p > .05$. A dash designation, (-), indicated not applicable.

Assumption six, homogeneity of variance, was the final assumption the researcher needed to examine through the data cleaning process to confirm results were valid. Laerd Statistics (2017) and Field (2013) highlighted the robustness of ANOVA in overcoming violations of variance, but also indicated the robustness diminishes when sample sizes are not relatively equal, which was the case with the data gathered for this study. To check for homogeneity of variance, the Levene's test was applied to the nine data groups for each ANOVA, the results are shown in Table 10 below.

Table 10

Levene's Test of Equality of Variance by Year and Grade

School Year	Grade	Levene's Statistic	*df1*	*df2*	Significance (*p*)
2016-2017	3	.1	2	333	.908
	4	.33	2	333	.722
	5	.45	2	329	.638
2017-2018	3	.71	2	583	.493
	4	.03	2	580	.966
	5	.33	2	572	.719
2018-2019	3	.9	2	577	.406
	4	.53	2	581	.591
	5	2.05	2	571	.13

Note. Levene's test supports equality of variance if $p > .05$.

Presentation of Findings, Tables, and Statistical Conclusions

RQ1. What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

H_o1. Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP.

The researcher examined research question one (RQ1) through the utilization of one-way ANOVAs for the three related sub-questions. To compare the differences in advanced and mastery level third grade math student achievement scores for 2016-2017 (RQ1a), 2017-2018 (RQ1b), and 2018-2019 (RQ1c) between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices, ANOVAs were conducted for each individual research sub-question. Supporting details and results of the three different one-way ANOVAs are presented in the below section, along with rejection or failure to reject the null hypotheses for 2016-2017 (H_o1a), 2017-2018 (H_o1b), and 2018-2019 (H_o1c). The indication of statistical significance for each ANOVA was set at a p value of .05, thus a $p < .05$ to reject the null hypothesis for all research questions and sub-questions. To assure validity of ANOVA results, an examination of outlier data points, normality of distribution, and homogeneity of variance were scrutinized. Outlier inspection was done by boxplots and comparison of ANOVA results with and without outliers. Normality enquiry was initiated with the Shapiro-Wilk test of normality, which suggests normal distribution if the null hypothesis is not rejected ($p > .05$), followed by additional examination of skewness and kurtosis values and histograms and Q Q-Plots. Finally, the assumption of variance was analyzed with Levene's Test for Equality of Variance which supports normality when the null hypothesis is not rejected ($p > .05$).

RQ1a: What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

H₀1a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2016-2017.

Table 11

2016-2017 Grade 3 Descriptive Statistics With and Without Outliers

Outliers Included	Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
Yes	1	274	42.28	20.07	1.21	39.9	44.67	2	100
	3	59	41.47	21.01	2.74	36	46.95	3	94
	4	3	44	26.06	15.04	-20.73	108.73	17	69
	Total	336	42.16	20.22	1.1	39.99	44.33	2	100
No	1	272	41.86	19.52	1.18	39.53	44.19	2	98
	3	59	41.47	21.01	2.74	36	46.95	3	94
	4	3	44	26.06	15.04	-20.73	108.73	17	69
	Total	334	41.81	19.78	1.08	39.68	43.94	2	98

Table 12

One-Way ANOVA of Third Grade 2016-2017 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	F	Significance (*p*)	Partial Eta Squared
Yes	Between groups	42.13	2	21.07	.05	.95	.00
	Within groups	136,944.51	333	411.25			
	Total	136,986.64	333				
No	Between groups	21.71	2	10.86	.03	.973	.00
	Within groups	130,233.4	331	393.45			
	Total	130,255.12	333				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

For RQ1a, a one-way ANOVA was conducted to determine the difference in third grade student achievement scores (based on advanced and mastery achievement percentages on the 2016-2017 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 274$), Tier 3 ($n = 59$), and non-tiered ($n = 3$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 11). There were no outliers within Tier 3 or the non-tiered data groups, but outliers were present within Tier 1 (see Table 6; Figure H1), as assessed by boxplots; data groups were normally distributed for the Tier 3 and non-tiered groups, as assessed by Shapiro-Wilk's test ($p > .05$), with Tier 1 indicating non-normality ($p < .05$; see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .908$; see Table 10). The LEAP scores increased from Tier 3 ($M = 41.47$, $SD = 21.01$) to Tier 1

(M = 42.28, SD = 20.07) to non-tiered (M = 44, S = 26.06), in that order, but the differences between these curriculum material groups were not statistically significant, $F(2, 333) = .05$, $p = .95$ (see Tables 11 and 12). The group means were not statistically significantly different ($p > .05$). Therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis.

Since all the data groups within RQ1a did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results. For data within RQ1a there were two outliers in Tier 1 data and none within Tier 3 or the non-tiered data (see Table 6; Figure H1), as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. After analysis of outlier values, the researcher concluded the outlier values were recorded accurately, contained no extreme outliers (greater than three box-lengths from the edge of the box), and did not influence the results in a substantial manner (Laerd Statistics, 2017), with ANOVA results containing outliers resulting in a significance of $p = .95$ and without outliers at a significance of $p = .973$ (see Table 12). As a result, the researcher continued investigating RQ1a with the data as the researcher originally collected it.

The LEAP scores within the RQ1a data groups were normally distributed for the Tier 3 and non-tiered groups, as assessed by Shapiro-Wilk's test ($p > .05$) but showed Tier 1 departed from normality (see Table 9). The result for Tier 1 was not surprising because the sample size with that tier was greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5) and visual examination of histograms and Q Q-Plots (see Appendix K) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 skewness and kurtosis values (absolute values) were within ranges that supported normality with skewness values for Tier 1 and Tier 3 similarly skewed. While non-tiered data was skewed differently, the low sample size of that group limited the comparison reliability to the other groups. Following visual examination, Tier 1 data was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of the histogram. Likewise, data for Tier 1 was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot (see Figure K1 and K2). Given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could be *reliably considered* for RQ1a even though distribution was not perfectly normalized, with additional backing from the above numerical evaluation methods and visual examinations to *reliably support* the RQ1a results.

RQ1b: What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

H_o1b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will

have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2017-2018.

Table 13

Descriptive Statistics for Third Grade 2017-2018 With and Without Outliers

Outliers Included	Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
Yes	1	471	40.74	19.47	0.9	38.98	42.5	4	100
	3	92	35.88	22.12	2.31	31.30	40.46	4	100
	4	23	36.13	19.78	4.12	27.58	44.68	14	100
	Total	586	39.8	19.97	0.83	38.17	41.42	4	100
No	1	467	40.24	18.77	0.87	38.53	41.94	4	97
	3	89	33.85	19.44	2.06	29.76	37.95	4	81
	4	22	33.23	14.38	3.07	26.85	39.6	14	73
	Total	578	38.99	18.88	0.79	37.44	40.53	4	97

Table 14

One-Way ANOVA of Third Grade 2017-2018 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	2,138.25	2	1069.13	2.7	.068	.009
	Within groups	231,217.17	583	396.6			
	Total	233,355.43	585				
No	Between groups	3,802.84	2	1901.42	5.42	.005	.019
	Within groups	201,755.06	575	350.88			
	Total	205,557.89	577				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

For RQ1b, a one-way ANOVA was conducted to determine the difference in third grade student achievement scores (based on advanced and mastery achievement percentages on the 2017-2018 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 471$), Tier 3 ($n = 92$), and non-tiered ($n = 23$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 13). There were outliers present within Tier 1, Tier 3, and the non-tiered group (see Table 7; Figure H2), as assessed by boxplots; data groups were not normally distributed for the Tier 1, Tier 3, and non-tiered groups, as assessed by Shapiro-Wilk's test ($p < .05$; see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .493$; see Table 10). The LEAP scores increased from Tier 3 ($M = 35.88$, $SD = 22.12$) to Tier 4 ($M = 36.13$, $SD = 19.78$) to Tier 1 ($M = 40.74$, $SD = 19.47$), in that order, but the differences between these curriculum material groups were not statistically significant, $F(2, 583) = 2.696$, $p = .068$ (see Tables 13 and 14). The group means were not statistically significantly different ($p > .05$). Therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis.

Since all the data groups within RQ1b did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results. For data within RQ1b there were four outliers in Tier 1, three outliers for Tier 3, and one outlier for the non-tiered group (see Table 7; Figure H2), as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. After analysis of outlier values, the researcher concluded the outlier values were recorded accurately, contained no extreme outliers (greater than three box-lengths from the edge of the box), but did influence the results in a substantial manner (Laerd Statistics, 2017), with ANOVA results containing outliers resulting in a significance of $p = .068$ and without outliers at a significance of $p = .005$ (see Table 14). As a result, the researcher continued investigating RQ1b with the data as the researcher originally collected it but also repeated all assumption tests without outliers since the ANOVA results with outliers were nearly statistically significant and the results without outliers were statistically significant. More discussion related to the RQ1b data without outliers will take place after the presentation of the data with outliers.

The LEAP scores within the RQ1b data groups were non-normally distributed for the Tier 1, Tier 3, and non-tiered groups, as assessed by Shapiro-Wilk's test ($p < .05$; see Table 9). The result for Tier 1 and Tier 3 were not surprising because the sample size with those tiers were greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 and Tier 3, along with the non-tiered group, included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix K), with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 and Tier 3 skewness and kurtosis values (absolute values) were within ranges that supported normality, which was not the case for the non-tiered group; but skewness values for Tier 1, Tier 3, and non-tiered data groups were similarly skewed. Following the visual examination, Tier 1 and Tier 3 were not perfectly normal, but Tier 1 was approximately normally distributed, with Tier 3 moderately reflecting normality, as assessed by visual inspection of their histograms, with the non-tiered group not indicating normality. Likewise, data for Tier 1 and Tier 3 were not perfectly normal, but Tier 1 was approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot, with Tier 3 moderately reflecting normality, with the non-tiered group not indicating normality (see Figures K3-K8). Given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could be *cautiously considered* for RQ1b even though distribution was not perfectly normalized, with additional backing from the above numerical evaluation methods and visual examinations to *cautiously support* the RQ1b results.

As mentioned above, the outliers within RQ1b did impact the results in a substantial manner, which prompted the researcher to examine and share ANOVA and assumption results that did not include the outliers to ensure validity and transparency of the results (Laerd Statistics, 2017; Liao, et al., 2016). Additionally, the researcher reexamined Tier 1 and Tier 3 results with outliers, excluding the non-normalized non-tiered group, with an independent *t*-test, which is the traditional test when only two independent variables are utilized (Laerd Statistics, 2017). Since Tier 1 and Tier 3 data met the Central Limit Theorem threshold and other assumption criteria for approximately normal distribution in addition to the other ANOVA assumptions, which mirror the required assumptions for the independent *t*-test (Laerd Statistics, 2017; Pelham, 2013), the transition to independent *t*-test analysis presented a swift opportunity to gain additional insight into the relationship between curriculum material choice and student achievement.

For the independent *t*-test, including outliers, there were 471 Tier 1 scores and 92 Tier 3 scores (see Table 13). An independent-samples *t*-test was run to determine if there were differences in student achievement between Tier 1 and Tier 3 curriculum materials. There were outliers present within Tier 1 and Tier 3 (see Table 7; Figure H2), as assessed by inspection of boxplots. Student achievement scores for both Tier 1 and Tier 3 were not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$; see Table 9) but had acceptable normality for analysis through the criteria outlined above and support from Appendix N, and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .348$; see Appendix O). The student achievement scores were greater in Tier 1 ($M = 40.74$, $SD = 19.47$) than Tier 3 scores ($M = 35.88$, $SD = 22.12$), a statistically significant difference, M = 4.86, 95% CI [0.39, 9.32], $t(561) = 2.14$, $p = .033$, *d* = .24 (see Appendix O). There was a statistically significant difference between means ($p < .05$), and therefore, we can reject the null hypothesis and accept the alternative hypothesis.

Once RQ1b outliers were removed, a one-way ANOVA was conducted to determine the difference in third grade student achievement scores (based on advanced and mastery achievement percentages on the 2017-2018 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 467$), Tier 3 ($n = 89$), and non-tiered ($n = 22$), with non-tiered designated as Tier 4 when numerical association was needed. There were no outliers within Tier 1 and Tier 3, but one new outlier developed within the non-tiered group (see Appendix P), as assessed by boxplots; the non-tiered group was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), with Tier 1 and Tier 3 indicating non-normality ($p < .05$) but revealing approximately normal distribution through other normality methods (see Appendix Q) with additional support from Appendix N; and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .233$; see Appendix Q). Data is presented as mean ± standard deviation. Scores from the LEAP were statistically significantly different between different curriculum material groups, $F(2, 575) = 5.42$, $p = .005$, $\eta_p^2 = .019$ (see Appendix R). The LEAP scores increased from non-tiered ($M = 33.23$, $SD = 14.38$) to Tier 3 ($M = 33.85$, $SD = 19.44$) to Tier 1 ($M = 40.24$, $SD = 18.77$), in that order. Tukey post hoc analysis revealed that the mean increase from Tier 3 to Tier 1 (6.38, 95% CI [1.29, 11.47]) was statistically significant ($p = .009$), but no other group differences were statistically significant (see Appendix R). The group means were statistically significantly different ($p < .05$). Therefore, we can reject the null hypothesis and accept the alternative hypothesis.

RQ1c: What are the differences in advanced and mastery level third grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

H_o1c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2018-2019.

Table 15

Descriptive Statistics for Third Grade 2018-2019 With and Without Outliers

Outliers Included	Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
Yes	1	569	41.04	20.02	0.82	39.39	42.69	2	101
	3	3	45	13.45	7.77	11.58	78.42	34	60
	4	8	45.5	26.38	9.33	23.45	67.55	22	97
	Total	580	41.12	20.07	0.83	39.48	42.76	2	101
No	1	561	40.21	18.91	0.8	38.64	41.77	2	95
	3	3	45	13.45	7.77	11.58	78.42	34	60
	4	8	45.5	26.38	9.33	23.45	67.55	22	97
	Total	572	40.31	19.98	0.79	38.75	41.86	2	97

Table 16

One-Way ANOVA of Third Grade 2018-2019 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	202.57	2	101.28	.25	.778	.001
	Within groups	232,924.23	577	403.68			
	Total	233,126.79	579				
No	Between groups	287.45	2	143.72	.4	.672	.001
	Within groups	205,418.01	569	361.02			
	Total	205,705.46	571				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

For RQ1c, a one-way ANOVA was conducted to determine the difference in third grade student achievement scores (based on advanced and mastery achievement percentages on the 2018-2019 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 569$), Tier 3 ($n = 3$), and non-tiered ($n = 8$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 15). There were no outliers within Tier 3, or the non-tiered data groups, but outliers were present within Tier 1 (see Table 8; Figure H3), as assessed by boxplots; data groups were normally distributed for the Tier 3 and non-tiered groups, as assessed by Shapiro-Wilk's test ($p > .05$), with Tier 1 indicating non-normality ($p < .05$; see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .406$; see Table 10). Scores from the LEAP increased from Tier 1 ($M = 41.04$, $SD = 20.02$) to Tier 3

(M = 45, SD = 13.45) to non-tiered (M = 45.5, SD = 26.38), in that order, but the differences between these curriculum material groups were not statistically significant, $F(2, 577) = .25$, $p = .778$ (see Tables 15 and 16). The group means were not statistically significantly different ($p > .05$). Therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis.

Since all the data groups within RQ1c did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results. For data within RQ1c there were eight outliers in Tier 1 data and none within Tier 3 or the non-tiered data (see Table 8; Figure H3), as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. After analysis of outlier values, the researcher concluded the outlier values were recorded accurately, contained no extreme outliers (greater than three box-lengths from the edge of the box), and did not influence the results in a substantial manner (Laerd Statistics, 2017), with ANOVA results containing outliers resulting in a significance of $p = .778$ and without outliers at a significance of $p = .672$ (see Table 16). As a result, the researcher continued investigating RQ1c with the data as the researcher originally collected it.

The LEAP scores within the RQ1c data groups were normally distributed for the Tier 3 and non-tiered groups, as assessed by Shapiro-Wilk's test ($p > .05$), but showed Tier 1 departed from normality (see Table 9). The result for Tier 1 was not surprising because the sample size with that tier was greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix K) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 skewness and kurtosis values (absolute values) were within ranges that supported normality with skewness values for Tier 1, Tier 3, and non-tiered data groups similarly skewed, although the low sample size of the Tier 3 and non-tiered groups limited the comparison reliability. Following visual examination, Tier 1 data was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of the histograms. Likewise, data for Tier 1 was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot (see Figures K9-K10). Despite the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could *not even be cautiously considered* for RQ1c with approximately normal distribution for Tier 1 and additional backing from the above numerical evaluation methods and visual examinations, because small sample sizes for Tier 3 and the non-tiered group *prevent even cautious support* for RQ1c results.

RQ2. What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

H_o2. Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP.

The researcher examined research question two (RQ2) through the utilization of one-way ANOVAs for the three related sub-questions. To compare the differences in advanced and mastery level fourth grade math student achievement scores for 2016-2017 (RQ2a), 2017-2018 (RQ2b), and 2018-2019 (RQ2c) between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices, ANOVAs were conducted for each individual research sub-question. Supporting details and results of the three different one-way ANOVAs are presented in the below section, along with rejection or failure to reject the null hypotheses for 2016-2017 (H_o2a), 2017-2018 (H_o2b), and 2018-2019 (H_o2c). The indication of statistical significance for each ANOVA was set at a *p* value of .05, thus a $p < .05$ to reject the null hypothesis for all research questions and sub-questions. To assure validity of ANOVA results, an examination of outlier data points, normality of distribution, and homogeneity of variance were scrutinized. Outlier inspection was done by boxplots and comparison of ANOVA results with and without outliers. Normality enquiry was initiated with the Shapiro-Wilk test of normality, which suggests normal distribution if the null hypothesis is not rejected ($p > .05$), followed by additional examination of skewness and kurtosis values and histograms and Q Q-Plots. Finally, the assumption of variance was analyzed with Levene's Test for Equality of Variance which supports normality when the null hypothesis is not rejected ($p > .05$).

RQ2a: What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

H_o2a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2016-2017.

Table 17

Descriptive Statistics for Fourth Grade 2016-2017

Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
1	273	37.82	21.54	1.30	35.25	40.38	2	98
3	60	37.42	22.33	2.88	31.65	43.19	3	93
4	3	45.33	32.52	18.77	-35.44	126.11	10	74
Total	336	37.81	21.71	1.18	35.48	40.14	2	96

Table 18

One-Way ANOVA of Fourth Grade 2016-2017 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	179.1	2	89.55	.19	.828	.001
	Within groups	157,710.09	333	473.6			
	Total	157,889.19	335				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

For RQ2a, a one-way ANOVA was conducted to determine the difference in fourth grade student achievement scores (based on advanced and mastery achievement percentages on the 2016-2017 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 274$), Tier 3 ($n = 60$), and non-tiered ($n = 3$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 17). There were no outliers within Tier 3 (see Table 6; Figure I1), as assessed by boxplots; data groups were normally distributed for the non-tiered group, as assessed by Shapiro-Wilk's test ($p > .05$), with Tier 1 and Tier 3 indicating non-normality ($p < .05$; see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .722$; see Table 10). Scores from the LEAP increased from Tier 3 ($M = 37.42$, $SD = 22.331$) to Tier 1 ($M = 37.82$, $SD = 21.54$) to non-tiered ($M = 45.33$, $SD = 32.52$), in that order, but the differences between these curriculum material groups were not statistically significant, $F(2, 333) = .19$, $p = .828$ (see Tables 17 and 18). The group means were not statistically significantly different ($p > .05$). Therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis. Since all the data groups within RQ2a did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results.

The LEAP scores within the RQ2a data groups were normally distributed for the non-tiered group, as assessed by Shapiro-Wilk's test ($p > .05$) but showed Tier 1 and Tier 3 departed from normality (see Table 9). The result for Tier 1 and Tier 3 were not surprising because the sample sizes with those tiers were greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 and Tier 3 included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix L) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 and Tier 3 skewness and kurtosis values (absolute values) were within ranges that supported normality with skewness values for Tier 1 and Tier 3 similarly skewed. While non-tiered data was skewed differently, the low sample size of that group limited the comparison reliability to the other groups. Following the visual examination, Tier 1 and Tier 3 were not perfectly normal, but Tier 1 was approximately normally distributed, with Tier 3 moderately reflecting normality, as assessed by visual inspection of their histograms. Likewise, data for Tier 1 and Tier 3 were not perfectly normal, but Tier 1 was approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot, with Tier 3 moderately reflecting normality (see Figure L1-L4). Given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could be *reliably considered* for RQ2a even though distribution was not perfectly normalized, with additional backing from the above numerical evaluation methods and visual examinations to *reliably support* the RQ2a results.

RQ2b: What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

H_o2b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2017-2018.

Table 19

Descriptive Statistics for Fourth Grade 2017-2018 With and Without Outliers

Outliers Included	Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
Yes	1	468	37.67	20.65	0.95	35.79	39.54	2	100
	3	92	32.67	21.37	2.23	28.25	37.1	2	90
	4	23	30.91	21.84	4.55	21.47	40.36	2	90
	Total	583	36.61	20.88	0.87	34.92	38.31	2	100
No	1	465	37.27	20.1	0.93	35.44	39.1	2	97
	3	92	32.67	21.37	2.23	28.25	37.1	2	90
	4	22	28.23	18.05	3.85	20.22	36.23	2	69
	Total	579	36.2	20.33	0.85	34.54	37.85	2	97

Table 20

One-Way ANOVA of Fourth Grade 2017-2018 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	2,696.46	2	1,348.23	3.11	.045	.011
	Within groups	251,111.708	580	432.95			
	Total	253,808.165	582				
No	Between groups	3,073.47	2	1,536.73	3.75	.024	.013
	Within groups	235,855.48	576	409.47			
	Total	238,928.95	578				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

Table 21

Tukey Multiple Comparison Post Hoc Analysis of Fourth Grade 2017-2018 Data

Outliers Included	(I) Tier	(J) Tier	Mean Difference (I-J)	*SE*	Significance (*p*)	95% *LL*	95% *UL*
Yes	1	3	4.99	2.37	.090	-0.58	10.57
		4	6.76	4.44	.282	-3.69	17.20
	3	1	-4.99	2.37	.09	-10.57	0.58
		4	1.76	4.85	.930	-9.64	13.16
	4	1	-6.76	4.44	.282	-17.20	3.69
		3	-1.76	4.85	.930	-13.16	9.64
No	1	3	4.6	2.31	.116	-0.83	10.02
		4	9.04	4.42	.102	-1.33	19.42
	3	1	-4.56	2.31	.116	-10.02	0.83
		4	4.45	4.8	.624	-6.84	15.73
	4	1	-9.04	4.42	.102	-19.42	1.33
		3	-4.45	4.8	.624	-15.73	6.84

Note. Tukey reflects statistical significance when $p < .05$ between groups.

For RQ2b, a one-way ANOVA was conducted to determine the difference in fourth grade student achievement scores (based on advanced and mastery achievement percentages on the 2017-2018 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 468$), Tier 3 ($n = 92$), and non-tiered ($n = 23$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 19). There were no outliers within Tier 3, but Tier 1 and the non-tiered group contained outliers (see Table 7; Figure I2), as assessed by boxplots; the non-tiered group was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), with Tier 1 and Tier 3 indicating non-normality ($p < .05$; see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .966$; see Table 10). Data is presented as mean ± standard deviation. Scores from the LEAP were statistically significantly different between different curriculum material groups, $F(2, 580) = 3.11$, $p = .045$, $\eta_p^2 = .011$ (see Table 20). The LEAP scores increased from non-tiered ($M = 30.91$, $SD = 21.84$) to Tier 3 ($M = 32.67$, $SD = 21.37$) and Tier 1 ($M = 37.67$, $SD = 20.65$), in that order (see Table 19). Tukey post hoc analysis revealed no group differences were statistically significant (see Tables 21). The group means were statistically significantly different ($p < .05$). Therefore, we can reject the null hypothesis and accept the alternative hypothesis. To support the validity of these results, assumptions for the results for the ANOVA were assessed.

Since all the data groups within RQ2b did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results. For data within RQ2b there were three outliers in Tier 1 and one in the non-tiered data (see Table 7; Figure I2), as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. After analysis of outlier values, the researcher concluded the outlier values were recorded accurately, contained no extreme outliers (greater than three box-lengths from the edge of the box), and did not influence the results in a substantial manner (Laerd Statistics, 2017), with ANOVA results containing outliers resulting in a significance of $p = .045$ and without outliers at a significance of $p = .024$ (see

Table 21). As a result, the researcher continued investigating RQ2b with the data as the researcher originally collected it.

The LEAP scores within the RQ2b data groups were normally distributed for the non-tiered groups, as assessed by Shapiro-Wilk's test ($p > .05$), but showed Tier 1 and Tier 3 departed from normality (see Table 9). The result for Tier 1 and Tier 3 were not surprising because the sample size with those tiers were greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 and Tier 3 included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix L) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 and Tier 3 skewness and kurtosis values (absolute values) were within ranges that supported normality with skewness values for Tier 1 and Tier 3, and non-tiered data groups similarly skewed. Following the visual examination, Tier 1 and Tier 3 data were not perfectly normal, but Tier 1 was approximately normally distributed, as assessed by visual inspection of their histograms, with Tier 3 data not reflecting normality. Likewise, data for Tier 1 and Tier 3 were not perfectly normal, but Tier 1 was approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot, with Tier 3 moderately reflecting normality (see Figures L5-L8). Given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could be *considered with slight caution* for RQ2b even though distribution was not perfectly normalized, with additional backing from the above numerical evaluation methods and visual examinations to *support with slight caution* the RQ2b results.

As mentioned above, the outliers within RQ2b did not impact the results in a substantial manner and results with and without outliers were shared to ensure validity and transparency of the results (Laerd Statistics, 2017; Liao, et al., 2016). However, slight concern related to the low sample size of the non-tiered group, consequently not meeting the Central Limit Theorem threshold, while Tier 1 and Tier 3 data met the threshold with much larger sample sizes, prompted the researcher to reexamine Tier 1 and Tier 3 with an independent *t*-test. This revised approach for analysis presented a timely opportunity to gain additional insight into the relationship between curriculum material choice and student achievement through an independent t-test, the traditional test when only two independent variables are utilized, that also mirrors the required ANOVA assumptions (Laerd Statistics, 2017; Pelham, 2013).

For the independent *t*-test, including outliers, there were 468 Tier 1 scores and 92 Tier 3 scores (see Table 19). An independent-samples *t*-test was run to determine if there were differences in student achievement between Tier 1 and Tier 3 curriculum materials. There were outliers present within Tier 1 and Tier 3 (see Table 7; Figure I2), as assessed by inspection of boxplots. Student achievement scores for both Tier 1 and Tier 3 were not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$; see Table 9) but had acceptable normality for analysis through the criteria outlined above and support from Appendix N, and there was homogeneity of variances, as assessed by Levene's test for equality of variances (*p* = .966; see Appendix S). The student achievement scores were greater in Tier 1 (*M* = 37.67, *SD* = 20.65) than Tier 3 scores (*M* = 32.67, *SD* = 21.37), a statistically significant difference, M = 4.99, 95% CI [0.34, 9.65], *t*(558) = 2.11, *p* = .035, *d* = .24 (see Appendix S). There was a statistically significant difference between means ($p < .05$), and therefore, we can *with slight caution* reject the null hypothesis and accept the alternative hypothesis.

RQ2c: What are the differences in advanced and mastery level fourth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

H_o2c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2018-2019.

Table 22

Descriptive Statistics for Fourth Grade 2018-2019

Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
1	572	38.7	21.63	.9	36.93	40.48	2	100
3	4	49	15.98	7.99	23.57	74.43	33	68
4	8	39.38	27.21	9.62	16.63	62.12	16	94
Total	584	38.78	21.66	0.9	37.02	40.54	2	100

Table 23

One-Way ANOVA of Fourth Grade 2018-2019 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	4,24.03	2	212.02	.45	.637	.002
	Within groups	273,067.35	581	469.99			
	Total	273,491.38	583				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

For RQ2c, a one-way ANOVA was conducted to determine the difference in fourth grade student achievement scores (based on advanced and mastery achievement percentages on the 2018-2019 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 572$), Tier 3 ($n = 4$), and non-tiered ($n = 8$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 22). There were no outliers (see Table 8; Figure I3), as assessed by boxplots; data groups were normally distributed for Tier 3, as assessed by Shapiro-Wilk's test ($p > .05$), with Tier 1 and the non-tiered group indicating non-normality ($p < .05$; see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .591$; see Table 10). Scores from the LEAP increased from Tier 1 ($M = 38.7$, $SD = 21.63$) to the non-tiered group ($M = 39.38$, $SD = 27.21$) to Tier 3 ($M = 49$, $SD = 15.98$), in that order, but the differences between these curriculum material groups were not statistically significant, $F(2, 581) = .451$, $p = .637$ (see Tables 22 and 23). The group means were not statistically significantly different ($p > .05$). Therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis. Since all the data groups within RQ2c did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results.

The LEAP scores within the RQ2c data groups were normally distributed for Tier 3, as assessed by Shapiro-Wilk's test ($p > .05$) but showed Tier 1 and the non-tiered group departed from normality (see Table 9). The result for Tier 1 was not surprising because the sample size with that tier was greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 and the non-tiered group included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix L) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 and non-tiered skewness and kurtosis values (absolute values) were within ranges that supported normality with skewness values for Tier 1, Tier 3, and non-tiered data groups similarly skewed, although the low sample size of the Tier 3 and non-tiered groups limited the comparison reliability. Following visual examination, Tier 1 data was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of their histograms, with the small sample size for the non-tiered group making it impractical to assess normality with a histogram. Likewise, data for Tier 1 was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot, with the small sample size for the non-tiered group making it impractical to assess normality with a Q Q-Plot (see Figures L9-L12). Despite the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results *could not even be cautiously considered* for RQ2c with approximately normal distribution for Tier 1 and additional backing from the above numerical evaluation methods and visual examinations, because small sample sizes for Tier 3 and the non-tiered group *prevent even cautious support* for RQ2c results.

RQ3. What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices?

H_o3. Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP.

The researcher examined research question three (RQ3) through the utilization of one-way ANOVAs for the three related sub-questions. To compare the differences in advanced and mastery level fifth grade math student achievement scores for 2016-2017 (RQ3a), 2017-2018 (RQ3b), and 2018-2019 (RQ3c) between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices, ANOVAs were conducted for each individual research sub-question. Supporting details and results of the three different one-way ANOVAs are presented in the below section, along with rejection or failure to reject the null hypotheses for 2016-2017 (H_o3a), 2017-2018 (H_o3b), and 2018-2019 (H_o3c). The indication of statistical significance for each ANOVA was set at a p value of .05, thus a $p < .05$ to reject the null hypothesis for all research questions and sub-questions. To assure validity of ANOVA results, an examination of outlier data points, normality of distribution, and homogeneity of variance were scrutinized. Outlier inspection was done by boxplots and comparison of ANOVA results with and without outliers. Normality enquiry was initiated with the Shapiro-Wilk test of normality, which suggests normal distribution if the null hypothesis is not rejected ($p > .05$), followed by additional examination of skewness and kurtosis values and histograms and Q Q-Plots. Finally, the assumption of variance was analyzed with Levene's Test for Equality of Variance which supports normality when the null hypothesis is not rejected ($p > .05$).

RQ3a: What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2016-2017?

H_o3a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2016-2017.

Table 24

Descriptive Statistics for Fifth Grade 2016-2017 With and Without Outliers

Outliers Included	Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
Yes	1	271	31.18	20.65	1.26	28.71	33.65	2	100
	3	59	28.83	20.5	2.67	23.49	34.17	2	91
	4	2	19.5	16.26	11.5	-126.62	165.62	8	31
	Total	332	30.7	20.59	1.13	28.47	32.92	2	100
No	1	266	29.99	18.88	1.16	27.71	32.27	2	83
	3	57	26.72	17.35	2.3	22.11	31.32	2	75
	4	2	19.5	16.24	11.5	-126.62	165.62	8	31
	Total	325	29.35	18.62	1.03	27.32	31.38	2	83

Table 25

One-Way ANOVA of Fifth Grade 2016-2017 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	520.69	2	260.35	.61	.543	.004
	Within groups	139,803.6	329	424.94			
	Total	140,324.27	331				
No	Between groups	697.04	2	348.52	1	.367	.006
	Within groups	111,570.98	322	346.49			
	Total	112,268.01	324				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

For RQ3a, a one-way ANOVA was conducted to determine the difference in fifth grade student achievement scores (based on advanced and mastery achievement percentages on the 2016-2017 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 271$), Tier 3 ($n = 59$), and non-tiered ($n = 2$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 24). There were no outliers within the non-tiered group but Tier 1 and Tier 3 did have outliers (see Table 6; Figure J1), as assessed by boxplots; data for Tier 1 and Tier 3 were not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$), with the non-tiered group containing a sample size of two, which was too small for calculation (see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .638$; see Table 10). Scores from the LEAP increased from non-tiered

($M = 19.5$, $SD = 16.26$) to Tier 3 ($M = 28.83$, $SD = 20.5$) to Tier 1 ($M = 31.18$, $SD = 20.65$), in that order, but the differences between these curriculum material groups were not statistically significant, $F(2, 329) = .61$, $p = .54$ (see Tables 24 and 25). The group means were not statistically significantly different ($p > .05$). Therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis.

Since all the data groups within RQ3a did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results. For data within RQ3a there were five outliers in Tier 1, two within Tier 3, and none for the non-tiered data (see Table 6; Figure J1), as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. After analysis of outlier values, the researcher concluded the outlier values were recorded accurately, contained no extreme outliers (greater than three box-lengths from the edge of the box), and did not influence the results in a substantial manner (Laerd Statistics, 2017), with ANOVA results containing outliers resulting in a significance of $p = .543$ and without outliers at a significance of $p = .367$ (see Table 25). As a result, the researcher continued investigating RQ3a with the data as the researcher originally collected it.

The LEAP scores within the RQ3a data groups were not normally distributed for Tier 1 and Tier 3, as assessed by Shapiro-Wilk's test ($p < .05$), with the non-tiered group containing a sample size of two, which was too small for calculation (see Table 9). The result for Tier 1 was not surprising because the sample size with that tier was greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017).

Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 and Tier 3 included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix M) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 and Tier 3 skewness and kurtosis values (absolute values) were within ranges that supported normality with skewness values for Tier 1 and Tier 3 similarly skewed. The non-tiered data was not available for comparison because of that group's low sample size. Following the visual examination, Tier 1 and Tier 3 data were not perfectly normal, but Tier 1 was approximately normally distributed, as assessed by visual inspection of their histograms, with Tier 3 not approximately normal but not entirely non-normal. Likewise, data for Tier 1 and Tier 3 were not perfectly normal, but Tier 1 was approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot, with Tier 3 moderately reflecting normality (see Figures M1-M4). Given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could be *cautiously considered* for RQ3a even though distribution was not perfectly normalized, with additional backing from the above numerical evaluation methods and visual examinations to *cautiously support* the RQ3a results.

RQ3b: What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2017-2018?

Ho3b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2017-2018.

Table 26

Descriptive Statistics for Fifth Grade 2017-2018 With and Without Outliers

Outliers Included	Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
Yes	1	463	29.4	18.95	0.88	27.67	31.13	2	100
	3	90	25.07	19.7	2.08	20.94	29.19	2	88
	4	22	21.41	17.99	3.84	13.43	29.39	2	78
	Total	575	28.42	19.12	0.8	26.85	29.98	2	100
No	1	454	28.2	17.09	0.8	26.63	29.78	2	77
	3	87	22.97	16.36	1.75	19.48	26.45	2	76
	4	21	18.71	13.13	2.87	12.74	24.69	2	46
	Total	562	27.04	17	0.72	26.63	28.45	2	77

Table 27

One-Way ANOVA of Fifth Grade 2017-2018 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	2,537.66	2	1,268.83	3.5	.031	.012
	Within groups	207,188	572	362.217			
	Total	209,725.66	574				
No	Between groups	3,516.01	2	1758	6.19	.002	.022
	Within groups	559	559	283.9			
	Total	561	561				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

Table 28

Tukey Multiple Comparison Post Hoc Analysis of Fifth Grade 2017-2018 Data

Outliers Included	(I) Tier	(J) Tier	Mean Difference (I-J)	*SE*	Significance (*p*)	95% *LL*	95% *UL*
Yes	1	3	4.33	2.19	.119	-0.82	9.48
		4	7.99	4.15	.133	-1.77	17.75
	3	1	-4.33	2.19	.119	-9.48	0.82
		4	3.66	4.53	.698	-6.98	14.29
	4	1	-7.99	4.15	.133	-17.75	1.77
		3	-3.66	4.53	.698	-14.29	6.98
No	1	3	5.24	1.97	.022	0.61	9.87
		4	9.49	3.76	.032	0.65	18.33
	3	1	-5.24	1.97	.022	-9.87	-0.61
		4	4.25	4.1	.553	-5.38	13.88
	4	1	-9.49	3.76	.032	-18.33	-0.65
		3	-4.25	4.1	.553	-13.88	5.38

Note. Tukey reflects statistical significance when $p < .05$ between groups.

For RQ3b, a one-way ANOVA was conducted to determine the difference in fifth grade student achievement scores (based on advanced and mastery achievement percentages on the 2017-2018 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 468$), Tier 3 ($n = 92$), and non-tiered ($n = 23$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 26). There were outliers present within Tier 1, Tier 3, and the non-tiered group (see Table 7; Figure J2), as assessed by boxplots; data groups were not normally distributed for Tier 1, Tier 3, and the non-tiered group, as assessed by Shapiro-Wilk's test ($p < .05$; see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .719$; see Table 10). Data is presented as mean ± standard deviation. Scores from the LEAP were statistically significantly different between different curriculum material groups, $F(2, 572) = 3.5$, $p = .031$, $\eta_p^2 = .012$ (see Table 27). The LEAP scores increased from non-tiered ($M = 21.41$, $SD = 17.99$) to Tier 3 ($M = 25.07$, $SD = 19.7$) and Tier 1 ($M = 29.4$, $SD = 18.95$), in that order (see Table 26). Tukey post hoc analysis revealed no group differences were statistically significant (see Table 28). The group means were statistically significantly different ($p < .05$). Therefore, we can reject the null hypothesis and accept the alternative hypothesis. To support the validity of these results, assumptions for the results for the ANOVA were assessed.

Since all the data groups within RQ3b did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results. For data within RQ3b there were nine outliers in Tier 1, three in Tier 3, and one in the non-tiered data group (see Table 7; Figure J2), as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. After analysis of outlier values, the researcher concluded the outlier values were recorded accurately, contained no extreme outliers (greater than three box-lengths from the edge of the box), and did influence the results in a substantial manner (Laerd Statistics, 2017), with ANOVA results containing outliers resulting in a significance of $p = .031$ and without outliers at a significance of

$p = .002$ (see Table 27). As a result, the researcher continued investigating RQ3b with the data as the researcher originally collected it but also repeated all assumption tests without outliers since the ANOVA results were statistically significant with outliers, $p = .031$, and the results without outliers were statistically significant at the greater level of $p = .002$. More discussion related to the RQ3b data without outliers will take place after the presentation of the data with outliers.

The LEAP scores within the RQ3b data groups were non-normally distributed for Tier 1, Tier 3, and the non-tiered groups, as assessed by Shapiro-Wilk's test ($p < .05$; see Table 9). The results for Tier 1 and Tier 3 were not surprising because the sample sizes with those tiers were greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 and Tier 3, along with the non-tiered group, included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix M) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 and Tier 3 skewness and kurtosis values (absolute values) were within ranges that supported normality, but were not in the non-tiered group, with skewness values for Tier 1, Tier 3, and non-tiered data groups similarly skewed. Following the visual examination, Tier 1 was not perfectly normal, but did appear approximately normally distributed, Tier 3 data was not approximately normal but not entirely non-normal, and the non-tiered group did not appear to reflect normality, as assessed by visual inspection of their histograms. Likewise, collected data for Tier 1 was not perfectly normal, but did appear approximately normally distributed, Tier 3 data was not approximately normal but not entirely non-normal, and the non-tiered group did not appear to reflect normality, as assessed by visual inspection of a Normal Q-Q Plot (see Figures M5-M10). Given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could be *cautiously considered* for RQ3b even though distribution was not perfectly normalized, with additional backing from the above numerical evaluation methods and visual examinations to *cautiously support* the RQ3b results.

As mentioned above, the outliers within RQ3b did impact the results in a substantial manner, which prompted the researcher to examine and share ANOVA and assumption results that did not include the outliers to ensure validity and transparency of the results (Laerd Statistics, 2017; Liao, et al., 2016). Additionally, the researcher reexamined Tier 1 and Tier 3 results with outliers, excluding the non-normalized non-tiered group, with an independent *t*-test, which is the traditional test when only two independent variables are utilized (Laerd Statistics, 2017). While Tier 1 and Tier 3 data did not meet the assumption criteria for perfectly normal distribution, it was not entirely non-normally distributed and met the Central Limit Theorem threshold, along with the other ANOVA assumptions reliably, which mirror the required assumptions for the independent *t*-test (Laerd Statistics, 2017; Pelham, 2013), thus a transition to an independent *t*-test analysis for Tier 1 and Tier 3 presented a swift opportunity to gain additional insight into the relationship between curriculum material choice and student achievement.

For the independent *t*-test, including outliers, there were 463 Tier 1 scores and 90 Tier 3 scores (see Table 26). An independent-samples *t*-test was run to determine if there were differences in student achievement between Tier 1 and Tier 3 curriculum materials. There were outliers present within Tier 1 and Tier 3 (see Table 7; Figure J2), as assessed by inspection of boxplots. Student achievement scores for both tiers were not normally distributed, as assessed by Shapiro-Wilk's test ($p < .05$; see Table 9) but had acceptable normality for analysis through the criteria outlined above and support from Appendix N, and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .793$; see Appendix T). The student achievement scores were greater in Tier 1 ($M = 29.44$, $SD = 18.95$) than Tier 3 scores ($M = 25.07$, $SD = 25.07$), a statistically significant difference, M = 4.33, 95% CI [0.02, 8.65], $t(551) = 1.97$, $p = .049$, $d = .23$ (see Appendix T). There was a statistically significant difference between means ($p < .05$), and therefore, we can reject the null hypothesis and accept the alternative hypothesis with only *slight caution*.

Once RQ3b outliers were removed, a one-way ANOVA was conducted to determine the difference in fifth grade student achievement scores (based on advanced and mastery achievement percentages on the 2017-2018 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 454$), Tier 3 ($n = 87$), and non-tiered ($n = 21$), with non-tiered designated as Tier 4 when numerical association was needed. There were no outliers within Tier 1 and the non-tiered group, but one new outlier developed for Tier 3 (see Appendix U), as assessed by boxplots; the non-tiered group was normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$) with Tier 1 and Tier 3, indicating non-normality ($p < .05$) but revealing approximately normal distribution through other normality methods once the original outliers were removed (see Appendix V) with additional support from Appendix N; and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances (p = .291; see Appendix V). Data is presented as mean ± standard deviation. Scores from the LEAP were statistically significantly different between different curriculum material groups, $F(2, 559) = 6.19$, $p = .002$, $\eta_p^2 = .022$ (see Appendix W). The LEAP scores increased from non-tiered (*M* = 18.71, *SD* = 13.13) to Tier 3 (*M* = 22.97, *SD* = 16.36) to Tier 1 (*M* = 28.2, *SD* = 17.09), in that order. Tukey post hoc analysis revealed that the mean increase from Tier 3 to Tier 1 (5.24, 95% CI [0.61, 9.9]) was statistically significant (p = .022), and non-tiered (Tier 4) to Tier 1 (9.49, 95% CI [0.65, 18.33]) was statistically significant (p = .032), but no other group differences were statistically significant (see Appendix W). The group means were statistically significantly different ($p < .05$). Therefore, we can reject the null hypothesis and accept the alternative hypothesis.

RQ3c: What are the differences in advanced and mastery level fifth grade math student achievement in the state of Louisiana based on LEAP 2025 scores between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for 2018-2019?

H_o3c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2018-2019.

Table 29

Descriptive Statistics for Fifth Grade 2018-2019 With and Without Outliers

Outliers Included	Tier	*n*	Mean	*SD*	*SE*	95% *LL*	95% *UL*	Min.	Max.
Yes	1	568	32.7	20.72	0.87	30.99	34.4	2	100
	3	4	48.5	26.74	13.37	5.95	91.05	22	85
	4	2	58	48.08	34	-374.01	490.01	24	92
	Total	574	32.89	20.89	0.87	31.18	34.61	2	100
No	1	559	31.7	19.33	0.82	30.1	33.31	2	88
	3	4	48.5	26.74	13.37	5.95	91.05	22	85
	4	2	58	48.08	34	-374.01	490.01	24	92
	Total	565	31.92	19.55	0.82	30.3	33.53	2	92

Table 30

One-Way ANOVA of Fifth Grade 2018-2019 Achievement by Outlier Inclusion

Outliers Included	Source	Sum of Squares	*df*	Mean of Squares	*F*	Significance (*p*)	Partial Eta Squared
Yes	Between groups	2,257.21	2	1,128.61	2.6	.075	.009
	Within groups	24,867.31	571	434.09			
	Total	250,124.52	573				
No	Between groups	2,486.22	2	1,243.12	3.28	.038	.012
	Within groups	213,025.71	562	379.05			
	Total	215,511.92	564				

Note. ANOVA reflects statistically significant results and rejects null hypothesis if $p < .05$.

For RQ3c, a one-way ANOVA was conducted to determine the difference in fifth grade student achievement scores (based on advanced and mastery achievement percentages on the 2018-2019 LEAP) for schools with different curriculum materials. Participants were classified into three groups: Tier 1 ($n = 568$), Tier 3 ($n = 4$), and non-tiered ($n = 2$), with non-tiered designated as Tier 4 when numerical association was needed (see Table 29). There were no outliers within Tier 3 and the non-tiered group, but Tier 1 had outliers (see Table 8; Figure J3), as assessed by boxplots; data was normally distributed for Tier 3, as assessed by Shapiro-Wilk's test ($p > .05$), and Tier 1 was not normally distributed ($p < .05$), with the non-tiered group containing a sample size of two, which was too small for calculation (see Table 9); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p = .13$; see Table 10). Scores from the LEAP increased from Tier 1

($M = 32.7$, $SD = 20.72$) to Tier 3 ($M = 48.5$, $SD = 26.74$) to the non-tiered group ($M = 58$, $SD = 48.08$), in that order, but the differences between these curriculum material groups were not statistically significant, $F(2, 571) = 2.6$, $p = .75$ (see Tables 29 and 30). The group means were not statistically significantly different ($p > .05$). Therefore, we cannot reject the null hypothesis and we cannot accept the alternative hypothesis.

Since all the data groups within RQ3c did not satisfy all the ANOVA related assumptions based on the original guidelines outlined within Chapter Three, additional analysis and reporting was needed to support the validity and provide transparency of these results. For data within RQ3c there were nine outliers in the Tier 1 data and none within Tier 3 or the non-tiered data (see Table 8; Figure J3), as assessed by inspection of a boxplot for values greater than 1.5 box-lengths from the edge of the box. After analysis of outlier values, the researcher concluded the outlier values were recorded accurately, contained no extreme outliers (greater than three box-lengths from the edge of the box), and did influence the results in a substantial manner (Laerd Statistics, 2017), with ANOVA results containing outliers resulting in a significance of $p = .075$ and without outliers at a significance of $p = .038$ (see Table 30). However, the small samples sizes of Tier 3 (n =4) and non-tiered ($n = 2$) did not warrant additional scrutiny, and the researcher continued investigating RQ3c with the data as the researcher originally collected it.

The LEAP scores within the RQ3c data groups were normally distributed for Tier 3, as assessed by Shapiro-Wilk's test ($p > .05$), but showed Tier 1 departed from normality, and the non-tiered group, containing a sample size of two, was too small for calculation (see Table 9). The result for Tier 1 was not surprising because the sample size with that tier was greater than 50 (see Table 9), which can cause the Shapiro-Wilk test to fail the assumption of normality even if normality was present (George & Mallery, 2005; Field, 2013; Laerd Statistics, 2017). Additional methods for normality conducted by the researcher to either affirm or reject normality for Tier 1 included values of skewness and kurtosis thresholds that would affirm or reject normality (see Table N5), visual examination of histograms and Q Q-Plots (see Appendix M) with additional discussion within Appendix N.

Figures from Table N5 showed Tier 1 skewness and kurtosis values (absolute values) were within ranges that supported normality with skewness values for Tier 1 and Tier 3 similarly skewed. The non-tiered data was not available for calculation and comparison because of the group's low sample size. Following visual examination, Tier 1 data was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of their histograms. Likewise, data for Tier 1 was not perfectly normal, but approximately normally distributed, as assessed by visual inspection of a Normal Q-Q Plot (see Figures M11-M12). Despite the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017), the ANOVA results could *not even be cautiously considered* for RQ3c with approximately normal distribution for Tier 1 and additional backing from the above numerical evaluation methods and visual examinations, because small sample sizes for Tier 3 and the non-tiered group *prevent even cautious support* for RQ3c results.

Summary

The focus of Chapter Four was to provide an analysis of the data. To support that goal, the research questions and null hypotheses were outlined, along with relevant details related to the data cleaning process to support transparent and valid results. A one-way

ANOVA, with a post hoc Tukey test or independent *t*-test where applicable, to address each research question and null hypothesis. An examination of required one-way ANOVA assumptions to examine the reliability of results also served a crucial part the data analysis within Chapter Four. Once the one-way ANOVA was calculated, each null hypothesis was rejected or failed to be rejected based on the statistical significance of the results. Chapter Five will expand on the findings within Chapter Four with additional commentary on the results, implications of the results, recommendations for future research, and it will conclude this study.

Table 31

Null Hypothesis Summary

Null Hypothesis	Test	Sig. (*p*)	Decision
Ho1a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2016-2017.	ANOVA	.95	Failed to reject the null hypothesis.
Ho1b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2017-2018.	ANOVA	.068	Failed to reject the null hypothesis.
Ho1c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for third grade students as measured by the LEAP in 2018-2019.	ANOVA	.778	Failed to reject the null hypothesis.

Null Hypothesis	Test	Sig. (*p*)	Decision
Ho2a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2016-2017.	ANOVA	.828	Failed to reject the null hypothesis.
Ho2b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2017-2018.	ANOVA	.045	Reject the null hypothesis.
Ho2c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fourth grade students as measured by the LEAP in 2018-2019.	ANOVA	.637	Failed to reject the null hypothesis.
Ho3a: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2016-2017.	ANOVA	.543	Failed to reject the null hypothesis.
Ho3b: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2017-2018.	ANOVA	.031	Reject the null hypothesis.
Ho3c: Mathematics curriculum material choice from Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices from the Louisiana Department of Education will have no statistically significant difference on mathematic student achievement at the advanced and mastery level for fifth grade students as measured by the LEAP in 2018-2019.	ANOVA	.075	Failed to reject the null hypothesis.

Note. Samples sizes for *Ho1c, Ho2c, and Ho3c* were too small to be reliably considered.

CHAPTER FIVE

CONCLUSION

Introduction

The purpose of this causal-comparative study was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) to compare math curricula choice to LEAP mathematics achievement for third, fourth, and fifth grade students in public schools within the state of Louisiana. Curriculum materials, especially during mathematics instruction, have been a driving force behind the instruction that takes place within the classroom (Arican, 2018; Ball & Cohen, 1996; Bellens et al., 2020; Monaghan, 2013; Remillard, 2005; Remillard & Kim, 2020; Ruggeri, 2021). However, the lack of information related to the effectiveness of different curriculum materials has hindered school leaders from making curriculum material decisions based on optimal student performance (Blazar et al., 2020; Chingos & Whitehurts, 2012; Polikoff, Campbell, et al., 2020; Ruggeri, 2021). Considering an entire school building, and possibly school district in many cases, is impacted by curriculum material choice, providing more relevant information on curriculum materials is a critical step educational stakeholders must take (Koedel et al., 2017; Polikoff, 2018, 2021). While investigations in this area of study have been done before, an emphasis on examining what impact curriculum materials have or have not had on student learning has been more recent (Solomon et al., 2019). To both to contribute and explore this area of study, the researcher selected the state of Louisiana, in part, because of the curriculum material tier system already in place to inform educational leaders on curriculum material alignment to the state standards as an avenue for student achievement (Kaufman et al., 2018).

Summary of Findings

The predominant question within this inquiry was, what is the difference in student achievement and curriculum material choice? To answer that question, three main research questions were developed to examine the difference in student achievement for third, fourth, and fifth grades, with one main research question for each grade level. Three sub-questions for each main question were developed to encompass a three-year range for the 2016-2017, 2017-2018, and 2018-2019 school years. Initially, the researcher planned to use data from more recent school years, but COVID-19 interference with 2019-2020 testing and beyond made that problematic so the earlier year of 2016-2017 was added to the data set to replace the 2019-2020 school year. Student achievement was based on the percentage of students who scored in the top two LEAP performance tiers, advanced and mastery, which was consistent with similar types of studies within the field (Kodel et al., 2017; Monaghan, 2013; Solomon et al., 2019; Walsh, 2009). One-way ANOVAs were performed for each sub-question to determine if there was a statistically significant difference ($p < .05$) between curriculum material choice and student achievement.

The examination of RQ1, the difference between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for third grade advanced and mastery level math student achievement scores for 2016-2017 (RQ1a), 2017-2018 (RQ1b), and 2018-2019 (RQ1c) produced varied results. Both RQ1a and RQ1c showed no statistically significant results with $p = .95$ and $p = .778$ (see Tables 20 and 24). It should also be noted Tier 3 and the non-tiered group within RQ1c were both below 10 data points, while Tier 1 had over 550 data points, which makes the ANOVA results from RQ1c somewhat irrelevant. Meanwhile, RQ1b showed nearly statistically significant results, $p = .068$ (see Table 14). However, once four outliers from the 471 data points within Tier 1, three outliers from the 92 data points within Tier 3, and the one outlier from the 23 data points within the non-tiered group were removed and the ANOVA was repeated, an ANOVA result of $p = .005$ and effect size of .019 indicated statistically significant results. The Tukey post hoc test result of $p = .009$ between Tier 1 and Tier 3 showed a statistically significant result (see Appendix R).

Additionally, the analysis for the assumption of normality for RQ1b related to the ANOVA showed both Tier 1 and Tier 3 had acceptable normality for analysis given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017) and support from Appendix N and Appendix K (Figures K3-K8). However, the non-tiered group did not meet the skewness or kurtosis normality guidelines for any of the above-mentioned methods that limited the results to only be *cautiously supported*. Furthermore, the number of data points for the non-tiered group was only 23, compared to much higher numbers of data points for Tier 1 and Tier 3.

As a result, the researcher examined the implications of the removal of the non-tiered group, leaving only the two groups for comparison and transitioning to an independent *t*-test. The independent *t*-test is the traditional method used when only two sample groups are used in a comparison of difference (Laerd Statistics, 2017). The results revealed $p = .033$ and Cohen's *d* effect size of .24, which indicated statistically significant results between Tier 1 and Tier 3 (see Appendix O). While the use of an independent *t*-test deviated from the initial research design, which must be acknowledged, the opportunity to gain additional insight into the under studied area of curriculum material use and student achievement, and then apply methodology that more closely matched the needs of the data, was an opportunity the researcher did not want to miss.

The examination of RQ2, the difference between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for fourth grade advanced and mastery level math student achievement scores for 2016-2017 (RQ2a), 2017-2018 (RQ2b), and 2018-2019 (RQ2c) produced varied results, similar to RQ1 results. Both RQ2a and RQ2c showed no statistically significant results with $p = .828$ and $p = .637$ (see Tables 26 and 31). The RQ2c data from Tier 3 and the non-tiered group were both below 10 data points, while Tier 1 had over 550 data points, which made the ANOVA results from RQ1c somewhat extraneous. Meanwhile, RQ2b showed statistically significant results, $p = .045$ (see Table 20). The Tukey post hoc test lacked statistically significant results (see Table 21). To ensure validity and transparency related to outliers, as suggested by Laerd Statistics (2017) and Liao et al. (2016), examination and presentation of the data was conducted with and without outliers. Once the three outliers from the 468 data points within Tier 1 and the one outlier from the 23 data points within the non-tiered group were removed and the ANOVA was repeated, results showed $p = .024$ and effect size of .013 (see Table 20), with a Tukey post hoc that lacked statistically significant results (see Table 29).

As before, the researcher examined the implications with the non-tiered group removed from the 2017-2018 data, leaving only the two groups for comparison, and transitioning to an independent *t*-test. The independent *t*-test is the traditional method used when only two sample groups are used in a comparison (Laerd Statistics, 2017). The results of the *t*-test showed $p = .035$ and Cohen's *d* effect size of .24, which indicated statistically significant results between Tier 1 and Tier 3 (see Appendix S). While the use of an independent *t*-test deviated from the initial research design, which must be acknowledged, the opportunity to gain additional insight did not want to be missed by the researcher. Even with the *t*-tests robustness to overcome issues of normality (Laerd Statistics, 2017), outcomes should be interpreted with *slight caution*, as the RQ2b distribution for Tier 1 approximately normalized and Tier 3 only moderately reflecting normality.

The examination of RQ3, the difference between Tier 1, Tier 2, Tier 3, and other non-tiered curriculum material choices for fifth grade advanced and mastery level math student achievement scores for 2016-2017 (RQ3a), 2017-2018 (RQ3b), and 2018-2019 (RQ3c) produced mixed results, similar to RQ1 and RQ2. Again, the first (RQ3a) and third (RQ3c) sub-questions showed no statistically significant results with $p = .543$ and $p = .75$ (see Tables 33 and 38). The data from Tier 3 and the non-tiered group was even lower than the third grade (RQ1c) and fourth grade (RQ2c) before, with fifth grade (RQ3c) containing fewer than 5 data points for both groups, while Tier 1 had over 550 data points. Thus, the ANOVA results from RQ3c were essentially irrelevant. Meanwhile, RQ3b showed statistically significant results, $p = .031$ (see Table 27), with a Tukey post hoc that lacked statistically significant results (see Table 28).

Like before, to ensure validity and transparency related to outliers, as suggested by Laerd Statistics (2017) and Liao et al. (2016), examination and presentation of the data was conducted with and without outliers. Once the three outliers from the 468 data points within Tier 1 and the one outlier from the 23 data points within the non-tiered group were removed and the ANOVA was repeated, the results showed $p = .002$ and effect size of .022 (see Table 27). The Tukey post hoc test showed a statistically significant result between Tier 1 and Tier 3, $p = .049$, and Tier 1 and the non-tiered group, $p = .032$ (see Table 28).

Analysis for the assumption of normality related to the ANOVA for RQ3b showed both Tier 1 and Tier 3 had adequate normality for analysis given the robustness of ANOVA to overcome violations of normality (Judd, et al., 2017; Field, 2013; Laerd Statistics, 2017) and based on the assumption of normality derived from the Central Limit Theorem's affirmation of normality with larger sample sizes (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017) and support from Appendix N . However, the non-tiered group did not meet the skewness or kurtosis normality guidelines for any of the above-mentioned methods that limited the results to only be *cautiously supported* for RQ3b. Furthermore, the number of data points for the non-tiered group was only 23, compared to much higher numbers of data points for Tier 1 and Tier 3.

As a result, the researcher examined the implications if the non-tiered group was removed, leaving only the two groups for comparison, and transitioning to an independent *t*-test. The independent *t*-test is the traditional method used when only two sample groups are used in a comparison (Laerd Statistics, 2017). The results of the *t*-test showed $p = .049$ and Cohen's *d* effect size of .23, which indicated statistically significant results between Tier 1 and Tier 3 (see Appendix T). While the use of an independent *t*-test deviated from the initial research design, which must be acknowledged, the opportunity to gain additional insight did not want to be missed by the researcher. Even with the *t*-tests robustness to overcome issues of normality (Laerd Statistics, 2017), outcomes should be interpreted with only *slight caution*, as the RQ3b distribution for T1 and T3 only appeared to be somewhat to approximately normalized.

A quick summary of the above findings essentially indicated three of the nine research sub-questions had samples sizes too small for proper ANOVA comparison, which was the 2018-2019 data, sub-questions “c” within each main research question. Of the remaining six subgroups, three were not statistically significant, the 2016-2017 data, sub-questions “a” within each main research question. Of the remaining three subgroups, all 2017-2018 data and sub-questions “b” within each main research question, all showed statistically significant results in at least one or more ways either through the ANOVA within the original research design, an ANOVA after the removal of outliers, or the utilization of a *t*-test when only two samples within the subgroup met the assumption and sample size standards. But even in those instances, the effect size based on Partial Eta Squared was typically small, near .01, for ANOVA results both with and without outliers from 2017-2018 (Cohen, 1988). Likewise, the *t*-test results were all near .2 utilizing Cohen’s *d*, which also indicated a small effect size (Cohen, 1988). While the low effect size might cause the results to be viewed as less meaningful, which would not be altogether unreasonable, Polikoff’s (2021) contention that the selection of one curriculum material over another when selection of new or updated curriculum materials occurs, even if the effect is small, could be viewed as significant because it would reach the entire population with minimal additional effort and cost by the school. Additional perspective and patterns that materialized from the results of this study to provide additional context for the findings are described in the discussion section below.

Discussion

As detailed above, the data within this study produced mixed results with several different patterns and observations of note from either the data gathering process or statistical results. During the data gathering process, it quickly became clear there was heavy migration toward Tier 1 curriculum from the other tiers. This migration could have been somewhat predicted based on the emphasis of the LDOE to support and provide information for highly aligned curriculum materials (Kaufman et al., 2018; Walker, 2022), the highest of which is Tier 1. The lack of information overall on what curriculum materials are being used (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018) would make such a conjecture difficult to support before information was initially gathered, which was why control of curriculum material choice by schools was listed as a limitation within this study. However, both the LDOE emphasis on quality curriculum materials and the migration of schools toward Tier 1 materials was not misplaced when viewed through the lens of Gagne's cumulative learning theory as the LDOE sought to leverage materials to help support instruction that is aligned to standards, correctly build up to scaffolded instruction, work toward the final desired task, and investigate whether the overall sequence of instruction optimized student learning.

For the first two years of the study, 2016-2017 and 2017-2018, use of Tier 3 curriculum materials was prevalent enough across all three grades for analysis, but that shifted the final year when Tier 3 curriculum materials, as well as non-tiered curriculum, drastically declined by schools within the sample groups (see Table 9). This heavy migration toward Tier 1 materials in 2018-2019 made the ANOVA results for that year insignificant, since sample sizes for Tier 3 and the non-tiered group were too small to reliably compare for each grade. As a result, the non-tiered group will usually be mentioned only for the 2017-2018 school year, when the sample size for each grade was 20 or higher, and results from the 2018-2019 school year for all grades will typically be left out or limited within discussion since the Tier 3 and the non-tiered groups were too small for a valid comparison.

Another observation involved the sample size additions from 2016-2017 to 2017-2018 and the resulting mean score changes for Tier 1 and Tier 3. For the 2016-2017 school year, the difference between Tier 1 and Tier 3 means for third and fourth grade was less than one percent, with fifth grade a bit further apart at 2.4%. Tier 1 samples represented the greater mean in all cases. However, the sample results for Tier 1 and Tier 3 the next school year, 2017-2018, showed a difference of about four to five percentage points between Tier 1 and Tier 3, and approximately seven percentage points between Tier 1 and the non-tiered group, with Tier 1 again demonstrating the superior scores in all cases. Also of note, the difference between student performance for the state of Louisiana compared to Tier 1 for each of nine samples, three for each grade each year, was always 2.3 percent or less and one percent or less for four out of the nine samples, with the state average always slightly greater than Tier 1 (see Table 32).

Table 32

Average Group LEAP Scores with Sample Size by Year and Grade

Year	Grade	*State of Louisiana*		Tier 1		*Tier 3*		*Non-Tiered*	
		N	*M*	*n*	*M*	*n*	*M*	*n*	*M*
2016-2017	3	786	42	274	42.28	59	41.47	3	44
	4	778	39	273	37.82	60	37.42	3	45.33
	5	757	32	266	29.99	57	26.72	2	19.5
2017-2018	3	789	42	471	40.74	92	35.88	23	36.13
	4	782	38	468	37.67	92	32.67	23	30.91
	5	764	30	463	29.4	90	25.07	22	21.41
2018-2019	3	761	42	569	41.04	3	45	8	45.5
	4	770	41	572	38.7	4	49	8	39.38
	5	751	34	568	32.7	4	48.5	2	58

Note. Data was from LDOE (2019b).

Upon reflection, perhaps a deeper examination related to the schools that were added to Tier 3 from 2016-2017 to 2017-2018 might be warranted to account for the decrease in mean average from year one to year two compared to the more stable Tier 1 results despite a high increase in data samples during the same time. However, one of the delimitations of this study was to specifically focus on the variables within the research questions, curriculum materials and student achievement, and to discover if a difference exists, not to discover *why* a difference does or does not exist. Furthermore, the ability to intentionally design a study from the onset to account for these other variances was complicated since there was so little information on what curriculum materials are used, not only within Louisiana, but across the country (Blazar et al., 2020; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018).

The combination of statistically insignificant findings for the 2016-2017 year for each grade, followed by statistically significant or nearly statistically significant results for each grade in 2017-2018, and closing with the 2018-2019 school year that only had one sample size large enough for reliable comparison, left the researcher to conclude the overall results in the most simplistic term were mixed." The mixed nature of the results seemed to be supported by research on all sides of this field of study. Results from the 2017-2018 school year of this study affirmed the work of those who advocated for and concluded that curriculum materials have an impact on student achievement (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Chingos & Whitehurts, 2012; Cress, 2019; Koedel et al., 2017; Lein Authement, 2022; Polikoff, Petrilli, et al., 2020; Steiner, 2017; Walsh, 2009; White, 2018). As for those who maintain or discovered curriculum materials do not impact student achievement (Kendrick et al., 2020; Nargi, 2018; Womersley, 2020), that point of view was supported by the 2016-2017 findings. But the group the results of this study most closely support might not be those who endorse the viewpoint that a difference in curriculum materials impacts student achievement or those who do not view curriculum materials as a differentiator of student achievement, but rather those who advocate for more overall research based on the inconclusive studies in this area (Blazer et al., 2020; Koedel et al., 2017; Monagham, 2013; Nargi, 2018; Solomon et al., 2019).

The three tenets of Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) basically summarize what curriculum materials overall are designed to do, as they help the teacher scaffold learning through instruction, monitor progress through learning activities, and provide sequenced learning with clear assessment over the established learning goals (Machalow, 2020; Remillard & Kim, 2020; Stein et al., 2007; Suppa 2018). So, could the mixed results simply be the byproduct of a curriculum material being in place, essentially any somewhat credible curriculum material being in place? If that were the case, could one conclude that the majority of the benefits of cumulative learning theory on student learning can be obtained simply by having a somewhat credible curriculum material in place, with differences between curriculum materials too inconsequential to be of concern? After all, Marzano (2003) noted the importance of a guaranteed and viable curriculum, but without detailed lists of what curriculum materials met this standard. Rather, Marzano emphasized the importance of clear expectations for what to teach, the guaranteed part, and providing student learning opportunities based on those expectations, the viable aspect (Marzano, 2003), which conceptually align with the above cumulative learning theory principles. But if that were the case, why did this study, at least in part, and other studies find statistically significant results (Agodini et al., 2010; Bhatt & Koedel, 2012; Bhatt et al., 2013; Cress, 2019; Koedel et al., 2017; Lein Authement, 2022; Polikoff, Petrilli, et al., 2020; Walsh, 2009; White, 2018)? While the low effect size and varying statistically significant results from this study suggest mixed findings, the ability for Tier 1 curriculum materials to outperform the other tiers could yield a bigger effect if Polikoff's (2021) insight that a small positive effect generated by the selection of one curriculum material over another can actually be largely impactful when utilized by school, school district, or state populations because the effect would be applied to the entire group. The limited significantly significant results from this study and other research in this field of study continue to emphasize the need for researchers, state institutions, and other educational stakeholders to dedicate more resources to determine the correlation more clearly between curriculum material choice and student achievement using new, improved research design methods in order to overcome the limitations of this study and gain more definitive results.

Professional Implications

The limited volume of research related to what curriculum materials are used in schools is routinely a major obstruction when it comes to assessing curriculum material impact on student achievement (Blazar et al., 2020; Cummins-Colburn, 2007; Hutt & Polikoff, 2018; Kane, 2016; Koedel et al., 2017; Polikoff, 2018). That obstruction, along with the limited number of studies related to curriculum materials and student achievement, were not only issues for this study but also for educational leaders as they look for avenues to improve student achievement (Blazar et al., 2020; Chingos & Whitehurts, 2012; Koedel et al., 2017; Polikoff, Campbell, et al., 2020; Ruggeri, 2021). Given the lower than desired or stagnant student achievement levels in mathematics over the last decade (OECD, 2018; NCES, 2019; Schleicher, 2019) and increasing achievement gap (Oglesby-Phelps, 2022; Sparks, 2018), educational stakeholders must be diligent in the exploration of paths that bring about improvement. Despite the under studied influence of curriculum materials on student achievement, educational leaders have still consistently looked to curriculum materials as a pathway to increase student achievement (Koedel et al., 2017; Reys et al., 2003; Solomon et al., 2019; Superfine et al., 2010). This study looked to address both of those issues by adding to the limited information on what curriculum materials have been used and exploring the difference between curriculum materials and student achievement.

The overall mixed results of this research support the need for continued diligence within this area of study. Specifically, more investigation should determine whether the Tier 1 curriculum material's high performance, seen through the statistically significant results for 2017-2018, and sustaining the highest average scores for all comparisons with reliable sample sizes, is a consistent theme found by recent reliable studies or an anomaly. Furthermore, the 2017-2018 results are now five years old, prior to the COVID-19 pandemic, and since then, changes to the educational landscape may further obscure how to best apply the results from this to delineate the current relationship between curriculum materials and student achievement. Until additional research takes place, educational decision makers within the state of Louisiana can look to evaluate the existing curricula in use and how well it matches Tier 1 standards and whether adjustments in curriculum materials need to take place. Likewise, other areas of the country with similar standards to the state of Louisiana could use the existing LDOE tier system to aid in the process of curriculum materials choices or take a more comprehensive step and begin to develop a curriculum material evaluation system like the LDOE system to tap the potential of curriculum materials to enhance student learning. The next step, at either the state or national level, would naturally be for educational stakeholders to investigate the possibility of developing and implementing transparent curriculum material evaluation systems to provide school leaders within each state the information needed to make informed decisions when selecting curriculum materials. Until then, either improving or initiating better mechanisms for collecting data on what curriculum materials are in use to facilitate more research should be widely considered given the lack of quality avenues currently in place to collect that information (Kane, 2016; Koedel et al., 2017).

The importance and awareness of the above initiatives are amplified after the realization a healthy mathematical understanding is a vital element in ensuring an individual's success in education and future vocation; consequently, quality math instruction is crucial to the development of a quality labor force within a society (Daro & Asturias, 2019; Denton, 2021; Hiser, 2016; National Council of Teachers of Mathematics [NCTM], 2018; White, 2018). Despite the influence of quality math education on individuals and society, the lack of success linked to properly motivating students during mathematical learning opportunities has continued to be a concern (Lee et al., 2021).

The alarm for current issues within math education was only magnified in light of Carunungan's (2022) assertion that part of the issue elementary students encountered with mathematics was the elementary teacher, since elementary teachers were often uneasy teaching mathematics and therefore resorted to teaching in ways that were not always best for students. When viewed through Gagne's cumulative learning theory, using teaching methods or motivational techniques that do not allow students to make mathematical connections has the potential to not only hamper scaffolding rigorous course content (tenet one) but also, impact the competence at which students learn and progress through component tasks and formative assessments (tenet two).

As a result, education leaders finding and leveraging quality curriculum materials that can help both teachers teach the content (Arican, 2018; Carunungan, 2022; Machalow, 2020; Polikoff, 2021; Ruggeri, 2021) and potentially raise student achievement must be diligently explored. The potential for this influence to manifest was partially shown and supported within this study as Tier 1 curriculum materials had the highest average scores within each grade and year with reliable samples sizes. Further support was found through the statistically significant results in 2017-2018, as Tier 1 data showed higher averages than Tier 3 or the non-tiered group by four to eight percent for each grade. The implications of more informed selection, utilization, and evaluation of curriculum materials within the education system should not be overlooked.

Recommendations for Future Research

As continually addressed by other researchers who explored the difference between curriculum materials and student achievement, and reinforced by the researcher within this study, more research and facilitation of research related to curriculum materials is needed. However, initiating future research in this area will be impeded if based off a simplistic approach or belief that researchers simply need to seize the initiative because clear limitations currently curb the practical expansion of research. While research in this area can be done, as evident by this study and others, reduction of the current limitations and effort required could awaken this area of research. Until then, generating an increase in curiosity within the educational community related to the difference between curriculum material choice and student achievement is needed by the few directly studying the difference, coupled with similar efforts by those studying either curriculum materials or student achievement in other capacities viewing curriculum materials as a variable of emphasis. Finding ways to increase awareness is needed, but an equally crucial undertaking for progress is to discover how to either widen the use of, or expand the creation of, mechanisms to document curriculum material use within the education system in a way that is accurate and accessible for research. While the LDOE has already implemented a system at the state level that provides a starting point for curriculum material data collection by listing how different curriculum materials align to the state standards, it does not actually track what is used.

Additional exploration of the impact of the LDOE tier system by other researchers and other educational systems presents an opportunity for the advancement of research in this field and for other states to begin examining if a similar system, possibly one that documents curriculum material use, would be beneficial in the quest to raise student achievement. If through an increase in attention to this area of study, the improvement of current mechanisms for curriculum material selection or evaluation, or expansion in the creation of new tools by education stakeholders, a true opportunity for researchers to seize the initiative and inject new findings regarding curriculum materials and student achievement could take place.

Future opportunities for study, hopefully with access to larger populations from which to draw samples through improved documentation related to curriculum material use if the above suggestions manifest, could include the examination of the difference in curriculum material choice and student achievement longitudinally. With more transparency of curriculum material use, the ability for researchers to examine how either the use of one curriculum material, several materials, or even the transition between materials over a three-year, five-year, or longer time period becomes much more feasible. Limitations within this study related to the unknown migration of schools to Tier 1 curriculum materials, lack of knowledge of which materials were in use, and the need for a reasonable timeframe of only three years all impacted the research design that future researchers now have an opportunity to account for and overcome. Likewise, more targeted studies related to fidelity of curriculum material implementation, teacher training associated with the curriculum prior to or during the curriculum materials utilization, teacher experience, in-person compared to online curriculum materials, and socioeconomic status associated with curriculum materials and student achievement, are all needed areas of future study; all could be more accessible by either the ability to better account for existing limitations or through an increase in information on curriculum material utilization for the desired research population.

Given the current ambiguity surrounding curriculum material utilization, perhaps more research directly related to the limitations within this field of study and how to best account for them is of the most practical use. Even though the researcher in this study attempted to overcome some of those limitations, the researcher was left with 2018-2019 data too misaligned for proper analysis and produced minimal return on investment. If investigation into the consistent issues within this field of study and how to best compensate, researchers might be empowered to generate not only more research, but more meaningful research.

The analysis of curriculum materials and student performance shifts could also be another point more closely studied. Within this study the 2016-2017 data was not statistically significant and had similar average LEAP scores between Tier 1 and Tier 3. Yet, the next year in 2017-2018 a statistically significant difference was seen, with average scores on the LEAP between Tier 1 and Tier 3 and the non-tiered group four or more percentage points apart for each grade level. Understanding why that changed occurred, and others like, it could provide meaningful insights. The researcher could make a conjecture that Tier 3 performance dropped with the increase in sample size because traditionally lower achieving schools were mostly added the second year, or that a disproportionate number of high achieving schools dominated the population the first year for Tier 3, but this would on be conjecture because the demographics of the schools were a limitation within this study.

Future studies, however, might be able to put systems in place to better examine those nuances, thus providing a better understanding for how curriculum materials and student achievement interact. Moreover, since Tier 1 curriculum materials overshadowed the populations of the other tiers, especially for 2018-2019, contingencies to separate that tier into individual curriculums, instead of an entire Tier 1 group, to examine the difference in student achievement between each specific curriculum could also yield beneficial results.

Conclusion

While the difficulties currently associated with curriculum material research may or may not also contribute to the lack of activity in this area of study, the existing research directly indicates there is a gap in literature between curriculum material and student achievement that needs to be better understood (Blazar et al., 2020; Bhatt & Koedel, 2012; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Kane, 2016; Koedel et al., 2017; Polikoff, 2018; Polikoff, Campbell, et al., 2020; Remillard, 2005; Steiner, 2017). To address the shortage, this study attempted to provide two things: a detailed analysis of existing research within the literature review and research conducted to identify the difference between curriculum material choice and student achievement. One of the critiques related to studies on the effectiveness of curriculum materials and student achievement was that they typically do not pass high standards of scrutiny (Chingos & Whitehurts, 2012; Solomon et al., 2019). For that reason, the researcher intentionally utilized a research design that was consistent with appropriate research practices and other studies within the field that received publication. While the results of this study did indicate curriculum material choice could positively impact student learning, the overall lack of information related to curriculum materials and student achievement continues to leave educational decision makers without clear information on how to, or if, they can leverage curriculum materials to improve student achievement, unless more research is generated (Blazar et al., 2020; Chingos & Whitehurts, 2012; Cummins-Colburn, 2007; Koedel et al., 2017; Polikoff, Campbell, et al., 2020; Ruggeri, 2021).

If the warning by Dr. Caballero (1989) that the doors of future possibilities for students close when progress in mathematics stops is accurate, it is paramount for the educational system to find ways to keep students engaged and successful in mathematics at both the individual and societal levels (Daro & Asturias, 2019; Denton, 2021; Hiser, 2016; NCTM, 2018; White, 2018). Thus, the chance to study whether curriculum materials could help keep doors open for more students provided the setting for this causal-comparative study, the purpose of which was to use Gagne's cumulative learning theory (Gagne, 1962a, 1962b, 1965) to determine if there is a relationship between math curricula choice and math achievement. As previously stated, the findings from this research study were mixed, but not without relevance. Those mixed results reinforced two essential things: first, the ability of curriculum materials to have a statistically significant difference on student achievement, even if inconsistently, second, the need for more research in this area so educators can gain a better understanding and come to definitive answers on the relationship between curriculum material and student achievement. Currently research does not have a clear answer, which has been underscored by the mixed results in this study and others.

Within the mixed results, there were still positive outcomes; even a contribution as simple as providing clear information detailing what curriculum materials are in use, can provide a starting point for future research since knowing what curriculum materials are in use has been a well-documented obstacle, as demonstrated within this study. The unknown and undocumented mass migration of school utilization of Tier 1 curriculum materials presented both an issue and a benefit. The issue was that it limited the size of sample groups for comparison, but it also provided insight into the potential benefits of agencies, such as the LDOE, to help schools make empowered, evidence-based decisions when selecting curriculum materials. Even though Tier 1 curriculum materials did outperform the others within this study, a standalone study such as this cannot definitively conclude material choice makes a difference when viewed in light of the lack of current available research revealing enough evidence to show such an impact on student learning.

For that reason, this study was designed to discover *if* there was a difference between curriculum material choice and student achievement, not *why* there was or was not a difference in curriculum material choice and student achievement. This was partially why many of the concepts within the recommendations for future research were not included within this study. Hopefully, through the tireless work of researchers to overcome the obstacles related to the uncertainty of which curriculum materials are in use and gathering the individual data to generate studies, a better understanding may emerge about whether a difference truly exists. Improvements in the development and use of mechanisms to accurately gather and share data related to what curriculum materials are in use would make research more manageable, as could the adoption and wider use of curriculum evaluation systems similar to the tier system generated by the LDOE, providing more of a starting point for comparison to determine if the potential in this area can be operationalized, or if it can be dismissed. More research is needed to provide clarity regarding *if* there is a difference between curriculum materials and student achievement. Once educational stakeholders know with a high degree of certainty *if* curriculum materials do or do not have the potential to make a difference on student achievement, a shift to more research related to *why* a difference does or does not exist can take place, empowering educational stakeholders to leverage school resources based on evidence for maximizing student learning.

REFERENCE LIST

Ager-Sharrieff, L. (2022). *Perceptions of elementary math teachers grades K-5 on increasing student engagement through small group instruction* (Order No. 29260667). Available from ProQuest Dissertations & Theses Global. (2704363374). Retrieved from https://www.proquest.com/dissertations-theses/perceptions-elementary-math-teachers-grades-k-5/docview/2704363374/se-2

Agodini, R., Harris, B., Atkins-Burnett, S., Heaviside, S., & Novak, T. (2010). *Achievement Effects of Four Early Elementary School Math Curricula: Findings for First and Second Graders*. National Center for Education Evaluation and Regional Assistance, U.S. Department of Education, Institute of Education Sciences. NCEE 2011-4001.

Albert, M. D. (2020). *Functional behavioral assessment: Putting the function into functional behavioral assessments* (Order No. 27837022). Available from ProQuest Dissertations & Theses Global. (2395804465). Retrieved from https://www.proquest.com/dissertations-theses/functional-behavioral-assessment-putting-function/docview/2395804465/se-2?accountid=14196

Aldamen, H., Duncan, K., & Ziegelmayer, J. L. (2018). Cumulative learning and sustained engagement in an introduction to accounting course. *Asian Review of Accounting, 26*(1), 19-38. doi:http://dx.doi.org/10.1108/ARA-03-2017-0036

Allensworth, E., Cashdollar, S., & Gwynne, J. (2021). Improvements in math instruction and student achievement through professional learning around the Common Core State Standards in Chicago. *AERA Open, 7*. https://doi.org/10.1177/2332858420986872

Anastasiou, P. (2022). *Using digital storytelling in science: Meaning making with students aged 10-12 years old* (Order No. 29133355). Available from ProQuest Dissertations

& Theses Global. (2644358362). Retrieved from https://www.proquest.com/dissertations-theses/using-digital-storytelling-science-meaning-making/docview/2644358362/se-2?accountid=14196

Arican, M. (2018). Preservice middle and high school mathematics teachers' strategies when solving proportion problems. *International Journal of Science and Mathematics Education, 16*(2), 315-335.

Arizona Department of Education. (2018). *2018 Technical Report of AIMS*. [PDF file]. https://www.azed.gov/sites/default/files/2020/04/AIMS_Tech_Report_2018_508compliant.pdf?id=5ea872bf03e2b3106450cb18

Ball, D. L., & Cohen, D. K. (1996). Reform by the book: What is-or might be- the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher, 25*(9), 6-8, 14.

Bandalos, D. (2018). *Measurement theory and applications for social sciences*. Guilford.

Barthelemy, J. A. (2022). *An analysis of the every student succeeds act (ESSA) in the advent of the covid 19 pandemic: Policy implication* (Order No. 29162623). Available from ProQuest Dissertations & Theses Global. (2655223437). Retrieved from https://www.proquest.com/dissertations-theses/analysis-every-student-succeeds-act-essa-advent/docview/2655223437/se-2?accountid=14196

Bates, B. (2019). *Learning theories simplified*. Los Angeles, CA: SAGE Publications Ltd.

Beatty, A., & Pritchett, L. (2012). "The negative consequences of overambitious curricula in developing countries", CDG Working Paper No. 293, Center for Global Development, Washington, DC.

Berber Sardinha, T. & Veirano Pinto, M. (2019). *Multi-Dimensional Analysis: Research Methods and Current Issues*. Bloomsbury Academic.

Bellens, K., Van den Noortgate, W., & Van Damme, J. (2020). The informed choice: mathematics textbook assessment in light of educational freedom, effectiveness, and improvement in primary education. *School Effectiveness & School Improvement, 31*(2), 192–211. https://doi.org/10.1080/09243453.2019.1642215

Benton, R. (2022). Teachers' ability to foster student creativity in the high stakes testing environment (Order No. 29066941). Available from ProQuest Dissertations & Theses Global. (2662750128). Retrieved from https://www.proquest.com/dissertations-theses/teachers-ability-foster-student-creativity-high/docview/2662750128/se-2?accountid=14196
Berber Sardinha, T. & Veirano Pinto, M. (2019). *Multi-Dimensional Analysis: Research Methods and Current Issues*. Bloomsbury Academic.
Berrett, A. N., & Carter, N. J. (2018). Imagine math facts improves multiplication fact fluency in third grade students. Journal of Behavioral Education, 27(2), 223-239. doi:10.1007/s10864-017-9288-1

Bhatt, R., & Koedel, C. (2012). "Large-Scale Evaluations of Curricular Effectiveness: The Case of Elementary Mathematics in Indiana," *Educational Evaluation and Policy Analysis 34*(4): 391-412.

Bhatt, R., Koedel, C., & Lehmann, D. (2013). "Is Curriculum Quality Uniform? Evidence from Florida," *Economics of Education Review 34*(1): 107-121.

Blazar, D., Heller, B., Kane, T., Polikoff, M., Staiger, D., Carrell, S., Goldhaber, D., Harris, D., Hitch, R., Holden, K., & Kurlaender, M. (2020). Curriculum Reform in The

Common Core Era: Evaluating Elementary Math Textbooks Across Six U.S. States. *Journal of Policy Analysis & Management, 39*(4), 966. https://doi.org/10.1002/pam.22257

Bloom, A. (2020). *Effect of an active note restructuring intervention on developmental mathematics students' test performance* (Order No. 28390807). Available from ProQuest Dissertations & Theses Global. (2515681773). Retrieved from https://www.proquest.com/dissertations-theses/effect-active-note-restructuring-intervention-on/docview/2515681773/se-2

Boaler, J., Dieckmann, J., Perez-Nunez, G., Sun, K., & Williams, C. (2018). Changing students minds and achievement in mathematics: The impact of a free online student course. *Frontiers in Education, 3*. https://doi.org/10.3389/feduc.2018.00026

Brock, K. (2021). *The perspectives of elementary special education teachers on the implementation of antecedent interventions towards decreasing disruptive behaviors* (Order No. 28863506). Available from ProQuest Dissertations & Theses Global. (2626888754). Retrieved from https://www.proquest.com/dissertations-theses/perspectives-elementary-special-education/docview/2626888754/se-2?accountid=14196

Brodesky, A., Fagan, E., & MacVicar, T. (2021). Strengthen mathematics intervention to promote student success. *Principal Leadership, 21*(6), 46-49.

Bruner, J. S. (1966). Toward a theory of instruction. New York: W.W. Norton.

Burnette, D. (2018). States still strain to find one path on accountability. Education Week, 37(35), 1-17. Retrieved from https://www.edweek.org/

Burns McOmber, E. (2020). *Exploring the predictability power of arizona's college and career readiness indicators on college enrollment rates: A multiple regression analysis* (Order No. 27736674). Available from ProQuest Dissertations & Theses Global. (2358737479). Retrieved from https://www.proquest.com/dissertations-theses/exploring-predictability-power-arizonas-college/docview/2358737479/se-2

Buzick, H. M., Rhoad-Drogalis, A., Laitusis, C. C., & King, T. C. (2019). Teachers' Views of Their Practices Related to Common Core State Standards-Aligned Assessments. Research Report. ETS RR-19-40. *ETS Research Report Series*.

Caballero, J. (1989). Everybody a mathematician? *CAIP Quarterly 2*, Fall.

Cage, G. (2021). *An investigation examining the closing of the achievement gap in louisiana with the NGSS* (Order No. 29111381). Available from ProQuest Dissertations & Theses Global. (2665131845). Retrieved from https://www.proquest.com/dissertations-theses/investigation-examining-closing-achievement-gap/docview/2665131845/se-2?accountid=14196

Carunungan, M. (2022). *An exploratory study on how math stories engage young learners in mathematical sense-making* (Order No. 29318973). Available from ProQuest Dissertations & Theses Global. (2707990276). Retrieved from https://www.proquest.com/dissertations-theses/exploratory-study-on-how-math-stories-engage/docview/2707990276/se-2

Chastain, C. (2021). *The effect of high school schedules on ninth grade student achievement indicators and overall school performance measures: A mixed methods study* (Order No. 28970699). Available from ProQuest Dissertations & Theses Global.

(2652879567). Retrieved from https://www.proquest.com/dissertations-theses/effect-high-school-schedules-on-ninth-grade/docview/2652879567/se-2

Chen, Y. (2021). *Self-regulation and academic learning in preschoolers with autism spectrum disorder: Individual differences and links to executive function, effortful control, reward sensitivity, school engagement, and adaptive behavior* (Order No. 28419731). Available from ProQuest Dissertations & Theses Global. (2524893701). Retrieved from https://www.proquest.com/dissertations-theses/self-regulation-academic-learning-preschoolers/docview/2524893701/se-2?accountid=14196

Chingos, M. M., & Whitehurts, J. (2012). Choosing blindly: Instructional materials, teacher effectiveness, and the common core. *Brookings Institute Reports,* retrieved November 9, 2018, from: https://www.brookings.edu/wp-content/uploads/2016/06/0410_curriculum_chingos_whitehurst.pdf

Christensen, L. B., Johnson, R. B., & Turner, L. A. (2014). *Research methods, design, and analysis* (12th ed.). Boston, MA: Allyn & Bacon.

Chu, Y. (2019). What Are They Talking about When They Talk about Equity? A Content Analysis of Equity Principles and Provisions in State Every Student Succeeds Act Plans. *Education Policy Analysis Archives, 27*(158).

Clark, J. (2021). *Case study of growth mindset and the impact on math anxiety* (Order No. 28647353). Available from ProQuest Dissertations & Theses Global. (2572612113). Retrieved from https://www.proquest.com/dissertations-theses/case-study-growth-mindset-impact-on-math-anxiety/docview/2572612113/se-2

Clark, K. (2018). Learning theories: Behaviorism. *Radiologic Technology, 90*(2), 172–175.

Clemons, S. (2022). *North carolina high school alternative school administrators' perceptions of school performance measures* (Order No. 29063289). Available from ProQuest Dissertations & Theses Global. (2658238505). Retrieved from https://www.proquest.com/dissertations-theses/north-carolina-high-school-alternative/docview/2658238505/se-2?accountid=14196

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

Common Core State Standards Initiative. (2021a). *Key shifts in mathematics*. http://www.corestandards.org/other-resources/key-shifts-in-mathematics/

Common Core State Standards Initiative. (2021b). *Mathematics standards*. http://www.corestandards.org/Math/

Common Core State Standards Initiative. (2022a). *About the standards*. http://www.corestandards.org/about-the-standards/

Common Core State Standards Initiative. (2022b). *Development Process*. http://www.corestandards.org/about-the-standards/development-process/

Conrad, Z. (2021). The validity of using diagnostic classification models for criterion-referenced interpretations (Order No. 28540266). Available from ProQuest Dissertations & Theses Global. (2558599300). Retrieved from https://www.proquest.com/dissertations-theses/validity-using-diagnostic-classification-models/docview/2558599300/se-2?accountid=14196

Cress, T. D. (2019). *Influence of the reflex math fact fluency program on math scores* (Order No. 27545430). Available from ProQuest Dissertations & Theses Global.

(2311974320). Retrieved from https://search.proquest.com/docview/2311974320?accountid=14196

Creswell, J. W. (2019). Educational research: *Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Thousand Oaks, CA: Sage Publications.

Creswell, J., & Poth, C. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE.

Cox, R. T. (2022). *The Impact of Physical Education on Students' Performance Outcomes in Mathematics* (Order No. 28861467). Available from ProQuest Dissertations & Theses Global. (2626029023). https://www.proquest.com/dissertations-theses/impact-physical-education-on-students-performance/docview/2626029023/se-2

Cruze, A. N. (2022). *Latinx high school students' Math/Science motivational beliefs predicting AP math and AP science enrollment and the role of educators* (Order No. 29169386). Available from ProQuest Dissertations & Theses Global. (2673005682). Retrieved from https://www.proquest.com/dissertations-theses/latinx-high-school-students-math-science/docview/2673005682/se-2?accountid=14196

Cummins-Colburn, B. (2007). *Differences between state -adopted textbooks and student outcomes on the texas assessment of knowledge and skills examination* (Order No. 3269605). Available from ProQuest Dissertations & Theses Global. (304701296). Retrieved from https://search.proquest.com/dissertations-theses/differences-between-state-adopted-textbooks/docview/304701296/se-2?accountid=14196

Daro, P., & Asturias, H. (2019). Branching out: designing high school math pathways for equity. (Just Equations, October 2019). Retrieved from: https://justequations.org/wpcontent/uploads/Just-Equations-2019-Report-Branching-Out-Digital.pd

Davis, K. (2022). *Teacher efficacy and student academic achievement in mathematics at the middle and high school levels* (Order No. 28966740). Available from ProQuest Dissertations & Theses Global. (2636380337). Retrieved from https://www.proquest.com/dissertations-theses/teacher-efficacy-student-academic-achievement/docview/2636380337/se-2?accountid=14196

Davis, L. (2019). Common Core and the Continued Socioeconomic Achievement Gap: How Can We Better Prepare Future Teachers? *Journal of Education and Learning, 8*(6), 1–14.

Delacre, M., Leys, C., Mora, Y. L., & Lakens, D. (2019). Taking parametric assumptions seriously: Arguments for the use of Welch's F-test instead of the classical F-test in one-way ANOVA. *International Review of Social Psychology, 32*(1), 1-12. https://doi.org/10.5334/irsp.198

Dellner, W. (1979). *Effects Of Instructional Material, Review Questions, And Sequence Upon Problem Solving Behavior* (Order No. 7925048). Available from ProQuest Dissertations & Theses Global. (303019538). Retrieved from https://www.proquest.com/dissertations-theses/effects-instructional-material-review-questions/docview/303019538/se-2?accountid=14196

Demir, S. (2022). Comparison of Normality Tests in Terms of Sample Sizes under Different Skewness and Kurtosis Coefficients. *International Journal of Assessment Tools in Education, 9*(2), 397-409.

Denton, G. (2021). *Teacher capacity to adopt and use student-centered instruction in secondary math classrooms* (Order No. 28644265). Available from ProQuest Dissertations & Theses Global. (2565158701). Retrieved from https://www.proquest.com/dissertations-theses/teacher-capacity-adopt-use-student-centered/docview/2565158701/se-2

Ebanks, K. (2020). *Promoting equity: The impact of reading comprehension strategies on mathematics achievement level* (Order No. 28151289). Available from ProQuest Dissertations & Theses Global. (2474837753). Retrieved from https://www.proquest.com/dissertations-theses/promoting-equity-impact-reading-comprehension/docview/2474837753/se-2?accountid=14196

Eisner, E. (1987). Why the textbook influences curriculum. *Curriculum Review, 26*, 11–13.

Elliott, D. (1990). Textbooks and the curriculum in the postwar era: 1950-1980. *Yearbook (National Society for the Study of Education), 89th*(1), 42–55.

English, R., & Reigeluth, C. (1996). Formative Research on Sequencing Instruction with the Elaboration Theory. *Educational Technology Research and Development, 44*(1), 23-42. Retrieved from http://www.jstor.org/stable/30220137

Ewen, R. (2010). An introduction to theories of personality (7th ed.). New York, NY: Psychology Press.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. Behavior Research Methods, 41, 1149-1160.

Fernandez, D. (2021). *Understanding elementary special education teachers' experiences teaching students with autism spectrum disorder* (Order No. 28650052). Available from ProQuest Dissertations & Theses Global. (2564987137). Retrieved from https://www.proquest.com/dissertations-theses/understanding-elementary-special-education/docview/2564987137/se-2?accountid=14196

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th Ed.). SAGE Publications.

Ford, A. (2018). *The impact of excessive high-stakes standardized tests on students and teachers* (Doctoral dissertation.). Retrieved from Retrieved from ProQuest Dissertations and Theses database. (UMI No. 10845854).

Fraenkel, J., Wallen, N., & Hyun, H. (2019). *How to design and evaluate research in education* (10th ed.). McGraw-Hill Education.

Frey, B. (2018). *The Sage encyclopedia of educational research, measurement, and evaluation*. SAGE Publications, Inc.

Friedberg, S., Barone, D., Belding, J., Chen, A., Dixon, L., Fennell, F., Fisher, D., Frey, N., Howe, R., Shanahan, T., & Thomas B. Fordham Institute. (2018). The State of State Standards Post-Common Core. In *Thomas B. Fordham Institute*. Thomas B. Fordham Institute.

Gagne, R. (1962a). The acquisition of knowledge. *Psychological Review, 69*(4), 355-365.

Gagne, R. (1962b). Military training and principles of learning. *American Psychologist, 17*(2), 83–91. https://doi.org/10.1037/h0048613

Gagne, R. (1965). *The conditions of learning*. Austin, TX: Holt, Rinehart, and Winston.

Gagné, R. M. (1968). Contributions of learning to human development. *Psychological Review, 75*(3), 177–191.

Gagne, R. (1974). Educational Technology and the Learning Process. *Educational Researcher, 3*(1), 3–8.

Gagne, R. (1985). *The conditions of learning (4th ed.)*. New York: Holt, Rinehart, and Winston.

Gagne, R., & Briggs, L. (1979). *Principles of instructional design (2nd ed.).* New York: Holt, Rinehart, and Winston.

Gagne, R., & Deci, E. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior, 26*(4), 331-362. https://doi.org/10.1002/job.322

Galoyan, T. (2019). *Multidisciplinary investigation into neurocognitive basis of problem-solving transfer and implications for pedagogy: A mixed-methods study* (Order No. 22583717). Available from ProQuest Dissertations & Theses Global. (2284204803). Retrieved from https://www.proquest.com/dissertations-theses/multidisciplinary-investigation-into/docview/2284204803/se-2

Gardner, M. (2022). Evaluating when subscores can have value in psychological and health applications (Order No. 28965039). Available from ProQuest Dissertations & Theses Global. (2669561999). Retrieved from https://www.proquest.com/dissertations-theses/evaluating-when-subscores-can-have-value/docview/2669561999/se-2

Gay, L., Mills, G., & Airasian, P. (2009). *Educational research: Competencies for analysis and applications (9th ed.)*. Upper Saddle River, NJ: Pearson Education, Inc

George, D., & Mallery, P. (2005). *SPSS for Windows step by step: a simple guide and reference, 12.0 update* (Fifth edition.). Pearson Education.

Gewertz, C. (2018). Math score slide to a 20-year low on ACT (American College Testing). *Education Week, 38*(10), 7.

Gleckler, M. (2021). *Faculty attitudes toward online education: Faculty profiles in the diffusion of innovation* (Order No. 29133371). Available from ProQuest Dissertations & Theses Global. (2646728496). Retrieved from https://www.proquest.com/dissertations-theses/faculty-attitudes-toward-online-education/docview/2646728496/se-2?accountid=14196

Gillespie, K. (2021). *Effects of number talks on number sense acquisition and mathematical discourse* (Order No. 28322814). Available from ProQuest Dissertations & Theses Global. (2531554462). Retrieved from https://www.proquest.com/dissertations-theses/effects-number-talks-on-sense-acquisition/docview/2531554462/se-2

Good, T., & Brophy, J. (1977). *Educational psychology: A realistic approach*. New York: Holt, Rinehart and Winston.

Goodwin, J. (2018, August 5). *Top five math strategies for struggling students*. Magoosh Math. https://magoosh.com/math/math-strategies/

Greathouse, J. (2021). *Effects of Digital Curriculum on Standardized Tests: A Quantitative Causal-Comparative Study* (Order No. 28546992). Available from ProQuest Dissertations & Theses Global. (2546642737).

https://www.proquest.com/dissertations-theses/effects-digital-curriculum-on-standardized-tests/docview/2546642737/se-2

Gyke, C. (2020). *Effectiveness of an interdependent group contingency with randomization using the iPad* (Order No. 28255827). Available from ProQuest Dissertations & Theses Global. (2484855036). Retrieved from https://www.proquest.com/dissertations-theses/effectiveness-interdependent-group-contingency/docview/2484855036/se-2?accountid=14196

Hageman, K. (2020). *Performance of 8th-grade algebra I students assessed with and without the spark testing method* (Order No. 27959055). Available from ProQuest Dissertations & Theses Global. (2435583667). Retrieved from https://www.proquest.com/dissertations-theses/performance-8-sup-th-grade-algebra-i-students/docview/2435583667/se-2?accountid=14196

Hanselman, P. (2018). Do School Learning Opportunities Compound or Compensate for Background Inequalities? Evidence from the Case of Assignment to Effective Teachers. *Sociology of Education, 91*(2), 132–158.

Hanushek, E., Peterson, P., Talpey, L., & Woessmann, L. (2019). The Unwavering SES Achievement Gap: Trends in US Student Performance (No. w25648). Washington, DC: National Bureau of Economic Research.

Harvey, L. (2021). “Bigger than the syllabus”: Exploring second language writing instructors’ curriculum knowledge (Order No. 28719184). Available from ProQuest Dissertations & Theses Global. (2572546491). Retrieved from https://www.proquest.com/dissertations-theses/bigger-than-syllabus-exploring-second-language/docview/2572546491/se-2?accountid=14196

Hatcher, L. (2018). *Case study: Changes in elementary student mindset after mathematics anxiety and growth mindset teacher training* (Order No. 10838898). Available from ProQuest Dissertations & Theses Global. (2081135466). Retrieved from https://www.proquest.com/dissertations-theses/case-study-changes-elementary-student-mindset/docview/2081135466/se-2

Hawking, S. (2009). A stubbornly persistent illusion: The essential scientific works of albert einstein. New York: Running Press.

Heaster-Ekholm, K. L. (2020). Popular Instructional Design Models: Their Theoretical Roots and Cultural Considerations. *International Journal of Education & Development Using Information & Communication Technology, 16*(3), 50–65.

Heise, M. (2017). From No Child Left behind to Every Student Succeeds: Back to a Future for Education Federalism. *Columbia Law Review, 117*(7), 1859–1896.

Hiser, R. (2016). *What is the effect of the implementation of a common core aligned math program go math versus a math program not aligned to the common core saxon math on student achievement?* (Doctoral dissertation). Retrieved from ProQuest. (Order No. 10182416)

Hoffman, B. (2021). *Designing critical thinking experiences in online courses during a global pandemic: A narrative research study exploring community college instructors' approach to course design and promoting critical thinking skills* (Order No. 28412125). Available from ProQuest Dissertations & Theses Global. (2520822507). Retrieved from https://www.proquest.com/dissertations-theses/designing-critical-thinking-experiences-online/docview/2520822507/se-2

Hong, M. (2022). *Investigating reliabilities for multi-level data with missing values* (Order No. 28264142). Available from ProQuest Dissertations & Theses Global. (2682272650). Retrieved from https://www.proquest.com/dissertations-theses/investigating-reliabilities-multi-level-data-with/docview/2682272650/se-2

Horne Wooley, L. (2022). *Standardized testing: A comparison amongst magnet, stem, and traditional middle schools* (Order No. 29166546). Available from ProQuest Dissertations & Theses Global. (2663457774). Retrieved from https://www.proquest.com/dissertations-theses/standardized-testing-comparison-amongst-magnet/docview/2663457774/se-2?accountid=14196

Hudak, G. (2022). Understanding schools that were able to show sustained improvement within the renewal school program (Order No. 28969329). Available from ProQuest Dissertations & Theses Global. (2647707047). Retrieved from https://www.proquest.com/dissertations-theses/understanding-schools-that-were-able-show/docview/2647707047/se-2?accountid=14196

Hudson, M. (2021). *The interrelationship between the millennial Generation's aerospace work engagement and their employer retention* (Order No. 28647966). Available from ProQuest Dissertations & Theses Global. (2568519429). Retrieved from https://www.proquest.com/dissertations-theses/interrelationship-between-millennial-generation-s/docview/2568519429/se-2?accountid=14196

Hunsaker, M. R. (2019). *Graduate students perceptions of leadership theories and how they might relate to their successes in distance learning environments* (Order No. 27735662). Available from ProQuest Dissertations & Theses Global. (2381689613).

Retrieved from https://www.proquest.com/dissertations-theses/graduate-students-perceptions-leadership-theories/docview/2381689613/se-2

Hussar, B., Zhang, J., Hein, S., Wang, K., Roberts, A., Cui, J., Smith, M., Mann, F. B., Barmer, A., Dilig, R., Nachazel, T., Barnett, M., & Purcell, S. (2020). *The Condition of Education 2020* [NCES 2020-144]. https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2020144

Hutt, E., & Polikoff, M. (2018). Reasonable Expectations: A Reply to Elmendorf and Shanske 2018. *SSRN Electronic Journal* ; ISSN 1556-5068. https://doi.org/10.2139/ssrn.3185304

International Business Machines. (2021). *Summarize Statistics: Search in SPSS Statistics 25.0.0*. https://www.ibm.com/docs/en/spss-statistics/25.0.0?topic=summarize-statistics

Jad-Moussa, R. (2022). *A phenomenological study of exemplary corporate learning and development leaders leading from the heart* (Order No. 29064890). Available from ProQuest Dissertations & Theses Global. (2649039128). Retrieved from https://www.proquest.com/dissertations-theses/phenomenological-study-exemplary-corporate/docview/2649039128/se-2?accountid=14196

Jing, S. (2021). *Estimating psychometric properties of computerized multistage testing* (Order No. 28651781). Available from ProQuest Dissertations & Theses Global. (2572971681). Retrieved from https://www.proquest.com/dissertations-theses/estimating-psychometric-properties-computerized/docview/2572971681/se-2

Joyner, S. (2022). A qualitative study on the fidelity implementation of response to intervention for K-2 students in reading (Order No. 29060895). Available from

ProQuest Dissertations & Theses Global. (2644408399). Retrieved from https://www.proquest.com/dissertations-theses/qualitative-study-on-fidelity-implementation/docview/2644408399/se-2?accountid=14196

Judd, C. M., McClelland, G. H., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond* (3rd Ed.). Routledge.

Kane, T. (2016, March 3). *Never judge a book by its cover—use student achievement instead.* Retrieved from the Brookings Institution webpage: https://www.brookings.edu/research/never-judge-a-book-by-its-cover-use-student-achievement-instead/

Kannam, J., & Weiss, M. (2019, June). *Alternative education in ESSA state plans: A review of 38 states*, 1–27. https://www.aypf.org/wpcontent/uploads/2019/06/Alternative-Education-inESSA-State-Plans-A-Review-of-38-States.pdf

Kaufman, J., Cannon, J., Culbertson, S., Hannan, M. Q., Hamilton, L., & Meyers, S. (2018). *Raising the Bar: Louisiana's Strategies for Improving Student Outcomes*. Santa Monica, CA: RAND Corporation. Retrieved from: https://www.rand.org/pubs/research_reports/RR2303.html.

Kemler, J. (2022). *Politics and its role on ethical decision-making processes for elementary principals* (Order No. 28865574). Available from ProQuest Dissertations & Theses Global. (2626931601). Retrieved from https://www.proquest.com/dissertations-theses/politics-role-on-ethical-decision-making/docview/2626931601/se-2?accountid=14196

Kendrick, N., Glaze, N., & Jones, J. (2020). An analysis of saxon and glencoe math with junior kindergarten through grade 8 students (Order No. 28255453). Available from

ProQuest Dissertations & Theses Global. (2466732932). Retrieved from https://www.proquest.com/dissertations-theses/analysis-saxon-glencoe-math-with-junior/docview/2466732932/se-2?accountid=14196

Koedel, C., Li, D., Polikoff, M., Hardaway, T., & Wrabel, S. (2017). Mathematics curriculum effects on student achievement in california. *AERA Open, 3*(1), 233285841769051. doi:10.1177/2332858417690511

Koenig, K. (2022). *The achievement gap in missouri: How suburban students are progressing over time* (Order No. 29165222). Available from ProQuest Dissertations & Theses Global. (2661133812). Retrieved from https://www.proquest.com/dissertations-theses/achievement-gap-missouri-how-suburban-students/docview/2661133812/se-2?accountid=14196

Laerd Statistics (2017). One-way ANOVA using SPSS statistics. *Statistical tutorials and software guides*. https://statistics.laerd.com

LaFave, R. (2020). *A qualitative study of the preparedness of novice teachers to teach effectively in a blended learning environment* (Order No. 28150161). Available from ProQuest Dissertations & Theses Global; Publicly Available Content Database. (2451139811). Retrieved from https://www.proquest.com/dissertations-theses/qualitative-study-preparedness-novice-teachers/docview/2451139811/se-2

Lambert, R. (2018). "Indefensible, illogical, and unsupported"; Countering deficit mythologies about the potential of students with learning disabilities in mathematics. *Education Sciences, 8*(2), 72. Doi:10.3390/educsci8020072

Land, T., Bartell, T., Drake, C., Foote, M., Roth McDuffie, A., Turner, E., & Aguirre, J. (2019). *Curriculum spaces for connecting to children's multiple mathematical* *knowledge bases*. *Journal of Curriculum Studies, 51(*4), 471–493. https://doi.org/10.1080/00220272.2018.1428365

Lee, J., Lee, H., Song, J., & Bong, M. (2021). Enhancing children's math motivation with a joint intervention on mindset and gender stereotypes. *Learning and Instruction, 73*. https://doi.org/10.1016/j-learninstruc.2020.101416

Lein Authement, M. (2022). *A case study of four teachers' experiences while implementing the latest version of the eureka math curriculum in the state of louisiana* (Order No. 29167027). Available from ProQuest Dissertations & Theses Global. (2679776408). Retrieved from https://www.proquest.com/dissertations-theses/case-study-four-teachers-experiences-while/docview/2679776408/se-2?accountid=14196

Letwinsky, K., & Cavender, M. (2018). Shifting preservice teachers' beliefs and understandings to support pedagogical change in mathematics. *International Journal of Research in Education and Science, 4*(1), 106-120. https://files.eric.ed.gov/fulltext/EJ1169843.pdf

Lewis-Davis, C. (2022). *A descriptive qualitative study on algebra I teachers' perceptions of math literacy* (Order No. 29063641). Available from ProQuest Dissertations & Theses Global. (2649300386). Retrieved from https://www.proquest.com/dissertations-theses/descriptive-qualitative-study-on-algebra-i/docview/2649300386/se-2?accountid=14196

Liao, H., Li, Y., & Brooks, G. (2016). Outlier impact and accommodation methods: Multiple comparisons of Type I error rates. *Journal of Modern Applied Statistical Methods, 15*(1), 452-471. https://doi.org/10.22237/jmasm/1462076520

Lofland, K. (2021). *Conceptual and procedural instruction in mathematics: A content analysis from 1970 to 2020* (Order No. 28775621). Available from ProQuest Dissertations & Theses Global. (2624115619). Retrieved from https://www.proquest.com/dissertations-theses/conceptual-procedural-instruction-mathematics/docview/2624115619/se-2?accountid=14196

Louisiana Department of Education (2018). *2017 LEAP 2025 Operational Technical Report English Language Arts and Mathematics*. [PDF file]. https://www.louisianabelieves.com/docs/default-source/assessment/2017-leap-2025-ela-and-math-operational-technical-report.pdf?sfvrsn=1fac911f_2

Louisiana Department of Education (2019a). *2018 LEAP 2025 Grades 3-8 Operational Technical Report English Language Arts and Mathematics*. [PDF file]. https://www.louisianabelieves.com/docs/default-source/assessment/2018-leap-2025-grade-3-8-ela-and-math-technical-report.pdf?sfvrsn=c9e99d1f_2

Louisiana Department of Education. (2019b) *2019-school-leap-2025-achievement-level-summary* [Data set]. https://www.louisianabelieves.com/resources/library/elementary-and-middle-school-performance

Louisiana Department of Education (2019c). *Feb 2019 multi stats (total by school system)* [Data set]. https://www.louisianabelieves.com/resources/library/student-attributes

Louisiana Department of Education (2020). *2019 LEAP 2025 Grades 3-8 Operational Technical Report English Language Arts and Mathematics*. [PDF file]. https://www.louisianabelieves.com/docs/default-source/assessment/2019-leap-2025-grade-3-8-technical-report.pdf?sfvrsn=db2981f_4

Louisiana Department of Education. (2022). *Instructional materials review*. https://www.louisianabelieves.com/academics/ONLINE-INSTRUCTIONAL-MATERIALS-REVIEWS/curricular-resources-annotated-reviews

Luck, S., Stewart, A., Simmons, A., & Rhemtulla, M. (2021). Standardized measurement error: A universal metric of data quality for averaged event-related potentials. *Psychophysiology, 58*(6), 1–15.

Machalow, R. (2020). *Rethinking standards-textbook alignment: How elementary math textbooks are interpreting and enacting the common core state standards* (Order No. 28031191). Available from ProQuest Dissertations & Theses Global. (2454693464). Retrieved from https://search.proquest.com/dissertations-theses/rethinking-standards-textbook-alignment-how/docview/2454693464/se-2?accountid=14196

Maggio, S. A. (2021). *Value of interactivity in online training on knowledge, self- efficacy, intentions and self-reported behaviors: A case study of north carolina small dairy processors* (Order No. 28484305). Available from ProQuest Dissertations & Theses Global. (2606850138). Retrieved from https://www.proquest.com/dissertations-theses/value-interactivity-online-training-on-knowledge/docview/2606850138/se-2?accountid=14196

Mangram, C., & Solis Metz, M. (2018). Partnering for improved parent mathematics engagement. *School Community Journal, 28*(1), 273–294. https://files.eric.ed.gov/fulltext/EJ1184914.pdf

Marzano, R. (2003). *What works in schools: Translating research into action*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mayers, D. E. (2021). *Qualitative case study on the influence of online personalized problem-based learning on student engagement* (Order No. 28499521). Available from ProQuest Dissertations & Theses Global. (2564491225). Retrieved from https://www.proquest.com/dissertations-theses/qualitative-case-study-on-influence-online/docview/2564491225/se-2

McBride, B. (1996). *A convergent and discriminant validity study of several instruments used to measure and predict performance in formal geometry* (Order No. 9621878). Available from ProQuest Dissertations & Theses Global. (304264097). Retrieved from https://www.proquest.com/dissertations-theses/convergent-discriminant-validity-study-several/docview/304264097/se-2?accountid=14196

McConnell, C., Conrad, B., & Uhrmacher, P. (2020). *Lesson Planning with Purpose: Five Approaches to Curriculum Design*. Teachers College Press.

McCrae, R., & Mottus, R. (2019). A new psychometrics: What personality scales measure, with implications for theory and assessment. *Current Directions in Psychological Science, 28*, 415-420.

McDonald, M. (2018). *The nurse educator's guide to assessing learning outcomes* (Fourth edition.). Jones & Bartlett Learning.

McLaughlin, K. (2021). *Development and initial validation of an instrument measuring role-specific knowledge of academic nurse educators* (Order No. 28866318). Available from ProQuest Dissertations & Theses Global. (2609392715). Retrieved from https://www.proquest.com/dissertations-theses/development-initial-validation-instrument/docview/2609392715/se-2

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144

Metzger, S. (2018). *Elementary school children's math utility conceptions: Associations with parents' conceptions, math achievement, and home math engagement* (Order No. 10934402). Available from ProQuest Dissertations & Theses Global. (2171851970). Retrieved from https://search.proquest.com/docview/2171851970?accountid=14196

Mitra, D. (2018). *Educational change and the political process*. Routledge.

Mishra, P., Pandey, C., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). *Descriptive statistics and normality tests for statistical data. Annals of Cardiac Anaesthesia, 22*(1), 67–72. https://doi.org/10.4103/aca.ACA_157_18

Monaghan, S. (2013). *Textbooks, teachers, and middle school mathematics student achievement* (Order No. 3603466). Available from ProQuest Dissertations & Theses Global. (1469609858). Retrieved from https://search.proquest.com/docview/1469609858?accountid=14196

Morales, K. (2021). Evaluating visual and verbal prompts: Acquisition of grocery shopping skills in adolescents with autism (Order No. 28315958). Available from ProQuest Dissertations & Theses Global. (2509634560). Retrieved from https://www.proquest.com/dissertations-theses/evaluating-visual-verbal-prompts-acquisition/docview/2509634560/se-2?accountid=14196

Nargi, R. (2018). Mathematics curricula and their impact on students' mathematic achievement (Order No. 10742281). Available from ProQuest Dissertations & Theses Global. (1972637884). Retrieved from https://www.proquest.com/dissertations-theses/mathematics-curricula-their-impact-on-students/docview/1972637884/se-2?accountid=14196

National Association of Educational Progress (NAEP). (2019). *National achievement-level results: Overall NAEP achievement levels. The Nation's Report Card*. https://www.nationsreportcard.gov/mathematics/nation/achievement/?grade=4

National Center for Educational Statistics (NCES). (2019). *TIMSS 2019 U.S. Results* Retrieved from: https://nces.ed.gov/timss/results19/index.asp#/math/intlcompare

National Center for Educational Statistics (NCES). (2020). *Participating in the NAEP?* Retrieved from https://nces.ed.gov/nationsreportcard/participating

National Council of Teachers of Mathematics (NCTM). (2014). *Principles to actions: Ensuring mathematical success for all.* Reston, VA: The National Council of Teachers of Mathematics, Inc.

National Council of Teachers of Mathematics (2018). *Catalyzing change in high school mathematics: Initiating critical conversations*. Reston, VA: The National Council of Teachers of Mathematics, Inc.

Nelson, C. (2021). *Identifying as a mathematician: How modes of belonging in the math community construct an elementary student's math identity* (Order No. 28319890). Available from ProQuest Dissertations & Theses Global. (2572604842). Retrieved from https://www.proquest.com/dissertations-theses/identifying-as-mathematician-how-modes-belonging/docview/2572604842/se-2

New York State Education Department. (2019). *New York State Regents Examination in Algebra I 2018 Technical Report*. New York: Pearson.

Nimtz, J. L. (2018). *Case studies of undergraduate student interactions with an online computer adaptive instruction intermediate algebra course* (Order No. 10931353). Available from ProQuest Dissertations & Theses Global. (2102516701). Retrieved from https://www.proquest.com/dissertations-theses/case-studies-undergraduate-student-interactions/docview/2102516701/se-2

Oberhausen, M. (1975). *Curriculum And Instruction: A Conceptual Analysis Of Their Relationship In Theory And In Practice* (Order No. 7602684). Available from ProQuest Dissertations & Theses Global. (302780863). Retrieved from https://www.proquest.com/dissertations-theses/curriculum-instruction-conceptual-analysis-their/docview/302780863/se-2?accountid=14196

Oglesby-Phelps, S. (2022). *The effects of do the math intervention on tier II elementary students* (Order No. 29161924). Available from ProQuest Dissertations & Theses Global. (2688090788). Retrieved from https://www.proquest.com/dissertations-theses/effects-do-math-intervention-on-tier-ii/docview/2688090788/se-2

Okey, J. & Gagne, R. (1970), "Revision of a science topic using evidence of performance on subordinate skills", *Journal of Research in Science Teaching*, Vol. 7 No. 4, pp. 321-325.

Organization for Economic Co-operation and Development. (2018). *PISA 2015: Results in Focus.* https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

Organization for Economic Cooperation and Development. (2020). *Mathematics performance. (PISA)*. https://data.oecd.org/pisa/mathematics-performance-pisa.htm

Owens, A. (2018). Income segregation between school districts and inequality in students' achievement. Sociology of Education, 91(1), 1–27. https://doi.org/10.1177/0038040717741180

Padilla-Hernandez, R. (2020). *The impact that a new co-requisite model for entry-level college courses is having on students' success in mathematics* (Order No. 28263640). Available from ProQuest Dissertations & Theses Global. (2476174483). Retrieved from https://www.proquest.com/dissertations-theses/impact-that-new-co-requisite-model-entry-level/docview/2476174483/se-2

Pardini, P. (2022). *The impact of the success for all: Reading edge program on student achievement at a title I middle school* (Order No. 29067998). Available from ProQuest Dissertations & Theses Global. (2667749499). Retrieved from https://www.proquest.com/dissertations-theses/impact-success-all-reading-edge-program-on/docview/2667749499/se-2?accountid=14196

Patton, W., & Sickles, R., Jr. (2021). *Predictors of student success on the wisconsin forward exam at urban charter network* (Order No. 28869224). Available from ProQuest Dissertations & Theses Global. (2616950462). Retrieved from https://www.proquest.com/dissertations-theses/predictors-student-success-on-wisconsin-forward/docview/2616950462/se-2

Pelham, B. (2013). *Intermediate statistics: A conceptual course*. Los Angeles: SAGE.

Pennington, K. (2022). *"New math" challenging parental involvement with homework: A qualitative descriptive study* (Order No. 28966186). Available from ProQuest Dissertations & Theses Global. (2635491821). Retrieved from https://www.proquest.com/dissertations-theses/new-math-challenging-parental-involvement-with/docview/2635491821/se-2?accountid=14196

Perry, C., Jr. (2021). How do school leaders in high-poverty high-performing elementary schools use data to increase student achievement? a multi-case study (Order No. 28410805). Available from ProQuest Dissertations & Theses Global. (2519028644). Retrieved from https://www.proquest.com/dissertations-theses/how-do-school-leaders-high-poverty-performing/docview/2519028644/se-2

Petrilli, M. (2017). A common core curriculum quandary. *Education Next, 17*(3).

Polczynski, A. (2021). *Exploring the relationship between technology use and student performance in developmental mathematics* (Order No. 28801913). Available from ProQuest Dissertations & Theses Global. (2572559715). https://www.proquest.com/dissertations-theses/exploring-relationship-between-technology-use/docview/2572559715/se-2?accountid=14196

Polikoff, M. (2015). How well aligned are textbooks to the common core standards in mathematics? *American Education Research Journal, 52*(6), 1185.

Polikoff, M. (2018). The Challenges of Curriculum Materials as a Reform Lever. Evidence Speaks Reports, Vol 2, #58. In *Center on Children and Families at Brookings*. Center on Children and Families at Brookings.

Polikoff, M. (2021). Beyond Standards: The Fragmentation of Education Governance and the Promise of Curriculum Reform. In *Harvard Education Press*. Harvard Education Press.

Polikoff, M., Campbell, S., Rabovsky, S., Koedel, C., Le, Q., Hardaway, T., & Gasparian, H. (2020). The formalized processes districts use to evaluate mathematics textbooks. *Journal of Curriculum Studies, 52,* 451–477.

Polikoff, M., Petrilli, M., & Loveless, T. (2020). A Decade On, Has Common Core Failed? Assessing the impact of national standards. *Education Next, 20*(2), 72-81.

Prigodich, K. (2021). *Early implementation of a standards-based mathematics curriculum: Understanding teacher perspectives and concerns* (Order No. 28768332). Available from ProQuest Dissertations & Theses Global. (2620821664). Retrieved from https://www.proquest.com/dissertations-theses/early-implementation-standards-based-mathematics/docview/2620821664/se-2

Rahman, Z. (2018). *Adjunct instructors' opportunities for learning through engagement with a research-based mathematics curriculum* (Order No. 10822003). Available from ProQuest Dissertations & Theses Global. (2043559370). Retrieved from https://www.proquest.com/dissertations-theses/adjunct-instructors-opportunities-learning/docview/2043559370/se-2?accountid=14196

Rebarber, T., & McCluskey, N. (2018). Common core, school choice & rethinking standards-based reform (ED593778). ERIC. https://files.eric.ed.gov/fulltext/ED593778.pdf

Regenstein, E., Boer, B., Zavitkovsky, P., & Advance Illinois. (2018). Establishing Achievable Goals: Recommendations for Improved Goal-Setting Under the Every Student Succeeds Act. In *Advance Illinois*. Advance Illinois.

Reigeluth, C. (1992). Commentary: Elaborating the Elaboration Theory. *Educational Technology Research and Development,40*(3), 80-86. Retrieved from http://www.jstor.org/stable/30220015

Remillard, J. (2005). Examining Key Concepts in Research on Teachers' Use of Mathematics Curricula. *Review of Educational Research*, 75(2), 211-246.

Remillard, J. (2018). Mapping the relationship between written and enacted curriculum: Examining teachers' decision making. In G. Kaiser, H. Forgasz, M. Graven, A. Kuzniak, E. Simmt, & B. Xu (Eds.), *Invited Lectures from the 13th International Congress on Mathematical Education*. Springer.

Remillard, J., & Kim, O. (2020). *Elementary Mathematics Curriculum Materials; Designs for Student Learning and Teacher Enactment*. Springer International Publishing. https://doi.org/10.1007/978-3-030-38588-0

Reys, R., Reys, B., Lapan, R., Holliday, G., & Wasman, D. (2003). Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education, 34*(1), 74-95. Retrieved from https://search.proquest.com/docview/223488193?accountid=12570

Richman, A. (2021). *What are we really doing here? exploring aims for school mathematics in curricular systems* (Order No. 28769090). Available from ProQuest Dissertations & Theses Global. (2586467985). Retrieved from https://www.proquest.com/dissertations-theses/what-are-we-really-doing-here-exploring-aims/docview/2586467985/se-2

Rivero, S. (2022). *The effect of applied behavior analysis and acceptance and commitment therapy on parental involvement* (Order No. 28971486). Available from ProQuest Dissertations & Theses Global. (2642347908). Retrieved from https://www.proquest.com/dissertations-theses/effect-applied-behavior-analysis-acceptance/docview/2642347908/se-2?accountid=14196

Robertson, A. (2022). The impact of analyzing science standards on elementary teacher science self-efficacy (Order No. 29169330). Available from ProQuest Dissertations & Theses Global. (2679755392). Retrieved from https://www.proquest.com/dissertations-theses/impact-analyzing-science-standards-on-elementary/docview/2679755392/se-2?accountid=14196

Roeder, M. L. (2020). *A community of learning in an elementary school mathematics classroom* (Order No. 28095288). Available from ProQuest Dissertations & Theses Global. (2451362936). Retrieved from https://www.proquest.com/dissertations-theses/community-learning-elementary-school-mathematics/docview/2451362936/se-2

Rotermund, S., & Burke, A. (2021) *Elementary and secondary STEM education*. National Center for Science and Engineering Statistics/National Science Foundation. Retrieved from http://ncses.nsf.gov/pubs/nsb20211

Ruggeri, C. (2021). *Textbook analysis on proportional reasoning in middle school textbooks* (Order No. 28315110). Available from ProQuest Dissertations & Theses Global. (2497192152). Retrieved from https://www.proquest.com/dissertations-theses/textbook-analysis-on-proportional-reasoning/docview/2497192152/se-2

Ruk, J. (2021). *Maintenance of cognitive demand during repeated task enactments using a teaching practice that builds on student thinking* (Order No. 29150738). Available from ProQuest Dissertations & Theses Global. (2645887743). Retrieved from https://www.proquest.com/dissertations-theses/maintenance-cognitive-demand-during-repeated-task/docview/2645887743/se-2

Russell, A. (2020). *An exploration of radiologic technology faculty characteristics and the use of pedagogical practices in the classroom* (Order No. 28153250). Available from ProQuest Dissertations & Theses Global. (2475113158). Retrieved from https://www.proquest.com/dissertations-theses/exploration-radiologic-technology-faculty/docview/2475113158/se-2?accountid=14196

Salkind, N. (2010). Encyclopedia of research design (Vol. 1). Los Angeles, CA: Sage. Retrieved from http://web.utk.edu/~ewbrewer/pdf/encylopedia/Encyclopedia of Research Design_Volume 1.pdf

Schleicher, A. (2019). PISA 2018: Insights and Interpretations. *Organization for Economic Co-operation and Development*. Retrieved from: https://www.oecd.org/pisa/PISA%202018%20Insights%20and%20Interpretations%20FINAL%20PDF.pdf

Schoenfeld, A. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educational Researcher*, *43*(8), 404–412.

Schunk, D. (2018). *Learning theories. An educational perspective* (Rev. ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

Simmons, A. K. (2021). *Development and implementation of computer-based learning modules on alternative communication for pediatric nurses and nursing students* (Order No. 28319925). Available from ProQuest Dissertations & Theses Global. (2516848931). Retrieved from https://www.proquest.com/dissertations-theses/development-implementation-computer-based/docview/2516848931/se-2

Sisson, J. N. (2020). *The effect of utilizing eye closure during immediate auditory memory recall of reading passage content in third graders* (Order No. 28095750). Available from ProQuest Dissertations & Theses Global. (2449469207). Retrieved from https://www.proquest.com/dissertations-theses/effect-utilizing-eye-closure-during-immediate/docview/2449469207/se-2

Solomon, T., Dupuis, A., O'Hara, A., Hockenberry, M., Lam, J., Goco, G., Ferguson, B., & Tannock, R. (2019). A cluster-randomized controlled trial of the effectiveness of the JUMP Math program of math instruction for improving elementary math achievement. *PLoS ONE*, 14(10), 1–36. https://doi.org/10.1371/journal.pone.0223049

Sparks, S. (2018). NAEP: Gaps widen between high fliers and low scorers. *Education Week, 37*(28).

Spaull, N. and Kotze, J. (2015), "Starting behind and staying behind in South Africa: the case of insurmountable learning deficits in mathematics", International Journal of Educational Development, Vol. 41, pp. 13-24, doi: https://doi.org/10.1016/j.ijedudev.2015.01.002.

Stalbert, M. (2022). *Attention-driven design: How instructional designers design to capture the Learner's attention* (Order No. 29212963). Available from ProQuest Dissertations & Theses Global; Publicly Available Content Database. (2679205813). Retrieved from https://www.proquest.com/dissertations-theses/attention-driven-design-how-instructional/docview/2679205813/se-2

Stein, M., & Kim, G. (2009). The role of mathematics curriculum materials in large scale urban reform: An analysis of demands and opportunities for teacher learning. 135 In J.T. Remillard, B.A. Herbel-Eisenmann, & G.M. Lloyd (Eds). *Mathematics teachers* *at work: Connecting curriculum materials and classroom instruction* (37- 55). New York, NY: Routledge.

Stein, M., Remillard, J., & Smith, M. (2007). How curriculum influences student learning. In F. K. Lester (Ed.), Second Handbook of Research on Mathematics Teaching and Learning (Vol. 1, pp. 319–370). Charlotte, NC: Information Age Publishing. Retrieved from http://www.citeulike.org/group

Steiner, D. (2017). Curriculum research: What we know and where we need to go. *StandardsWork*. Available at https://standardswork.org/wp-content/ uploads/2017/03/sw-curriculum-research-report-fnl.pdf.

Sterner, G., Wolff, U., & Helenius, O. (2019). Reasoning about representations: Effects of an early math intervention. *Scandinavian Journal of Educational Research, 6*(5), 782-800

Stralek, L. (2018). *Politics and education reform from 1965 to 2016* (Order No. 10787064). Available from ProQuest Dissertations & Theses Global. (2054024760). Retrieved from https://search.proquest.com/dissertations-theses/politics-education-reform-1965-2016/docview/2054024760/se-2?accountid=14196

Superfine, A., Kelso, C., & Beal, S. (2010). Examining the process of developing a research-based mathematics curriculum and its policy implications. *Educational Policy, 24*(6), 908-934. doi:10.1177/0895904809351690

Suppa, S. (2018). *Supporting novice mathematics teacher educators to teach ambitiously via continuously improved curriculum materials* (Order No. 10844882). Available from ProQuest Central. (2130966972). Retrieved from https://search.proquest.com/docview/2130966972?accountid=12570

Tanck, H. (2021). *Mapping middle school mathematics teachers' curriculum assemblages: A post-qualitative inquiry* (Order No. 28416682). Available from ProQuest Dissertations & Theses Global. (2540502668). Retrieved from https://www.proquest.com/dissertations-theses/mapping-middle-school-mathematics-teachers/docview/2540502668/se-2?accountid=14196

Troppe, P., Milanowski, A., Heid, C., Gill, B., Ross, C., & Johnson, E. (2017, May). *Implementation of Title I and Title II-A program initiatives: Results from 2013-14 executive summary*. Retrieved from https://ies.ed.gov/ncee/pubs/20174014/pdf/20174015.pdf

Tyson, H., & Woodward, A. (1989). Why Students Aren't Learning Very Much from Textbooks. *Educational Leadership, 47*(3), 14.

Udensi, C. (2019). Nursing students' contextual perspectives in aComputer-aided simulated environment: Qualitative descriptive case study (Order No. 27665179). Available from ProQuest Dissertations & Theses Global. (2322785417). Retrieved from https://www.proquest.com/dissertations-theses/nursing-students-contextual-perspectives/docview/2322785417/se-2?accountid=14196

Uibelhoer, D. (2020). *Practicing what they preach: A case study exploring the experiences of instructional designers as educators of an online teaching certificate program* (Order No. 28258574). Available from ProQuest Dissertations & Theses Global. (2469533673). Retrieved from https://www.proquest.com/dissertations-theses/practicing-what-they-preach-case-study-exploring/docview/2469533673/se-2?accountid=14196

United States Department of Education. (2021). [Every Student Succeeds Act]. ed.gov/essa?src=rn

United States Department of Education, National Center for Education Statistics (NCES). (2019). *The Condition of Education 2019* (NCES 2019-144)

Vareberg, K. (2021). *"Cause you Don't really need a teacher to learn stuff": Theorizing a 'Lanes of learning' model of informal, self-directed learning* (Order No. 28647125). Available from ProQuest Dissertations & Theses Global. (2572606721). Retrieved from https://www.proquest.com/dissertations-theses/cause-you-don-t-really-need-teacher-learn-stuff/docview/2572606721/se-2?accountid=14196

Walker, E. (2022). *Examining the relationship between leadership actions, teacher efficacy and student outcomes: An exploration of high, moderate, and low-achieving schools* (Order No. 29165536). Available from ProQuest Dissertations & Theses Global. (2659542166). Retrieved from https://www.proquest.com/dissertations-theses/examining-relationship-between-leadership-actions/docview/2659542166/se-2?accountid=14196

Walker, J. (2021). *Suspension or intervention: An exploration of administrators' attitudes toward exclusionary discipline practices of African American males in Louisiana middle schools* (Order No. 28410267). Available from ProQuest Dissertations & Theses Global.

Waller, A. (2021). *The effects of departmentalization on elementary students' academic achievement in mathematics and science* (Order No. 28548069). Available from ProQuest Dissertations & Theses Global. (2558435299). Retrieved from https://www.proquest.com/dissertations-theses/effects-departmentalization-on-elementary/docview/2558435299/se-2?accountid=14196

Walsh, T. (2009). The effect of saxon math on student achievement of sixth-grade students (Order No. 3362003). Available from ProQuest Central. (305063525). Retrieved from https://search.proquest.com/docview/305063525?accountid=12570

Wang, Z., & McDougall, D. (2019). Curriculum matters: What we teach and what students gain. *International Journal of Science and Mathematics Education, 17*(6), 1129-1149.

Westbury, I. (1990). Textbooks, textbook publishers, and the quality of schooling. *Yearbook (National Society for the Study of Education), 89th*(1), 1–22.

Whilby, C. (2020). *Closing academic achievement gap between economically advantaged and economically disadvantaged students: An improvement study* (Order No. 28150005). Available from ProQuest Dissertations & Theses Global. (2464172209). Retrieved from https://search.proquest.com/dissertations-theses/closing-academic-achievement-gap-between/docview/2464172209/se-2?accountid=14196

White, C. (2018). *A study of the implementation of a middle school math program and student achievement* (Order No. 10977597). Available from ProQuest Dissertations & Theses Global. (2385696027). Retrieved from https://www.proquest.com/dissertations-theses/study-implementation-middle-school-math-program/docview/2385696027/se-2?accountid=14196

Williams, T. (2019). The influence teacher quality has on student performance on the LEAP (Order No. 27829708). Available from ProQuest Dissertations & Theses Global. (2391985778). Retrieved from https://www.proquest.com/dissertations-theses/influence-teacher-quality-has-on-student/docview/2391985778/se-2

Womersley, H. (2022). Attributes of differentiated instruction in southeastern united states catholic school math classrooms (Order No. 29061672). Available from ProQuest Dissertations & Theses Global. (2679117722). Retrieved from https://www.proquest.com/dissertations-theses/attributes-differentiated-instruction/docview/2679117722/se-2?accountid=14196

Zhou, B. (2020). *A study of task-based language instruction in flipped english as foreign language classrooms in china* (Order No. 28262590). Available from ProQuest Dissertations & Theses Global. (2495504337). Retrieved from https://www.proquest.com/dissertations-theses/study-task-based-language-instruction-flipped/docview/2495504337/se-2

APPENDICES

APPENDIX A:

Copy of: RRB Approval Letter Copy

Southwest Baptist
UNIVERSITY

COLLEGE OF PROFESSIONAL PROGRAMS
1600 University Avenue
Bolivar, Missouri 65613
(417) 328-2099

December 30, 2022

Re: Difference in Student Achievement Between Math Curriculum Choice in the State of Louisiana

Dear Mr. Neugebauer,

On December 30, 2022 a review of your application and supporting documents for the above named research proposal was completed. The Research Review Board (RRB) for Southwest Baptist University has determined that the proposed research project meets the criteria for Exempt status as per policy 1.15.3 (A.1) in the faculty guidelines. As per the above policy "If the project is certified exempt, the principle investigator need not resubmit the project for continuing RRB review as long as there are no modifications in the exempted procedures". The study has now been approved, therefore, work on the project may begin.
If any modifications to the exempted procedures are made, the RRB will need to complete a new review of the changes to determine if the project remains Exempt or if further review is necessary.

Congratulations on the approval of your project, we wish you well during its completion.

Sincerely,

Colleen Shuler

Colleen Shuler, Ed.D.
Chair, Research Review Board
Assistant Professor of Education

APPENDIX B**:**

Copy of Email: Initial Survey

Dear (**Insert administrator name**),

My name is Brian Neugebauer, and I am a doctoral candidate at Southwest Baptist University. As part of the requirements to complete my Ed.D. in Educational Leadership, I am conducting a research study on math curriculum material choice and student achievement. This is an area of study lacking in comprehensive research and the reason for my contact. I am evaluating the 2016-17, 2017-18 and 2018-2019 school years to attempt to prevent any influence from the COVID-19 pandemic on the data.

(Insert school name) has been selected for participation in this study because your current school building meets the specific criteria for my research. Your participation is essential for successful completion of my research, which will require numerous accurate responses to be shown as valid. Individual school data will not be reported, and additional informed consent details are on the first page of the survey link. The test group for this survey showed the entire survey took less than five minutes once the curriculum material choice(s) was known.

*You may find it beneficial to first discuss with a colleague what curriculum materials were in use at your building (usually a textbook or similarly purchased material) to accurately provide answers for the years above and then return to the survey within the two-week window before going further. *

Survey Link: (**Insert survey link here**)

As a participation incentive, you have the option of being included in a random drawing for one of five $20 Amazon gifts cards once the research goal is met. You may also request the results of this research once it is finished. If you have any questions, please don’t hesitate to email, call, or text using the contact information below:

Email(s): (**Insert work email here**) (work email) or (**Insert survey email here**) (survey email)

Personal cellphone: (**Insert cellphone number here**)

Thank you for your consideration and dedication to advancing the field of curriculum material research. I am hoping to conclude my data collection within two weeks.

With appreciation,

Brian D. Neugebauer

Doctoral Candidate, Southwest Baptist University

APPENDIX C:

Copy of Email: One Week Follow Up

Dear (**Insert administrator name**),

My name is Brian Neugebauer, I previously emailed you regarding a research survey for my doctoral dissertation last (**Insert date here**)

I realize life is busy, which is why I am attempting to reach out again. Your contribution to my research would be helpful as I will try to conclude my data collection in approximately one week. Thankfully I am (**Insert percent here**) toward reaching a high enough survey participant number to facilitate statistically valid and reliable research results, but I need your help to reach the goal. Your help is truly vital, each school counts.

Survey Link: (**Insert survey link here**)

With gratitude,

Brian D. Neugebauer

Doctoral Candidate, Southwest Baptist University

*I have included the original email below in case you wish to review it. *

My name is Brian Neugebauer, and I am a doctoral candidate at Southwest Baptist University. As part of the requirements to complete my Ed.D. in Educational Leadership, I am conducting a research study on math curriculum material choice and student achievement. This is an area of study lacking in comprehensive research and the reason for my contact. I am evaluating the 2016-17, 2017-18 and 2018-2019 school years to attempt to prevent any influence from the COVID-19 pandemic on the data.

(Insert school name) has been selected for participation in this study because your current school building meets the specific criteria for my research. Your participation is essential for successful completion of my research, which will require numerous accurate responses to be shown as valid. Individual school data will not be reported, and additional informed consent details are on the first page of the survey link. The test group for this survey showed the entire survey took less than five minutes once the curriculum material choice(s) was known.

*You may find it beneficial to first discuss with a colleague what curriculum materials were in use at your building (usually a textbook or similarly purchased material) to accurately provide answers for the years above and then return to the survey within the two-week window before going further. *

Survey Link: (**Insert survey link here**)

As a participation incentive, you have the option of being included in a random drawing for one of five $20 Amazon gifts cards once the research goal is met. You may also request the results of this research once it is finished. If you have any questions, please don't hesitate to email, call, or text using the contact information below:

Email(s): (**Insert work email here**) (work email) or (**Insert survey email here**) (survey email)

Personal cellphone: (**Insert cellphone number here**)

Thank you for your consideration and dedication to advancing the field of curriculum material research. I am hoping to conclude my data collection within two weeks.

APPENDIX D:

Copy of Email: End of Two Weeks

(**Insert administrator name**),

This is Brian Neugebauer, I previously emailed you about my research survey for my doctoral dissertation (**Insert date here**). My research is currently near (**Insert percent here**) of the survey participation goal and I can use your help to reach the needed level for valid and reliable research.

(Insert school name) can contribute to my research in a meaningful way if you are able to complete the five-minute survey related to math curriculum materials (usually a textbook or similar purchased program). All you need to complete the survey is the textbook/curriculum material for any combination of 3rd, 4th and 5th grade for the 2016-2017, 2017-18 and 2018-19 school years.

I do realize there has been faculty, administrator, and curriculum material turnover since the above school years. Sadly, to avoid the COVID-19 pandemic impacting the research, I needed to use those years. Even if you are not able to determine the curriculum materials from memory or are newer to the building, my hope is a veteran teacher in the building, a school secretary or someone similar with knowledge from those years would be able to help.

While full survey completion is desired, even partial completion with accurate information you find from those years would be better than none. I know your job duties keep you busy, for that reason why I am attempting another email. If you are not able to help with my research, I truly understand and will remove you from my email list provided by the LDOE upon your request.

If you would like me to resend the original email information, have questions, or need anything else simply let me know.

Survey Link: (**Insert survey link here**)

Thank you for your time and support,

Brian D. Neugebauer

Doctoral Candidate, Southwest Baptist University

APPENDIX E:

Copy of Email: Still in Need of More Data

Dear (**Insert administrator name**),

My name is Brian Neugebauer, I previously emailed you regarding my dissertation research. Sadly, I did not obtain the number of responses I needed over the past two weeks and am hopeful this last email provides you the opportunity to capitalize on the incentive opportunity outlined below and me with the survey data I need.

You have been selected for participation in this study because your current school building meets the specific criteria for my research and your participation is essential for successful completion of my research. The link to the 5-to-10-minute survey utilizes drop boxes for completion and is included below. Individual school data will not be reported.

Survey Link: (**Insert survey link here**)

As a participation incentive you have the option of being included in a random drawing for one of five $20 Amazon gifts cards once the research goal is met. You may also request the results of this research once it is finished. If you have any questions, please email, call, or text, using the contact information below:

Email(s): (**Insert work email here**) (work email) or (**Insert survey email here**) (survey email) Personal cellphone: (**Insert cellphone number here**)

Sincerely,

Brian D. Neugebauer

Doctoral Candidate, Southwest Baptist University

APPENDIX F:

Copy of: Phone Call Transcript

Introduction:

My name is Brian Neugebauer, and I am a doctoral candidate at the Southwest Baptist University. I previously attempted to email (**insert administrators name**) but have not heard back. Are they available?

If administrator is available:

Hello, this is Brian Neugebauer, and I am thankful to get a hold of you. I emailed you recently to gather curriculum material information about your school for my doctoral research. Sadly, I did not obtain the number of responses I needed over the past two weeks and am calling to gain more responses. I know it might have been filtered into a junk or spam folder, which is why I called. Would you have time in the next day or two to complete the survey from my email? I can even send the email again if you need.

(Wait for response to continue or conclude conversation)

Thank you, please call me back at this number if you have questions and remember there is a drawing for one of five $20 Amazon gift cards for each person who completes the survey.

If administrator is not available:

I understand they are busy. Sadly, I did not obtain the number of responses I needed over the past two weeks and am calling to gain more responses. I know the email I sent might have been filtered into a junk or spam folder, which is why I called. What might be the best email for (**insert administrators name)**?

(Wait for response to continue or conclude conversation)

Thank you, could you please let them know I called and would greatly value their participation. If they have any questions, my phone number is (**Insert cellphone number here**), again my name is Brian.

APPENDIX G:

Gathered Data Details by Year, Source, Tier, and Type

Table G1

Source Type of Sample Data by Year and Grade

School Year	Grade	LDOE Data	Principal Survey	Parish/District Level Contact	Collected Total	Possible Total
2016-2017	3	0	72	264	336	786
	4	0	73	263	336	778
	5	0	70	262	332	757
2017-2018	3	252	68	266	586	789
	4	247	70	266	583	782
	5	24	67	264	575	764
2018-2019	3	241	73	266	580	761
	4	243	75	267	585	770
	5	238	72	264	574	751

Table G2

Number of Sample Data Points per Curriculum Tier by Year and Grade

School Year	Grade	Tier 1	Tier 2	Tier 3	Non-Tiered	Total
2016-2017	3	274	0	59	3	336
	4	273	0	60	3	336
	5	271	0	59	2	332
2017-2018	3	471	0	92	23	586
	4	468	0	92	23	583
	5	463	0	90	22	575
2018-2019	3	569	0	3	8	580
	4	572	0	4	8	584
	5	568	0	4	2	574

Table G3

Number of Sample Data Points per Curriculum Material by Year and Grade within Tier 1

School Year	Grade	Bridges in Math	Eureka Math	Ready Louisiana Mathematics	Zearn Math	Tier 1 Combination
2016-2017	3	4	245	2	23	0
	4	4	246	2	21	0
	5	4	245	1	21	0
2017-2018	3	4	416	2	36	13
	4	4	410	2	39	13
	5	4	405	2	39	13
2018-2019	3	4	400	24	140	1
	4	5	402	24	141	1
	5	4	396	25	142	1

Note. Tier 1 Combination designates multiple materials report at a school for a specific year and grade, but still utilized by the researcher since both were within Tier 1.

Table G4

Number of Sample Data Points per Curriculum Material by Year and Grade within Tier 3

School Year	Grade	Common Core Math Benchmark	enVision Math and enVision Math 2.0	HMH Go Math	HMH Math in Focus	My Math
2016-2017	3	1	2	42	0	15
	4	1	2	42	0	15
	5	1	2	41	0	15
2017-2018	3	1	5	70	1	15
	4	1	5	70	1	15
	5	1	5	68	1	15
2018-2019	3	0	0	3	0	0
	4	0	0	4	0	0
	5	0	0	4	0	0

Note. In 2016-2017 and 2017-2018, one reported use of enVision 2.0 occurs for each grade.

Table G5

Number of Sample Data Points per Curriculum Material by Year and Grade within Non-Tiered

School Year	Grade	Non-Tiered
2016-2017	3	3
	4	3
	5	2
2017-2018	3	23
	4	23
	5	22
2018-2019	3	8
	4	8
	5	2

APPENDIX H:

Third Grade Outlier Boxplots for Each Year

Figure H1

2016-2017 Grade 3 Outlier Boxplot

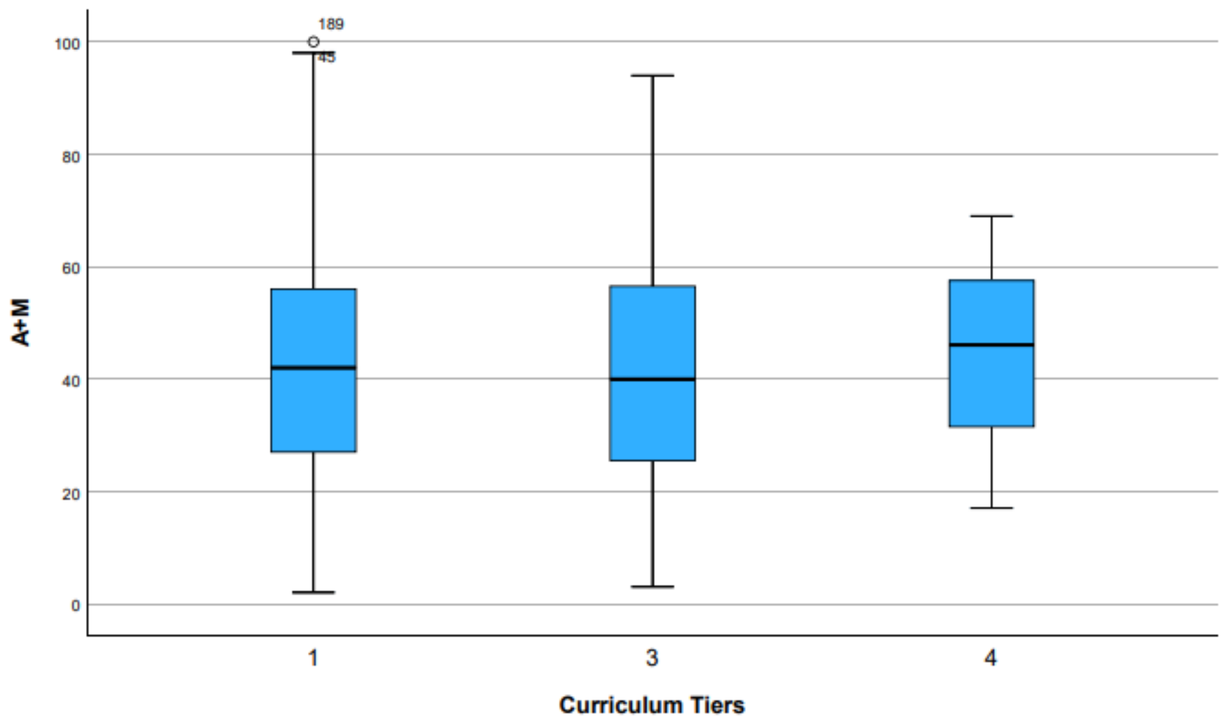


Figure H2

2017-2018 Grade 3 Outlier Boxplot

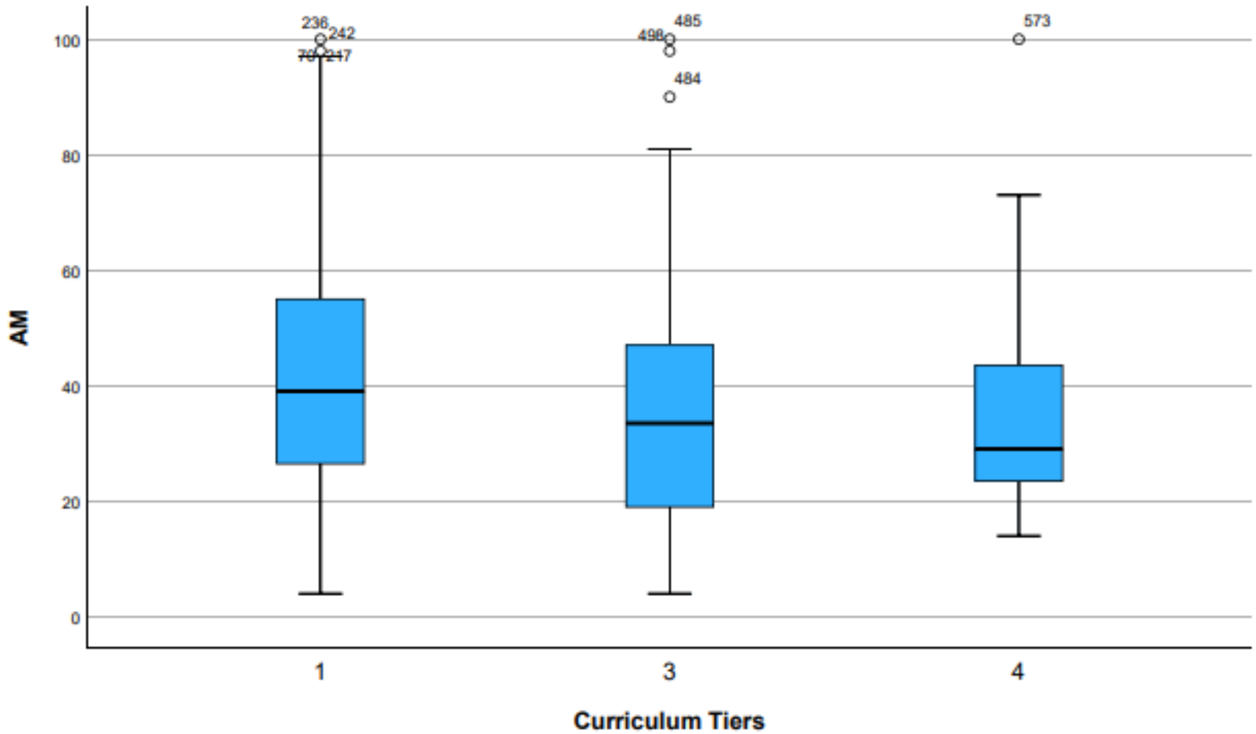
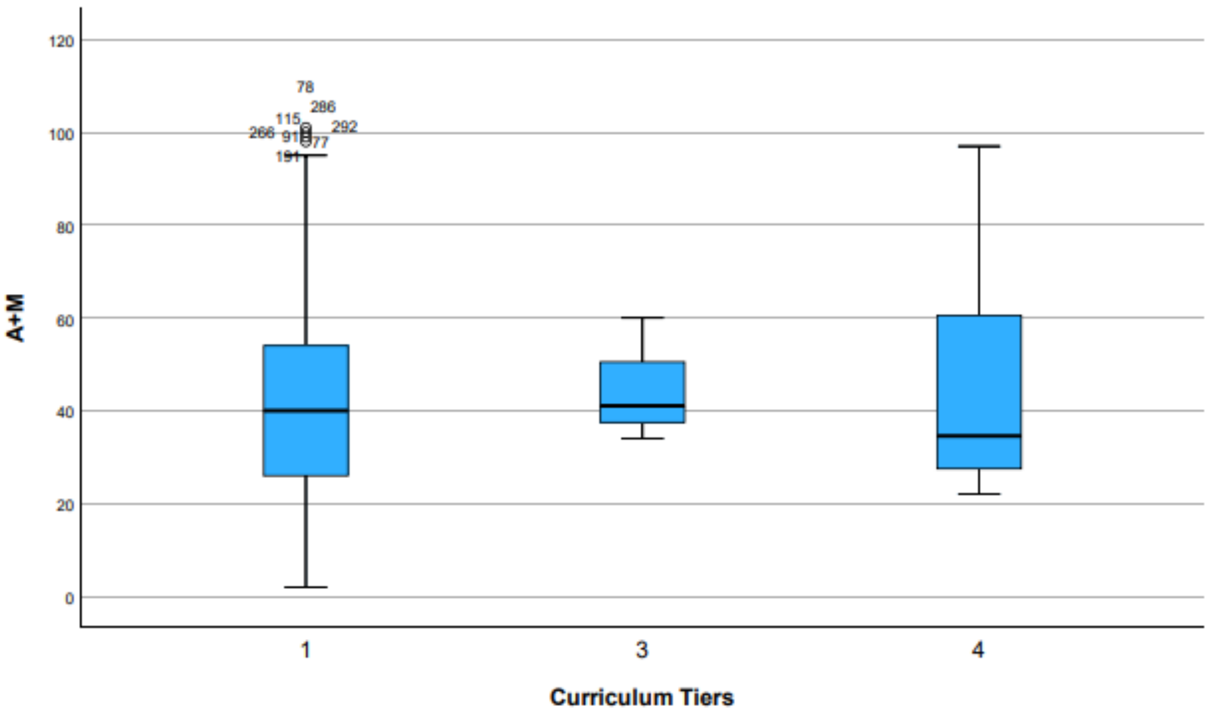


Figure H3

2018-2019 Grade 3 Outlier Boxplot



APPENDIX I:

Fourth Grade Outlier Boxplots for Each Year

Figure I1

2016-2017 Grade 4 Outlier Boxplot

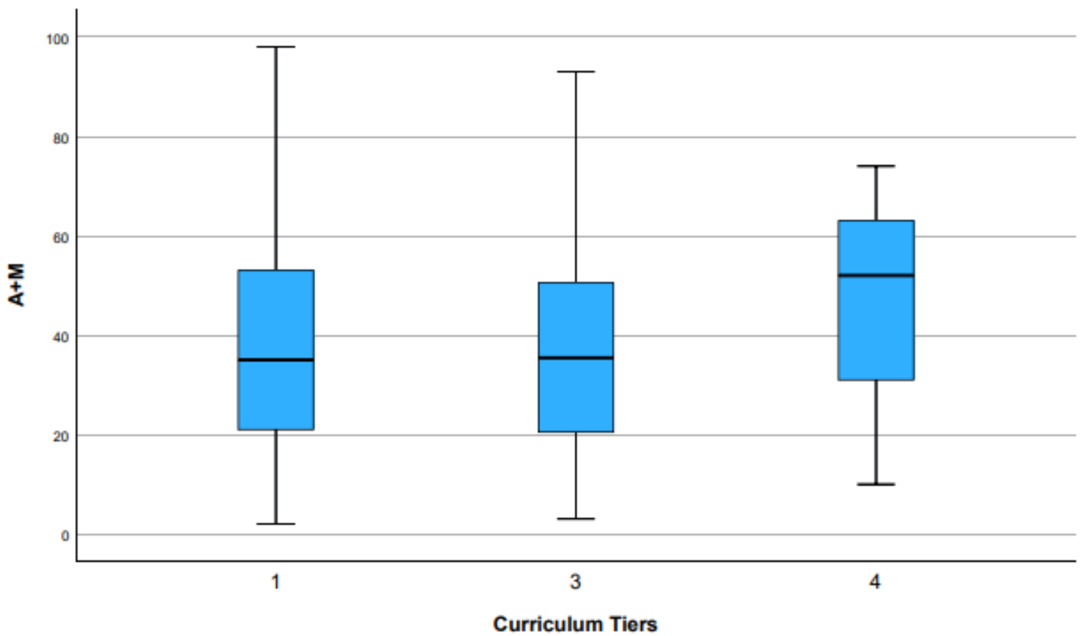


Figure I2

2017-2018 Grade 4 Outlier Boxplot

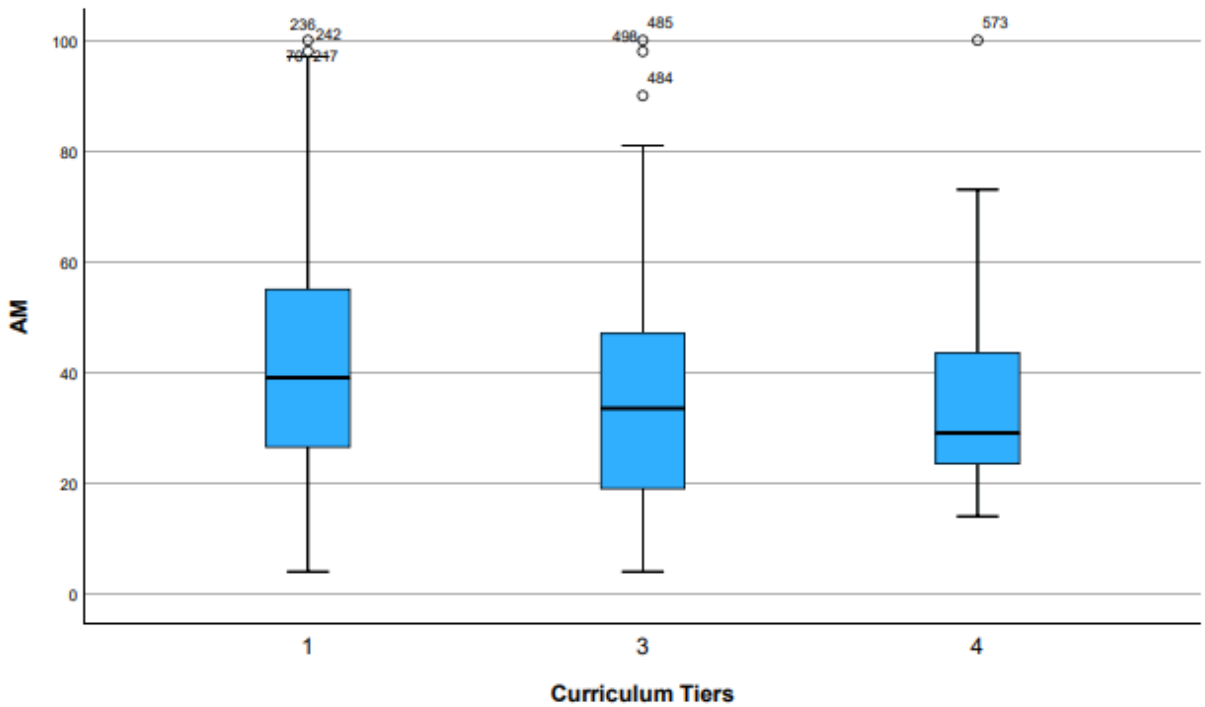
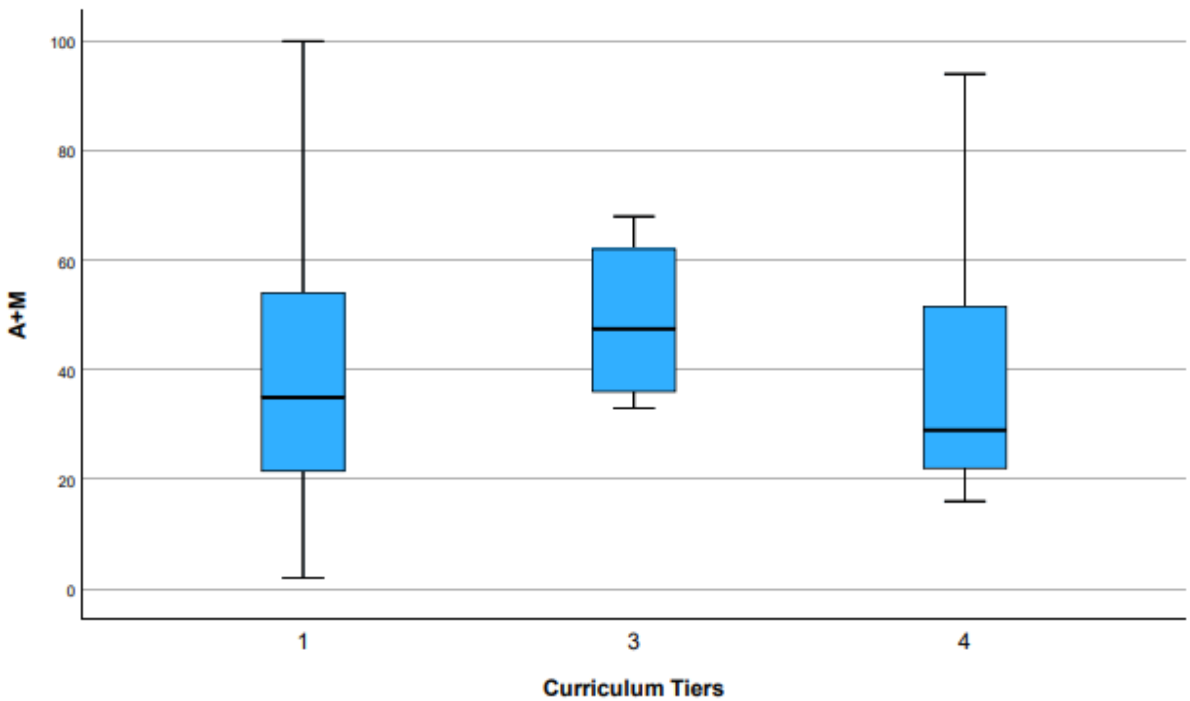


Figure I3

2018-2019 Grade 4 Outlier Boxplot



APPENDIX J:

Fifth Grade Outlier Boxplots for Each Year

Figure J1

2016-2017 Grade 5 Outlier Boxplot

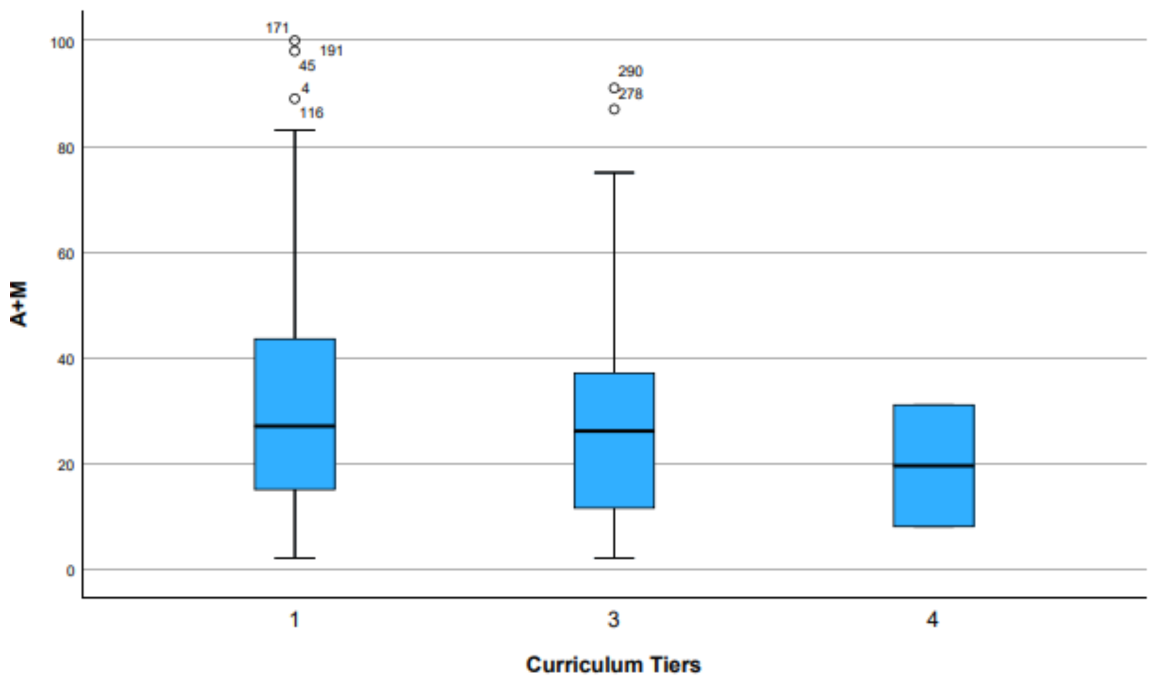


Figure J2

2017-2018 Grade 5 Outlier Boxplot

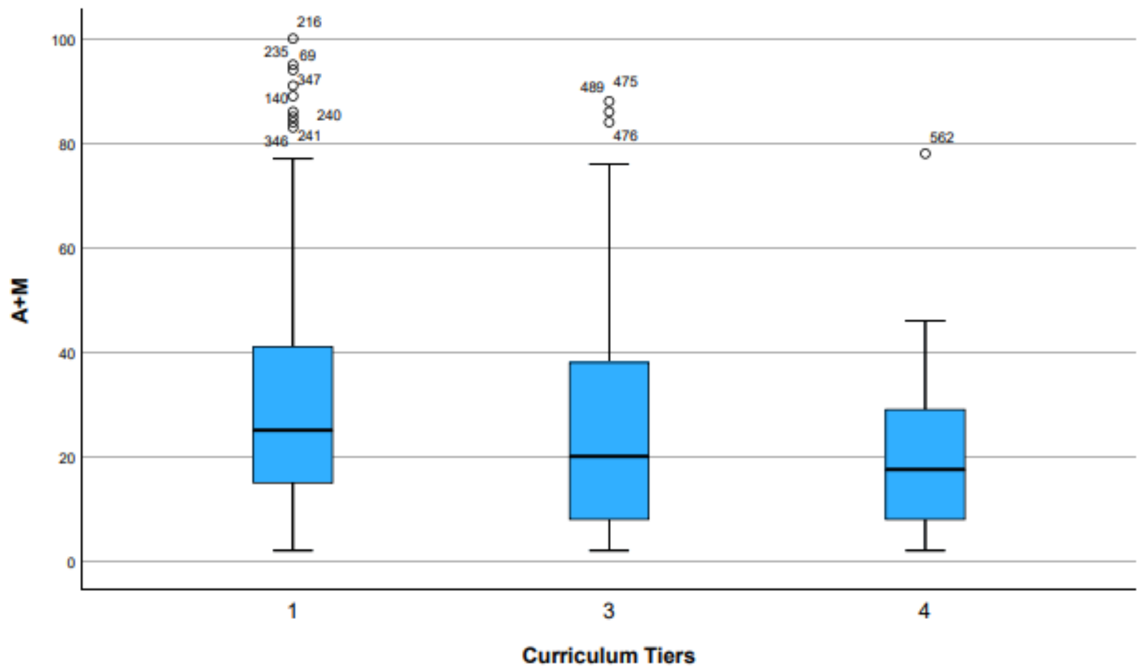
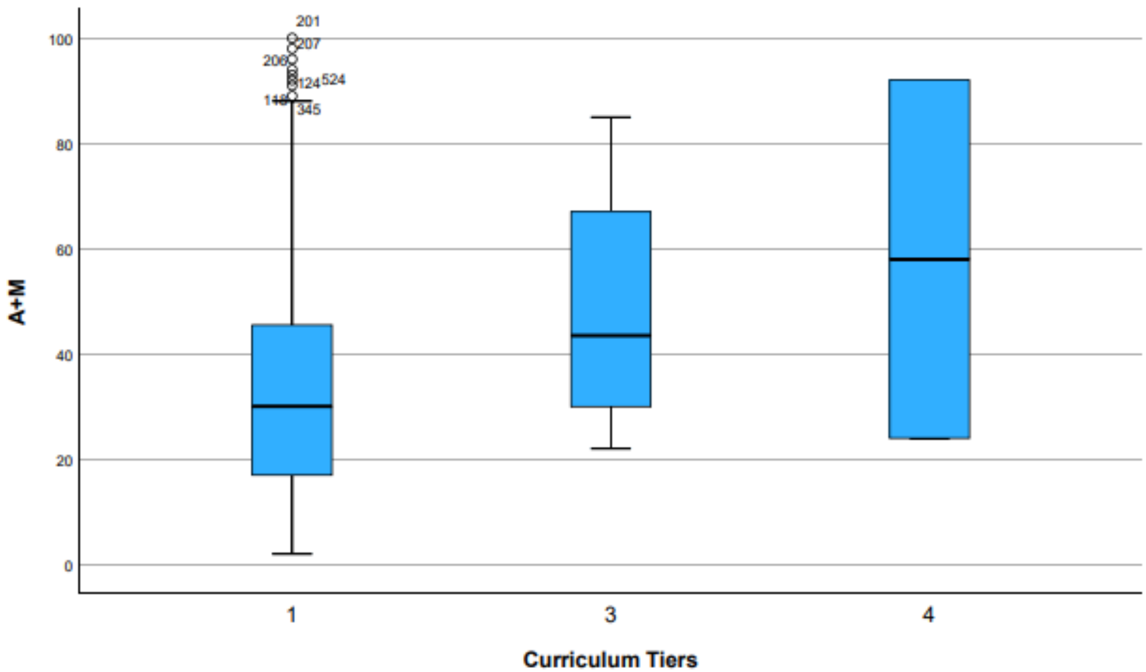


Figure J3

2018-2019 Grade 5 Outlier Boxplot



APPENDIX K:

Figure K1

2016-2017 Grade 3 Tier 1 Histogram and Q Q-Plot

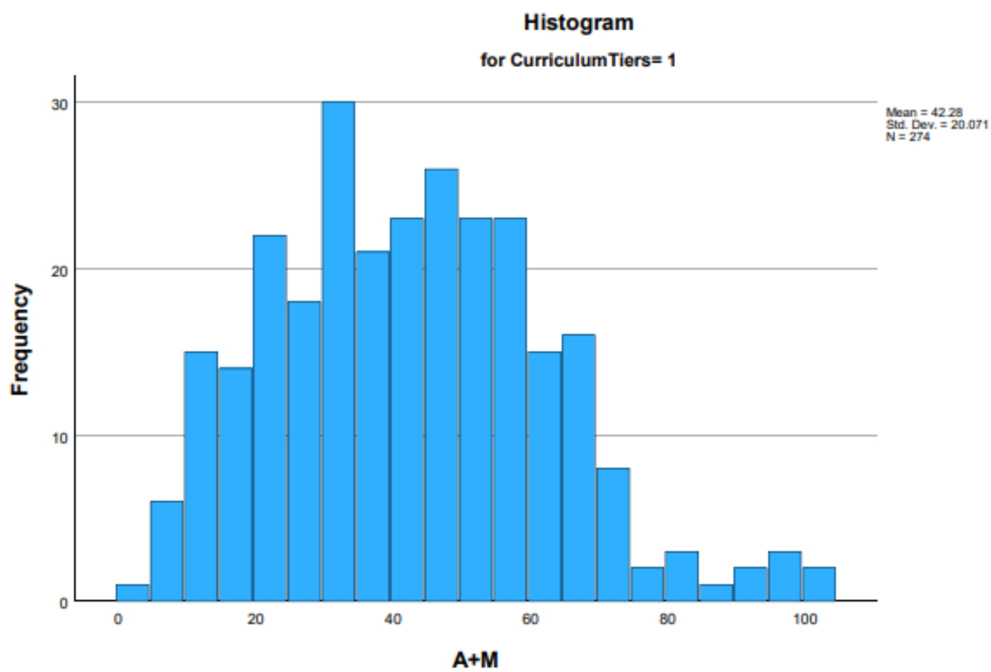


Figure K2

2016-2017 Grade 3 Tier 1 Q Q-Plot

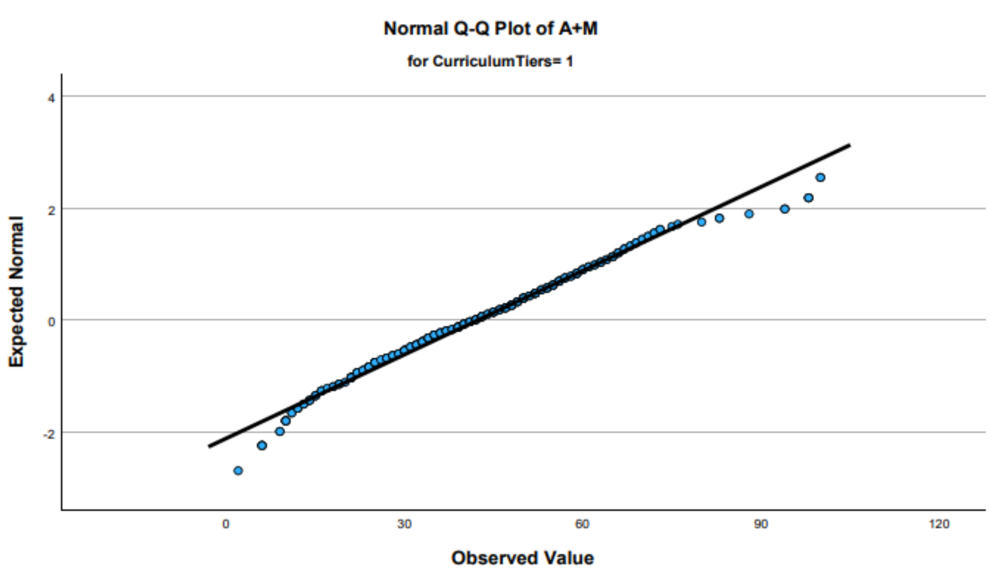


Figure K3

2017-2018 Grade 3 Tier 1 Histogram and Q Q-Plot

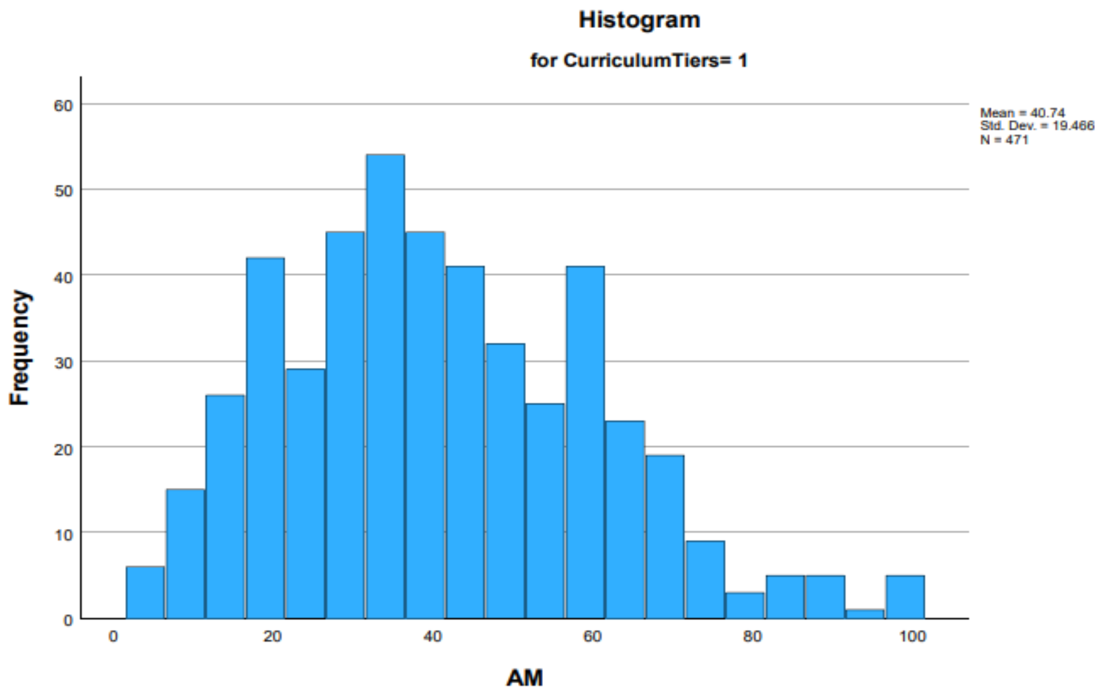


Figure K4

2017-2018 Grade 3 Tier 1 Q Q-Plot

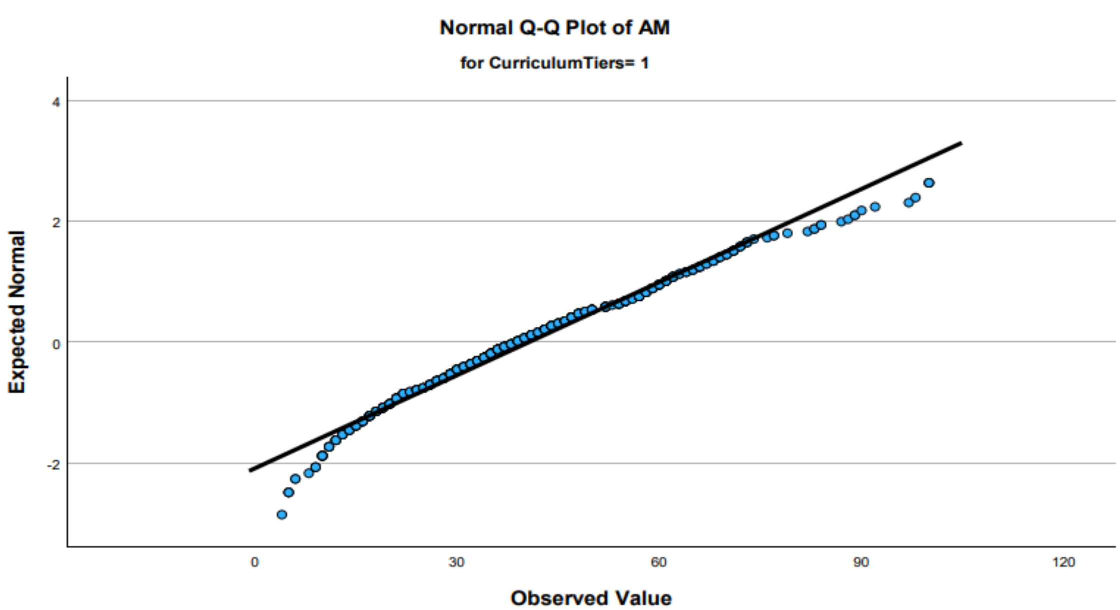


Figure K5

2017-2018 Grade 3 Tier 3 Histogram

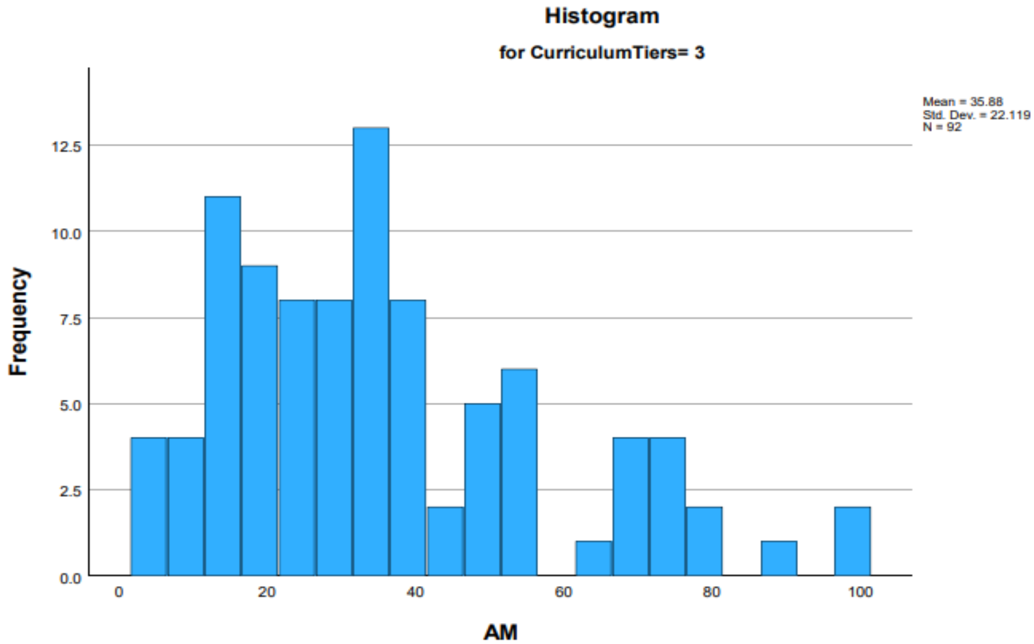


Figure K6

2017-2018 Grade 3 Tier 3 Q Q-Plot

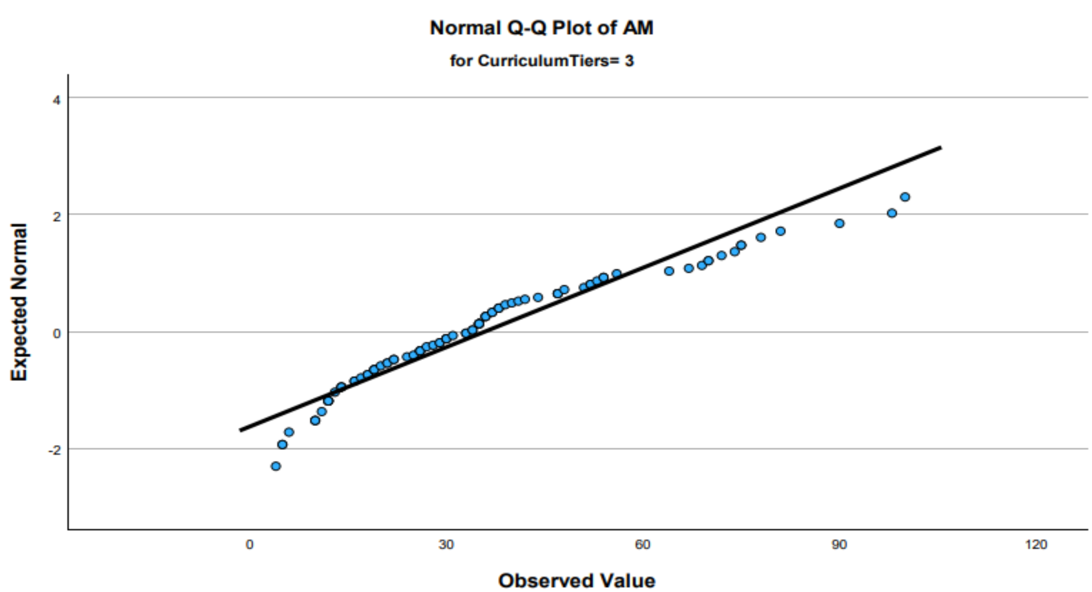


Figure K7

2017-2018 Grade 3 Non-tiered (Tier 4) Histogram

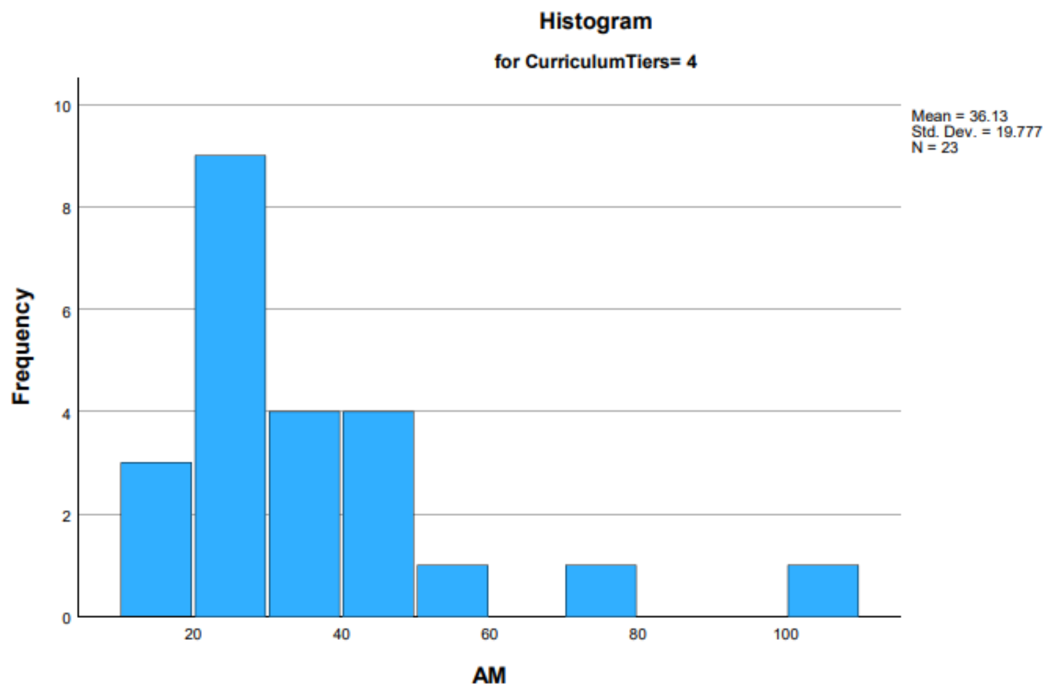


Figure K8

2017-2018 Grade 3 Non-tiered (Tier 4) Q Q-Plot

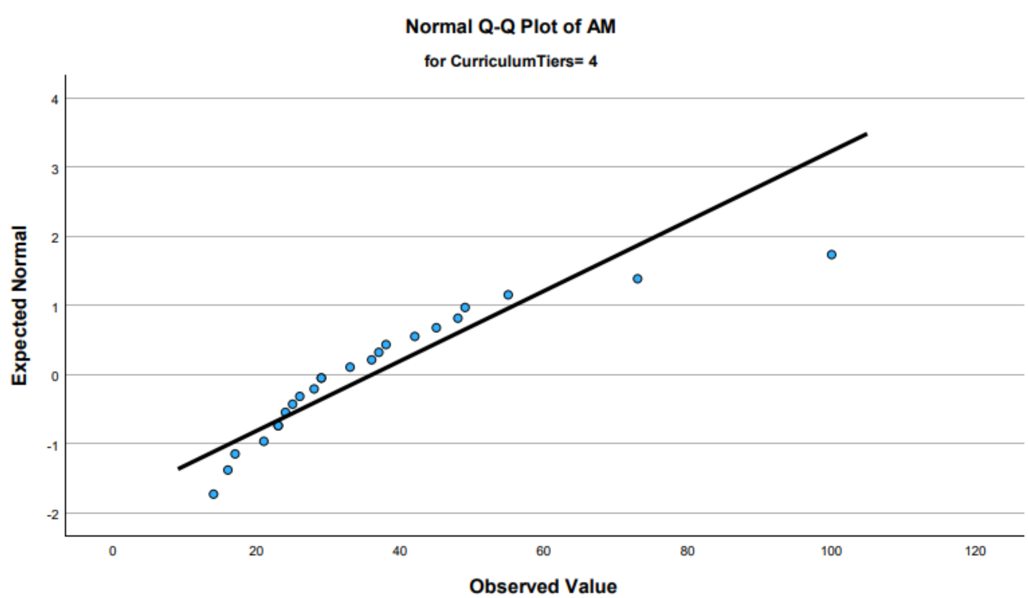


Figure K9

2018-2019 Grade 3 Tier 1 Histogram

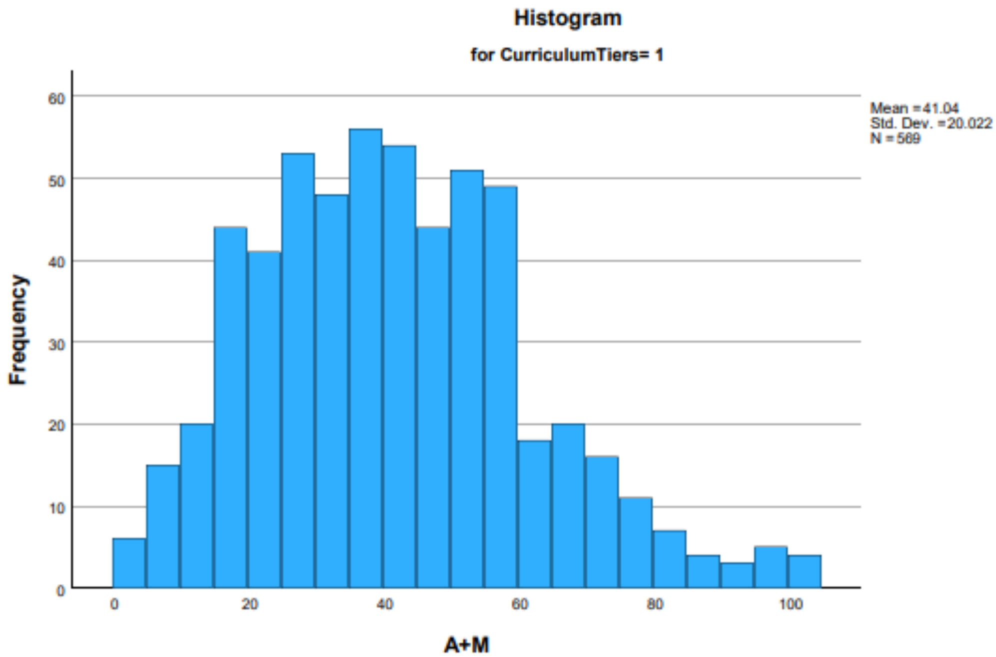
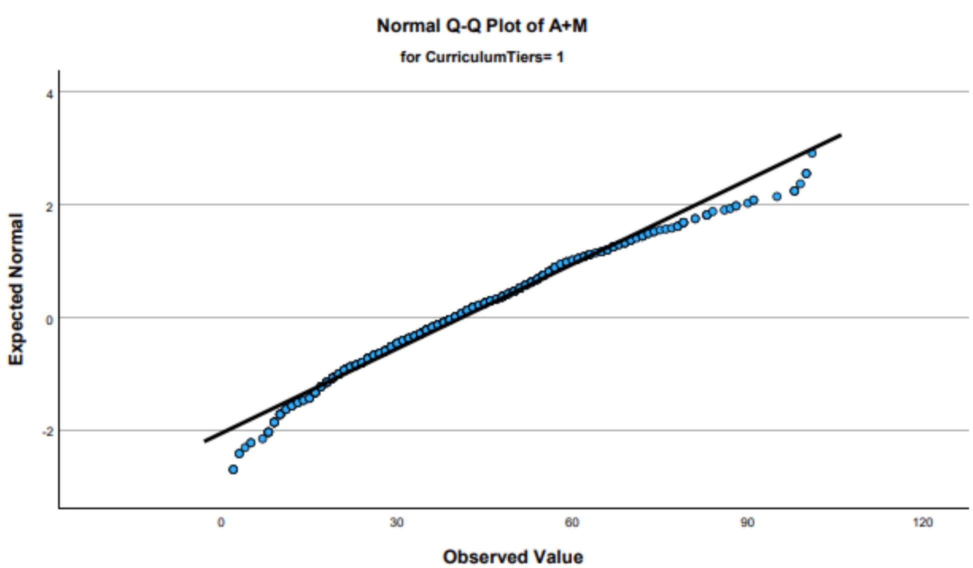


Figure K10

2018-2019 Grade 3 Tier 1 Q Q-Plot



APPENDIX L:

Fourth Grade Histograms and Q Q-Plots for Each Year

Figure L1

2016-2017 Grade 4 Tier 1 Histogram

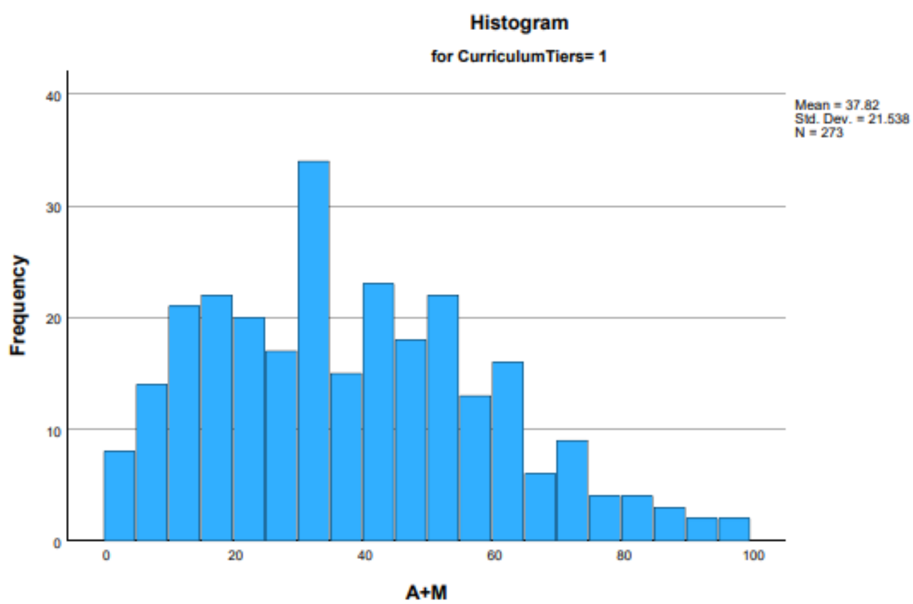


Figure L2

2016-2017 Grade 4 Tier 1 Q Q-Plot

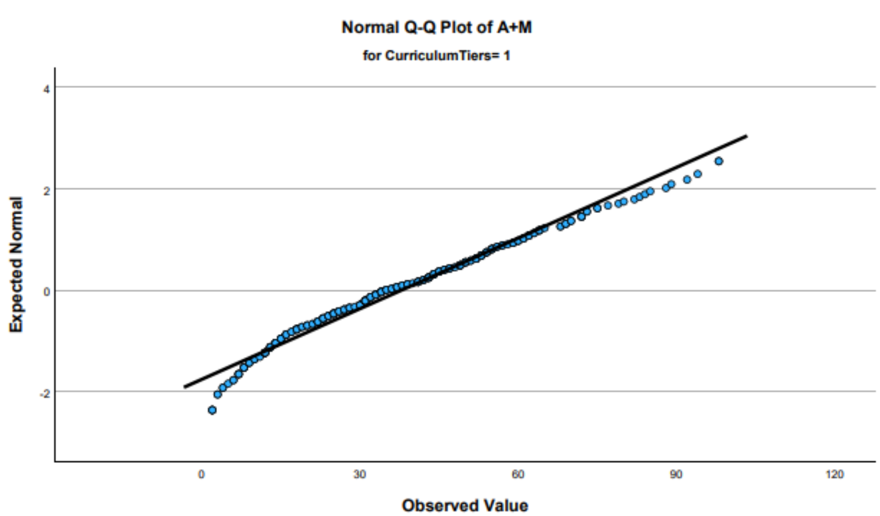


Figure L3

2016-2017 Grade 4 Tier 3 Histogram

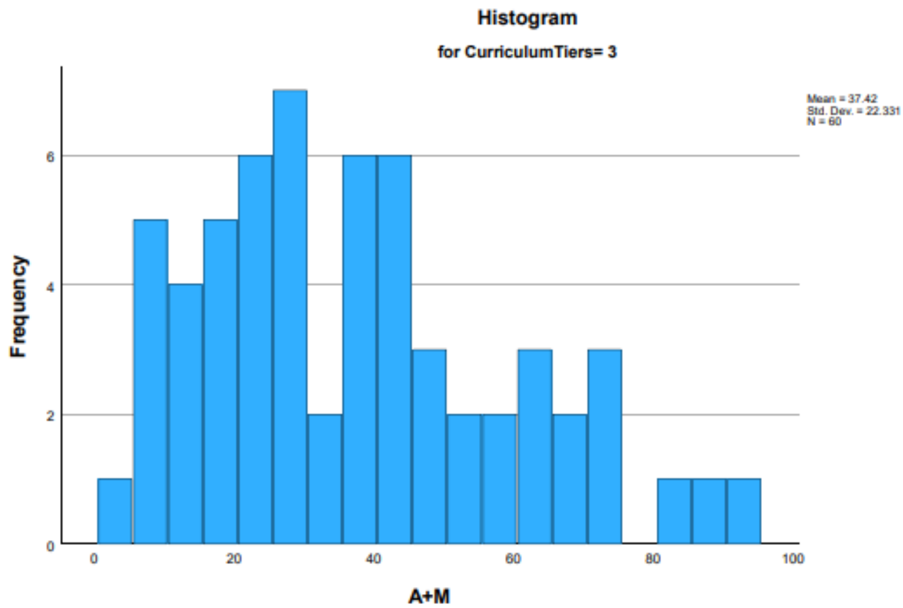


Figure L4

2016-2017 Grade 4 Tier 3 Q Q-Plot

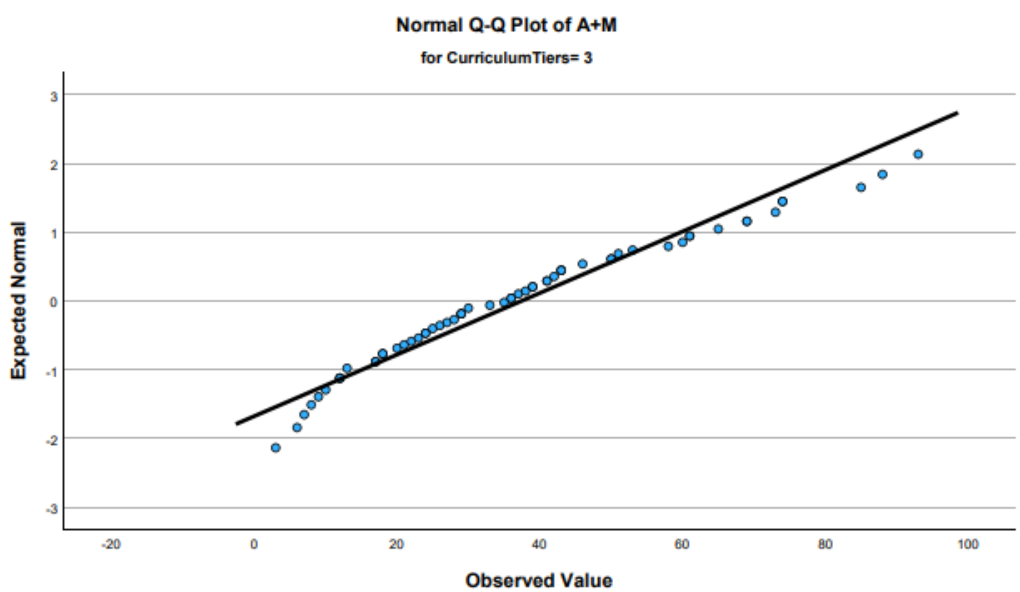


Figure L5

2017-2018 Grade 4 Tier 1 Histogram

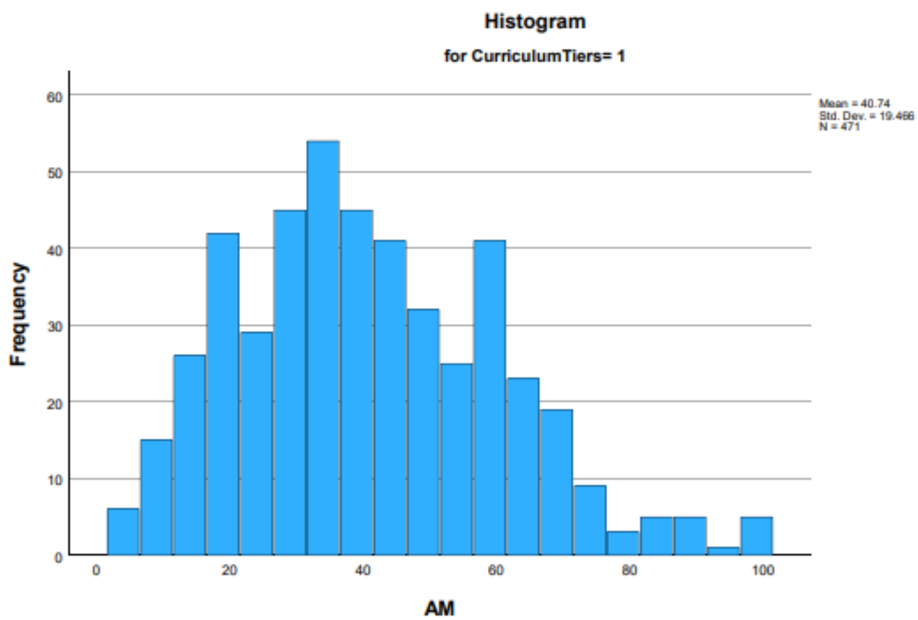


Figure L6

2017-2018 Grade 4 Tier 1 Q Q-Plot

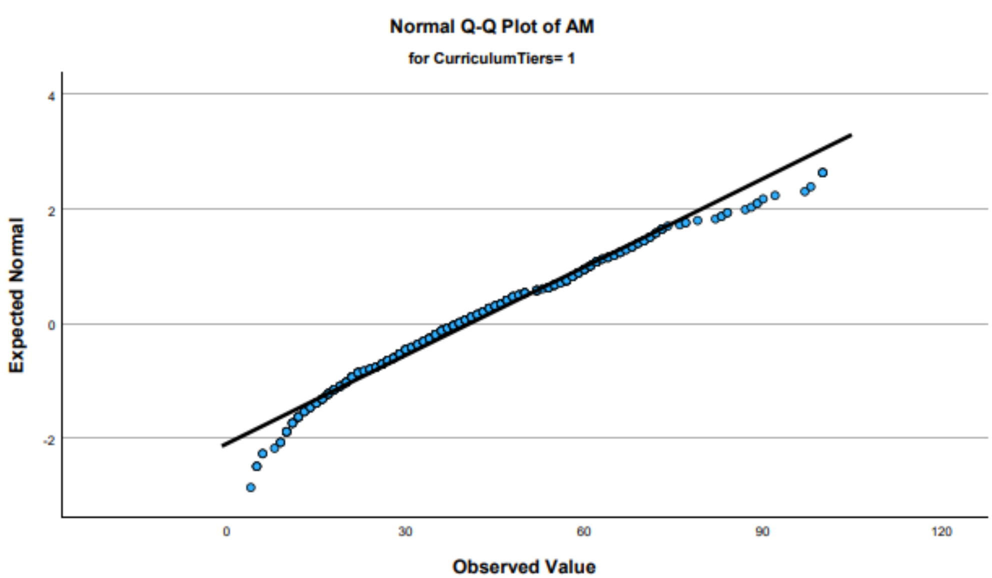


Figure L7

2017-2018 Grade 4 Tier 3 Histogram

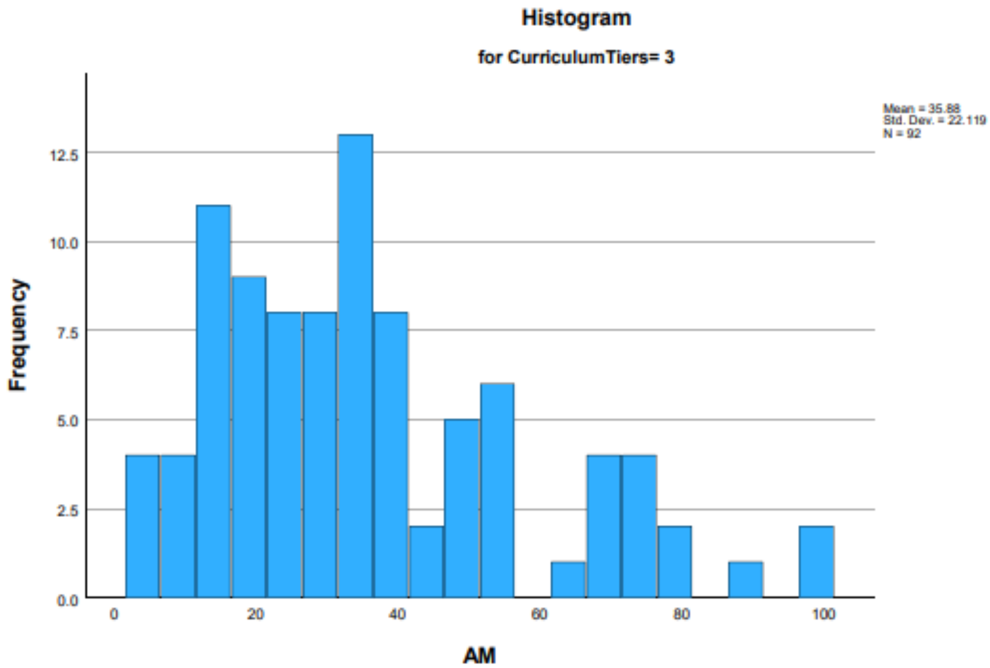


Figure L8

2017-2018 Grade 4 Tier 3 Q Q-Plot

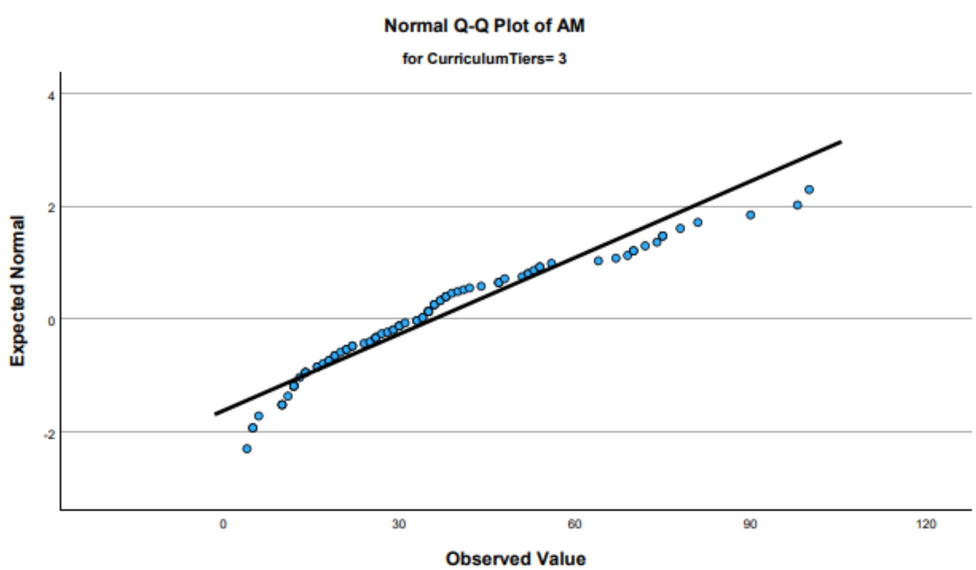


Figure L9

2018-2019 Grade 4 Tier 1 Histogram

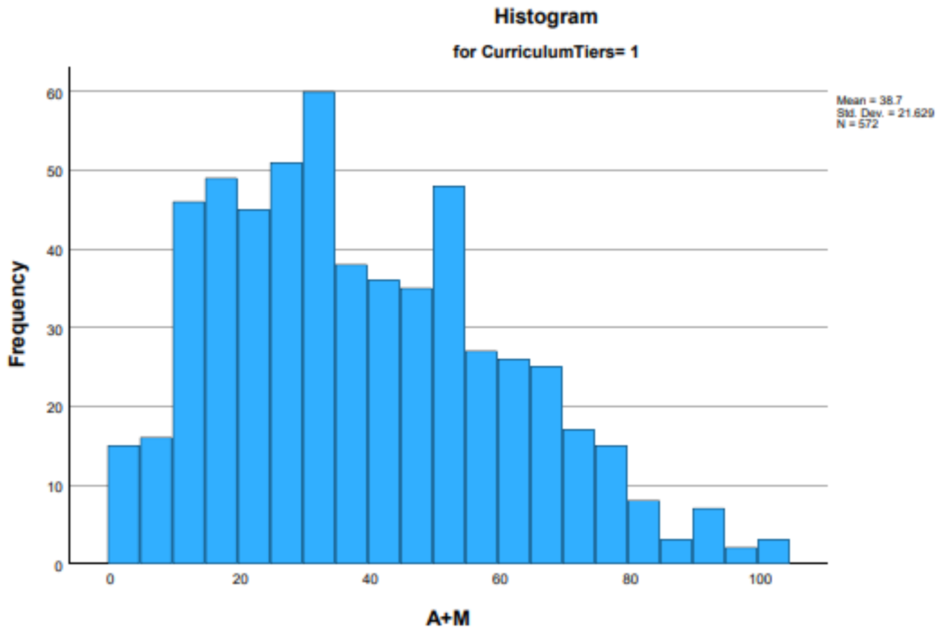


Figure L10

2018-2019 Grade 4 Tier 1 Q Q-Plot

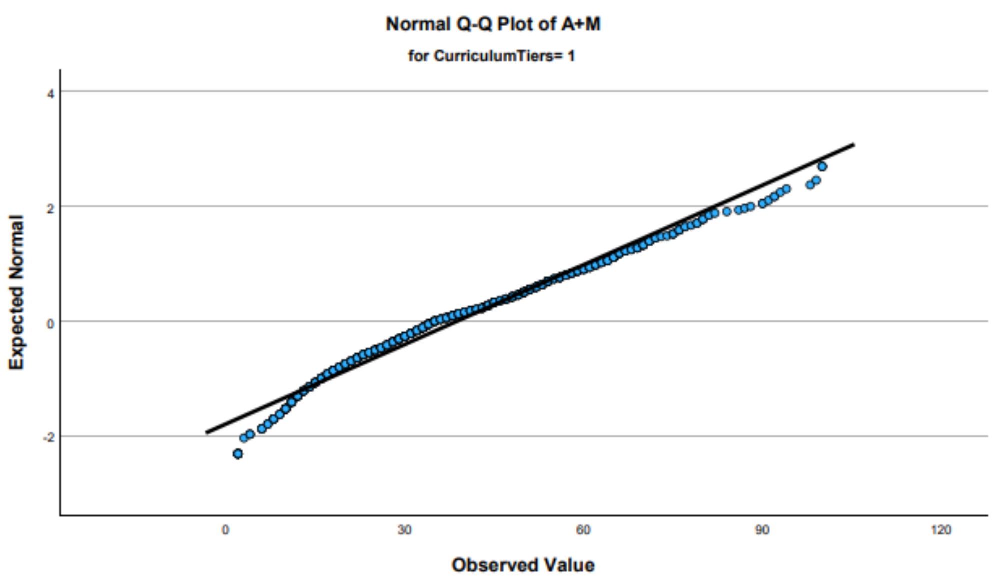


Figure L11

2018-2019 Grade 4 Non-tiered (Tier 4) Histogram

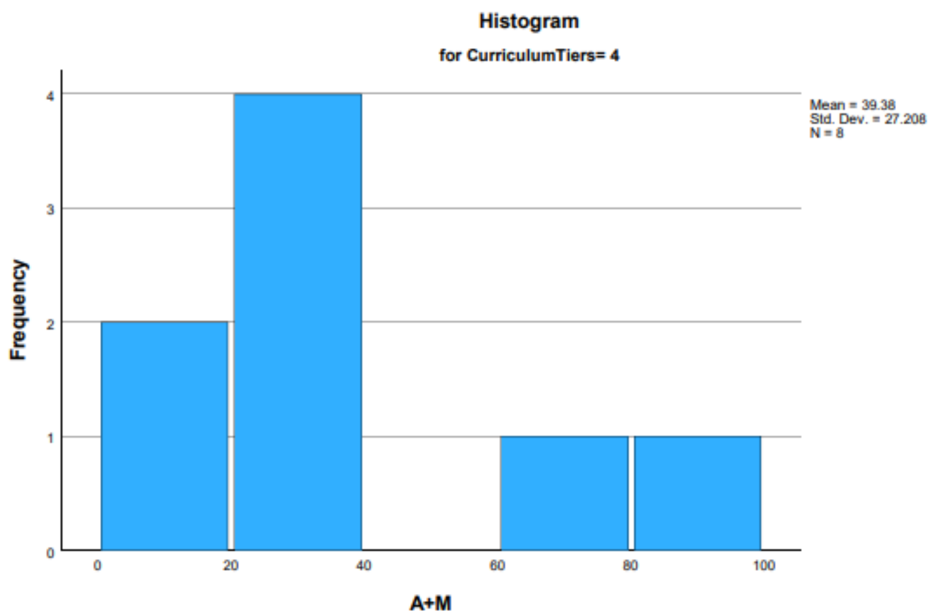
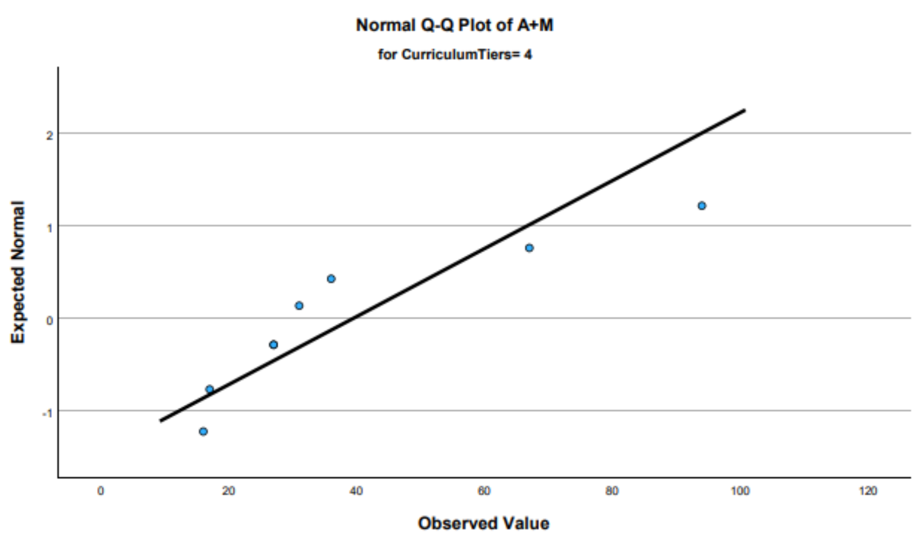


Figure L12

2018-2019 Grade 4 Non-tiered (Tier 4) Q Q-Plot



APPENDIX M:

Fifth Grade Histogram and Q Q-Plot for Each Year

Figure M1

2016-2017 Grade 5 Tier 1 Histogram

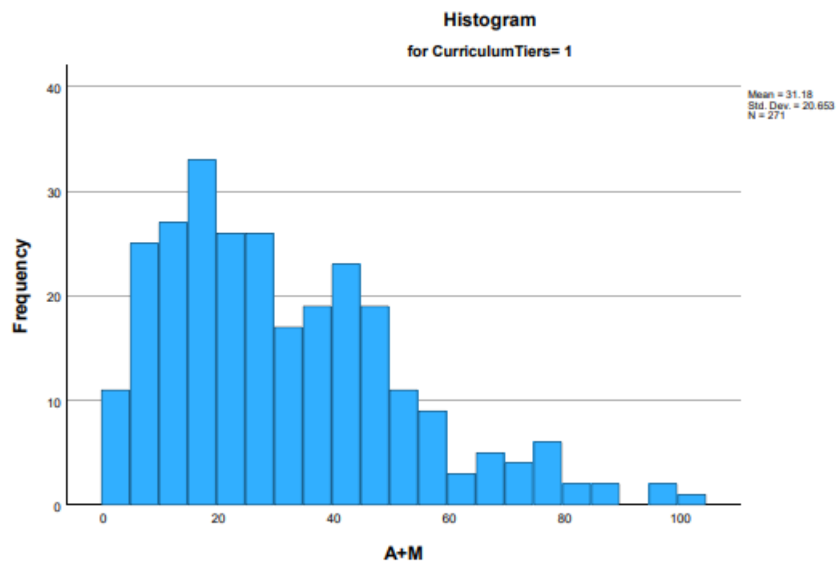


Figure M2

2016-2017 Grade 5 Tier 1 Q Q-Plot

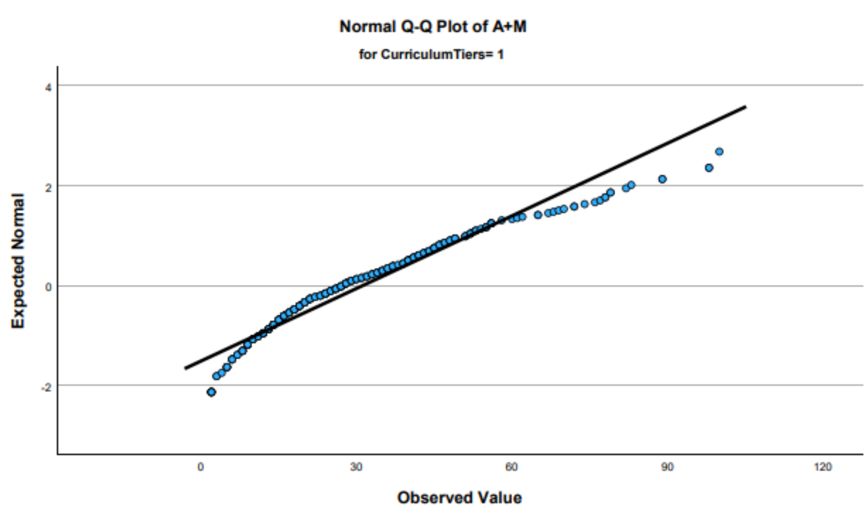


Figure M3

2016-2017 Grade 5 Tier 3 Histogram Q Q-Plot

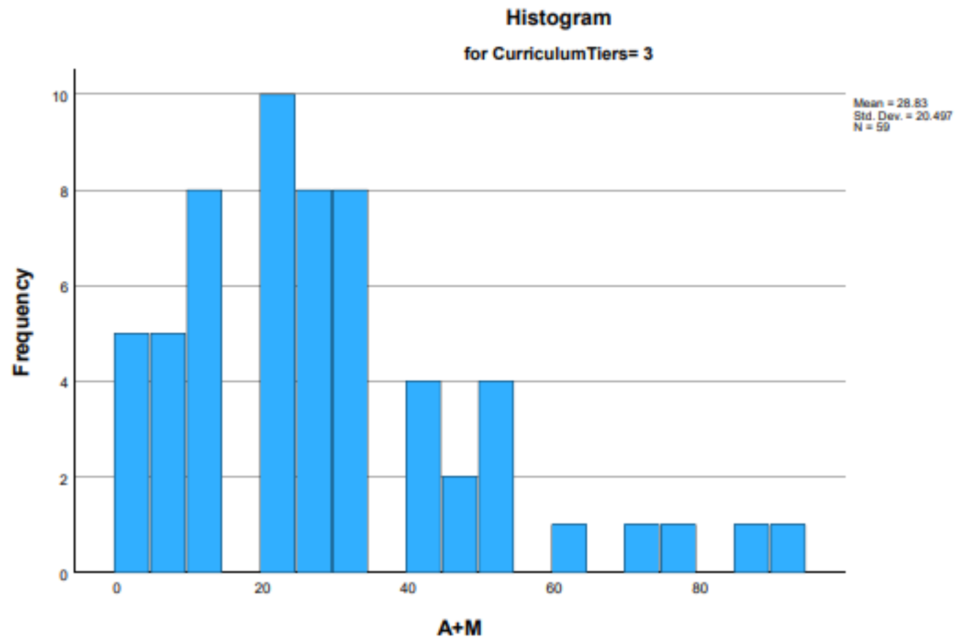


Figure M4

2016-2017 Grade 5 Tier 3 Histogram Q Q-Plot

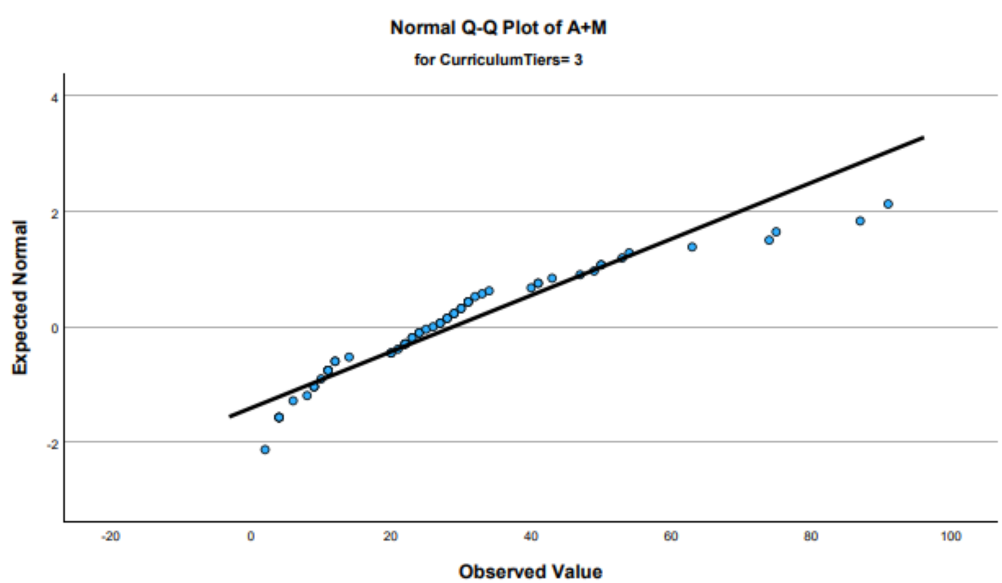


Figure M5

2017-2018 Grade 5 Tier 1 Histogram

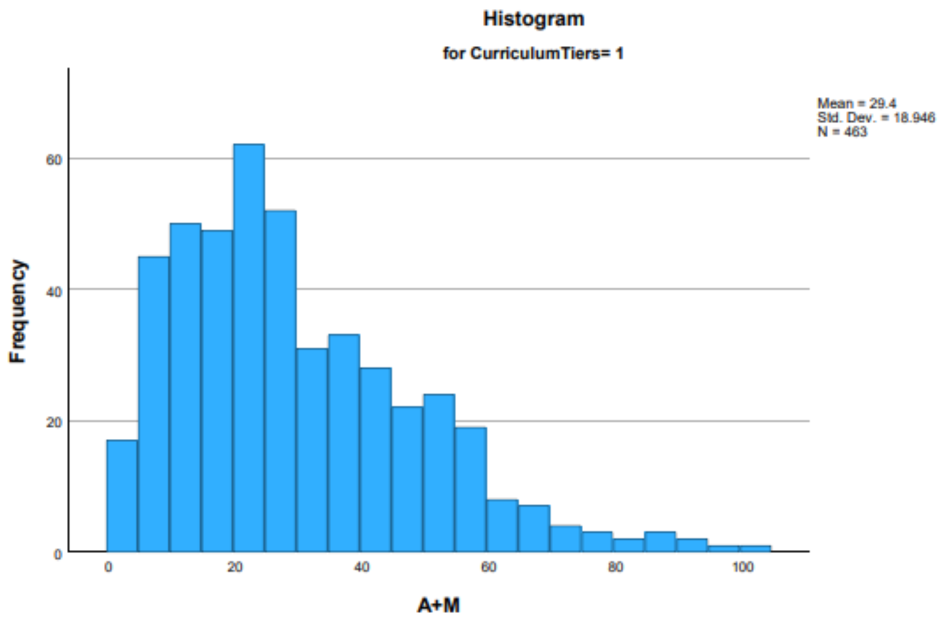


Figure M6

2017-2018 Grade 5 Tier 1 Q Q-Plot

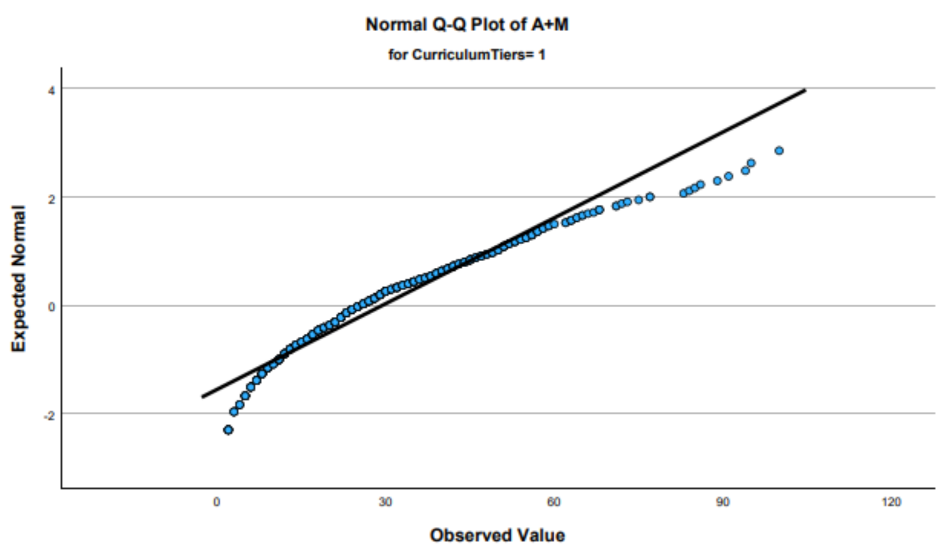


Figure M7

2017-2018 Grade 5 Tier 3 Histogram

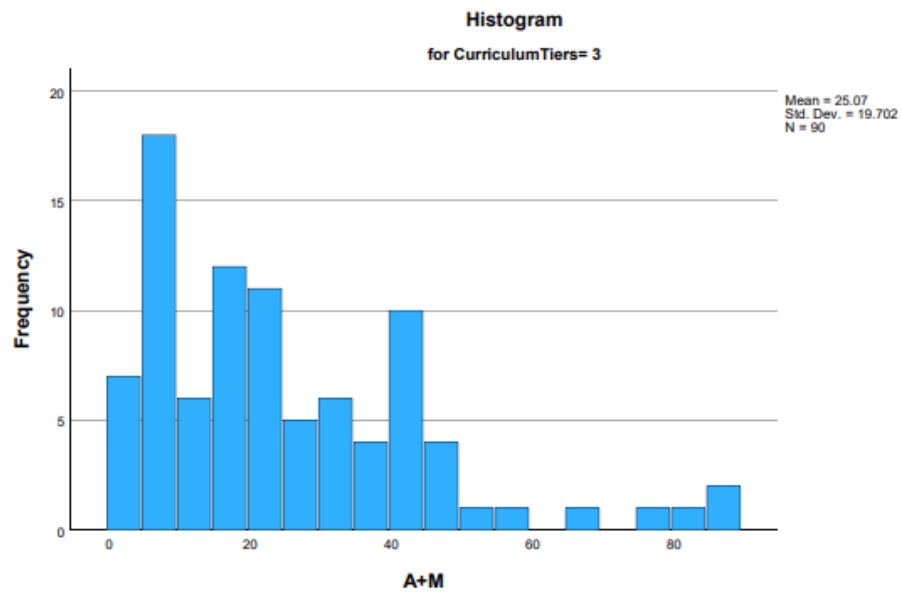


Figure M8

2017-2018 Grade 5 Tier 3 Q Q-Plot

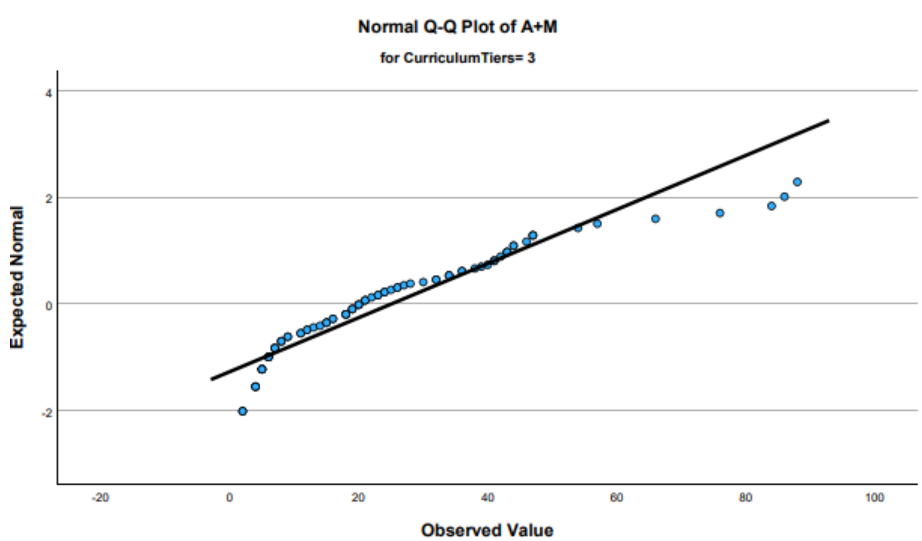


Figure M9

2018-2019 Grade 5 Non-tiered (Tier 4) Histogram

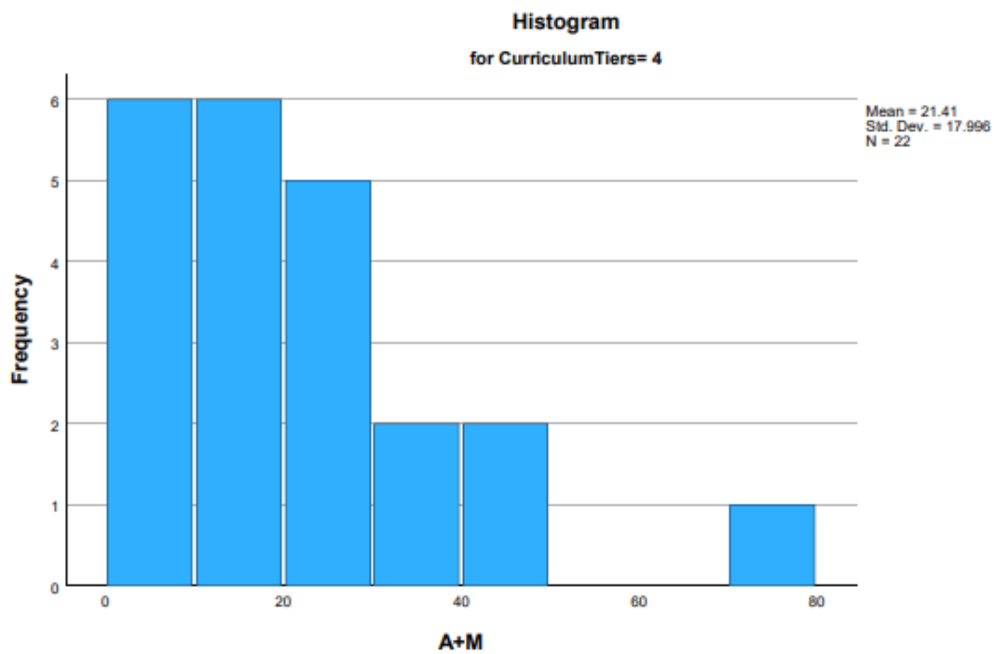


Figure M10

2018-2019 Grade 5 Non-tiered (Tier 4) Q Q-Plot

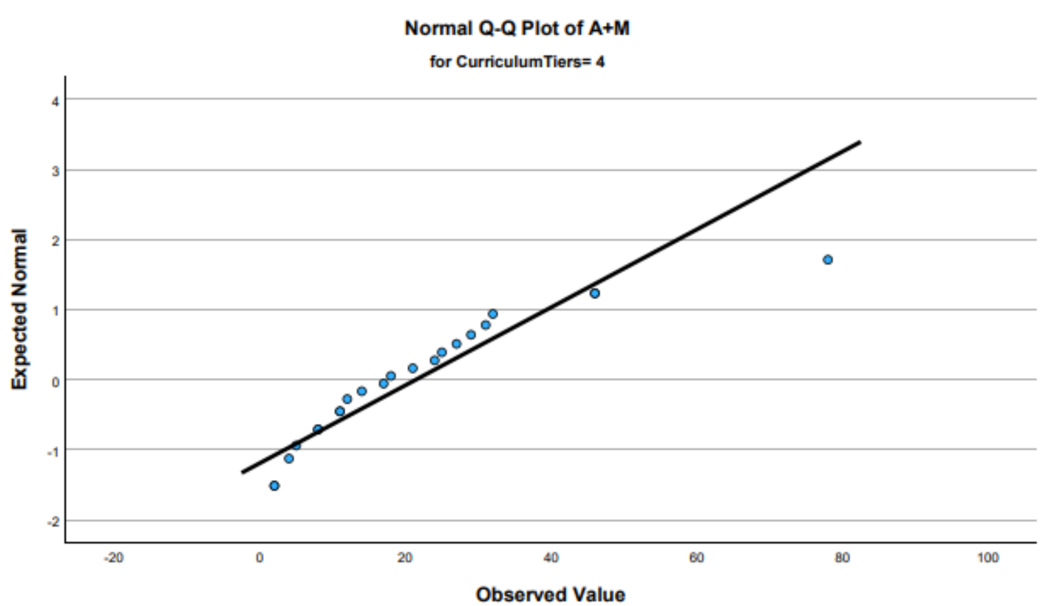


Figure M11

2018-2019 Grade 5 Tier 1 Histogram

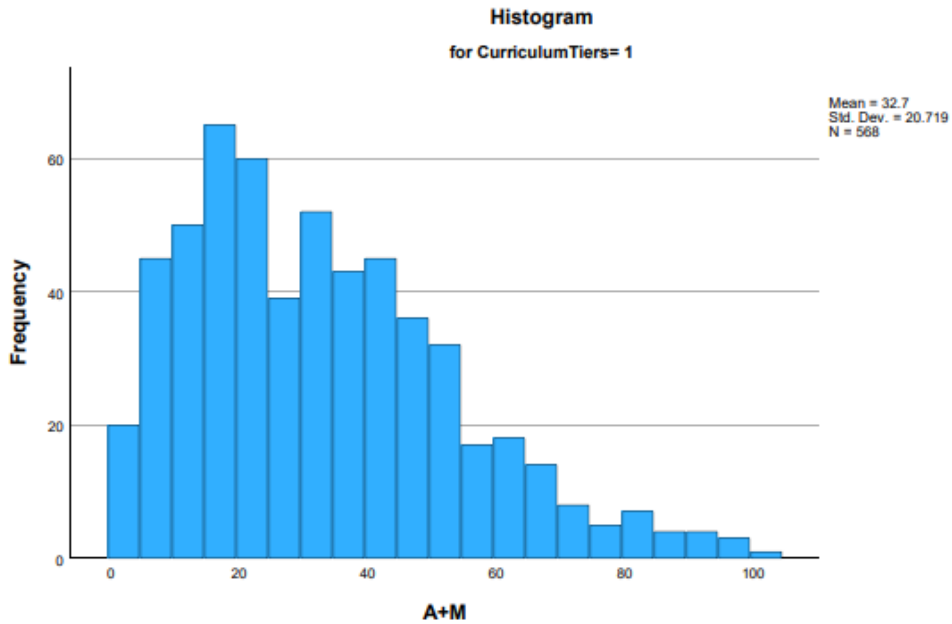
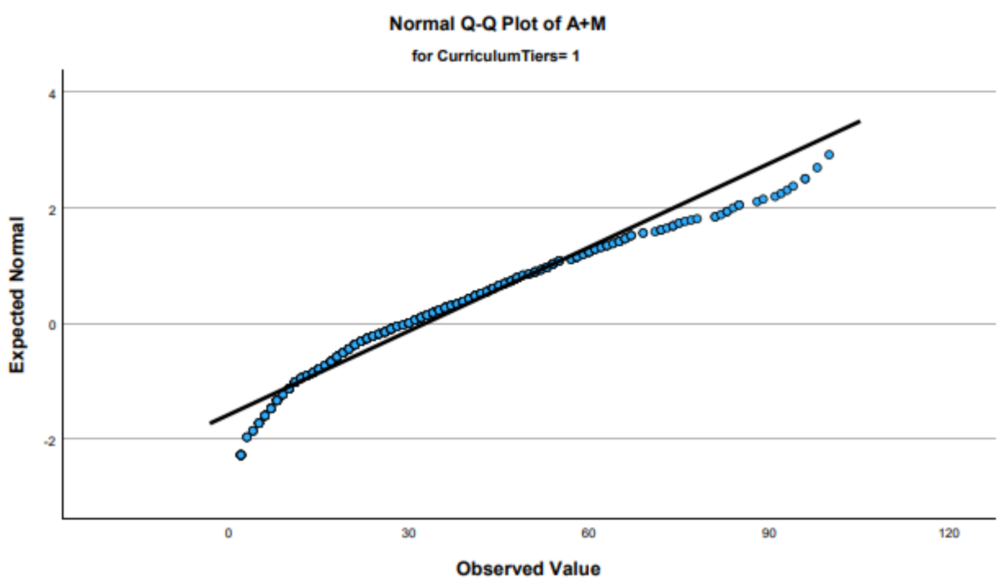


Figure M12

2018-2019 Grade 5 Tier 1 Q Q-Plot



APPENDIX N:

Additional Normality Analysis

Finding and utilizing alternate methods of analyzing normality more applicable to the sample subgroups with sample sizes either significantly less than or greater than 50, to replace the Shapiro-Wilk test, was an area of focus by the researcher. In fact, Laerd Statistics (2017) mentioned the use of multiple methods to detect and then assess normality to understand the data more rigorously and then assuming normality if the data meets a specific criterion, such as representing normality on two out of three methods.

An objective method for determining normality is the use of skewness and kurtosis ratios, referred to as *z*-test, calculated by dividing skewness or kurtosis values by their corresponding standard error (*SE*) (Demir, 2022; Field, 2013; Laerd Statistics, 2017; Mishra et al., 2019). Acceptable ranges for *z*-score results (from a *z*-test) vary from absolute values of 1.96 or approximately two (Demir, 2022; International Business Machines [IBM], 2021), 2.58 (Demis, 2022; Laerd Statistics, 2017), or any absolute value less than three (Bloom, 2020; Field, 2013).

Meanwhile, Field (2013) and Mishar et al. (2019) with similar but slightly different standards, base acceptable ratio results on sample size, with absolute values of 1.96 for sample sizes 50 or below, and either 2.58 (Demis, 2022; Field, 2013) or 3.29 (Demis, 2022; Field, 2013; Mishra, 2019) for samples above 50 since they generally have a lower standard error that impacts the results (Demir, 2020; Field, 2013; Flanagan, 2019). Since SPSS technically reports “excess” kurtosis, which establishes normality at zero, not kurtosis “proper” where normality is established at three, the researcher reported kurtosis related concepts and results to match “excess” kurtosis because SPSS was the tool utilized by the researcher (Flanagan, 2019; IBM, 2021).

SPSS descriptive statistics analysis for the data sets within this study reported skewness and kurtosis coefficients and corresponding standard errors for 22 of the 27 subgroups. Five of the subgroups had sample sizes too small for SPSS analysis. Of those 22 subgroups, seven had samples sizes below 50 and 15 had samples sizes at or above 50. For the seven sample subgroups below 50 data points, four of the seven had both skewness and kurtosis ratios with *z*-scores that supported normality at the 1.96 absolute value level (Demis, 2022; IBM, 2021), and five of the seven showed normality if the 2.58 (Demis, 2022; Laerd Statistics, 2017) level was applied.

For the 15 subgroups with samples sizes above 50, only five of the 15 subgroups had skewness and kurtosis ratios with an absolute value of 3.28 or less, which is not surprising given the large sample size of several of the subgroups (see Tables N1, N2 and N3). Demir (2022) and Field (2013) noted that sample sizes of 200 or more should not emphasize tests that calculate normality, especially as the only measure of normality, but rather examine the shape of the data or distribution along with the skewness and kurtosis values, not ratios, to identify normality.

This is somewhat echoed by Laerd Statistics (2017) which emphasized that steadfast coherence to only calculation methods does not always give the researcher a true understanding of the sample’s normality and that skewness and kurtosis values, graphical analysis, along with tests such as the Shapiro-Wilk should be considered before making judgments of normality. Specifically with large sample sizes, Fields (2013) stated the researcher can alleviate some concern if the skewness and kurtosis ratios are below an absolute value of 3.28, but not reject normality solely on those *z*-scores if the data was not below the desired threshold. As a result, the researcher calculated skewness and kurtosis ratios to gain *z*-scores for each sample but did not solely rely on those ratios when determining normality.

Table N1

2016-2017 Skewness and Kurtosis Values and z-scores by Grade within Tier

Grade	Tier	*n*	Skewness				Kurtosis			
			Value	*SE*	*z*-score	Supports Normality	Value	*SE*	*z*-score	Supports Normality
3rd	1	274	0.41	0.15	2.79	Yes	0.01	0.29	0.02	Yes
	3	59	0.48	0.31	1.56	Yes	-0.11	0.61	-0.18	Yes
	Non-tiered	3	-0.34	1.23	-0.28	Yes	-	-	-	-
4th	1	273	0.46	0.15	3.12	Yes	-0.36	0.29	-1.25	Yes
	3	60	0.63	0.31	2.04	Yes	-0.28	0.61	-0.48	Yes
	Non-tiered	3	-0.88	1.23	0.72	Yes	-	-	-	-
5th	1	271	0.93	0.15	6.28	No	0.65	0.3	2.2	Yes
	3	59	1.18	0.31	3.79	No	1.36	0.61	2.21	Yes
	Non-tiered	2	-	-	-	-	-	-	-	-

Note. Normality supported at ± 1.96 when sample size < *50,* ± 3.28 when sample size > 50. A dash designation, (-), indicated not applicable.

Table N2

2017-2018 Skewness and Kurtosis Values and z-scores by Grade within Tier

Grade	Tier	*n*	Skewness				Kurtosis			
			Value	*SE*	*z*-score	Supports Normality	Value	*SE*	*z*-score	Supports Normality
3rd	1	471	0.5	0.11	4.38	No	-0.06	0.23	-0.3	Yes
	3	92	0.93	0.25	3.69	No	0.41	0.5	0.82	Yes
	Non-tiered	23	1.82	0.48	3.78	No	4.11	0.94	4.4	No
4th	1	468	0.52	0.11	4.63	No	-0.15	0.23	-0.7	Yes
	3	92	0.79	0.25	3.16	Yes	0.05	0.5	0.09	Yes
	Non-tiered	23	1.03	0.48	2.14	No	1.05	0.94	1.12	Yes
5th	1	463	0.94	0.11	8.29	No	0.75	0.23	3.34	No
	3	90	1.25	0.25	4.91	No	1.58	0.5	3.14	Yes
	Non-tiered	22	1.63	0.49	3.31	No	3.51	0.95	3.68	No

Note. Normality supported ± 1.96 when sample size < *50,* ± 3.28 when sample size > 50. A dash designation, (-), indicated not applicable.

Table N3

2018-2019 Skewness and Kurtosis Values and z-scores by Grade within Tier

Grade	Tier	*n*	Skewness				Kurtosis			
			Value	*SE*	*z*-score	Supports Normality	Value	*SE*	*z*-score	Supports Normality
3rd	1	569	0.51	0.1	5.01	No	0.07	0.2	0.34	Yes
	3	3	1.22	1.23	0.99	Yes	-	-	-	-
	Non-tiered	8	1.35	0.75	1.79	Yes	0.88	1.48	0.59	Yes
4th	1	572	0.48	0.1	4.73	No	-0.39	0.2	1.92	Yes
	3	4	0.34	1.01	0.34	Yes	-2.85	2.62	1.09	Yes
	Non-tiered	8	1.47	0.75	1.96	Yes	1.41	1.48	0.95	Yes
5th	1	568	0.77	0.1	7.48	No	0.21	0.21	1.02	Yes
	3	4	1.01	1.01	0.99	Yes	1.37	2.62	0.52	Yes
	Non-tiered	2	-	-	-	-	-	-	-	-

Note. Normality supported ± 1.96 when sample size < *50,* ± 3.28 when sample size > 50. A dash designation, (-), indicated not applicable.

Another method comprised of the skewness or kurtosis values to support normality used the absolute value of each to determine if they were less than their corresponding standard error multiplied by two (George & Mallery, 2010; IBM, 2021). This method resulted in eight of 25 skewness subgroups and 16 of 22 kurtosis subgroups with a value less than their correlated standard error multiplied by two, thus supporting normality in those subgroups. Again, the total number of subgroups varied and was below the 27 subgroups listed above because of the lack of sample size associated with some subgroups, the outcomes are in Table N4 below.

The results of this type of analysis, like the *z*-score method listed above, involved the use of skewness and kurtosis standard errors, which Fields (2013) noted tend to be lower with larger samples. Thus, the potential for the above method involving the use of the standard error multiplied by two to reject normality even if it is present would increase for the large sample size subgroups (Field, 2013). Even though the above three calculation based normality examinations did not fully match the needs of all subgroup due to the varying sample sizes (Field, 2013; Laerd Statistics, 2017, Mishra et al., 2019), the methods did meet the needs of some of the subgroups and ultimately helped provide a more thorough understanding of the data, or what Laerd Statistics (2017) referred to as "a feel" for the data, before shifting the approach to non-calculation based methods to determine normality.

Table N4

Standard Error Multiplied by Two Compared to Original Value, by Year, Grade, and Tier

Year	Grade	Skewness *SE* Multiplied by Two (Greater than Skewness Value)			Kurtosis *SE* Multiplied by Two (Greater than Kurtosis Value)		
		Tier 1	Tier 3	Non-tiered	Tier 1	Tier 3	Non-tiered
2016-2017	3	0.29 (No)	0.62 (Yes)	2.45 (Yes)	0.59 (Yes)	1.23 (Yes)	-
	4	0.29 (No)	0.62 (No)	2.45 (Yes)	0.59 (Yes)	1.22 (Yes)	-
	5	0.3 (No)	0.62 (No)	-	0.59 (No)	1.23 (No)	-
2017-2018	3	0.23 (No)	0.5 (No)	0.96 (No)	0.45 (Yes)	0.99 (Yes)	1.87 (No)
	4	0.23 (No)	0.5 (No)	0.96 (No)	0.45 (Yes)	0.99 (Yes)	1.87 (Yes)
	5	0.23 (No)	0.51 (No)	0.98 (No)	0.45 (No)	1.01 (No)	1.91 (No)
2018-2019	3	0.2 (No)	2.45 (Yes)	1.5 (Yes)	0.41 (Yes)	-	2.96 (Yes)
	4	0.2 (No)	2.03 (Yes)	1.5 (Yes)	0.41 (Yes)	5.24 (Yes)	2.96 (Yes)
	5	0.21 (No)	2.03 (Yes)	-	0.41 (Yes)	5.24 (Yes)	-

Note. Normality supported if *SE* multiplied by two is greater than the original value, indicated with "(Yes)" below *SE* multiplied by two results; see Tables N1, N2, and N3 for original skewness values, kurtosis values, corresponding *SE* values and sample sizes. A dash designation, (-), indicated not applicable.

Since the original procedure to check normality outlined in Chapter Three, the Shapiro-Wilk test, and the additional calculation methods within Appendix N such as skewness and kurtosis values and their standard errors, were not entirely dependable for all sample sizes within this study, graphical analysis from histograms and Normal Q Q-Plots (see Appendices K, L and M), as well as skewness and kurtosis values that did not involve their standard errors, were examined. The utilization of graphical methods and skewness and kurtosis values to determine normality is not entirely uncommon, especially with large sample sizes, and has been used recently by other researchers (Bandalos, 2018; Berber Sardinha & Veirano Pinto, 2019; Burns McOmber, 2020; Chastain, 2021; Padilla-Hernandex, 2020; Patton & Sickles, 2021; Sisson, 2020). Like skewness and kurtosis *z*-scores calculations above, there is not a universally accepted practice for how skewness and kurtosis values, without their standard errors, should be evaluated to help determine normality. But there is consensual understanding that the data in the sample shifts away from normal distribution the further the values are from zero (Field, 2013).

Analysis of normality that utilizes skewness and kurtosis values vary, with recommended skewness absolute value ranges of one or less (Mishra et al., 2019; Sisson, 2020), 1.5 or less (Patton & Sickles, 2021) and two or less (Bandalos, 2018; Berber Sardinha & Veirano Pinto, 2019; George & Mallery, 2005), with values outside of those ranges potentially indicating non-normality. Kurtosis value boundaries applied by researchers have a larger range with limits as low as one (Mishra et al., 2019) or two (Bandalos, 2018; George & Mallery, 2005; Patton & Sickles, 2021, Sisson, 2020), and then less stringent boundaries of three (Berber Sardinha & Veirano Pinto, 2019), and then others proposing values up to seven (Bandalos, 2018) before concluding the kurtosis value could reflect non-normality.

Based on the above information, an absolute value of 1.5 for skewness and two for kurtosis was used by the researcher to examine whether the skewness and kurtosis values could be an indicator of normality. Under those conditions, 23 of the 25 subgroups had a skewness absolute value of 1.5 or less and 19 of 22 subgroups had kurtosis absolute values of two or less (see Table N5). The variety of methods outlined above allowed the researcher to examine assumption five, normality, for the different subgroups by applying methods that corresponded to the appropriate samples size of each subgroup, except for the situations where subgroup data sizes were too small for analysis.

Table N5

Skewness and Kurtosis Absolute Values by Year, Grade, and Tier

Year	Grade	Tier 1		Tier 3		Non-Tiered	
		Skewness	Kurtosis	Skewness	Kurtosis	Skewness	Kurtosis
2016-2017	3	0.41	0.01	0.48	0.11	0.34	-
	4	0.46	0.37	0.63	0.29	0.88	-
	5	0.93	0.65	1.18	1.36	-	-
2017-2018	3	0.5	0.07	0.93	0.41	1.82	4.11
	4	0.52	0.16	0.79	0.05	1.03	1.05
	5	0.94	0.75	1.25	1.58	1.63	3.51
2018-2019	3	0.51	0.07	1.22	N/A	1.35	0.88
	4	0.48	0.39	0.34	2.85	1.47	1.41
	5	0.77	0.21	1.01	1.37	-	-

Note. Absolute values of skewness less than 1.5 and kurtosis less than 2 support normality. A dash designation, (-), indicated not applicable.

In addition to the above normality tests and indicators, the researcher considered additional research before either affirming or rejecting normality of the data and determining if the data was valid based on analysis of assumption five of the data. Laerd Statistics (2017) discussed how the robustness of an ANOVA to non-normality can typically allow the use of *approximately* normal data, especially with larger samples sizes because of the Central Limit

Theorem (Delacre et al., 2019; Field, 2013; Laerd Statistics, 2017). Some researchers suggest a sample size as low as 25 or more could be an acceptable threshold (Bloom, 2020) while a more traditional standard of 30 to 40 (Field, 2013) for the Central Limit Theorem to support normality is held by others. Another consideration pointed out by Delacre et al. (2019) when examining data pertains to an increased awareness that measurement boundaries, such as percent scale from zero to 100 used within this study, could be a cause of non-normality, but even if that is the case a large enough sample size can help overcome this and allow non-normal data to still offer valid results (Laerd Statistics, 2017).

Delacre et al. (2019) and Judd et al. (2017) both specified that if the skewness value indicates non-normality the results can still be used if the related kurtosis value is normal in many situations. Likewise, Laerd Statistics (2017) asserted that skewed data that is similarly skewed, all negative or all positive, does not present as big of a problem to breaks in normality as when skewness within the sample is skewed differently, with some samples negatively skewed and others positively skewed, within the same data set. More details related to the normality of specific subgroups take place in the presentation of findings section of Chapter Four, along with individual discussion related to histogram distribution and Q Q-Plots for each subgroup.

APPENDIX O:

2017-2018 Grade 3 *t*-test Group Statistics and Results

Group Statistics

	Curriculum Tiers	N	Mean	Std. Deviation	Std. Error Mean
AM	1	471	40.74	19.466	.897
	3	92	35.88	22.119	2.306

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
AM	Equal variances assumed	.884	.348	2.140	561
	Equal variances not assumed			1.963	120.082

Independent Samples Test

		t-test for Equality of Means			
		Significance		Mean Difference	Std. Error Difference
		One-Sided p	Two-Sided p		
AM	Equal variances assumed	.016	.033	4.858	2.271
	Equal variances not assumed	.026	.052	4.858	2.474

Independent Samples Test

		t-test for Equality of Means 95% Confidence Interval of the Difference	
		Lower	Upper
AM	Equal variances assumed	.398	9.318
	Equal variances not assumed	-.041	9.758

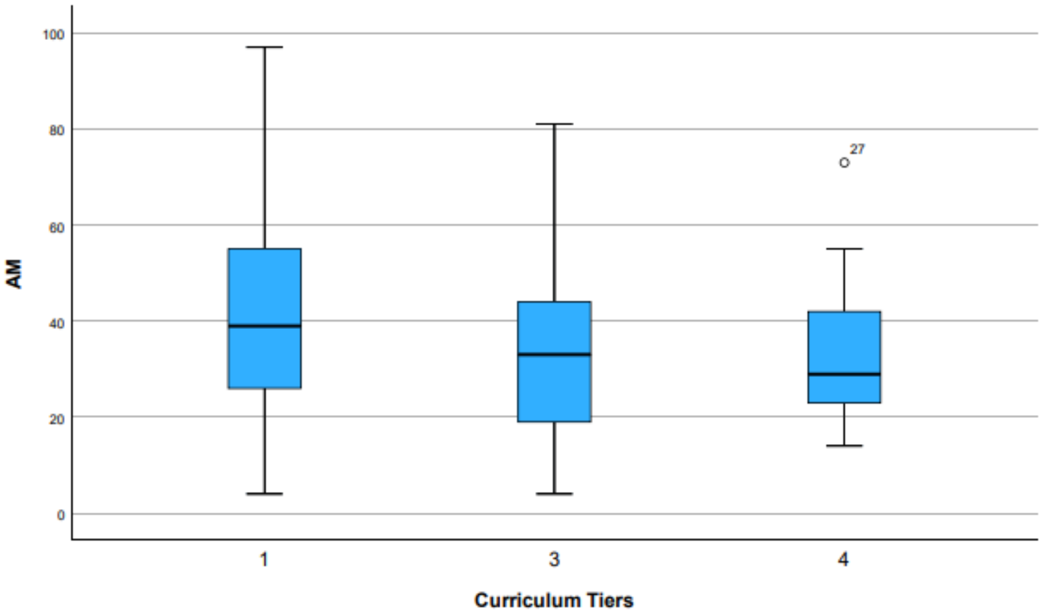
Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
AM	Cohen's d	19.920	.244	.020	.468
	Hedges' correction	19.947	.244	.020	.467
	Glass's delta	22.119	.220	-.007	.445

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

APPENDIX P:

2017-2018 Grade 3 Removed Original Outliers Boxplot



APPENDIX Q:

2017-2018 Grade 3 Removed Original Outliers Normality Methods and Levene's Test

Shapiro-Wilk test results:

Tests of Normality

	Curriculum Tiers	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
AM	1	.054	467	.002	.982	467	<.001
	3	.112	89	.008	.941	89	<.001
	4	.161	22	.142	.928	22	.111

a. Lilliefors Significance Correction

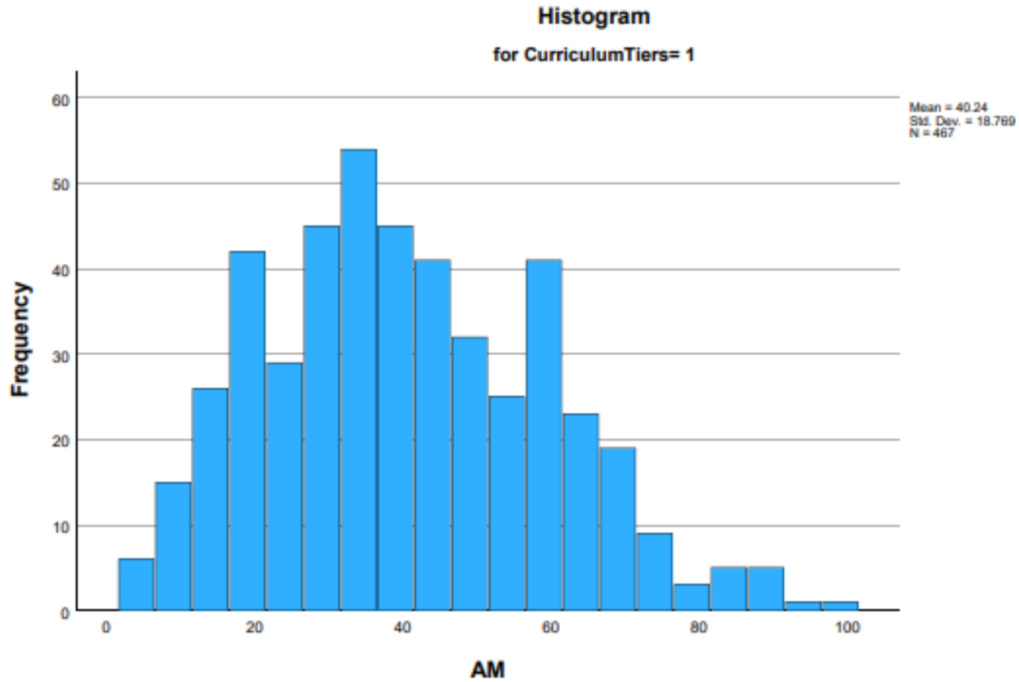
Additional normality methods:

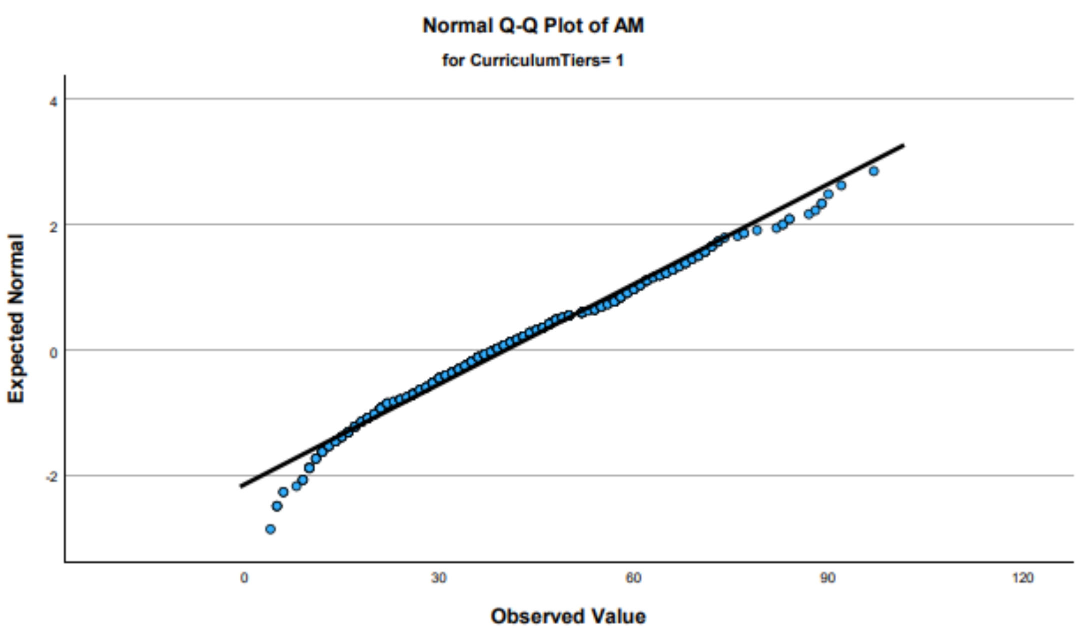
2017-2018 Grade 3 Skewness and Kurtosis Values and Normality Method by Tier

Tier	*n*	Skewness				Kurtosis			
		Value	*SE*	*z*-score	*SE* Multiplied by Two	Value	*SE*	*z*-score	*SE* Multiplied by Two
1	467	0.37	0.11	3.3	0.23	-0.37	0.23	-1.66	0.45
3	89	0.69	0.26	2.69	0.51	-0.22	0.51	-0.44	1.01
Non-tiered	22	1.07	0.49	2.17	0.98	1.29	0.95	1.36	1.91

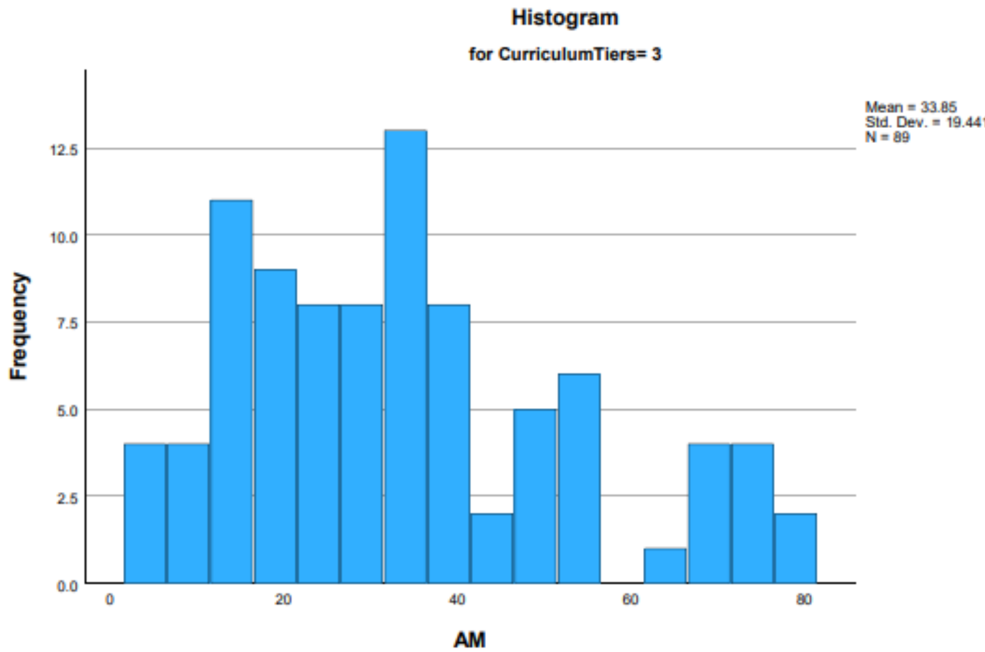
Note. Normality supported when: (1) Absolute values of skewness are less than 1.5 and kurtosis less than 2 support normality; (2) *z*-scores are ± 1.96 when sample size < *50,* ± 3.28 when sample size > 50; and (3) skewness or kurtosis *SE* multiplied by two is greater than the corresponding original value.

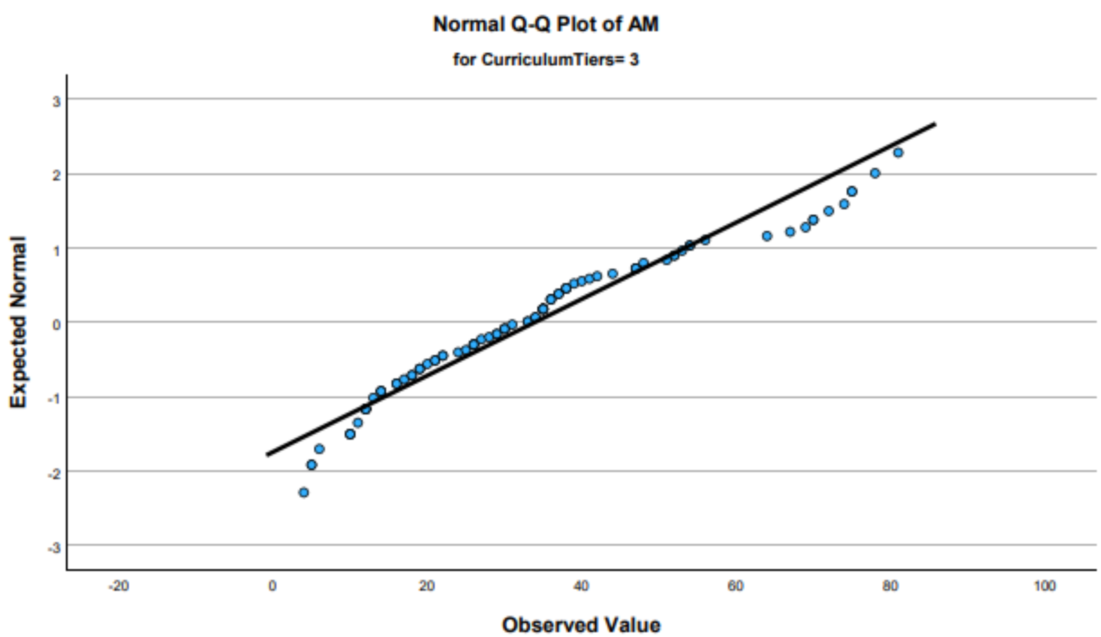
Tier 1 histogram and Q Q-Plot:



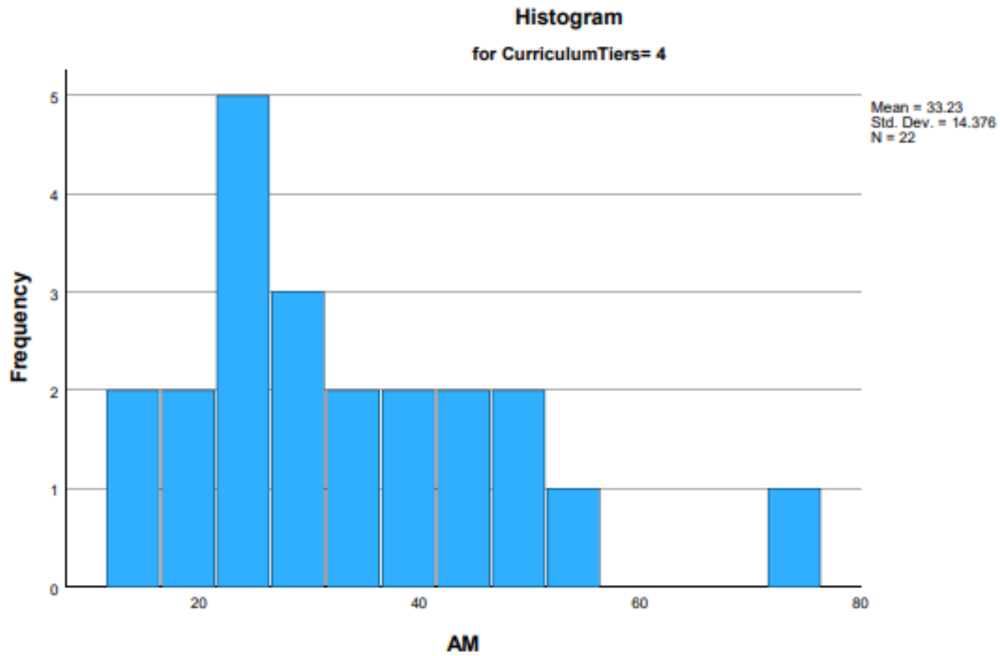


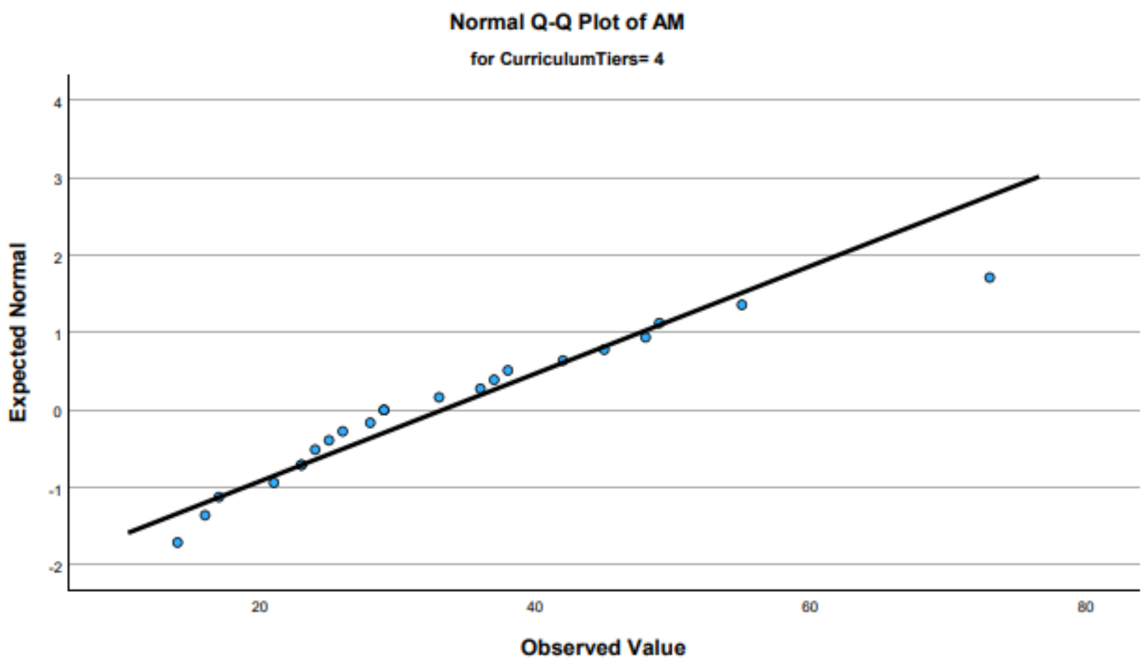
Tier 3 histogram and Q Q-Plot:





Non-tiered (Tier 4) histogram and Q Q-Plot:





Levene's Test of homogeneity of variances:

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
AM	Based on Mean	1.462	2	575	.233
	Based on Median	1.669	2	575	.189
	Based on Median and with adjusted df	1.669	2	571.759	.189
	Based on trimmed mean	1.508	2	575	.222

APPENDIX R:

2017-2018 Grade 3 Removed Original Outliers ANOVA Group Statistics and Results

Descriptive statistics:

Descriptives

AM

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum
1	467	40.24	18.769	.869	38.53	41.94	4
3	89	33.85	19.441	2.061	29.76	37.95	4
4	22	33.23	14.376	3.065	26.85	39.60	14
Total	578	38.99	18.875	.785	37.44	40.53	4

Descriptives

AM

	Maximum
1	97
3	81
4	73
Total	97

ANOVA results:

ANOVA

AM

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3802.835	2	1901.417	5.419	.005
Within Groups	201755.055	575	350.878		
Total	205557.889	577			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval Lower	Upper
AM	Eta-squared	.019	.002	.044
	Epsilon-squared	.015	-.002	.040
	Omega-squared Fixed-effect	.015	-.002	.040
	Omega-squared Random-effect	.008	-.001	.021

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Tukey post hoc test:

Post Hoc Tests

Multiple Comparisons

Dependent Variable: AM

	(I) Curriculum Tiers	(J) Curriculum Tiers	Mean Difference (I-J)	Std. Error	Sig.
Tukey HSD	1	3	6.382*	2.167	.009
		4	7.008	4.087	.200
	3	1	-6.382*	2.167	.009
		4	.627	4.460	.989
	4	1	-7.008	4.087	.200
		3	-.627	4.460	.989

Multiple Comparisons

Dependent Variable: AM

	(I) Curriculum Tiers	(J) Curriculum Tiers	95% Confidence Interval Lower Bound	Upper Bound
Tukey HSD	1	3	1.29	11.47
		4	-2.59	16.61
	3	1	-11.47	-1.29
		4	-9.85	11.11
	4	1	-16.61	2.59
		3	-11.11	9.85

APPENDIX S:

2017-2018 Grade 4 *t*-test Group Statistics and Results

Group Statistics

	Curriculum Tiers	N	Mean	Std. Deviation	Std. Error Mean
A+M	1	468	37.67	20.647	.954
	3	92	32.67	21.367	2.228

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
A+M	Equal variances assumed	.067	.796	2.109	558
	Equal variances not assumed			2.061	126.642

Independent Samples Test

		t-test for Equality of Means			
		Significance			
		One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference
A+M	Equal variances assumed	.018	.035	4.995	2.368
	Equal variances not assumed	.021	.041	4.995	2.423

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
A+M	Equal variances assumed	.343	9.647
	Equal variances not assumed	.199	9.791

Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
A+M	Cohen's d	20.766	.241	.016	.464
	Hedges' correction	20.794	.240	.016	.464
	Glass's delta	21.367	.234	.007	.459

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

APPENDIX T:

2017-2018 Grade 5 *t*-test Group Statistics and Results

Group Statistics

	Curriculum Tiers	N	Mean	Std. Deviation	Std. Error Mean
A+M	1	463	29.40	18.946	.881
	3	90	25.07	19.702	2.077

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
A+M	Equal variances assumed	.069	.793	1.972	551
	Equal variances not assumed			1.921	123.106

Independent Samples Test

		t-test for Equality of Means			
		Significance			
		One-Sided p	Two-Sided p	Mean Difference	Std. Error Difference
A+M	Equal variances assumed	.025	.049	4.333	2.197
	Equal variances not assumed	.029	.057	4.333	2.256

Independent Samples Test

		t-test for Equality of Means	
		95% Confidence Interval of the Difference	
		Lower	Upper
A+M	Equal variances assumed	.018	8.648
	Equal variances not assumed	-.132	8.798

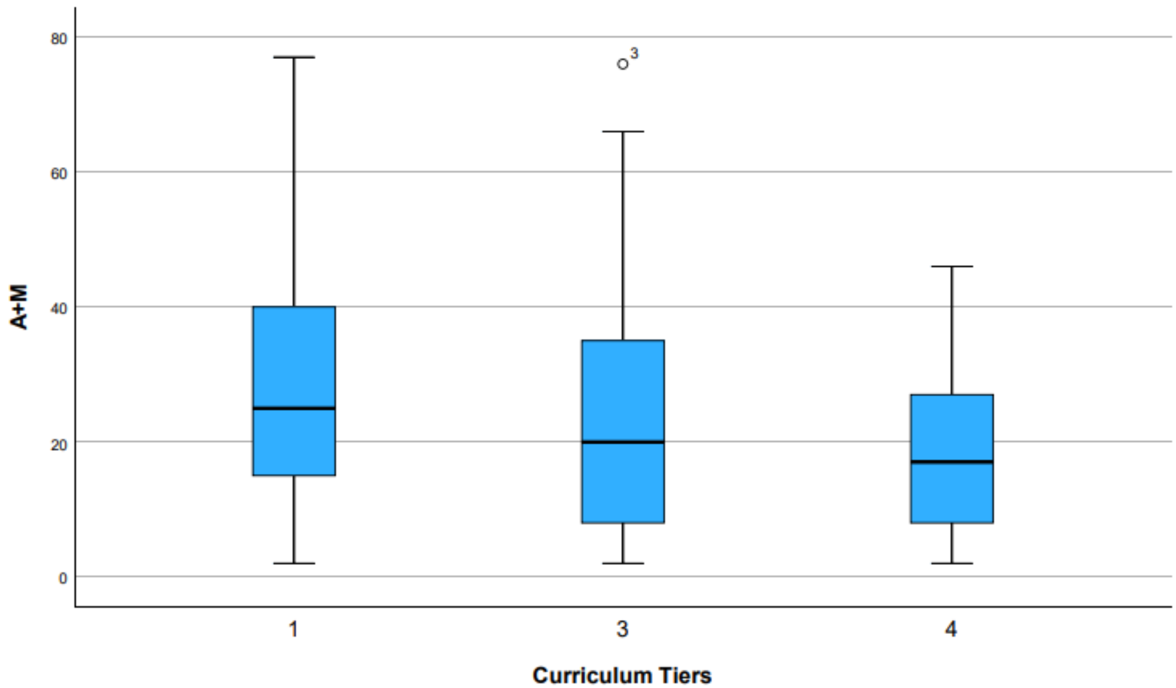
Independent Samples Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
A+M	Cohen's d	19.070	.227	.001	.453
	Hedges' correction	19.096	.227	.001	.453
	Glass's delta	19.702	.220	-.009	.447

a. The denominator used in estimating the effect sizes.
Cohen's d uses the pooled standard deviation.
Hedges' correction uses the pooled standard deviation, plus a correction factor.
Glass's delta uses the sample standard deviation of the control (i.e., the second) group.

APPENDIX U:

2017-2018 Grade 5 Removed Original Outliers Boxplot



APPENDIX V:

2017-2018 Grade 5 Removed Original Outliers Normality Methods and Levene's Test

Shapiro-Wilk test results:

Tests of Normality

	Curriculum Tiers	Kolmogorov-Smirnov[a] Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
A+M	1	.088	454	<.001	.956	454	<.001
	3	.111	87	.010	.925	87	<.001
	4	.124	21	.200*	.930	21	.137

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

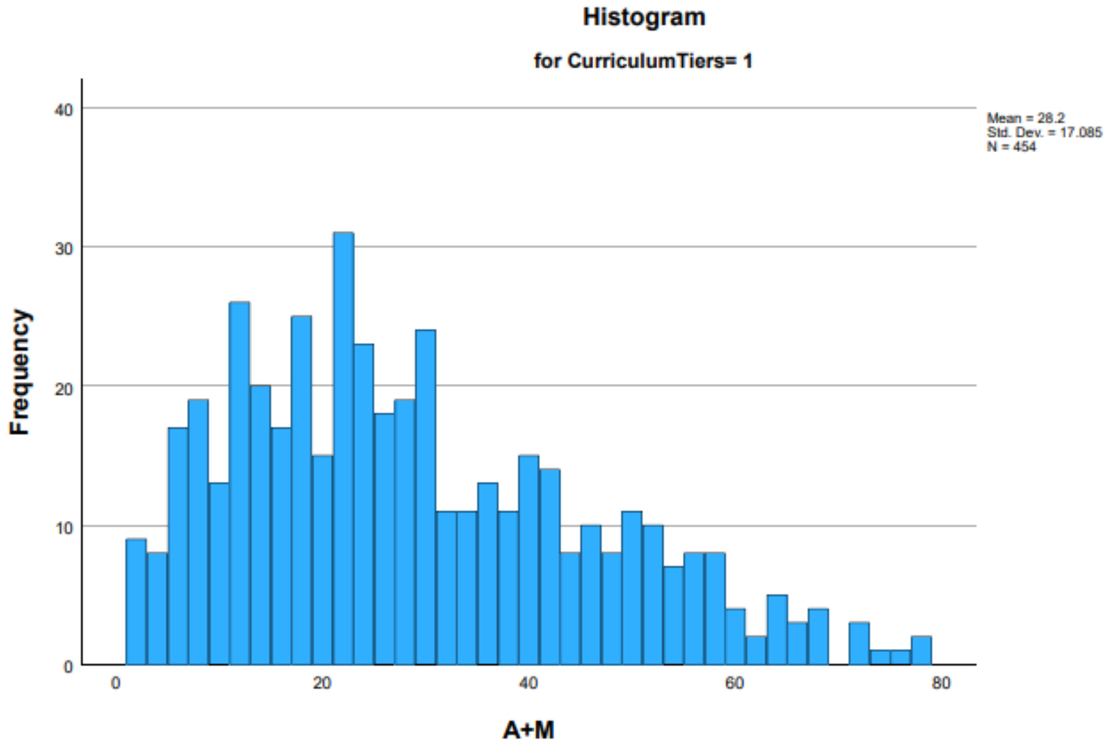
Additional normality methods:

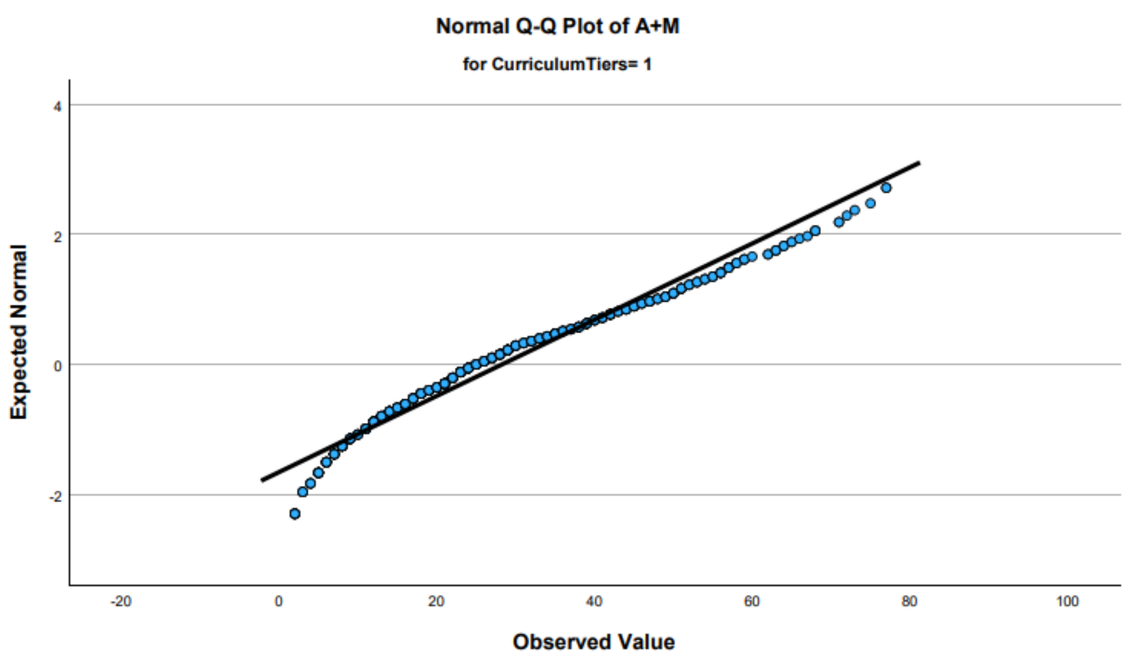
2017-2018 Grade 5 Skewness and Kurtosis Values and Normality Method by Tier

Tier	*n*	Skewness				Kurtosis			
		Value	*SE*	*z*-score	*SE* Multiplied by Two	Value	*SE*	*z*-score	*SE* Multiplied by Two
1	454	0.62	0.12	5.36	0.23	-0.33	0.23	-1.44	0.46
3	87	0.81	0.26	3.12	0.52	0.21	0.51	0.42	1.02

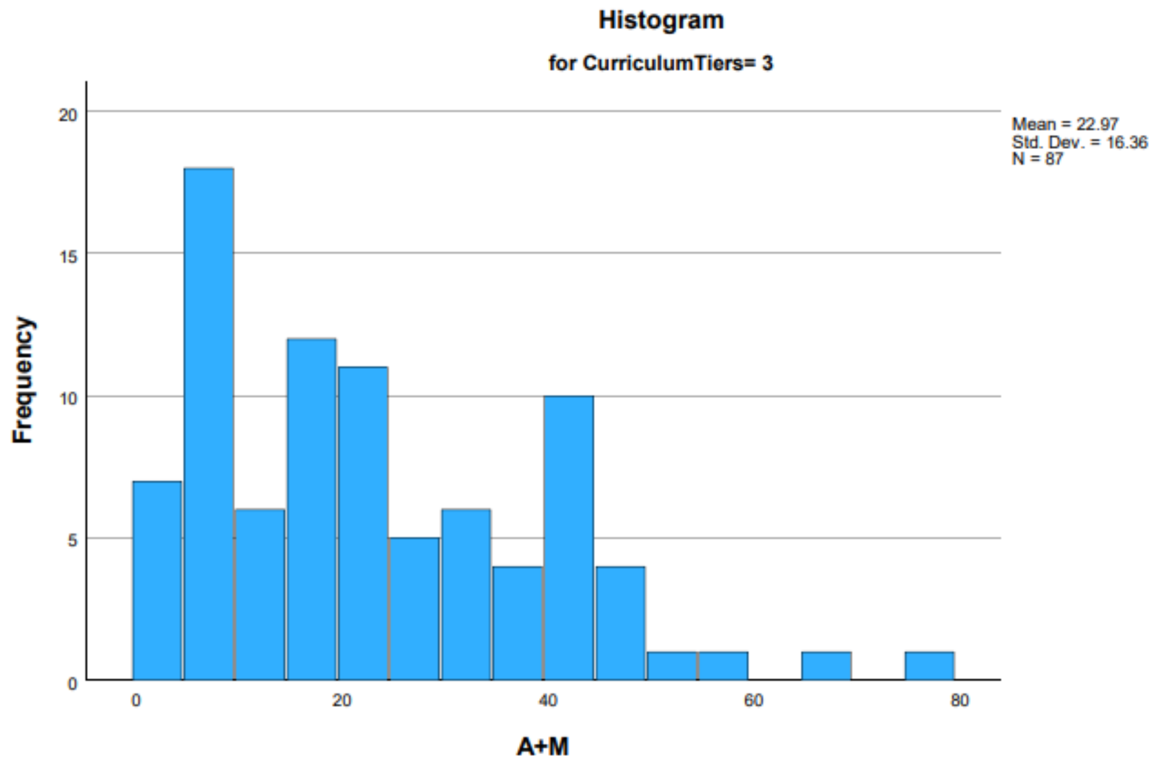
Note. Normality supported when: (1) Absolute values of skewness are less than 1.5 and kurtosis less than 2 support normality; (2) *z*-scores are ± 1.96 when sample size < *50,* ± 3.28 when sample size > 50; and (3) skewness or kurtosis *SE* multiplied by two is greater than the corresponding original value.

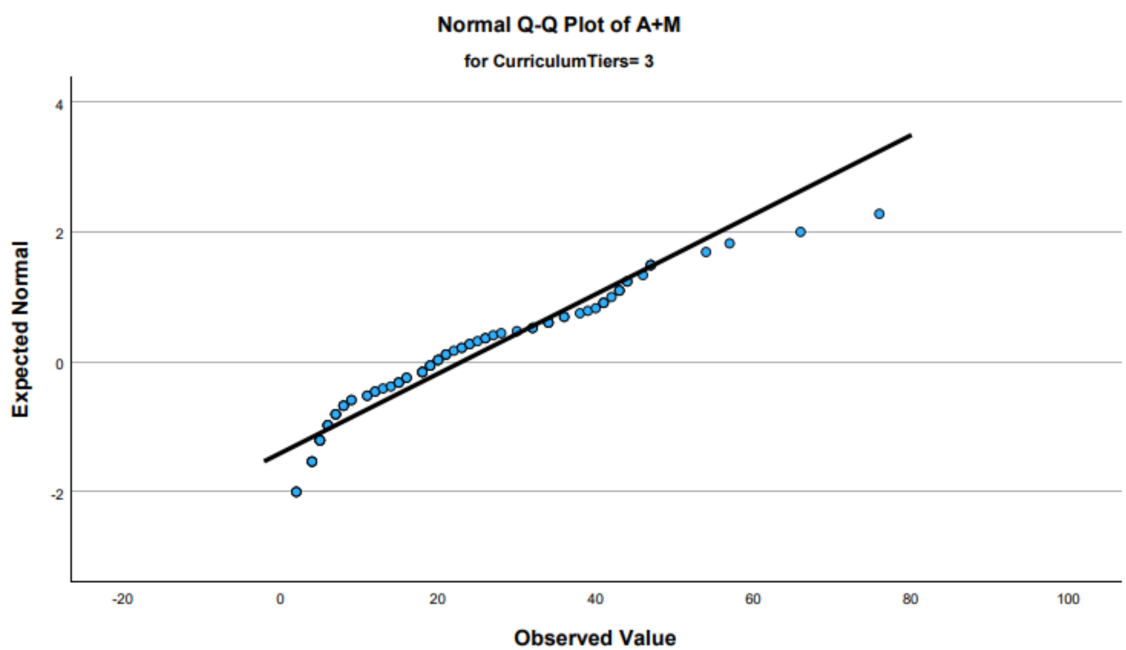
Tier 1 histogram and Q Q-Plot:





Tier 3 histogram and Q Q-Plot:





Levene's Test of homogeneity of variances:

Tests of Homogeneity of Variances

		Levene Statistic	df1	df2	Sig.
A+M	Based on Mean	1.237	2	559	.291
	Based on Median	1.048	2	559	.351
	Based on Median and with adjusted df	1.048	2	554.075	.351
	Based on trimmed mean	1.210	2	559	.299

APPENDIX W:

2017-2018 Grade 5 Removed Original Outliers ANOVA Group Statistics and Results

Descriptive statistics:

Descriptives

A+M

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
1	454	28.20	17.085	.802	26.63	29.78	2	77
3	87	22.97	16.360	1.754	19.48	26.45	2	76
4	21	18.71	13.127	2.865	12.74	24.69	2	46
Total	562	27.04	17.004	.717	25.63	28.45	2	77

ANOVA results:

ANOVA

A+M

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	3516.007	2	1758.004	6.192	.002
Within Groups	158697.132	559	283.895		
Total	162213.139	561			

ANOVA Effect Sizes[a,b]

		Point Estimate	95% Confidence Interval Lower	95% Confidence Interval Upper
A+M	Eta-squared	.022	.003	.049
	Epsilon-squared	.018	-.001	.045
	Omega-squared Fixed-effect	.018	-.001	.045
	Omega-squared Random-effect	.009	.000	.023

a. Eta-squared and Epsilon-squared are estimated based on the fixed-effect model.

b. Negative but less biased estimates are retained, not rounded to zero.

Tukey post hoc test:

Post Hoc Tests

Multiple Comparisons

Dependent Variable: A+M

	(I) Curriculum Tiers	(J) Curriculum Tiers	Mean Difference (I-J)	Std. Error	Sig.
Tukey HSD	1	3	5.239*	1.972	.022
		4	9.491*	3.761	.032
	3	1	-5.239*	1.972	.022
		4	4.251	4.097	.553
	4	1	-9.491*	3.761	.032
		3	-4.251	4.097	.553

Multiple Comparisons

Dependent Variable: A+M

	(I) Curriculum Tiers	(J) Curriculum Tiers	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	1	3	.61	9.87
		4	.65	18.33
	3	1	-9.87	-.61
		4	-5.38	13.88
	4	1	-18.33	-.65
		3	-13.88	5.38